

文章编号: 1000-5862(2019)04-0376-06

## 多级评分多策略认知诊断方法

黄 玉, 罗 芬, 熊建华, 丁树良, 甘登文\*

(江西师范大学计算机信息工程学院, 江西南昌 330022)

**摘要:** 多级评分比 2 级评分能提供更全面更丰富的诊断信息, 更好地反映被试对知识的掌握状况. 考虑到存在不同认知诊断策略对应不同属性层级关系的情形, 甚至在某种评分规则下, 存在同一个项目使用不同策略, 对应的满分值也不同的复杂情况, 该文利用项目反应理论等值的做法, 处理这种复杂情况. Monte Carlo 模拟实验表明: 对含有多策略的多级评分反应数据, 这种处理方法有助于提高认知诊断模型分类的准确性.

**关键词:** 认知诊断; 多级评分; 多策略认知诊断;  $Q$  矩阵

中图分类号: B 841.7; TP 301.6 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2019.04.08

### 0 引言

认知诊断(CD)将当代心理测量学与认知心理学的理论方法相融合,它不仅仅是停留在对个体(被试)的能力评价上,更主要是根据被试对项目的作答反应,诊断被试的知识状态(Knowledge State, KS),为对被试进行个性化补救提供依据.

在大多数认知诊断模型中,假定每个被试在解答同一个任务时的答题策略以及基本心理过程等均相同<sup>[1]</sup>,这可能与实际情形不相符,被试的年龄和工作记忆会对其采取的策略产生影响<sup>[2]</sup>.例如,小学算术运算题,国外小学生大多使用数手指、出声计算、直接提取和分解等策略<sup>[3]</sup>;我国小学生从 2 年级开始就会熟记九九乘法口诀表,对于简单的乘法运算无需计算就可以快速作答<sup>[4]</sup>.又如,对于问题“ $1+2+3+\cdots+100=?$ ”,低年级学生一般采用逐个直接相加的方法,高年级学生可能会使用“(1+100)+(2+99)+ $\cdots$ +(50+51)=5050”方法.显然,在答题时不同的被试可能采用不同的解题策略<sup>[5]</sup>,这种情况下如果仅仅使用第 1 种策略进行诊断,必然无法为使用第 2 种策略的被试进行有效诊断.在认知测试中,考虑答题策略的多样化能够更准确地判断被试的知识状态<sup>[6]</sup>.

随着考试形式的多样化,多级评分的试题在考试中占有举足轻重的比例,如计算题和论述题等,往往这类测试题更能反映出考生对知识的掌握状况.查阅国内外的研究文献,发现多数的认知诊断模型采用单策略 0-1 评分模式,而对于多策略的多级评分研究较少,因而有必要进行此类评分模式的认知诊断方法的研究.

### 1 多级评分多策略认知诊断方法的提出

#### 1.1 多策略认知诊断研究进展

在 2 级评分方式下,de la Torre 等<sup>[5]</sup>将单策略 DINA 扩展到多策略 DINA(MS-DINA)模型.涂冬波等<sup>[6]</sup>在此基础上结合广义距离判别法<sup>[7]</sup>以及修正的  $Q$  矩阵理论<sup>[8]</sup>,提出了 MSCD 方法.刘铁川<sup>[9]</sup>将混合 IRT 模型与 DINA 模型结合,提出了混合 DINA(Mix-DINA)模型.戴步云等<sup>[10]</sup>针对 CD-CAT 进行多策略研究,将 RRUM(缩减的再参数化统一模型)<sup>[11]</sup>拓展为 MS-RRUM(多策略 RRUM 模型).祝玉芳等<sup>[12]</sup>在文献[6]的基础上开发了一种多策略的多级评分认知诊断方法.

祝玉芳等<sup>[12]</sup>在利用多级评分的广义距离法<sup>[13]</sup>对某学校 750 名学生在“进位计数制”进行认知诊断过程中,通过对学生的访谈发现一些被试在某些

收稿日期: 2018-10-17

基金项目: 国家自然科学基金(31500909, 31360237, 31300876)和教育部人文社会科学研究青年基金(BYJC880060)资助项目.

