

文章编号: 1000-5862(2015)04-0371-06

多策略的多级评分认知诊断方法的开发

祝玉芳¹, 王黎华¹, 丁树良², 汪文义²

(1. 上饶师范学院小教分院, 江西 上饶 334000; 2. 江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 多级评分可以提供更多的认知诊断信息. 在对多级评分多策略实测数据进行分析的基础上提出多策略的多级评分认知诊断方法. 考虑到被试使用解题策略和评分方式无关, 给出一种先将多级评分转化为0-1评分, 再采用0-1评分方式对策略进行诊断, 最后使用多级评分诊断被试的属性掌握模式. 实验结果表明了这种方法的有效性.

关键词: 认知诊断; 多策略; 多级评分; Q 矩阵

中图分类号: B 841.7; TP 301.6 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.04.08

0 引言

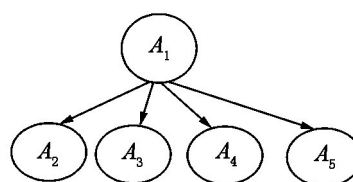
认知诊断是现代测量理论发展的重要领域. 根据被试对测试项目的作答情况, 诊断被试的知识状态, 即获得被试对诊断内容所涉及属性的掌握情况. 随着认知诊断研究的深入, 认知诊断的目的不仅要获得被试对认知属性的掌握情况, 还要诊断被试采用何种策略进行作答. 在大多数认知诊断方法中, 一般假设所有被试对诊断内容的作答采用相同的策略. 这种假设往往与现实情况不符. 不同的被试解决同一问题的心理加工特点和过程不尽相同, 对诊断内容的作答采用的策略也会不同. 在认知诊断中, 忽略被试采用策略的多样性及差异性势必会影响被试知识状态诊断的准确性; 而考虑解题策略的多样性可以提高对被试知识状态诊断的准确性^[1-2].

多级评分是考试的常见题型. 相同的测验题目, 采用多级评分可以比采用0-1评分提供更加丰富的被试信息, 特别是诊断信息. 然而在多级评分认知诊断测验中, 存在被试采用多种解题策略的现象. 事实上, 在实证研究中, 就遇到这个问题. 这样就形成了一种多策略的多级评分认知诊断测验情景. 查阅国内外的研究资料, 到目前为止, 尚未发现有适合于多策略的多级评分测验的认知诊断方法. 因此有必要研究多策略的多级评分的认知诊断的理论和方法.

1 多策略的多级评分认知诊断方法

1.1 问题的提出

在运用多级评分的广义距离法(GDD-P)^[3]对某中职学校750名学生在“进位计数制”这一内容进行认知诊断过程中, 通过对学生的访谈发现, 有被试在第1, 7, 14, 15题(记为 I_1, I_7, I_{14} 和 I_{15})使用不同的解题策略. 这个测验涉及的属性有5个, 分别为 A_1, A_2, A_3, A_4 和 A_5 . 属性间的层级关系及涵义如图1所示. 被试具体的策略使用情况见表1.



A_1 为进制的概念, A_2 为十进制转化成其它进制, A_3 为其它进制转换成十进制, A_4 为二进制转换成八进制或十六进制, A_5 为八进制或十六进制转换成二进制.

图1 属性涵义及属性间层级结构图

de la Torre等^[1]采用多个 Q 矩阵表征不同的解题策略, 每个 Q 矩阵中各列代表该策略下的认知技能, 同时也代表该题目对应的属性模式. 表2是“进位计数制”诊断测验的测验 Q 阵(Q_{14}), 如 I_1 列代表第1题的题目属性为(A_1, A_3), 如果被试使用的解

收稿日期: 2014-11-26

基金项目: 国家自然科学基金(30860084, 31160203, 31100756, 31360237), 教育部人文社会科学研究青年基金(13YJC880060), 江西省教育科学“十二五”规划2015年度重点课题(15ZD31045)和江西省教育厅科技计划(GJJ13207, GJJ13208, GJJ13209, GJJ13226, GJJ13227)资助项目.

