

基于 DINA 模型的 Q 矩阵修正方法*

涂冬波 蔡 艳 戴海琦

(江西师范大学心理学院, 南昌 330022)

摘 要 本研究开发了一种基于 DINA 模型的认知诊断测验 Q 矩阵修正方法—— γ 法, 为侦查并修正 Q 矩阵中的错误提供方法学支持, 从而为保证 Q 矩阵的合理性提供基础, 并为进一步提高认知诊断的准确率服务。本文采用 Monte Carlo 模拟及与国外同类研究相比较的方法进行, 研究发现: (1)不论在何种作答失误概率(5%, 10%, 15%)情况下, 当 s, g 临界值为 0.2, 0.25 或 0.3 时, 本研究提出的 γ 法均能有效地修正错误 Q 矩阵; 同时, 当 Q 矩阵无错误时, γ 法对该 Q 矩阵未做任何修改。表明 γ 法对 Q 矩阵是否存在错误具有较强的识别能力及修正能力。(2)与国外同类研究相比, 本研究提出的 γ 法具有较理想的修正率, 且与 de la Torre (2008) 提出 δ 法的修正效果相当。但相比较而言, γ 法较 δ 法更为简单。(3) γ 法不仅能有效地修正错误的 Q 矩阵, 而且还可以进一步提高认知诊断的正确率, 尤其是对模式判准确率(PMR), 诊断正确率的最高增幅高达 40%, 大大改善了认知诊断的准确率。

关键词 认知诊断; Q 矩阵修正; DINA 模型; γ 法

分类号 B841

1 引言

认知诊断(cognitive diagnosis)目前在国内外均引起研究者和应用者的广泛关注(Huebner & Wang, 2011; DeCarlo, 2010; de la Torre, 2009; de la Torre, Hong & Deng, 2010; 喻晓锋, 丁树良, 秦春影, 陆云娜, 2011; 涂冬波, 蔡艳, 戴海琦, 丁树良, 2010a, 2010b; 陈秋梅, 张敏强, 2010; 钱锦昕, 余嘉元, 2010; 李峰, 余娜, 辛涛, 2009)。认知诊断是一项复杂的工程, 它至少包括“Q 矩阵界定”和“诊断分类”两大部分(Tatsuoka, 2009)。“Q 矩阵界定”是认知诊断的基础, Q 矩阵是阐述认知属性与测验项目关系的唯一信息载体, 是进一步将被试诊断分类的重要依据。许多研究中, 一般均假设所构建的 Q 矩阵是正确的, 并在此基础上进行诊断分类。然而合理界定/标定 Q 矩阵不是一件易事, 最典型例子就是国外许多研究者对 Tatsuoka (1990)关于“分数减法”

属性界定争论了二十几年, 到目前为止仍未有定论。由此 DeCarlo (2010)曾感叹属性界定/标定的复杂性, 并认为它是导致认知诊断目前在实际应用中受限的主要原因之一。大量研究表明(Rupp & Templin, 2008; de la Torre, 2008), 如果测验 Q 矩阵被错误界定, 则对随后的认知诊断分析带来严重影响, 它会导致参数估计误差增大及诊断正确率大幅度下降(Rupp & Templin, 2008)。因此对于构建出的测验 Q 矩阵, 在结合实测数据基础上进行验证, 并对有误的 Q 矩阵进行修正就显得十分必要。

查阅国内外相关文献, 我们发现目前对于 Q 矩阵修正的研究非常少, 仅看到 Roussos 等(2007)和 de la Torre (2008)的研究。Roussos 等人(2007)基于融合模型(Fusion Model), 对 Q 矩阵修正方法提供了一些建议: 根据被试掌握属性和项目属性将被试分成三类, 对每个项目观察正确作答落入这三类人中的比例来判断项目的属性界定是否正确; de la

收稿日期: 2011-06-01

* 国家自然科学基金(31100756, 31160203, 30860084), 教育部人文社科项目(09YJCX012, 11YJC190002), 高等院校博士点基金项目(20103604120001), 江西省教育厅高校人文社科项目(XL1011), 江西省教育厅科技项目(GJJ10098), 江西师范大学青年英才培育资助计划。

通讯作者: 涂冬波, E-mail: tudongbo@yahoo.com.cn

Torre (2008)针对 DINA 模型提出了一种修正 Q 矩阵的经验法(δ 法), 研究发现 δ 法在一定程度上能有效地侦查并修正 Q 矩阵的错误。

鉴于 Q 矩阵在认知诊断中的重要性及 Q 矩阵标定的复杂性, 本研究在前人研究基础上, 拟进一步探讨 Q 矩阵的修正方法, 为侦查并修正 Q 矩阵中的错误, 为保证 Q 矩阵的合理性提供基础, 从而为提高认知诊断的准确率服务。

2 本研究 Q 矩阵修正的方法及思路

DINA (*deterministic inputs, noisy “and” gate*, 决定性输入噪音与门)模型是目前被广泛应用的认知诊断模型之一(DeCarlo, 2010), 该模型相对比较简单及诊断准确率较高(de la Torre & Douglas, 2004; Cheng, 2008)。DINA 模型的项目参数主要有两个, 分别是猜测参数(g)和失误参数(s)。 g 是指被试未掌握项目考核的所有属性, 但却答对项目的概率; s 指被试掌握了项目考核的所有属性, 但却答错的概率。参数 s 和 g 在一定程度上反应了诊断中的“噪音”(noisy), 其值过高不利于诊断分析。

