

一种多策略认知诊断方法：MSCD 方法的开发*

涂冬波 蔡 艳 戴海琦 丁树良

(江西师范大学心理学院, 南昌 330022)

摘 要 当前国内外开发的认知诊断模型基本上只能处理单策略的测验情景, 并假设所有被试均采用同一种加工策略/解题策略, 从而忽视了加工策略的多样性及差异性。本研究根据 de la Torre 和 Douglas (2008) 采用多个 Q 矩阵来表征多个加工策略的思想, 并结合使用丁树良等(2009)修正的 Q 矩阵理论及孙佳楠, 张淑梅、辛涛和包珏(2011)的广义距离判别法, 开发了一种新的多策略认知诊断方法—— MSCD 方法。Monte Carlo 模拟研究表明: 在单策略测验情景下, 传统的单策略认知诊断方法与采用 MSCD 方法的诊断正确率均比较理想, 且差异不大; 但在多策略测验情景时, 传统的单策略认知诊断方法诊断正确率较低, 而 MSCD 方法的诊断正确率却仍较理想; 当加工策略增至 5 种时, MSCD 方法仍有较高的边际判准率、模式判准率以及加工策略判准率。研究表明 MSCD 方法基本合理、可行。这为实现对加工策略的诊断提供了方法学支持, 有利于拓展认知诊断在实际中的应用。

关键词 认知诊断; 多策略认知诊断方法; 加工策略; Q 矩阵

分类号 B841

1 引言

认知诊断(Cognitive diagnosis)是认知心理学与心理计量学相结合的产物, 通常是在心理计量学模型中融合了相关认知变量来实现对被试的诊断及分类。目前心理测量学者们开发了 60 多种认知诊断计量模型(Fu & Li, 2008), 比较成熟的有 Tatsuoaka (1995)的规则空间模型(Rule Space Model), Junker 和 Sijtsma (2001)的 DINA 模型, de la Torre 和 Douglas (2004), de la Torre(2011)的 HO-DINA 和 GDINA 模型, Leighton, Gierl 和 Hunka (2004)以及 Leighton 和 Gierl (2007)的属性层次模型(Attribute Hierarchy Model), Hartz(2002)的融合模型(Fusion Model)等等。每种计量模型各具特点。但这些模型基本上只能处理单策略的测验情景, 一般都假设所有被试均采用同一种加工策略/解题策略, 从而忽视了加工策略的多样性及差异性。

加工策略/解题策略的多样性及差异性问题的向来是当前大多数认知诊断研究所回避的问题, 通常假定所有被试均采用相同的加工策略。人们渐渐发现这种假设往往与实际情况不符, 个体的心理加工特点不尽相同, 使用的加工策略/解题策略也各有不同, 从而解决同一问题的加工过程也不尽相同。因此开发多策略的认知诊断计量模型/方法(Multiple-Strategies Cognitive Diagnosis method)值得探讨, 它不仅需要诊断被试采用何种加工策略, 还要诊断被试对相应策略所涉及的认知属性的掌握情况, 因此较传统的认知诊断模型提供的信息更为丰富、也更具价值。

查阅国内外研究文献, 我们发现目前对于多策略认知诊断方法的研究非常少。de la Torre 和 Douglas (2008)构建了关于多策略 DINA 模型, 研究中采用多个 Q 矩阵来表征多种加工策略/解题策略, 并采用参数估计的方法估计被试选用每种策略的

收稿日期: 2011-09-23

* 国家自然科学基金(编号: 31100756, 31160203), 教育部人文社科项目(编号: 09YJCXLX012, 11YJC190002), 高等院校博士点基金项目(编号: 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 (编号: 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

概率。de la Torre 和 Douglas (2008)的研究结果表明对于两种加工策略/解题策略的测验情景,多策略 DINA 模型具有较好的性能,比较可惜的是作者并未报告其对于加工策略诊断的正确率,也未进一步充分考察该模型对于二种以上加工策略/解题策略的测验情景下的性能及表现。但 de la Torre 和 Douglas (2008)采用多个 Q 矩阵来表征多种加工策略/解题策略的思想值得借鉴。