作者简介: 祝玉芳(1982-), 女, 江西广丰人, 讲师, 主要从事计算机辅助教学及教育和心理测量方面的研究.

1.2 多策略的多级评分认知诊断方法的开发

本文在涂冬波等^[2]的基础上探讨开发多策略的多级评分认知诊断方法.用多个 Q 矩阵表征多种解题策略,每个 Q 矩阵中各列代表在该策略下项目的属性模式.本研究采用“进位计数制”测验中的 Q 矩阵,欲解决“进位计数制”测验中的多策略问题;诊断分类方法结合使用修正的 Q 矩阵理论^[5]及多级评分广义距离判别法(GDD-P)^[3].为了完整起见,对GDD-P简单介绍如下.

GDD-P是一种多级评分的认知诊断方法.该方法定义被试的观察反应模式和理想反应模式之间的广义距离 $d(Y_i, I_t)$ 为

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

其中

$d(Y_{ij}, I_j^{(t)}) = |Y_{ij} - I_j^{(t)}| \times (1 - P_{ij\theta}(\theta_i))$, (2)
 $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ij})$ 为被试 i 的观察作答模式,
 $I_t = (I_1^{(t)}, I_2^{(t)}, \dots, I_j^{(t)})$ 为第 t 种理想反应模式,
 $d(Y_{ij}, I_j^{(t)})$ 为项目 j 上被试 i 的观察反应 Y_{ij} 与项目 j 上第 t 种理想反应 $I_j^{(t)}$ 的广义距离, $d(Y_i, I_t)$ 为 Y_i 到 I_t 的广义距离,即为所有项目的广义距离之和,
 $P_{ij\theta}(\theta_i)$ 为能力水平为 θ_i 的被试 i 在项目 j 上得到反应为 $I_j^{(t)}$ 的概率, $P_{ij\theta}(\theta_i)$ 可以指定为一种多级评分IRT模型(如GRM或GPCM). (2)式中的被试 i 的能力参数 θ_i 的值用被试的观察作答模式的估计值 $\hat{\theta}_i$.将使 $d(Y_i, I_t)$ 值达到最小的理想反应模式 I_t 对应的属性掌握模式判为被试 i 的属性掌握模式^[3].

诊断过程分成2个步骤:(i)测验蓝图的制定,不同的策略对应不同的测验蓝图;(ii)对于不同的测验蓝图,分别做策略诊断和认知诊断.由于策略的使用不受评分方式的影响,而多级评分可以提供更加丰富的诊断信息,所以采用先将得分看成是0-1评分进行策略诊断,再做多级评分认知诊断的方式;由于使用不同策略解答同一个题目,可能满分值不同(如小学生计算从1累加到100,通常策略是逐一累加,而高斯却采用 $100+1$, $99+2$, $51+50$ 的策略计算,相信大部分老师都会认为这2种策略的满分应该不相同),对在不同测验蓝图中相同题目对应不同得分,要采取某种方式进行转换,否则,将不同测验矩阵对应的不同测验得分放在一起进行参数估计,由于同一个题目满分值不同而无法处理.因此在实际工作中,或者根据测验 Q 矩阵由评分员重新评分,或者采取某种方法由计算机根据某种法则补充

缺失的分数.下面结合上文提到的实例,对诊断过程具体化.

1.2.1 测验蓝图的制定 根据属性的层级结构(本文采用图1的属性层级结构)推导出可达阵 R ,通过扩张算法^[5-6]导出潜在 Q 阵(Q_r 阵), Q_r 阵再加一个零列,构成学生 Q 阵(Q_s 阵)^[7],它的列表示了被试知识状态的所有可能类.测验 Q 阵记为 Q_i 阵,不同的 Q_i 阵表示不同的解题策略,“进位计数制”测验的 Q_i 阵根据老师的解题策略采用 Q_{iA} (见表2),根据访谈结果又增加了2个 Q_i 阵,分别为 Q_{iB} (见表3)和 Q_{iC} (见表4).其中 Q_i 要满足的一个条件为包含可达阵 R (Q_{iA} 、 Q_{iB} 和 Q_{iC} 均包含 R),使得知识状态和理想反应模式一一对应^[8-9].