许多研究表明(Rupp & Templin, 2008; de la Torre, 2008;), DINA 模型的猜测参数(g)和失误参数(s)在一定程度上能反应测验 Q 矩阵的冗余与缺失。如果属性冗余, 则会导致参数 g 增大; 如果属性有缺失, 则会导致参数 s 增大。因此参数 s 和 g 在一定程度上反应了 Q 矩阵界定的正确性。

认知诊断中, 一般认为若被试 i 掌握了项目 j 考核的所有属性, 则被试倾向于答对该项目; 如果被试 i 未掌握项目 j 考核的所有属性, 则被试倾向于答错该项目。也就是说, 被试对项目考核属性的掌握与否会影响被试对项目的作答。本文的新方法也借用此思想, 如果项目 j 考核了属性 k , 则掌握属性 k 的被试组在项目 j 的得分应倾向高于未掌握属性 k 的被试组, 即掌握属性 k 与否会影响被试对项目 j (该项目考核了属性 k)的作答; 如果项目 j 未考核属性 k , 则掌握属性 k 的被试组在项目 j 的得分与未掌握属性 k 的被试组相当, 即掌握属性 k 与否不会影响被试对项目 j (该项目未考核属性 k)的作答。

因此本研究在结合 DINA 模型中项目参数(s, g)以及掌握属性与否是否影响对项目作答两个因素的基础上, 提出了 Q 矩阵的修正方法—— γ 法, 具体如下。

2.1 γ 法考察指标

首先考察认知诊断 DINA 模型下项目参数 s 和

g , 从而判断项目的属性界定是否有可能存在冗余或缺失等错误; 对可能存在属性错误界定的项目, 再分别考察被试对该项目每个属性的掌握与是否是否会显著影响其对该题的作答。

(1)如果某题的猜测参数 g 过大, 说明该题界定的属性可能冗余; 而又如果掌握属性 k 的组被试组与未掌握属性 k 的被试组对该题正确作答无实质性差异, 则该题很有可能未考察属性 k , 若其原始 Q 阵认定该项目考核了该属性, 则建议从该项目中删除该属性, 即 $1 \rightarrow 0$ 。

(2)如果某题的失误参数 s 过大, 说明该题界定的属性可能有漏或缺失; 而又如果掌握属性 k 的被试组与未掌握属性 k 的被试组对该题正确作答有实质性差异, 则该题很有可能考察了属性 k , 若其原始 Q 阵上未认定该项目考核了属性 k , 则建议在该项目中增加该属性, 即 $0 \rightarrow 1$ 。

(3)如果某项目猜测参数 g 和失误参数 s 均过大, 则说明该项目属性既有多余属性又缺少了已考属性, 修正方法与上面的(1)和(2)一致。比如项目 j 考核模式为(11000), 但却被错误界定为(10100), 则多了第三个属性、却少了第二个属性。

2.2 γ 法实现步骤

(1)DINA 模型参数估计。根据被试在测验上的原始得分数据及原始界定的 Q 矩阵(为行文方便, 我们均简称为“Q_{original}矩阵”), 采用 DINA 模型估计测验项目参数及被试对每个属性的掌握概率。

(2)根据 DINA 模型估计的项目参数 s 和 g , 查找 $g(j)$ 、 $s(j)$ 过大的项目, 即大于 s 和 g 临界值的题目。

(3)对任一属性 k 划分出掌握组被试与未掌握组被试

根据被试对属性掌握概率, 对任一属性 k , 将被试分成掌握组和未掌握组。Hartze (2002)指出若被试掌握属性 k 的概率在 0.6 以上, 则判被试掌握了属性 k ; 若掌握概率在 0.4 以下则判未掌握属性 k ; 若属性掌握概率在 0.4~0.6 之间则待判(即数据无足够证据进行判定)。本研究采用 Hartze (2002)的这一研究结论。即将掌握概率在 0.6 以上的视为“掌握组”, 0.4 以下的视为“未掌握组”, 0.4~0.6 之间的为“待判组”。

(4)掌握组被试与未掌握组被试在项目上得分的差异性大小检验

以效应大小(effect size, ES)作为差异显著性大小的实际指标。效应大小是基于统计检验力的指标

之一,它与 Z 检验、 t 检验、 F 检验等不尽相同,后者关注是否有差异,而效应大小则更关注观测到的差异是不是事实上的差异以及这种差异在实际中的程度如何(胡竹菁, 2010; Cohen, 1992)。

Cohen (1992)提出了均数差异的效应大小三个决断值: 0.2, 0.5 和 0.8, 分别代表较小差异、中度差异和较大差异,若小于 0.2 则说明在实际中无实质性差异。为了使新方法具有一定的灵敏度,我们采用 0.2 作为决断点,即效应大小小于 0.2 则说明无实质性差异,若大于 0.2 说明有实质性差异。

(5)对原始 Q 矩阵的修正

Q 矩阵修正公式见公式 1, 对符合公式 1 条件的 Q 矩阵中的元素进行修改, 其余元素不修改。

$$q(j,k)=$$

$$\begin{cases} 1 \rightarrow 0, & \text{if } g(j) > \text{临界值}, \text{and}, ES(j,k) < 0.2 \\ 0 \rightarrow 1, & \text{if } s(j) > \text{临界值}, \text{and}, ES(j,k) \geq 0.2 \end{cases}$$

(公式 1)

公式中“ $1 \rightarrow 0$ ”指将“1”修改为“0”, 即删除冗余属性; “ $0 \rightarrow 1$ ”指将“0”修改为“1”, 即增加缺失的属性; “ ES ”指效应大小。