鉴于当前国内外绝大多数认知诊断模型忽视加工策略多样性问题,本研究在前人研究基础上,拟进一步探讨多策略认知诊断方法的开发,为实现加工策略的诊断提供方法学支持,并为拓展认知诊断在实际中的应用服务。

2 多策略认知诊断方法的开发

本研究开发的 MSCD 方法沿用 de la Torre 和 Douglas (2008)的思想,采用多个 Q 矩阵来表征不同加工策略,每个 Q 矩阵中各列代表该策略下的认知技能;诊断分类方法结合使用丁树良、祝玉芳、林海菁和蔡艳(2009)修正的 Q 矩阵理论及孙佳楠、张淑梅、辛涛和包珏(2011)的广义距离判别法。MSCD 方法具体诊断过程如下(其诊断示意图见图 1):

(1)计算每种策略下被试理想掌握模式即知识状态(Knowledge States, KS)及在测验项目上的理想反应模式(Ideal Response Pattern, IRP) (计算过程可详见丁树良等, 2009)。如果测验含有 M 种解题策略,

则需计算 M 套理想掌握模式和在测验上的理想反应模式,也就是图 1 中间和右边部分。

(2)根据所有被试的观察作答模式(即测验得分数据),采用 IRT 模型估计被试的能力参数和项目参数。对于二值记分测验情景,主要的 IRT 模型有 1PLM、2PLM 和 3PLM,研究者可以根据数据的具体特征选用合适的模型。为了说明方便,此处沿用 GDD 方法中的 2PLM。

(3)采用孙佳楠等(2011)的广义距离判别法计算被试观察反应模式到每种策略下每种理想反应模式的广义距离,即 $d(Y_i, I_t^m)$ 。

$$d(Y_i, I_t^m) = \sum_{j=1}^J d(Y_{ij}, I_{j(t)}^m)$$

$$d(Y_{ij}, I_{j(t)}^m) = |Y_{ij} - I_{j(t)}^m| P_j(\theta_i)^{Y_{ij}} (1 - P_j(\theta_i))^{1-Y_{ij}}$$

J 是项目个数, $Y_i = (Y_{i1}, \dots, Y_{iJ})$ 表示被试 i 的观察反应模式, I_t^m 表示在第 m 种策略下第 t 种理想反应模式, $d(Y_{ij}, I_{j(t)}^m)$ 表示被试 i 在项目 j 上的观察反应与第 m 种策略下第 t 种理想掌握模式在项目 j 上理想反应间的广义距离。 $d(Y_i, I_t^m)$ 指被试 i 的观察反应模式与第 m 种策略下第 t 种理想反应模式间的广义距离。 θ_i 采用的是被试观察反应模式估计出的能力值(孙佳楠, 张淑梅, 辛涛, 包珏, 2011)。

(4)在所有加工策略下,取 $d(Y_i, I_t^m)$ 最小者,即 $\min\{d(Y_i, I_t^m), m=1, \dots, M\}$ 所对应 I_t^m 即作为被试的属性掌握模式, I_t^m 所对应的加工策略即为被试的加工策略或解题策略。