1.2.2 对不同的 Q_i 分别做认知诊断 Q_i 阵确定好以后,就可以编制测验进行实测,由此得到所有被试的观察作答模式和测验得分数据.评分方式可以是0-1评分,也可以是多级评分.多级评分数据可以方便地转换成为0-1评分,即满分对应1分,其他对应0分.对于0-1评分矩阵,应用文献[2]的方法,对解题策略进行诊断.由于被试策略的使用和评分方式无关,所以上面的方法实际上也得出了多级评分认知诊断测验的策略的诊断.

下面仅仅讨论知识状态的估计问题.同一个被试,基于不同的策略对应的测验 Q 矩阵,估计的知识状态可能不同,从而其理想得分(期望得分)可能不同,甚至因为评分规则的变化,导致其观察得分也不同.本文使用的多级评分的评分规则是每掌握项目中一个属性,得分增加1分^[3,10-11].在这种评分规则下,对于同一题目,使用不同解题策略涉及的属性个数会不同,使用本文多级评分的评分规则,同一题目因为属性个数不同会导致题目满分不同,如 Q_{iA} 和 Q_{iB} 中的第1题和第7题, Q_{iA} 在第1题和第7题的满分都是2分,而 Q_{iB} 在第1题和第7题的满分都是3分.

根据不同的评分方式,选择适合这种评分方式对应的认知诊断模型进行认知诊断分析.由于GDD-P诊断准确率高,本文采用GDD-P进行认知诊断.GDD-P是基于项目反应理论模型基础上开发的,要根据数据的具体特征选用合适的多级评分IRT模型(如GRM或GPCM等)估计被试的能力参数和项目参数.

1) 对于不同的策略,即不同的 Q_i ,如果每一题的满分均对应相同的话,即相同的题目包含的属性可以不同,但是所包含的属性数目相同.这时,就可以统一估计项目参数.诊断时,对每个 Q_i ,使用 Q_s 求

出理想反应模式,然后把所有的 Q_i 的理想反应模式集作为多策略的理想反应模式集,最后采用某一种多级评分认知诊断模型进行诊断分类. 本文采用 GDD-P^[3], 因此对于每一个观察反应模式(ORP), 均要计算 ORP 与理想反应模式(IRP)集中的每一个 IRP 的广义距离,选出广义距离值最小的 IRP,其对应的属性掌握模式即为该 ORP 的属性掌握模式.

2) 如果存在同一个题目在不同策略上满分不相同的情况,仍然按(i)方法进行诊断,由于同一个题目满分值不同而无法处理,这时按如下方法进行处理:把同一策略的 ORP 和相应的 IRP 放在一起,作为一个得分阵,采用一种多级评分 IRT 模型估计被试的能力参数和项目参数,如把被诊断为 A 策略的所有被试的观察作答模式和 A 策略的理想反应模式作为一个得分阵进行参数估计. 得到项目参数和能力参数后,在同一种策略下,采用 GDD-P 计算该策略下的被试观察作答模式到该策略下的每种理想反应模式的广义距离,广义距离达到最小的理想反应模式对应的属性掌握模式判为该被试的属性掌握模式^[3]. 这里需要注意的是:设计“进位计数制”测验时未考虑多策略的问题,先采用老师的解题策略,即 Q_{IA} ,通过访谈才增加了 B、C 策略,故所有的测验得分是根据 Q_{IA} 来评分的,即使用 A 策略来评分,那对被诊断为 B、C 策略的被试观察作答模式,此时如果与对应的理想反应模式一起参数估计的话,也会由于同一个题目满分值不同而无法估计,所以需要被判为 B、C 策略的被试重新用对应的测验 Q 阵来评分再估计参数.

2 多策略的多级评分认知诊断方法与单策略认知诊断方法的比较

2.1 实验设计

本文主要考察不同策略导致同一项目满分不同的情况. 通过模拟 3 种策略数据,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该数据进行分析 and 比较,以考察新开发的多策略认知诊断方法的可行性和科学性.