通信作者: 甘登文(1956-),男,江西奉新人,教授,主要从事智能教学软件和应用统计方面的研究. E-mail: gdw8120429@126.com

题目上使用了不同的解题策略,并在文献[6]的基础上提出了解决“进位计数制”测验中多策略问题的多策略多级评分认知诊断方法.该诊断过程分成2个步骤:(i)测验蓝图的制定,不同策略对应不同测验蓝图;(ii)对于不同测验蓝图,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.因此,先将得分看成0-1评分进行策略诊断,再进行多级评分认知诊断.由于使用不同策略解答同一个题目,可能满分值不同,对在不同测验蓝图中相同题目对应不同得分,需采取某种方式进行转换.

本文针对多级评分多策略认知诊断展开研究,采用基于IRT的认知诊断模型,通过测验观察得分直观、有效、稳定地反映被试的潜在特质以及项目的参数指标.考虑到祝玉芳等<sup>[12]</sup>采用先诊断被试答题使用的策略,再诊断其知识状态的2步诊断法;由于在诊断策略前的分值转换的处理方式可能造成信息的缺损,因此,本文提出了基于IRT的多级评分多策略的模式分类(认知诊断)方法,即若对同一个题目使用不同的策略,则规定的满分值会不同,采用IRT等值的做法,将不同 $Q$ 矩阵和不同知识状态对应的所有期望反应模式与所有观察反应模式放在一起做IRT参数估计,这时项目的参数就在相同量尺上,可以相互比较.

## 1.2 多级评分多策略认知诊断方法

本文在祝玉芳等<sup>[12]</sup>的基础上进一步对多级评分多策略认知诊断方法进行探讨.使用多个不同的 $Q$ 矩阵,每个 $Q$ 矩阵表示一种属性层级结构不同的策略,运用GDD-P(多级评分的广义距离法)<sup>[13]</sup>以及多级评分认知诊断测验蓝图的设计<sup>[14-15]</sup>.

众所周知,评分方式(即评分规则)会对被试的测验得分产生影响.本文的评分规则为:若被试每掌握项目所需的一个属性,则得分加1<sup>[13,16]</sup>.

$d(Y_i, I_t^m)$ 表示被试的测验得分向量与期望反应模式之间的GDD-P,公式如下:

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

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

其中 $J$ 为项目总数, $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ij})$ 为被试 $i$ 的测验得分向量, $I_t^m$ 为策略 $m$ 下第 $t$ 种期望反应, $d(Y_{ij}, I_{j(t)}^m)$ 为在项目 $j$ 上被试 $i$ 的测验得分与策略 $m$ 下第 $t$ 种期望反应之间的GDD-P值, $\theta_i$ 为根据观察作答反应估计出的被试 $i$ 的能力值.比较所有 $d(Y_i,$

$I_t^m)$ 的值,并将使 $d(Y_i, I_t^m)$ 值最小的 $I_t^m$ 对应的理想掌握模式作为被试的知识状态,对应的解题策略即为该被试的解题策略.

## 1.3 多级评分多策略认知诊断过程

具体的诊断过程如下:(i)根据不同策略的认知模型,计算所有可能的策略下被试理想的知识状态(理想掌握模式),进而获得每一种理想的知识状态在测试项目上的理想反应(期望反应);(ii)在所有被试都完成作答得到测验得分后,将每种策略下的期望反应模式与观察反应联合,使用适合的项目反应理论模型进行参数估计;(iii)使用认知诊断模型GDD-P<sup>[13]</sup>对每位被试进行诊断分类,分别计算被试的观察作答反应(测验得分向量)与每种策略下每种期望反应模式之间的GDD-P值,根据GDD-P的最小值确定被试的知识状态和答题策略.

## 2 多级评分多策略认知诊断方法与多级评分单策略认知诊断方法的比较

### 2.1 实验设计

情景1 多级评分单策略,所有被试均采用策略A答题.

情景2 多级评分多策略,一部分使用策略A答题,另一部分使用策略B答题.比较传统单策略认知诊断结果和本文提出方法的诊断结果,检验该方法的性能.