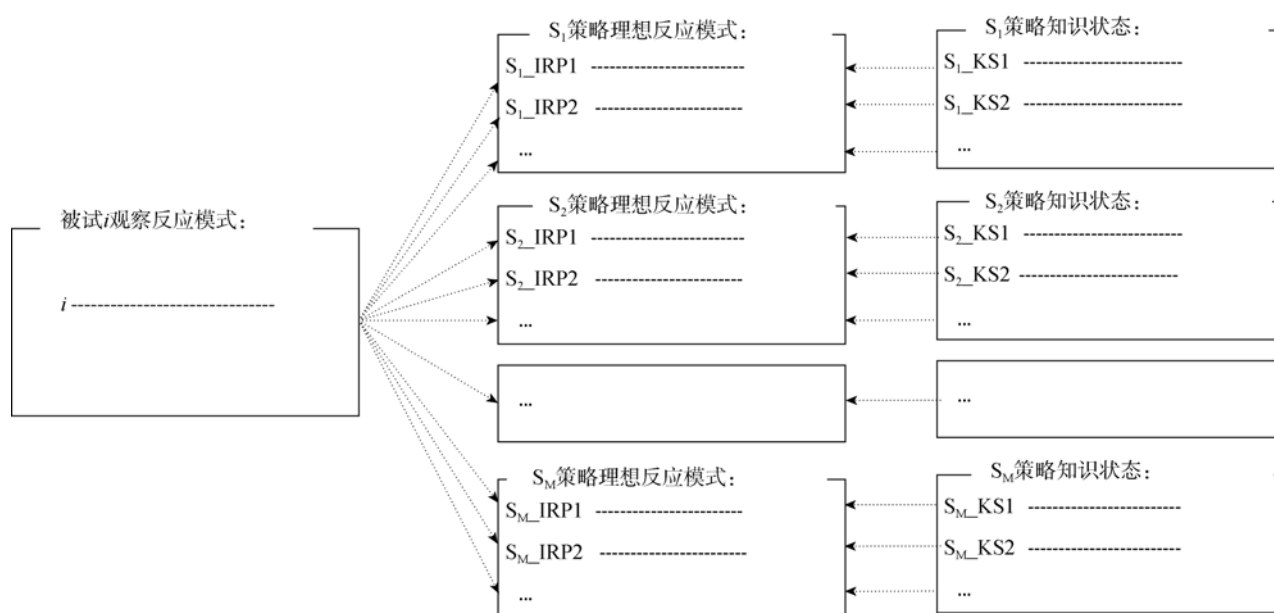


图 1 MSCD 方法的诊断示意图

值得注意的是:

(1)在进行项目参数和被试参数估计时, 仅使用到了被试的观察反应数据, 并未用到理想反应模式, 因此进行项目参数和被试参数的估计与加工策略及其理想反应模式无关; 从而广义距离 $d(Y_i, I_i^m)$ 的计算也不受加工策略的个数影响, 故 MSCD 方法不太受策略个数的影响, 理论上 MSCD 方法可以处理二种以上加工策略的测验情景。

(2)我们建议在采用 MSCD 方法时, 测验 Q 矩阵中应包含可达矩阵 R (Tatsuoka, 2009), 以实现理想掌握模式与理想反应模式的一一对应(丁树良, 杨淑群, 汪文义, 2011; 丁树良, 汪文义, 杨淑群, 2010)。

(3)对于在测验项目上全部答对的被试或全部答错的被试, 仅从反应模式本身而言, 我们是无法判断被试采用何种加工策略, 我们把这种情景视为“加工策略无法归类”或“待判”现象。

多策略认知诊断方法(MSCD 方法)具体要实现两个目标: 一是诊断被试采用何种加工策略; 二是诊断被试对相应加工策略中涉及的认知属性的掌握情况如何。这较传统的单策略模型提供的信息更为丰富、也更具价值。

3 研究一: MSCD 方法的判准率及与单策略认知诊断方法的比较

3.1 关于多策略数据的模拟

查阅国内外的相关研究文献, 目前我们只发现了一篇关于多策略认知诊断的模拟文献(见 de la Torre & Douglas, 2008), 其模拟中指定测验中含有二种加工策略/解题策略(即二个 Q 矩阵), 每种解题策略下均涉及 5 个属性, 详见表 1。

有研究(丁树良等 2010, 2011)表明, 测验 Q 矩阵若以可达矩阵 R 为子矩阵可使得知识状态与理想反应模式实现一一对应。由于 de la Torre & Douglas (2008)模拟的 20 个项目 Q 矩阵中不含相应的 R 阵, 因此本研究在 de la Torre & Douglas (2008)模拟的基础上, 分别增加 A 策略和 B 策略所对应的 R 阵试题(各 5 题, 见表 1 中有阴影部分的试题), 以实现知识状态与理想反应模式的一一对应。

3.2 实验设计

模拟两种类型测验情景, 一类测验情景是传统的单策略数据, 另一类测验情景是双策略数据, 分别采用传统的单策略认知诊断模型和新开发的多策略认知诊断方法对两类数据进行分析及比较, 以