2.2 Monte Carlo 模拟过程

根据上节中的(i)得到 3 种策略(A、B、C)的测验 Q 矩阵,分别为 Q_{IA} 、 Q_{IB} 和 Q_{IC} ,属性掌握模式(即知识状态)和理想反应模式也一一确定. 因此本部分只需模拟被试的属性掌握模式真值和被试的观察作答模式.

1) 被试属性掌握模式真值模拟. A、B、C 共 3 种策略下被试的属性掌握模式均有 17 种,每种属性掌握模式分配 30 名被试,即每种策略各有 510 名被试,共有 1 530 名被试.

2) 被试的观察作答模式的模拟. 在被试的属性掌握模式对应的理想反应模式基础上,模拟产生 2%、5%、10% 和 15% 的失误概率(即 slip)得到被试的观察作答模式. 如要模拟每个模式的每个项目的得分有 5% 的概率发生 slip 的情况,采用一个服从开区间(0, 1)上均匀分布 $U(0, 1)$ 的随机数 r ,若 $r > 0.925$ 且该得分不是满分,则该项目得分增加 1 分;若 $r < 0.025$ 且该项目得分不是 0 则该项目得分减 1 分,否则该项目得分不变,这样就模拟产生一个以 5% 概率发生 slip 的观察项目反应模式. 用相同的方法模拟产生 2%、10% 和 15% 失误概率的观察项目反应模式^[11].

3) 先把多级评分的观察作答模式和理想反应模式均转换为 0-1 评分;再计算观察作答模式和每个理想反应模式的 GDD 值,使得 GDD 值达最小的理想反应模式对应的策略即为被试的解答策略;把同一策略的观察作答模式和相应的理想反应模式,采用一种多级评分 IRT 模型(这里用 GRM)估计被试的能力参数和项目参数,然后用 GDD-P 确定被试的知识状态.

对于模拟数据的 ORP 矩阵,如原本用 B 策略评分的而被判为 A 策略的被试作答模式,不能像实测数据那样重新评分,所以在参数估计之前要进行分值转换,把不同策略间满分不同的项目进行分值转换. 分值转换的方法是:满分或者零分分别转换为相应的满分和零分,除此之外,采用抛随机数的方法进行转换. 如第 1 题如果使用 A 策略,项目属性向量为(1 0 1 0 0),项目满分为 2 分;如果使用 B 策略,项目属性向量为(1 1 0 0 1),项目满分为 3 分. 如果该题在 B 策略上得到 0 分或者 3 分,对应 A 策略 0 分或者 2 分;如果在 B 策略上得到 1 分或者 2 分,即 $x = 1$ 或者 2,应该对应 A 策略多少分? 本文利用一个相对简单的方法处理这个问题,即使用抛随机数的方式来解决这一问题. 若 $x = 1$,从(0, 1)上的均匀分布抽取随机数 u , $\mu < 0.5$,则对应 A 策略 0 分,否则对应 A 策略 1 分;若 $x = 2$,从(0, 1)上的均匀分布抽取随机数 u , $\mu < 0.2$,则对应 A 策略 0 分,否则对应 A 策略 1 分;第 7 题也做相同的分值转换. 分值转换时对原本用 B 策略评分而被判为 C 策略的被试作答模式的方法与原本用 B 策略而被判为 A 策略的相同;而原本用 A 或 C 策略而被判为 B 策略的则

在第 1 题和第 7 题上按如下分值转换: 如果这个题目在 *A* 或 *C* 策略上得到 0 分或者 2 分, 对应 *B* 策略 0 分或者 3 分; 如果此题 *A* 策略上得到 1 分, 如果 $u > 0.8$, 相当于 *B* 策略 2 分; 否则只相当于 1 分.

4) 每个试验均重复进行 20 次以减少误差.