由于 DINA 模型中的参数 s 和 g 是一种随机误差概率, 它们的取值越大说明随机误差的概率越大, 因此对于 s 和 g 的临界值如何取, 也是下面研究一欲重点回答的问题之一。当然公式 1 也表明, s 和 g 即使超过了临界值, 也并不一定就代表 Q 矩阵有问题而需修改。

3 研究一: γ 法中 s, g 临界值选择及 γ 法的可行性和准确性

为了考察本研究提出的 γ 法中 s, g 的临界值、 γ 法的可行性及其对 Q 矩阵修正的准确性, 采用三因素实验设计($3 \times 5 \times 6$), 分别为被试作答失误率(5%, 10%, 15%), s, g 的临界值(0.1, 0.15, 0.2, 0.25, 0.3)和 Q 矩阵错误率(0%, 5%, 10%, 15%, 20%, 25%)。其中“被试作答失误率”指在 Leighton 等人(2004)介绍的理想作答反应模式的情况下(即无猜测和失误), 根据一定的的失误比率模拟被试的作答反应数据矩阵, 用于被试得分矩阵的模拟。

实验中, 属性为 5 个, 属性间不可补偿、相互独立, 共 $2^5 - 1 = 31$ 种所有有可能的项目考核模式, 因此模拟 31 种考核的项目(即测验长度为 31 题), 被试 1000 人。

3.1 Monte Carlo 模拟过程

模拟过程分为两部分: 一是被式作答反应模拟

(含测验 Q 矩阵及被试掌握模式), 模拟方法完全采用 Leighton 等人(2004)所介绍的模拟方法; 另一是错误 Q 矩阵(Q_{original})的模拟。

3.1.1 被式作答反应模拟 (1)测验 Q 矩阵真值。即 $2^5 - 1 = 31$ 种考核模式的试题。

(2)被试属性掌握模式真值。采用 Leighton 等人(2004)的模拟方法, 被试掌握模式共 $2^5 = 32$ 种, 计算 32 种模式在 31 道试题上理想反应的总分, 将总分按正态分布进行划分, 以确定每种理想掌握模式的人数百分比, 共模拟 1000 人。

(3)被试作答反应矩阵。根据(1)和(2)模拟的测验 Q 矩阵真值及被试掌握模式真值, 在没有任何猜测及失误的情况下, 模拟被试在测验项目上的理想作答。采用 Leighton 等人(2004)的模拟方法, 在理想作答基础上, 模拟作答反应失误概率分别为 5%, 10%, 15% 的情况下被试的作答反应矩阵。

3.1.2 错误 Q 矩阵(Q_{original})的模拟 根据 3.1.1 部分中第(1)步模拟的 Q 矩阵真值, 分别模拟错误率为 0%, 5%, 10%, 15%, 20%, 25% 下的 Q_{original} 矩阵。这种错误类型有三种, 分别是属性冗余, 属性缺失, 属性既冗余又缺失。哪些项目的 Q 矩阵有误及是何类型错误均完全随机。

3.2 评价指标

采用模拟的得分矩阵及错误 Q 矩阵(即 Q_{original} 矩阵)进行 DINA 模型的诊断分析, 并根据诊断结果, 使用本研究提出的 γ 法对 Q_{original} 矩阵进行修正(我们将修正后的 Q 矩阵称为 Q_{modify} 矩阵), 分别将 Q_{original} 矩阵和 Q_{modify} 矩阵与 Q 矩阵真值进行比较, 分别计算 Q_{original} 矩阵和 Q_{modify} 的正确率, 从而判定 γ 法的修正效果。同时考察修正前后 Q 矩阵正确率的提高率、 Q 矩阵的修改率、正确修改率和错误修改率等指标, 全方位考察 γ 法的可行性及准确性。

3.3 结果

表 1、2、3 是在作答失误分别为 5%, 10% 和 15% 下 γ 法对 Q 矩阵的修正情况。表 1、2、3 中, “ Q_{bug} ”指模拟中 Q 矩阵的错误率(即 3.1 部分中的 Q_{original} 矩阵的错误率); “ Q_{original} 矩阵正确率”是指 Q_{original} 矩阵与真值 Q 矩阵的相同率; “ Q_{Modify} 矩阵正确率”指采用 γ 法对 Q_{original} 矩阵修正后的 Q 矩阵(即 Q_{Modify} 矩阵)与真值 Q 矩阵的相同率; “ Q 矩阵正确率的提高率”指修正后 Q 矩阵正确率与修正前 Q 矩阵正确率之差; “修改率”指 γ 法对 Q_{original} 矩阵的修改率, 即修改的元

素个数除以 $Q_original$ 矩阵元素总和; “正确修改率”指在修改元素中正确修改的比例; “错误修改率”指在修改元素中错误修改的比例。

从表 1、2、3 的结果, 我们可以发现不论在何种作答失误情况下, 当 s, g 临界值为 0.2、0.25 或 0.3 时, γ 法均能有效地改善 $Q_original$ 矩阵的正确率, 且修正前后 Q 矩阵正确率的提高明显; 且这时当 Q 矩阵无错误时(即 Q_bug 为 0), γ 法未对 Q 矩阵做任何修改, 表明 γ 法具有较强的错误识别能力。