考察新开发的多策略认知诊断方法的可行性及科学性。

表 1 结合 de la Torre & Douglas (2008)模拟的 Q 矩阵

Item	Attribute									
	Strategy A					Strategy B				
1	1	1	0	0	0	0	1	0	1	1
2	1	0	1	0	0	0	0	1	1	1
3	1	0	0	1	0	0	1	1	0	1
4	1	0	0	0	1	0	1	1	1	0
5	0	1	1	0	0	1	0	0	1	1
6	0	1	0	1	0	1	1	0	0	1
7	0	1	0	0	1	1	1	0	1	0
8	0	0	1	1	0	1	0	1	0	1
9	0	0	1	0	1	1	0	1	1	0
10	0	0	0	1	1	1	1	1	0	0
11	1	1	1	0	0	0	0	0	1	1
12	1	1	0	1	0	0	1	0	0	1
13	1	1	0	0	1	0	1	0	1	0
14	1	0	1	1	0	0	0	1	0	1
15	1	0	1	0	1	0	0	1	1	0
16	1	0	0	1	1	0	1	1	0	0
17	0	1	1	1	0	1	0	0	0	1
18	0	1	1	0	1	1	0	0	1	0
19	0	1	0	1	1	1	1	0	0	0
20	0	0	1	1	1	1	0	1	0	0
21	1	0	0	0	0	0	1	1	0	0
22	0	1	0	0	0	1	0	0	0	1
23	0	0	1	0	0	1	0	0	1	0
24	0	0	0	1	0	1	1	0	0	0
25	0	0	0	0	1	1	0	1	0	0
26	1	1	0	0	0	1	0	0	0	0
27	1	0	1	0	0	0	1	0	0	0
28	1	0	0	1	0	0	0	1	0	0
29	1	0	0	0	1	0	0	0	1	0
30	0	1	1	0	0	0	0	0	0	1

注: 上表前 20 题 Q 矩阵引自 de la Torre & Douglas (2008)第 607 页

第一类测验情景(Data_I): 所有被试均采用表 1 中 A 策略。

第二类测验情景(Data_II): 一部分被试采用表 1 中 A 策略, 另一部分被试采用表 1 中 B 策略。

孙佳楠等(2011)研究发现, 基于 Q 矩阵和广义距离的认知诊断方法总体上优于 DINA 模型、规则空间模型和属性层次模型, 因此本研究中采用基于 Q 矩阵和广义距离的认知诊断方法作为传统的单

策略认知诊断模型(下同),并与本研究开发的多策略认知诊断方法进行比较。

3.3 Monte Carlo 模拟过程

此处模拟固定测验 Q 矩阵、属性及解题策略,属性间相互独立,如表 1 所述,被试 1000 人,因此本部分只需模拟被试的掌握模式真值以及被试在测验项目上的作答反应。

(1)被试属性掌握模式真值模拟

A、B 两种策略下被试所有可能的属性掌握模式均有 $2^5=32$ 种,将这 32 种掌握模式平均分配给 1000 名被试,对于不能均分的,随机指派给被试(为了简化叙述,以下不能平均分配的,均随机指派给被试)。对于第一类单策略数据(Data_I),1000 名被试平均分配 A 策略中的 32 种掌握模式;对于第二类多策略数据(Data_II),500 名被试平均分配 A 策略中的 32 种掌握模式,另 500 名被试平均分配 B 策略中的 32 种掌握模式。但在实际诊断过程中,对于“加工策略无法归类”或“待判”的被试挑出即不进行诊断分析。

(2)被试作答反应的模拟

根据测验 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,即若被试掌握了项目考核的所有属性则答对该项目,若被试至少有一个项目考核属性未掌握则答错该项目。接着采用 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(即 slip)分别为 2%, 5%, 10% 的情况下被试的作答反应矩阵。现以一个例子说明被试作答反应的模拟:

若被试 i 使用的是 B 策略,其掌握模式为 B 策略下的(10100),该被试去做表 1 中的项目 2,项目 2

在 B 策略下考核模式为(00111),由于被试 i 未掌握项目 2 考核的属性 6 和属性 7,因此被试 i 在项目 2 上的理想作答为“0”即答错。再生成一随机数 r ,如果 r 小于等于随机失误概率(Slip),则被试 i 在项目 2 上观察作答为“1”,否则仍为“0”。其余依此类推。