2.3 研究结果

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) 和解题策略判准率(Strategy Match Ratio, SMR) [2]. 表 5 列出了单策略和多策略的多级评分认知诊断判准率的比较, 表 6 列出了多策略的

多级评分认知诊断的策略判准率. 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) 均不理想. *A* 种 slip 下的平均 AAMR 和 PMR 分别为 61.76% 和 31.40%; 而采用新开发的多策略多级评分认知诊断方法 *A* 种 slip 下的平均 AAMR 和 PMR 分别为 97.49% 和 90.32%, 且多策略的诊断正确率为 95.96%. 这表明对于多策略测验情景, 传统单策略多级评分认知诊断方法的诊断准确率非常低, 它不适用于多策略情景, 此时应采用多策略多级评分认知诊断方法, 且它的判准率均比较理想.

表 5 单策略和多策略的多级评分认知诊断判准率比较(20 次实验结果的平均值)

slip		单策略		多策略		<i>A</i> 策略		<i>B</i> 策略		<i>C</i> 策略	
		PMR	AAMR	PMR	AAMR	PMR	AAMR	PMR	AAMR	PMR	AAMR
15%	\bar{x}	0.310 6	0.629 9	0.881 8	0.974 0	0.893 8	0.977 0	0.920 6	0.982 1	0.830 8	0.963 0
	<i>s</i>	(0.003 9)	(0.001 8)	(0.009 6)	(0.002 6)	(0.014 3)	(0.003 1)	(0.014 7)	(0.005 9)	(0.016 1)	(0.003 7)
10%	\bar{x}	0.320 7	0.632 9	0.893 5	0.972 3	0.906 2	0.973 0	0.902 6	0.972 8	0.871 6	0.971 2
	<i>s</i>	(0.003 8)	(0.001 0)	(0.035 7)	(0.014 0)	(0.067 5)	(0.027 8)	(0.086 7)	(0.029 5)	(0.045 9)	(0.013 4)
5%	\bar{x}	0.300 3	0.596 8	0.888 3	0.967 6	0.861 2	0.959 6	0.899 5	0.968 6	0.904 3	0.974 7
	<i>s</i>	(0.018 2)	(0.028 4)	(0.023 3)	(0.006 6)	(0.050 9)	(0.014 7)	(0.040 7)	(0.012 7)	(0.038 7)	(0.013 2)
2%	\bar{x}	0.324 3	0.610 6	0.949 2	0.985 5	0.938 9	0.983 7	0.962 8	0.987 8	0.945 8	0.984 8
	<i>s</i>	(0.006 7)	(0.013 9)	(0.011 7)	(0.003 4)	(0.023 4)	(0.005 8)	(0.018 3)	(0.006 6)	(0.018 4)	(0.006 5)
	\bar{x}	0.314 0	0.617 6	0.903 2	0.974 9	0.900 0	0.973 3	0.921 4	0.977 8	0.888 1	0.973 4

表 6 多策略的多级评分认知诊断的策略判准率(20 次实验结果的平均值)

slip	多策略 SMR	<i>A</i> 策略 SMR	<i>B</i> 策略 SMR	<i>C</i> 策略 SMR
15%	0.924 889	0.898 725 49	0.931 960 784	0.936 470 588
10%	0.950 411	0.934 019 608	0.954 509 804	0.957 745 098
5%	0.973 624	0.963 039 216	0.975 098 039	0.980 098 039
2%	0.989 785	0.984 803 922	0.991 274 510	0.992 254 902
\bar{x}	0.959 677	0.945 147 059	0.963 210 784	0.966 642 157

3 讨论与展望

本文在“进位计数制”测验中遇到被试使用多策略, 及其不同策略导致同一项目的满分不同这一特殊情景上, 借鉴前人的研究, 开发了一种多策略的多级评分认知诊断方法. 研究结果表明: 在多策略测验情景时, 采用传统的单策略认知诊断方法诊断错误率较高, 而多策略的多级评分认知诊断方法的诊断准确率却较理想, 这表明多策略的多级评分认知诊断方法基本合理、可行.