结果还表明, 随着 s, g 临界值的减小, γ 法对 Q 矩阵的修改幅度(即“修改率”)整体上呈递增趋势,

也即 s, g 临界值越小, γ 法越灵敏, 但同样也会导致它的错误修改率增加(尤其当失误为 15% 时, 见表 3); s, g 临界值越大, 它会导致 γ 法不够灵敏, 即“修改率”会减少, “修正前后 Q 矩阵正确率的提高率”不高, 但错误修改率明显降低。因此 s, g 临界值的确定不宜过低也不宜过高, 过低易导致错误修改率增加, 过高易导致修正前后 Q 矩阵正确率的提高率偏低(相当于未修正或修正过少)。从表 1、2、3 试验结果来看, s, g 临界值为 0.2、0.25 和 0.3 相对比较适宜, 这几种情况下 γ 法均能有效地提高 Q 矩阵的正确率。

表 1 作答失误为 5% 的情况下 Q 矩阵的修正

Q_bug	s,g 临界值	Q_original 矩阵正确率	Q_Modify 矩阵正确率	Q 矩阵正确率 的提高率	修改率	正确修改率	错误修改率
0%	0.10	1	1	0	0	—	—
	0.15	1	1	0	0	—	—
	0.20	1	1	0	0	—	—
	0.25	1	1	0	0	—	—
	0.30	1	1	0	0	—	—
5%	0.10	0.9548	1	0.05	0.0452	1	0
	0.15	0.9548	0.9871	0.03	0.0323	1	0
	0.20	0.9548	0.9871	0.03	0.0323	1	0
	0.25	0.9548	0.9871	0.03	0.0323	1	0
	0.30	0.9548	0.9871	0.03	0.0323	1	0
	平均	0.9548	0.98968	0.0349	0.0349	1	0
10%	0.10	0.9032	0.9935	0.09	0.0903	1	0
	0.15	0.9032	0.9742	0.07	0.071	1	0
	0.20	0.9032	0.9613	0.06	0.0581	1	0
	0.25	0.9032	0.9613	0.06	0.0581	1	0
	0.30	0.9032	0.9613	0.06	0.0581	1	0
	平均	0.9032	0.97032	0.0671	0.0671	1	0
15%	0.10	0.8516	0.9806	0.13	0.129	1	0
	0.15	0.8516	0.9613	0.11	0.1097	1	0
	0.20	0.8516	0.9613	0.11	0.1097	1	0
	0.25	0.8516	0.9613	0.11	0.1097	1	0
	0.30	0.8516	0.9613	0.11	0.1097	1	0
	平均	0.8516	0.96516	0.1136	0.1136	1	0
20%	0.10	0.8	0.9871	0.19	0.1871	1	0
	0.15	0.8	0.9613	0.16	0.1613	1	0
	0.20	0.8	0.9355	0.14	0.1355	1	0
	0.25	0.8	0.929	0.13	0.129	1	0
	0.30	0.8	0.929	0.13	0.129	1	0
	平均	0.8	0.94838	0.1484	0.1484	1	0
25%	0.10	0.7548	0.9677	0.21	0.2129	1	0
	0.15	0.7548	0.9419	0.19	0.1871	1	0
	0.20	0.7548	0.929	0.17	0.1742	1	0
	0.25	0.7548	0.9097	0.15	0.1548	1	0
	0.30	0.7548	0.8903	0.14	0.1355	1	0
	平均	0.7548	0.92772	0.1729	0.1729	1	0

表 2 作答失误为 10%的情况下 Q 矩阵的修正

Q_bug	s,g 临界值	Q_original 矩阵 正确率	Q_Modify 矩阵正 确率	Q 矩阵正确率的 提高率	修改率	正确修改率	错误修改 率
0%	0.10	1	0.9355	-0.0645	0.0645	0	1
	0.15	1	1	0	0	—	—
	0.20	1	1	0	0	—	—
	0.25	1	1	0	0	—	—
	0.30	1	1	0	0	—	—
5%	0.10	0.9548	0.9742	0.02	0.071	0.636	0.364
	0.15	0.9548	1	0.05	0.0452	1	0
	0.20	0.9548	0.9935	0.04	0.0387	1	0
	0.25	0.9548	0.9806	0.03	0.0258	1	0
	0.30	0.9548	0.9806	0.03	0.0258	1	0
	平均	0.9548	0.98578	0.034	0.0413	0.927	0.073
10%	0.10	0.9032	0.9613	0.06	0.1226	0.737	0.263
	0.15	0.9032	0.9871	0.08	0.0839	1	0
	0.20	0.9032	0.9677	0.06	0.0645	1	0
	0.25	0.9032	0.9484	0.05	0.0452	1	0
	0.30	0.9032	0.9484	0.05	0.0452	1	0
	平均	0.9032	0.96258	0.06	0.07228	0.947	0.053
15%	0.10	0.8516	0.9677	0.12	0.129	0.950	0.050
	0.15	0.8516	0.9677	0.12	0.1161	1	0
	0.20	0.8516	0.9548	0.1	0.1032	1	0
	0.25	0.8516	0.9161	0.06	0.0645	1	0
	0.30	0.8516	0.9161	0.06	0.0645	1	0
	平均	0.8516	0.94448	0.092	0.09546	0.990	0.010
20%	0.10	0.8	0.9548	0.15	0.2065	0.875	0.125
	0.15	0.8	0.9548	0.15	0.1548	1	0
	0.20	0.8	0.9419	0.14	0.1419	1	0
	0.25	0.8	0.9226	0.12	0.1226	1	0
	0.30	0.8	0.9097	0.11	0.1097	1	0
	平均	0.8	0.9368	0.134	0.1471	0.975	0.025
25%	0.10	0.7548	0.9355	0.18	0.2452	0.868	0.132
	0.15	0.7548	0.9290	0.17	0.2000	0.936	0.065
	0.20	0.7548	0.9161	0.16	0.1613	1	0
	0.25	0.7548	0.8903	0.14	0.1355	1	0
	0.30	0.7548	0.8710	0.12	0.1161	1	0
	平均	0.7548	0.9084	0.154	0.17162	0.961	0.039