(3)每种测验情景均实验 30 次。

3.4 评价指标

采用属性边际判准率(Average Attribute Match Ratio, AAMR)、模式判准率(Pattern Match Ratio, PMR)以及解题策略判准率(Strategy Match Ratio, SMR)三个指标为评价指标:

$$PMR = \frac{\sum_{i=1}^N N_{i_correct}}{N \times K}$$

$$AAMR = \frac{\sum_{i=1}^N \sum_{k=1}^K N_{ik_correct}}{N \times K}$$

$$SMR = \frac{\sum_{i=1}^N N_{i_strategy_correct}}{N}$$

其中, N 为被试总数, $N_{i_correct}$ 表示被试 i 的整个属性掌握模式是否判对,判对为 1,判错为 0; K 为属性个数, $N_{ik_correct}$ 表示被试 i 的属性 k 是否判对,判对为 1,判错为 0; $N_{i_strategy_correct}$ 表示被试 i 的解题策略是否判对,判对为 1,判错为 0。

3.5 研究结果

表 2 是单策略和多策略认知诊断方法在不同测验情景下三个评价指标上的比较。

3.5.1 单策略测验情景 不论是采用传统的单策略认知诊断方法,还是采用新开发的多策略认知诊断方法,两者的属性边际判准率(AAMR)和模式判准率(PMR)均比较理想,而且基本一致;对于多策略模型,在三种 slip 平均情况下仅有不到 0.5% 的被试的解题策略被误判为 B 策略。这表明对于单策略

表 2 单策略和多策略认知诊断方法判准率比较(30 次实验结果的平均值)

测验情景	Slip	单策略认知诊断模型				多策略认知诊断方法			
		AAMR	PMR	A 策略 SMR	B 策略 SMR	AAMR	PMR	A 策略 SMR	B 策略 SMR
单策略 (1000 人 A 策略)	2%	0.9891	0.9561	1	—	0.9884	0.9532	0.9994	—
	5%	0.9645	0.8895	1	—	0.9704	0.8878	0.9968	—
	10%	0.9372	0.7849	1	—	0.9367	0.7808	0.9924	—
	Mean	0.9636	0.8768	1	—	0.9652	0.8739	0.9962	—
多策略 (500 人 A 策略, 500 人 B 策略)	2%	0.7142	0.5052	1	0	0.9872	0.9544	0.9888	0.9884
	5%	0.7195	0.4735	1	0	0.9622	0.876	0.9596	0.9636
	10%	0.7042	0.4085	1	0	0.9184	0.7538	0.923	0.9092
	Mean	0.7126	0.4624	1	0	0.9559	0.8614	0.9571	0.9537

表 3 不同测验情景的特征

测验情景	策略类型	每种策略使用人数	每种策略包含的属性数
三种解题策略	A,B,C	500	A-6, B-5, C-4, D-3, E-3 (共涉及 23 个认知属性)
四种解题策略	A,B,C,D	500	
五种解题策略	A,B,C,D,E	500	

测验情景,单策略和多策略认知诊断方法诊断的结果基本一致,多策略认知诊断方法仍然适用于单策略测验情景。

3.5.2 多策略测验情景 在多策略测验情景下,若采用传统的单策略认知诊断方法,被试属性边际判准率(AAMR)和模式判准率(PMR)均很不理想,三种 slip 下的平均的 AAMR 和 PMR 分别为 70.7% 和 45.8%;而采用新开发的多策略认知诊断方法,三种 slip 下的平均的 AAMR 和 PMR 分别为 95.6% 和 86.1%,且 A 策略和 B 策略的诊断正确率都在 95%以上。这表明对于多策略测验情景,传统单策略认知诊断方法的错判率非常高,它不适用于多策略测验情景;这于这类数据应采用多策略认知诊断方法,且它的属性边际判准率、模式判准率以及解题策略判准率均比较理想。