限于时间及条件, 该方法还有许多地方需要进

一步完善, 如多种试验条件下多策略的多级评分认知诊断方法的性能及表现, 如在“进位计数制”测验中, 老师讲解的解题策略是 *A* 策略, 所以大部分人会使用 *A* 策略, 而使用 *B* 策略和 *C* 策略的人数只是小部分, 此时的多策略比单策略的优越性差距不会减少; 更值得进一步研究的问题是多策略的多级评分认知诊断方法在实际应用中的效果; 本文采用掌握题目中 1 个属性便增加 1 分的多级评分的评分方式, 探讨更合适的多级评分的评分方式也很有价值. 特别是 1.2 节中的(ii) 中的同一个被试在同一个题目的得分, 因为使用策略的不同而不同, 并且既不是 0 分, 也不是满分, 应该如何转换, 文中上述想

法,即对于那些在第 j 题上既不是 0 分又不是满分的被试,在这个题目上的得分的随机成分比较大,所以使用抛随机数的方法有一定的道理,但是十分粗糙.如果使用被试在这个解题策略下该题得分概率和某一个 $[0, 1]$ 上的随机数比较,或许效果会更好.

4 参考文献

- [1] de la Torre J, Douglas J A. Model evaluation and multiple strategies in cognitive diagnosis: an analysis of fraction subtraction data [J]. Psychometrika, 2008, 73(4): 595-624.
- [2] 涂冬波, 蔡艳, 戴海琦, 等. 一种多策略认知诊断方法: MSCD 方法的开发 [J]. 心理学报, 2012, 44(11): 1547-1553.
- [3] Sun Jia'nan, Xin Tao, Zhang Shumei, et al. Apolytomous extension of the generalized distance discriminating method [J]. Applied Psychological Measurement, 2013, 37(7): 503-521.
- [4] 李瑜. 多选题认知诊断测验编制及多策略的多选题认知诊断模型的开发 [D]. 南昌: 江西师范大学, 2014.
- [5] 丁树良, 祝玉芳, 林海菁, 等. Tatsuoaka Q 矩阵理论的修正 [J]. 心理学报, 2009, 41(2): 175-181.
- [6] Ding Shuliang, Luo Fen, Cai Yan, et al. Complement to Tatsuoaka's Q matrix theory [C]. Shingemasu K, Okada A, Imaizumi T, et al. New trends in psychometrics. Tokyo: Universal Academy, 2008: 417-423.
- [7] 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.
- [8] 丁树良, 杨淑群, 汪文义. 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.
- [9] 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, 33(5): 396-401.
- [10] 田伟, 辛涛. 基于等级反应模型的规则空间方法 [J]. 心理学报, 2012, 44(1): 249-262.
- [11] 祝玉芳, 丁树良. 基于等级反应模型的属性层级方法 [J]. 心理学报, 2009, 41(3): 267-275.

The Development of Multiple – Strategies Cognitive Diagnosis with Polytomous Response

ZHU Yufang¹, WANG Lihua¹, DING Shuliang², WANG Wenyi²

(1. Shangrao Normal School, Shangrao Jiangxi 334000, China;

2. Computer Information Engineering College, Jiangxi Normal University, Nanchang Jiangxi 330027, China)

Abstract: Polytomous models can give more information for score than dichotomous. Based on the application of the cognitive diagnosis with polytomous response, the current study developed a multiple-strategies cognitive diagnosis with polytomous response. Examinees using the problem-solving strategies and score independent are considered, the paper adopts the following methods for multiple-strategies cognitive diagnosis with polytomous response. Firstly, transform the score of polytomous to the score of dichotomous. Secondly, use the multiple-strategies cognitive diagnosis with dichotomous response to find the problem-solving strategies of examinees. Finally, employ the multiple-strategies cognitive diagnosis with polytomous response to diagnose students' attribute mastery patterns. Monte Carlo method was employed here to the feasibility of the multiple-strategies cognitive diagnosis with polytomous response, the results show that the method is effective and reasonable.

Key words: cognitive diagnosis; multiple strategies; polytomous response; Q -matrix

(责任编辑: 冉小晓)