因此总体来看, γ 法对 Q 矩阵修正基本可行且能有效地改善 Q 矩阵的正确率, 在实践中不失为一种可供借鉴的方法。

4 研究二: γ 法与 δ 法(de la Torre, 2008)的比较

为了进一步考察 γ 法的可行性及其对 Q 矩阵修正的准确性, 我们将 γ 法与 de la Torre (2008)提

出的 Q 矩阵修正的 δ 法进行比较。为了使研究结果间具有可比性, 研究二中的所有实验条件均与 dela Torre (2008)的保持一致。属性 5 个, 项目 30 个(见表 4), 被试 5000 人, 所有项目的 s,g 参数真值均为 0.2。根据模拟真值采用 DINA 模型的项目反应函数模拟被试的得分矩阵。Q 矩阵的真值见表 4。对表 4 中项目 1、项目 11 和项目 21 三题模拟有错误的 Q 矩阵, 共 11 种错误情况, 见表 5。

表 3 作答失误为 15%的情况下 Q 矩阵的修正

Q_bug	s,g 临界值	Q_original 矩阵 正确率	Q_Modify 矩阵正 确率	Q 矩阵正确率的 提高率	修改率	正确修改率	错误修改 率
0%	0.10	1	0.8903	-0.1097	0.1097	0	1
	0.15	1	0.9871	-0.0129	0.0129	0	1
	0.20	1	1	0	0	—	—
	0.25	1	1	0	0	—	—
	0.30	1	1	0	0	—	—
5%	0.10	0.9548	0.8774	-0.08	0.1548	0.25	0.75
	0.15	0.9548	0.929	-0.03	0.1032	0.375	0.625
	0.20	0.9548	0.9935	0.04	0.0387	1	0
	0.25	0.9548	0.9806	0.03	0.0258	1	0
	0.30	0.9548	0.9806	0.03	0.0258	1	0
	平均	0.9548	0.95222	-0.002	0.06966	0.725	0.275
10%	0.10	0.9032	0.9355	0.04	0.1419	0.6364	0.3636
	0.15	0.9032	0.9677	0.07	0.1097	0.8235	0.1765
	0.20	0.9032	0.9677	0.07	0.071	1	0
	0.25	0.9032	0.9419	0.05	0.0452	1	0
	0.30	0.9032	0.9419	0.05	0.0452	1	0
	平均	0.9032	0.95094	0.056	0.0826	0.89198	0.10802
15%	0.10	0.8516	0.8903	0.04	0.1677	0.6154	0.3846
	0.15	0.8516	0.9161	0.06	0.1419	0.7273	0.2727
	0.20	0.8516	0.9355	0.08	0.1097	0.8824	0.1176
	0.25	0.8516	0.9226	0.07	0.0968	0.8667	0.1333
	0.30	0.8516	0.9161	0.06	0.0645	1	0
	平均	0.8516	0.91612	0.062	0.11612	0.81836	0.18164
20%	0.10	0.8	0.8516	0.05	0.1935	0.6333	0.3667
	0.15	0.8	0.8839	0.08	0.1613	0.76	0.24
	0.20	0.8	0.9097	0.11	0.1355	0.9048	0.0952
	0.25	0.8	0.8903	0.09	0.1161	0.8889	0.1111
	0.30	0.8	0.8774	0.08	0.0774	1	0
	平均	0.8	0.88258	0.082	0.13676	0.8374	0.1626
25%	0.10	0.7548	0.8903	0.14	0.2194	0.8235	0.1765
	0.15	0.7548	0.9161	0.17	0.1935	0.9333	0.0667
	0.20	0.7548	0.9032	0.15	0.1677	0.9615	0.0385
	0.25	0.7548	0.8839	0.14	0.1484	0.9565	0.0435
	0.30	0.7548	0.8645	0.12	0.129	0.95	0.05
	平均	0.7548	0.8916	0.144	0.1716	0.92496	0.07504

表 6 是 γ 法对 Q 矩阵的修正情况。从表 6 我们可看出, 当 s,g 临界值为 0.15 和 0.2 时, γ 法均能有效地修正 11 种实验情况下的所有错误的 Q 矩阵, 且无一处错误修改。但当 s,g 临界值为 0.25 时, 我们发现第 8、9 和 11 种实验条件下, 第 21 题修正不够彻

底, 且都是对属性 4 不能进行很好的修正, 后来进一步检查这三种情况下项目 21 的项目参数估计结果, 发现该项目参数 g 值均小于 0.25, 也就是说当 s,g 临界值越大, γ 法的灵敏度就显得不够。虽然对项目 21 修正不够彻底, 但也均未有错误修改情况, 只是有的

错误未全修改到。总体来看在 de la Torre (2008)的实验条件下, γ 法对错误 Q 矩阵的修正情况比较理想, 且与 de la Torre (2008)采用的 δ 法报告的结果(详见

de la Torre, 2008 文献)基本相当。 δ 法在 DINA 模型分析的基础上还需经过复杂的 E-M 迭代运算, 因此相比较而言, γ 法较 δ 法显得更为简单。