4 研究二：多策略认知诊断方法的性能

研究一中,测验 Q 矩阵固定,解题策略最多为两种,且两种策略包含属性的个数也相同。为了使研究更具一般性,以及更为全面、充分考察新模型的性能,本部分主要模拟多种试验条件下多策略认知诊断方法(MSCD 方法)性能及表现。本部分涉及的解题策略最多达五种,每种解题策略下所包含的认知属性个数不尽相同,且每次测验 Q 矩阵是根据相应的测验情景随机生成。具体三种测验情景见表 3。

4.1 Monte Carlo 模拟过程

模拟如表 3 三种测验情景,所有测验情景下测验项目数仍固定为 30 题,属性间不可补偿相互独立。

(1)测验 Q 矩阵模拟

每种策略下测验 Q 矩阵随机生成,但为了实现理想掌握模式与理想反应模式的一一对应,模拟时所生成的 Q 矩阵包含可达矩阵 R。

(2)被试属性掌握模式真值模拟

与研究一的模拟相一致。每种解题策略均模拟 500 人,对于有五种解题策略的测验情景,则需模拟 2500 名被试的属性掌握模式真值。

(3)被试作答反应的模拟

与研究一的模拟相一致。

(4)每种测验情景均实验 30 次。

4.2 评价指标

同样采用属性边际判准率(AAMR)、模式判准率(PMR)以及解题策略判准率(SMR)三个指标为评价指标。

4.3 研究结果

表 4 表明在二种以上加工策略测验情景下,多策略认知诊断方法的属性边际判准率和策略判准率均约 95%,模式判准率也约 90%,总体来讲比较理想,说明 MSCD 方法适用于二种以上的加工策略。而且不论是三种解题策略还是四种或五种解题策略,其属性边际判准率和策略判准没有实质性差异。但在模式判准率上,随着解题策略的增加,模式判准率也略有提高(但提高的幅度很小),这可能与测验人数有关。本模拟中,策略越多的测验情景被试量越多,IRT 参数估计越准,当然其原因还有待进一步探讨。表 4 综合反映了 MSCD 方法能较好地处理二种以上加工策略的测验情景(多达五种),且诊断的正确率较高,同时该方法诊断的准确性受加工策略个数的影响较小,值得借鉴和推广。

5 研究结论与讨论

5.1 研究结论

加工策略/解题策略的多样性及差异性问题向来是大多数认知诊断研究所回避的问题,通常假定所有被试均采用相同的加工策略。本研究在前人研究基础上,开发了一种多策略的认知诊断方法(MSCD 方法),该方法不仅诊断被试的加工策略,还诊断相应策略下的相关认知属性的掌握情况。

研究结果表明:在单策略测验情景下,采用传统的认知诊断方法与采用 MSCD 方法的诊断正确率均比较理想,且差异不大;但在多策略测验情景时,传统的认知诊断方法诊断正确较低,而 MSCD 方法的诊断正确率却仍较理想;当加工策略增至 5 种时, MSCD 方法仍有较高的边际判准率、模式判准率以及加工策略判准率,这表明 MSCD 方法基本合理、可行。

表 4 不同加工策略下模型判准率比较(30 次实验结果的平均值)

测验 情景	Slip	多策略认知诊断方法						
		AAMR	PMR	A 策略 SMR	B 策略 SMR	C 策略 SMR	D 策略 SMR	E 策略 SMR
三种解题 策略	2%	0.9834	0.9573	0.9951	0.9938	0.9899	—	—
	5%	0.9578	0.8853	0.9841	0.9781	0.9628	—	—
	10%	0.9045	0.7788	0.9611	0.9423	0.915	—	—
	Mean	0.9486	0.8738	0.9801	0.9714	0.9559	—	—
四种解题 策略	2%	0.9855	0.9623	0.9936	0.9937	0.9875	0.9878	—
	5%	0.9582	0.9056	0.9835	0.9779	0.9633	0.9659	—
	10%	0.9035	0.8034	0.9426	0.9444	0.9105	0.9068	—
	Mean	0.949	0.8904	0.9732	0.972	0.9538	0.9535	—
五种解题 策略	2%	0.9855	0.9615	0.9947	0.9961	0.9889	0.9848	0.9863
	5%	0.9588	0.9091	0.9858	0.9826	0.9628	0.9576	0.9644
	10%	0.8982	0.8139	0.9591	0.9448	0.9098	0.9002	0.8985
	Mean	0.9475	0.8948	0.9799	0.9745	0.9538	0.9475	0.9497