### 2.2 策略数据( $Q$ 矩阵)的模拟

多数研究实际上隐含了一个假设,即不同策略的属性对应相同的层级关系,只要考察它们使用的 $Q$ 矩阵就可以证实,因为 $Q$ 矩阵是属性层级结构的数学表示.也可能存在不同策略对应的属性层级不同的情况.由此,假设测验中存在2种答题策略,各自对应的5个属性的层级结构如图1所示(策略A的 $A_1 \sim A_5$ 与策略B的 $A_1 \sim A_5$ 相同).

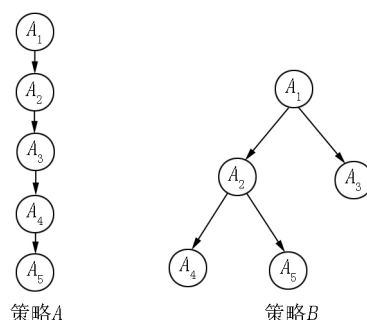


图1 2种策略对应属性间的层级结构图

由此可导出每种策略的可达矩阵  $R$ 、潜在  $Q$  矩阵( $Q_r$  阵)、学生  $Q$  矩阵( $Q_s$  阵)(详细计算方法见文献[8]).对每种策略的  $Q_r$  阵重复若干次得到相应的测验  $Q$  矩阵,测验  $Q$  阵涵盖了各自的  $R$  阵,因此理想掌握模式与期望反应模式一一对应<sup>[17-18]</sup>.测验  $Q$  阵设计如表 1 所示.

表 1 模拟的  $Q$  矩阵

项目	策略											
	A						B					
	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$f_{jA}$	$A_1$	$A_2$	$A_3$	$A_4$	$A_5$	$f_{jB}$
1	1	0	0	0	0	1	1	0	0	0	0	1
2	1	1	0	0	0	2	1	1	0	0	0	2
3	1	1	1	0	0	3	1	0	1	0	0	2
4	1	1	1	1	0	4	1	1	0	1	0	3
5	1	1	1	1	1	5	1	1	0	0	1	3
6	1	0	0	0	0	1	1	1	1	0	0	3
7	1	1	0	0	0	2	1	1	1	1	0	4
8	1	1	1	0	0	3	1	1	1	0	1	4
9	1	1	1	1	0	4	1	1	0	1	1	4
10	1	1	1	1	1	5	1	1	1	1	1	5
11	1	0	0	0	0	1	1	0	0	0	0	1
12	1	1	0	0	0	2	1	1	0	0	0	2
13	1	1	1	0	0	3	1	0	1	0	0	2
14	1	1	1	1	0	4	1	1	0	1	0	3
15	1	1	1	1	1	5	1	1	0	0	1	3
16	1	0	0	0	0	1	1	1	1	0	0	3
17	1	1	0	0	0	2	1	1	1	1	0	4
18	1	1	1	0	0	3	1	1	1	0	1	4
19	1	1	1	1	0	4	1	1	0	1	1	4
20	1	1	1	1	1	5	1	1	1	1	1	5
21	1	0	0	0	0	1	1	0	0	0	0	1
22	1	1	0	0	0	2	1	1	0	0	0	2
23	1	1	1	0	0	3	1	0	1	0	0	2
24	1	1	1	1	0	4	1	1	0	1	0	3
25	1	1	1	1	1	5	1	1	0	0	1	3
26	1	0	0	0	0	1	1	1	1	0	0	3
27	1	1	0	0	0	2	1	1	1	1	0	4
28	1	1	1	0	0	3	1	1	1	0	1	4
29	1	1	1	1	0	4	1	1	0	1	1	4
30	1	1	1	1	1	5	1	1	1	1	1	5

注:  $f_{jA}$  和  $f_{jB}$  分别为策略 A 和策略 B 的项目满分值.

### 2.3 Monte Carlo 模拟过程

根据上节的方法固定属性、策略以及测验  $Q$  阵. 以下是多级评分多策略的模式分类方法具体过程.

(i) 被试知识状态真值模拟. 由每个策略对应的属性层级可知, 策略 A 的理想掌握模式有 6 种, 策略 B 有 11 种. 这 17 种属性模式各分派 100 名被试, 即共 1 700 名被试, 其中 600 人使用策略 A, 1 100 人

使用策略 B. 对于知识状态为全 0 的被试, 无法诊断其采用的答题策略, 因此在实际的诊断测验中, 不对这种类型的被试进行诊断分析.