表 4 de la Torre (2008)模拟的 Q 矩阵真值

item	α_1	α_2	α_3	α_4	α_5	item	α_1	α_2	α_3	α_4	α_5
1	1	0	0	0	0	16	0	1	0	1	0
2	0	1	0	0	0	17	0	1	0	0	1
3	0	0	1	0	0	18	0	0	1	1	0
4	0	0	0	1	0	19	0	0	1	0	1
5	0	0	0	0	1	20	0	0	0	1	1
6	1	0	0	0	0	21	1	1	1	0	0
7	0	1	0	0	0	22	1	1	0	1	0
8	0	0	1	0	0	23	1	1	0	0	1
9	0	0	0	1	0	24	1	0	1	1	0
10	0	0	0	0	1	25	1	0	1	0	1
11	1	1	0	0	0	26	1	0	0	1	1
12	1	0	1	0	0	27	0	1	1	1	0
13	1	0	0	1	0	28	0	1	1	0	1
14	1	0	0	0	1	29	0	1	0	1	1
15	0	1	1	0	0	30	0	0	1	1	1

资料来源: de la Torre (2008)

表 5 de la Torre (2008)模拟的错误 Q 矩阵情况

实验条件	改变的项目	改变类型
1	1	$\alpha_1 \rightarrow 0, \alpha_2 \rightarrow 1$
2	1	$\alpha_2 \rightarrow 1$
3	11	$\alpha_1 \rightarrow 0, \alpha_3 \rightarrow 1$
4	11	$\alpha_1 \rightarrow 0$
5	11	$\alpha_3 \rightarrow 1$
6	21	$\alpha_1 \rightarrow 0$
7	21	$\alpha_1 \rightarrow 0, \alpha_2 \rightarrow 0$
8	21	$\alpha_1 \rightarrow 0, \alpha_4 \rightarrow 1$
9	21	$\alpha_1 \rightarrow 0, \alpha_2 \rightarrow 0, \alpha_4 \rightarrow 1$
10	21	$\alpha_1 \rightarrow 0, \alpha_2 \rightarrow 0, \alpha_4 \rightarrow 1, \alpha_5 \rightarrow 1$
11	1	$\alpha_2 \rightarrow 1$
	11	$\alpha_2 \rightarrow 0, \alpha_3 \rightarrow 0$
	21	$\alpha_1 \rightarrow 0, \alpha_4 \rightarrow 1$

资料来源: de la Torre (2008)

5 研究三: γ 法下 Q 矩阵修正对诊断正确率的影响

本部分主要在研究一实验基础上, 进一步考察本文提出的 Q 矩阵修正方法—— γ 法对诊断正确率的影响, 突出考察 Q 矩阵修正前后诊断正确率的变

化。为了便于说明问题及节省文章篇幅, 此处只报告当 s, g 临界值为 0.2 情况下 Q 矩阵修正前后诊断正确率。结果见表 7 和图 1。

诊断正确率分别采用边际判准率(Marginal Match Ratio, MMR)和模式判准率(Pattern Match Ratio, PMR)两个指标。

表 6 不同实验条件下 γ 法 Q 矩阵的修正情况

实验 条件	失误 Q 阵 题项	Q 矩阵 真值	错误 Q 矩阵	s,g 临界值	修正后的 Q 矩阵
1	1	10000	01000	0.15	10000
				0.20	10000
				0.25	10000
2	1	10000	11000	0.15	10000
				0.20	10000
				0.25	10000
3	11	11000	01100	0.15	11000
				0.20	11000
				0.25	11000
4	11	11000	01000	0.15	11000
				0.20	11000
				0.25	11000
5	11	11000	11100	0.15	11000
				0.20	11000
				0.25	11000
6	21	11100	01100	0.15	11100
				0.20	11100
				0.25	11100
7	21	11100	00100	0.15	11100
				0.20	11100
				0.25	11100
8	21	11100	01110	0.15	11100
				0.20	11100
				0.25	11110
9	21	11100	00110	0.15	11100
				0.20	11100
				0.25	11110
10	21	11100	00111	0.15	11100
				0.20	11100
				0.25	11100
11	1, 11, 21	10000	11000	0.15	10000, 11000, 11100
		11000	10100	0.20	10000, 11000, 11100
		11100	01110	0.25	10000, 11000, 11110

表 7 和图 1 可以看出, 采用 γ 法修正后的 Q 矩阵进行诊断分析, 其诊断的准确率(含边际判准率和模式判准率)明显高于修正前 Q 矩阵诊断结果, γ 法能有效改善诊断的精度; 同时, 随着 Q 矩阵错误率(Q_bug)的增加(图 1), 修正前后 Q 矩阵的边际判准率(MMR)和模式判准率(PMR)的提高率也不断增加, 尤其是对模式判准率(PMR), 最高增幅高达 40% (表 7), 大大改善了诊断的准确性。

因此 γ 法不仅能效地修正 Q 矩阵, 而且还可以进一步提高认知诊断的准确率, 这对于推动认知诊

断在实际的运用具有重要的借鉴意义。

6 小结及讨论

6.1 结论

本研究开发了一种基于 DINA 模型的 Q 矩阵修改方法—— γ 法, 并采用 Monte Carlo 模拟及与国外同类研究相比较的方法, 进一步验证 γ 法的可行性及准确性。通过本文的三个研究可以发现:

(1)不论在何种作答失误概率(5%, 10%, 15%)情况下, 当 s,g 临界值为 0.2, 0.25 或 0.3 时, 本研究

提出的 γ 法均能有效地修正错误Q矩阵;同时,当Q矩阵无错误时, γ 法对该Q矩阵未做任何修改。

表明 γ 法对Q矩阵是否存在错误具有较强的识别能力及修正能力。

表 7 s, g 临界值为 0.2 时 Q 矩阵修正前后诊断正确率

Q_bug	作答失误差	MMR-边际判准率			PMR-模式判准率		
		修正前	修正后	提高率	修正前	修正后	提高率
5%	5%	0.9726	0.9806	0.0080	0.8890	0.9140	0.0250
	10%	0.9524	0.9530	0.0006	0.8240	0.8250	0.0010
	15%	0.9022	0.9112	0.0090	0.6630	0.6790	0.0160
	平均	0.9424	0.9483	0.0059	0.7920	0.8060	0.0140
10%	5%	0.9322	0.9706	0.0384	0.7600	0.8840	0.1240
	10%	0.9064	0.9544	0.0480	0.6910	0.8160	0.1250
	15%	0.8496	0.8920	0.0424	0.5580	0.6630	0.1050
	平均	0.8961	0.9390	0.0429	0.6697	0.7877	0.1180
15%	5%	0.9458	0.9800	0.0342	0.8030	0.9100	0.1070
	10%	0.8554	0.9366	0.0812	0.5310	0.7890	0.2580
	15%	0.7912	0.8816	0.0904	0.4150	0.6470	0.2320
	平均	0.8641	0.9327	0.0686	0.5830	0.7820	0.1990
20%	5%	0.9206	0.9802	0.0596	0.6960	0.9140	0.2180
	10%	0.8706	0.9402	0.0696	0.5630	0.7910	0.2280
	15%	0.7494	0.8488	0.0994	0.3260	0.5450	0.2190
	平均	0.8469	0.9231	0.0762	0.5283	0.7500	0.2217
25%	5%	0.8428	0.9646	0.1218	0.5480	0.8630	0.3150
	10%	0.8268	0.9524	0.1256	0.4120	0.8150	0.4030
	15%	0.7460	0.8706	0.1246	0.3360	0.5800	0.2440
	平均	0.8052	0.9292	0.1240	0.4320	0.7527	0.3207

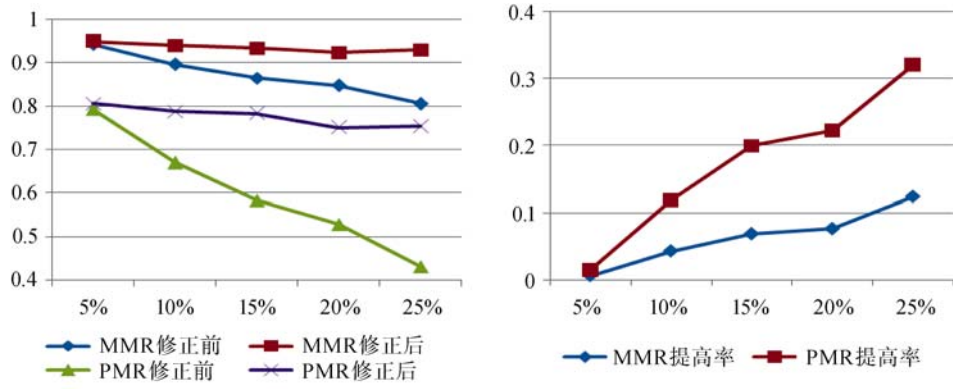


图 1 γ 法下 Q 矩阵修正对诊断正确率的影响(s, g 临界值为 0.2)

(2)与国外同类研究相比,本研究提出的 γ 法具有较理想的修正率,且与 de la Torre (2008)提出 δ 法的修正率相当。但相比较而言, γ 法较 δ 法更为简单。

(3) γ 法不仅能有效地修正错误的Q矩阵,而且还可以进一步提高认知诊断的正确率,尤其是对模式判准率(PMR),诊断正确率的最高增幅高达

40%,大大改善了认知诊断的准确率。这对于推动认知诊断在实际的运用具有重要的借鉴意义。

6.2 讨论

(1)关于 s, g 的临界值问题

研究一表明,随着 s, g 临界值的减小, γ 法对Q矩阵的修改幅度(即“修改率”)整体上呈递增趋势,也即 s, g 临界值越小, γ 法越灵敏,但同样也会导致

它的错误修改率增加(尤其当失误为 15%时); s, g 临界值越大, 它会导致 γ 法不够灵敏, 即“修改率”会减少, “修正前后 Q 矩阵正确率的提高率”不高, 但错误修改率明显偏低。因此 s, g 临界值的确定不宜过低也不宜过高, 过低易导致错误修改率增加, 过高易导致修正前后 Q 矩阵正确率的提高率偏低(相当于未修正或修正过少)。结合本文研究结果及 DINA 模型中参数 s, g 的含义, 我们建议在实际工作中 s, g 的临界值应在 0.2 及以上, 0.5 以下(这种反应“噪音”的失误或猜对的概率不宜超过 50%)。当然, 未来应进一步开展大量研究, 以进一步深清 s, g 的临界值对 Q 矩阵的修正影响, 从而更好的确定 s, g 在实际应用中的临界值。当然, 在 γ 法中, 即使项目参数 s, g 超过了临界值, 也并不一定就代表 Q 矩阵有误。

(2)关于 Q 矩阵的修正方法与专家意见的充分结合

本研究提出 Q 矩阵修正方法, 主要目的是为了在实际工作中辅助人们进行 Q 矩阵界定, 为 Q 矩阵界定提供辅助方法支持, 以尽量保证 Q 矩阵的相对合理性, 从而提高认知诊断的准确率。当然, 即使对 Q 矩阵进行了修改, 也还需进一步结合相关学科领域专家的意见, 以确定这种修改的可解释性及合理性, 从而最终确认是否要进行修改。因此, 我们认为只有将 Q 矩阵的修正方法与专家意见的充分结合, 方能更为有效地保证 Q 矩阵的合理性。