5.2 讨论与展望

5.2.1 关于 MSCD 方法与 de la Torre 和 Douglas (2008)方法的比较 本研究中, 由于在计算广义距离过程中, 项目参数和被试参数的估计均是基于被试观察的作答反应数据, 而这与测验涉及的策略多少没有本质关系。因此在 MSCD 方法中, 策略的多少对于计算广义距离并没有实质性的影响; 研究二也表明三种、四种和五种解题策略下的属性判准率及策略判准率基本相当, 这也进一步证实一定范围内, 策略的多少对于 MSCD 方法诊断正确率没有实质性的影响。而这一点与 de la Torre 和 Douglas (2008)提出的多策略的 DINA 模型不尽相同, 由于其是对被试选用每种策略的概率进行参数估计, 并且所有策略下的所有属性均同时估计, 这必然出现策略越多待估参数(策略参数和属性参数)也越多、估计也就越难的困境。显然 de la Torre 和 Douglas (2008)的多策略的 DINA 模型的诊断效果会受策略个数的影响和制约。同时, 由于 de la Torre 和 Douglas (2008)采用复杂的 EM 算法估计被试选用策略的概率, 而本研究根据广义距离进行判别, 因此相比较而言本研究开发的 MSCD 方法显得更为简单, 从而在实践中更易操作。但 MSCD 方法在灵活性上还有待进一步完善, 比如允许被试在不同题目上采用不同的策略。

5.2.2 关于“加工策略无法归类”或“待判”现象 对于在测验项目上全部答对的被试和全部答错的被试, 仅从观察反应模式本身而言, 我们是无法判断被试采用何种加工策略, 也即出现“加工策略无法归类”或“待判”现象。这种情况下我们一般不判

或待判, 即作答数据无充分信息判断被试的加工策略, 需结合其它相关资料作进一步分析。当然这应是今后研究有待深入考虑的问题。

5.2.3 关于 MSCD 方法在规则空间模型(RSM)及属性层级模型(AHM)上的推广 MSCD 方法的原理同样适用于规则空间模型及属性层级模型。对于规则空间模型, 一个测验 Q 矩阵对应一组纯规则点, 这些纯规则点构成一个规则空间; 若采用多个 Q 矩阵来表征测验的多策略, 则每一个 Q 矩阵我们均可以构造出一个规则空间, 当然每一个规则空间也由相应的多个纯规则点组成。这样一种策略对应一个规则空间, 每个规则空间均由多个纯规则点组成。接着采用规则空间模型方法计算每个被试观察的序偶 $\{(\theta', \zeta')\}$ 与每种策略下所有纯规则点的序偶 $\{(\theta, \zeta)\}$ 间的马氏距离, 我们同样取马氏距离最小者(或者取马氏距离最小和次小者, 再经贝叶斯判别, 取后验概率最大者)所对应的策略为被试的加工策略, 马氏距离最小者(或后验概率最大者)所对应的知识状态或理想掌握模式为被试的掌握模式。

属性层次模型中, 一个 Q 矩阵对应一组理想反应模式, 若采用多个 Q 矩阵来表征测验的多策略, 则每个 Q 矩阵对应一组理想反应模式, 多个 Q 矩阵就对应多组理想反应模式, 而每组理想反应模式对应于一种 Q 矩阵即一种策略; 接着可以通过属性层次模型的 A 方法或 B 方法(A method or B method, Leighton et al., 2004)计算每个被试观察反应模式与每组理想反应模式中任一理想反应模式间的相似率(或似然比), 取相似率最大者所对应的掌握模式作为被试的掌握模式, 其最大者所对应的策略即为