(ii) 计算每种策略下的期望反应模式. 每种策略下的被试在不发生失误和猜测答题情形下, 根据测验的评分方式得到其期望反应. 例如, 策略 A 下属性掌握模式为 (1 1 1 1 0) 的被试, 在 1 ~ 6 个项目上的期望反应为 (1 2 3 4 4 1); 策略 B 下属性掌握模式为 (1 1 1 0 1) 的被试, 在 1 ~ 6 个项目上的期望反应则为 (1 2 2 2 3 3). 策略 A 下属性的理想掌握模式为 (1 1 1 1 0), 在前 6 个项目上的期望反应为 (4 4 4 3 2 3); 策略 B 下同样的理想掌握模式, 在前 6 个项目上的期望反应则为 (2 2 2 3 1 2).

(iii) 模拟被试的测验得分. 将每种期望反应在给定  $s_{lip}$  (即失误概率) 分别为 2%、5%、10% 和 15% 下波动 100 次得到 (具体做法见文献[19]).

(iv) 估计项目参数与被试能力值. 本文使用 F. Samejima 等级反应模型 (GRM)<sup>[20]</sup> 拟合数据, 其原因在于 GRM 是适用于多级评分并且不包含猜测参数的简单模型. 利用多序列相关系数法<sup>[21]</sup> 估计项目参数和被试能力. 将所有被试的测验得分矩阵  $O_{RP}$  与所有的期望反应模式矩阵  $I_{RP}$  联合估计项目参数, 并采用最大后验估计被试的能力.

(v) 诊断被试的知识状态和答题策略. 计算被试的测验得分向量与每种策略下每种期望反应之间的 GDD-P 值, 使得 GDD-P 值最小的期望反应对应的理想掌握模式和策略即为被试的知识状态和使用的答题策略.

(vi) 为了降低实验偏差, 每种测验均重复实验 30 次.

### 2.4 评价指标

针对诊断结果计算模式判准率 (Pattern Match Ratio, PMR)、属性边际判准率 (Average Attribute Match Ratio, AAMR)、策略判准率 (Strategy Match Ratio, SMR), 计算公式分别为

$$P_{MR} = \left( \sum_{i=1}^N N_i \right) / (NK),$$

$$A_{AMR} = \left( \sum_{i=1}^N \sum_{k=1}^K N_{ik} \right) / (NK),$$

$$S_{MR} = \left( \sum_{i=1}^N N_{i \rightarrow \text{strategy}} \right) / N,$$

其中  $N$  为被试总数,  $K$  为属性总数,  $N_i$  为是否判对第  $i$  个被试的整个知识状态, 对为 1, 错为 0;  $N_{ik}$  为是否判对第  $i$  个被试的第  $k$  个属性, 对为 1, 错为 0;  $N_{i-strategy}$  为是否判对第  $i$  个被试的答题策略, 对为 1, 错为 0.

2.5 实验结果

表 2 为单策略和多策略认知诊断方法在 2 种测验情景下关于 3 个评价指标的比较.

表 2 表明: 作答数据含有多策略反应后, 使用多策略的诊断结果要明显好于单策略的诊断结果, 其结果的模式判准率 ( $P_{MR}$ ) 在 4 种失误差率下的平均值

为 97.40%, 当失误差率为 2% 时, 其模式判断准确率高达 99.78%; 属性边际判准率 ( $A_{AMR}$ ) 在 4 种失误差率下的平均值为 99.35%, 且  $S_{MR}$  的平均值为 93.25%; 其中, 使用策略 A 的被试诊断结果的  $P_{MR}$ 、 $A_{AMR}$  和  $S_{MR}$  在 4 种失误差率下的平均值分别为 94.74%、98.61% 和 99.98%, 使用策略 B 的被试诊断结果的  $P_{MR}$ 、 $A_{AMR}$  和  $S_{MR}$  在 4 种失误差率下的平均值分别为 98.73%、99.72% 和 89.89%. 若采用传统的单策略的诊断方法进行诊断, 则发现  $P_{MR}$  和  $A_{AMR}$  在 4 种失误差率下的平均值分别为 45.28% 和 82.08%.