限于篇幅及时间, 本研究还有需多有待进一步完善的地方, 如 s, g 的临界值问题; 研究中属性的个数相对不多; 属性间的关系也只考察了独立型等等。这些问题都有待进一步深入。

总之, 本研究为 Q 矩阵的合理界定提供了一种辅助方法, 我们也希望借此抛砖引玉, 以推动认知诊断在我国的进一步发展。

参 考 文 献

- Chen, Q. M., & Zhang, M. Q. (2010). Development of cognitive diagnosis models and their application methods. *Advances in Psychological Science*, 18(3), 522-529.
- [陈秋梅, 张敏强. (2010). 认知诊断模型发展及其应用方法述评. *心理科学进展*, 18(3), 522-529.]
- Cheng, Y. (2008). *Computerized adaptive testing: New developments and applications*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- DeCarlo, L. T. (2010). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, 35(1), 8-24.
- de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45(4), 346-362.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- de la Torre, J., Hong, Y., & Deng, W. L. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.
- Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Hu, Z. J. (2010). The principle and method of estimating the statistical power and effect size when make Z test. *Psychological Exploration*, 30(1), 68-73.
- [胡竹菁. (2010). 平均数差异显著性检验统计检验力和效果大小的估计原理与方法. *心理学探新*, 30(1), 68-73.]
- Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, 71(2), 407-419.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoaka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205-236.
- Li, F., Yu, N., & Xin, T. (2009). Development of diagnostic math test for grader 4 and grader 5 based on the rule space model. *Psychological Development and Education*, 25, 113-118.
- [李峰, 余娜, 辛涛. (2009). 小学四、五年级数学诊断性测验的编制: 基于规则空间模型的方法. *心理发展与教育*, 25, 113-118.]
- Qian, J. X., & Yu, J. Y. (2010). The neural network-based PSP method in cognitive diagnosis. *Psychological Science*, 33(4), 915-917.
- [钱锦昕, 余嘉元. (2010). 认知诊断中基于神经网络的 PSP 方法. *心理科学*, 33(4), 915-917.]
- Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnostic system. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive Diagnostic Assessment for Education: Theory and Applications* (pp. 242-274). Cambridge: Cambridge University Press.
- Rupp, A. A., & Templin, J. L. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68(1), 78-96.
- Tatsuoka, K. K. (1990). Toward an integration of Item Response Theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). New Jersey: Lawrence Erlbaum Associates Publishers.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction*

- to the Rule Space Method. New York: Routledge, Taylor & Francis Group.
- Tu, D. B., Cai, Y., Dai, H. Q., & Ding, S. L. (2010a). A polytomous cognitive diagnosis model: P-DINA Model. *Acta Psychologica Sinica*, 42(10), 1011-1020.
- [涂冬波, 蔡艳, 戴海琦, 丁树良. (2010a). 一种多级评分的认知诊断模型: P-DINA 模型的开发. *心理学报*, 42(10), 1011-1020.]
- Tu, D. B., Cai, Y., Dai, H. Q., & Ding, S. L. (2010b). Cognitive diagnosis about primary school children's math problem-solving. *Psychological Science*, 33(6), 1461-1466.
- [涂冬波, 蔡艳, 戴海琦, 丁树良. (2010b). 小学儿童数学问题解决的认知诊断. *心理科学*, 33(6), 1461-1466.]
- Yu, X. F., Ding, S. L., Qin, C. Y., & Lu, Y. N. (2010). Application of bayesian networks to identify hierarchical relation among attributes in cognitive diagnosis. *Acta Psychologica Sinica*, 43(3), 338-346.
- [喻晓锋, 丁树良, 秦春影, 陆云娜. (2011). 贝叶斯网在认知诊断属性层级结构确定中的应用. *心理学报*, 43(3), 338-346.]

A New Method of Q-matrix Validation Based on DINA Model

TU Dong-Bo; CAI Yan; DAI Hai-Qi

(School of Psychology, Jiangxi Normal University, Nanchang 330022, China)

Abstract

Recently more and more attentions were paid on cognitive diagnosis (CD). The recognition of Q matrix was the basis and proposition of CD, and it was the only information media about cognitive attributes and test items. Only if it was recognized correctly, then the following cognitive analysis might be reliable.

This paper developed a new modification method of Q matrix (called γ method) based on DINA model. The method could detect and modify the mistakes on Q matrix that could make the Q matrix rationally and promote the correct match ratio of CD.

Monte Carlo simulation method was use here, and some comparisons with western similar studies were done. Findings showed:

(1) Under any response slip probability (5%, 10%, 15%) context, when the critical value of parameters s and g was set to be 0.2, 0.25 or 0.3, γ method could modify the mistakes of Q matrix efficiently. When Q matrix was recognized correctly, no modification will be done by γ method. These indicated that γ method could work well on recognition and modification on Q matrix whether it has mistakes or not.

(2) Compared to the similar western studies, the correct modification ratio of γ method was relatively great and similar with the consequences of δ method suggested by de la Tarre (2008). But γ method was simpler than δ method.

(3) The γ method could not only rectify the mistakes of Q matrix efficiently but also promote the correct match ratio of CD. Especially the correct match ratio of PMR increased greater than 40% which was a wonderful work.

Key words cognitive diagnosis; Q-matrix validation; DINA model; the γ method