被试加工策略。

本文对 0-1 评分的多策略认知诊断方法进行了研究, 开发了适用多策略测验情景的认知诊断新方法——MSCD 方法。限于时间及条件, 该方法还有许多地方值得进一步完善, 比如允许被试在不同的题目上采用不同的加工策略, 以增加 MSCD 方法的灵活性; 若测验在只有少数/个别试题存在多策略情景下, MSCD 方法的敏感性会降低等等; 当然未来研究还可以进一步考虑拓展到多级评分情况; 另外本研究中的“加工策略无法归类”或“待判”现象也待进一步深入; 同时进一步考察不同属性阶层关系的测验情景下 MSCD 方法的诊断效果也是很有意思的研究; MSCD 方法在规则空间模型及属性层级模型上的推广效果如何, 也可进一步研究; 更值得进一步探讨的问题是 MSCD 诊断方法在实际应用中的效果。

参 考 文 献

- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179–199.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333–353.
- de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, 73(4), 595–624.
- Ding, S. L., Wang, W. Y., & Yang, S. Q. (2011). The design of cognitive diagnosis test blueprints. *Journal of Psychological Science*, 34(2), 258–265.
- [丁树良, 汪文义, 杨淑群. (2011). 认知诊断测验蓝图的设计. *心理科学*, 34(2), 258–265.]
- Ding, S. L., Yang, S. Q., & Wang, W. Y. (2010). The importance of reachability matrix in constructing cognitively diagnostic testing. *Journal of Jiangxi Normal University (Natural Science Edition)*, 34(5), 490–494.
- [丁树良, 杨淑群, 汪文义. (2010). 可达矩阵在认知诊断测验编制中的重要作用. *江西师范大学学报(自然科学版)*, 34(5), 490–494.]
- Ding, S. L., Zhu, Y. F., Lin, H. Q., & Cai, Y. (2009). Modification of Tatsuoka's Q matrix theory. *Acta Psychologica Sinica*, 41(2), 175–181.
- [丁树良, 祝玉芳, 林海菁, 蔡艳. (2009). Tatsuoka Q 矩阵理论的修正. *心理学报*, 41(2), 175–181.]
- Fu, J., & Li, Y. (2008). Cognitively diagnostic psychometric models: An integrative review. ETS research report.
- Hartz, S. M. (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.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258–272.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205–237.
- Leighton, J. P., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications* (pp. 242–274). Cambridge, UK: Cambridge University Press.
- Sun, J. N., Zhang, S. M., Xin, T., & Bao, Y. (2011). A cognitive diagnosis method based on Q-matrix and generalized distance. *Acta Psychologica Sinica*, 43(9), 1095–1102.
- [孙佳楠, 张淑梅, 辛涛, 包珏. (2011). 基于 Q 矩阵和广义距离的认知诊断方法. *心理学报*, 43(9), 1095–1102.]
- Tatsuoka, K. K. (1995). Architecture of knowledge structure and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 327–361). Hillsdale, NJ: Erlbaum.
- Tatsuoka, K. K. (2009). *Cognitive Assessment: An introduction to the rule space method*. New York, NY: Routledge Academic.

A New Multiple-Strategies Cognitive Diagnosis Model: the MSCD Method

TU Dong-Bo; CAI Yan; DAI Hai-Qi; DING Shu-Liang

(Psychology College of Jiangxi Normal University, Nanchang 330027, China)

Abstract

Almost all of the current cognitive diagnosis models allow for single-strategy of problem solving, and assume all examinees use the same processing strategy. Based on the extant studies, the current study developed a new multiple-strategies cognitive diagnosis model, called MSCD method.

Monte Carlo method was employed here to explore the feasibility of the MSCD method and to examine the estimation precision as well as the properties of the MSCD method.

The findings were presented: (1) Under the single-strategy tests or data, the average attribute match ration (AAMR), pattern match ration (PMR) and strategy match ration (SMR) of the MSCD method were as high as those of single-strategy cognitive diagnosis model (single-strategy model). (2) Under multiple-strategy tests or data, the MSCD method also had high AAMR, PMR and SMR, but single-strategy model did not. (3) When the number of strategy reaching 5, the MSCD method also had very high AAMR, PMR and SMR. All in all, the new developing model, the MSCD method, was reasonable and could be accepted.

Key words cognitive diagnosis; multiple-strategies cognitive diagnosis model; processing strategy; Q-matrix