表 2 单策略和多策略认知诊断方法的判准率比较(实验 30 次的平均值)

情景	$s_{lip}$	单策略 A			多策略			策略 A			策略 B		
		$P_{MR}$	$A_{AMR}$	$S_{MR}$	$P_{MR}$	$A_{AMR}$	$S_{MR}$	$P_{MR}$	$A_{AMR}$	$S_{MR}$	$P_{MR}$	$A_{AMR}$	$S_{MR}$
单策略	2%	0.999 7	0.999 9	—	—	—	—	0.999 7	0.999 9	1.000 0	—	—	—
	5%	0.998 4	0.999 7	—	—	—	—	0.998 4	0.999 7	1.000 0	—	—	—
	10%	0.992 6	0.998 5	—	—	—	—	0.992 6	0.998 5	1.000 0	—	—	—
	15%	0.983 8	0.996 7	—	—	—	—	0.983 4	0.996 6	0.999 7	—	—	—
	平均值	0.993 6	0.998 7	—	—	—	—	0.993 5	0.998 7	0.999 9	—	—	—
多策略	2%	0.464 5	0.834 1	0.997 8	0.999 6	0.933 3	0.993 5	0.998 7	1.000 0	0.999 9	1.000 0	0.900 0	0.899 8
	5%	0.461 9	0.826 8	0.994 4	0.998 9	0.933 2	0.984 5	0.996 9	1.000 0	0.999 4	0.999 9	0.899 9	0.899 8
	10%	0.452 9	0.818 3	0.978 2	0.995 3	0.932 2	0.946 5	0.988 6	0.999 9	0.994 0	0.998 7	0.898 4	0.898 4
	15%	0.431 9	0.804 0	0.925 6	0.980 1	0.931 4	0.865 0	0.960 0	0.999 1	0.956 0	0.990 2	0.897 5	0.897 5
	平均值	0.452 8	0.820 8	0.974 0	0.993 5	0.932 5	0.947 4	0.986 1	0.999 8	0.987 3	0.997 2	0.898 9	0.898 9

3 多级评分多策略认知诊断方法的性能

3.1 实验设计

在同样的多策略实验情形下, 采用祝玉芳等<sup>[12]</sup>的多策略的多级评分认知诊断方法与本文提出的新方法分别进行诊断分析, 并比较 2 种方法的诊断结果. 策略数据及评价指标与第 2 节相同.

3.2 Monte Carlo 模拟过程

本文方法模拟过程已详细介绍, 以下为祝玉芳等<sup>[12]</sup>的方法模拟过程, 其中步骤 (i) ~ (iii) 模拟方法与第 2 节一致.

(iv) 确定被试的策略参数. 将所有被试多级评分的观察作答反应以及每个策略下的所有理想反应模式全部转换成 0-1 评分. 将所有被试 0-1 评分的测验得分与期望反应模式拼接进行参数估计, 利用 GDD 值确定被试的答题策略.

(v) 估计项目参数与被试能力值. 将被诊断出

使用相同策略的被试的测验得分与相应策略的期望反应模式组合估计项目参数和被试的能力参数.

(vi) 诊断被试的属性掌握模式. 分别计算被试的测验得分向量与判定的策略下的每种期望反应模式之间的 GDD-P 值, 诊断被试的知识状态(属性掌握模式).

(vii) 重复实验 30 次.

3.3 实验结果

表 3 为在多级评分多种策略测验中分别采用本文提出的认知诊断方法和祝玉芳等<sup>[12]</sup>的认知诊断方法进行诊断, 并在 3 个评价指标上进行比较.

表 3 表明: 在多策略多级评分测验情形中, 本文提出的诊断方法在各个  $s_{lip}$  下对于被试的模式判准率分别为 99.78%、99.44%、97.82% 和 92.56%, 平均值为 97.40%, 它们均高于祝玉芳等<sup>[12]</sup>方法的模式判准率 93.10%、92.63%、90.48% 和 84.70% 及平均值 90.23%. 实验结果表明祝玉芳等<sup>[12]</sup>的分值转换损失了部分信息.

表3 本文的方法和文献[12]的方法判准率比较(实验30次的平均值)

$s_{ip}$	方法	多策略测验情形								
		$P_{MR}$	$A_{AMR}$	$S_{MR}$	策略A			策略B		
					$P_{MR}$	$A_{AMR}$	$S_{MR}$	$P_{MR}$	$A_{AMR}$	$S_{MR}$
2%	本文	0.997 8	0.999 6	0.933 3	0.993 5	0.998 7	1.000 0	0.999 9	1.000 0	0.900 0
	文献[12]	0.931 0	0.972 9	0.931 3	0.993 5	0.998 7	0.994 5	0.899 7	0.959 9	0.899 7
5%	本文	0.994 4	0.998 9	0.933 2	0.984 5	0.996 9	1.000 0	0.999 4	0.999 9	0.899 8
	文献[12]	0.926 3	0.971 9	0.926 1	0.984 8	0.996 9	0.983 8	0.897 1	0.959 4	0.897 3
10%	本文	0.978 2	0.995 3	0.932 2	0.946 5	0.988 6	0.999 9	0.994 0	0.998 7	0.898 4
	文献[12]	0.904 8	0.966 8	0.907 0	0.966 3	0.992 9	0.962 7	0.874 0	0.953 8	0.879 2
15%	本文	0.925 6	0.980 1	0.931 4	0.865 0	0.960 0	0.999 1	0.956 0	0.990 2	0.897 5
	文献[12]	0.847 0	0.951 1	0.861 0	0.937 8	0.985 3	0.943 9	0.801 6	0.934 0	0.819 5
平均值	本文	0.974 0	0.993 5	0.932 5	0.947 4	0.986 1	0.999 8	0.987 3	0.997 2	0.898 9
	文献[12]	0.902 3	0.965 7	0.906 4	0.970 6	0.993 5	0.971 2	0.868 1	0.951 8	0.873 9

#### 4 结论、讨论与展望

本文在祝玉芳等<sup>[12]</sup>的研究基础上提出了一种新的基于IRT的多级评分多策略认知诊断数据处理方法,利用GDD-P诊断被试的知识状态和答题策略.研究表明在多级评分的测验中答题策略不唯一时:(i)本文方法的诊断结果要明显优于单策略诊断方法;(ii)祝玉芳等<sup>[12]</sup>的分值转换造成了信息的缺损.模拟实验结果表明,新方法能提高分类的准确性.

本文在模拟研究的过程中,通过计算被试的测验得分向量与每种策略下每种期望反应模式之间的GDD-P值来确定被试的知识状态及其答题策略.然而,这与祝玉芳等<sup>[12]</sup>的方法有所不同,他们的研究对于每个被试先判定策略再诊断知识状态,提到由于评分规则不会对答题策略的采用产生影响,故先将被试的测验得分和所有的期望反应均转化成0-1评分,随后用GDD值确定被试在作答时的策略,最后用GDD-P诊断被试的知识状态.

值得注意的是,由于本文的评分方式会导致同一个项目采用不同的答题策略涉及到的属性内容和个数以及满分值有所不同.故项目的满分值取所有被试在该项目上测验得分中最高得分.对此,祝玉芳等<sup>[12]</sup>指出在估计参数之前需要进行不同策略之间的分值转换.由于其研究的项目满分最多为3分,分值转换相对简单,若满分值过大,则这种分值转换方法无法推广,且随机性大.因此,本文在模拟祝玉芳等<sup>[12]</sup>的诊断方法时没有进行分值转换.

在认知测试中,属性层级关系反映被试对属性认知过程的层次性<sup>[22]</sup>.被试在作答每个项目时使用不同策略所需的属性内容、属性个数以及对于这些

属性的认知过程可能有所不同.因此,不同的策略所对应的属性层级也可能有差异.注意到属性层级关系可以使用可达阵(或者必要Q矩阵)表达,不同的策略对应的可达阵可能不同<sup>[23]</sup>.因此本文探索在这种情景下的认知诊断问题,先假设测验中2种不同解题策略分别对应不同的属性层级结构.当然,这样的假设的合理性还有待深入研究.

#### 5 参考文献

- [1] Rupp A A, Mislevy R J. Cognitive foundations of structured item response models [C]. New York: Cambridge University Press, 2007.
- [2] 丁晓, 吕娜, 杨雅琳, 等. 工作记忆成分的年龄相关差异对算术策略运用的预测效应 [J]. 心理学报, 2017, 49(6): 759-770.
- [3] Geary D C, Hoard M K, Byrd-Craven J, et al. Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability [J]. Child Development, 2007, 78(4): 1343-1359.
- [4] Zhou Xinlin, Chen Chuansheng, Dong Qi, et al. Event-related potentials of single-digit addition, subtraction, and multiplication [J]. Neuropsychologia, 2006, 44(12): 2500-2507.
- [5] de la Torre J, Douglas J. Model evaluation and multiple strategies in cognitive diagnosis: an analysis of fraction subtraction data [J]. Psychometrika, 2008, 73(4): 595-624.
- [6] 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.
- [7] 孙佳楠, 张淑梅, 辛涛, 等. 基于Q矩阵和广义距离的认知诊断方法 [J]. 心理学报, 2011, 43(9): 1095-1102.

- [8] 丁树良, 祝玉芳, 林海菁, 等. Tatsuo-ka  $Q$  矩阵理论的修正 [J]. 心理学报, 2009, 41(2): 175-181.
- [9] 刘铁川. Mix-DINA 模型功能开发及其与 DINA、MS-DINA 模型的模拟与实证比较 [D]. 南昌: 江西师范大学, 2012.
- [10] 戴步云, 张敏强, 焦璨, 等. 基于 CD-CAT 的多策略 RRUM 模型及其选题方法开发 [J]. 心理学报, 2015, 47(12): 1511-1519.
- [11] Hartz S M. A bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality [D]. Urbana: University of Illinois at Urbana-Champaign, 2002.
- [12] 祝玉芳, 王黎华, 丁树良, 等. 多策略的多级评分认知诊断方法的开发 [J]. 江西师范大学学报: 自然科学版, 2015, 39(4): 371-376.
- [13] Sun Jianan, Xin Tao, Zhang Shumei, et al. A polytomous extension of the generalized distance discriminating method [J]. Applied Psychological Measurement, 2013, 37(7): 503-521.
- [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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] Leighton J P, Gierl M J, Hunka S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuo-ka's rule-space approach [J]. Journal of Educational Measurement, 2004, 41(3): 205-237.
- [20] Samejima F. Estimation of latent ability using a response pattern of graded scores [J]. Psychometrika, 1970, 35(1): 139.
- [21] 陈青, 丁树良, 朱隆尹, 等. 3 参数等级反应模型及其参数估计 [J]. 江西师范大学学报: 自然科学版, 2010, 34(2): 117-122.
- [22] Leighton J P, Gierl M J. Cognitive diagnostic assessment for education: theory and applications [J]. Journal of Qingdao Technical College, 2007, 45(4): 407-411.
- [23] 丁树良, 罗芬, 汪文义, 等.  $Q$  矩阵探微 [J]. 江西师范大学学报: 哲学社会科学版, 2017, 50(1): 71-79.

## The Multiple-Strategy Cognitive Diagnosis Method with Polytomous Scoring

HUANG Yu, LUO Fen, XIONG Jianhua, DING Shuliang, GAN Dengwen\*

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

**Abstract:** Cognitive diagnosis (CD) is the product of the combination of modern psychometrics and cognitive psychology. Polytomous scoring items can provide more comprehensive information and better reflect the KS of the examinees. Taking into account the existence of different cognitive strategies corresponding to different attribute hierarchies, even under some scoring rules, there is a complex situation in which the same item uses different problem-solving strategies maybe corresponding to different values, borrowing the equivalent approach of project response theory to deal with this complex situation. Monte Carlo simulation experiments show that in the polytomous scoring test when the response data of examinees contained multiple-strategy responses: The diagnosis results using multiple-strategy CD method pattern match ration (PMR) better than the single-strategy CD method. The method, the PMR in various slip all higher than other methods, their method the score conversion caused the lossing of information.

**Key words:** cognitive diagnosis; multiple-strategy cognitive diagnosis; polytomous;  $Q$  matrix

(责任编辑: 冉小晓)