

李俊杰,郑慧婧,牛忆哲,等.融入选项信息的多项选择题认知诊断测验分析方法[J].江西师范大学学报(自然科学版),2023,47(4):368-376.

LI Junjie, ZHENG Huijing, NIU Yizhe, et al. The analytical method of multiple choice and multiple answer cognitive diagnostic tests with option information [J]. Journal of Jiangxi Normal University (Natural Science) 2023, 47(4) : 368-376.

文章编号:1000-5862(2023)04-0368-09

融入选项信息的多项选择题认知诊断测验分析方法

李俊杰^{1,2}, 郑慧婧¹, 牛忆哲¹, 康春花¹, 曾平飞^{1*}

(1. 浙江师范大学基础教育质量监测中心, 浙江 金华 321004);

2. 北京师范大学中国基础教育质量监测协同创新中心, 北京 100875)

摘要: 多项选择题作为认知诊断测验中一种重要题型,具有诸多优点的同时,也具有多个正确答案的特性。目前大多数认知诊断模型在处理多项选择题反应数据时采用二级评分或多级计分方法来估计考生的知识状态,但这2种计分方法的计分仍停留在题目层面上,并未深入到具体的选项层面,不能充分利用多项选择题中的选项信息。因此,该文开发了选项向量计分和空间向量距离判别法对多项选择题认知诊断测验分析方法,并通过模拟研究和实证研究来验证选项向量计分和空间向量距离判别法分析多项选择题认知诊断测验的性能。

关键词: 多项选择题; 认知诊断测验; 计分方法; 判别方法

中图分类号: B 841 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2023.04.06

0 引言

认知诊断评估也称认知诊断,是对个体认知过程、加工技能或知识结构的诊断评估^[1-3],教师可以根据诊断评估结果给予学生有效的指导以实现个性化教学。多项选择题作为在认知诊断测验中一种重要题型,具有一系列优点^[4],逐渐受到测验编制者和使用者的青睐。除具有一般选择题具有的知识容量大、答案简明、评分客观的功能外,多项选择题还具有不同于其他类型选择题的特点:1)它可以测量多个层次的知识 and 能力,即识记、理解、分析、综合等;2)它可以控制猜测行为,与单选题相比较,多项选择题答案不具有唯一性,正确选项的多样性和由此产生的选择方式的多样性,使得考生靠猜测得分的可能性较小,考生只有具备一定的分析、比较、辨别、理解的综合能力才能保证选择的准确性,以提高解题质量和能力;3)易于实现计算机化测验。目前也有研究者提出了一些认知诊断模型和判别

方法用于处理选择题,如 MC-DINA 模型^[5]、NR-LCDM 模型^[6]、结构化 MC-DINA 模型 1 和结构化 MC-DINA 模型 2^[7]、 $d_{h,MC}$ 距离法^[8],但是,这些模型和判别方法均只能处理正确选项只有 1 个的选择题。而在中国各类考试(如高考的数学、物理科目,研究生入学考试中的政治科目,心理咨询师等级考试)中的多项选择题测试形式是多个正确选项的,因此上述认知诊断模型和判别方法不适用于这类多项选择题。本文所指的多项选择题是指选项多、正确选项也多的题型,具有中国考试特色的真正意义上的多项选择题,它突出的难度就是正确选项“不止 1 个”。因此,在编制多项选择题认知诊断测验时,不仅可以对题目进行编码,还可以对选项进行编码。在这种编码条件下,不仅题目可以提供诊断信息,选项中包含的信息还可用于诊断分类。若可以充分利用多项选择题的规则,使用合适的计分方法记录选项的诊断信息,采用合适的判别方法提取更多的测验信息,则可以提高诊断的准确率,可更加细致地诊断学生的知识掌握情况,为学生提供

收稿日期:2023-03-09

基金项目:教育部人文社会科学课题(22YJA190005)资助项目。

通信作者:曾平飞(1963—),男,广西荔浦人,教授,博士,主要从事心理测量与评价方面的研究。E-mail:zpf@zjnu.edu.cn

全面且细致的反馈.

多项选择题有别于其他题型,其主要特点就是多项选择题具有多个正确答案.测验项目的计分方式是影响测验信息利用率的重要因素之一,采用不同的计分方式,同一测验得到的诊断信息量不同,诊断结果也就不同.目前,大多数认知诊断模型在处理多项选择题反应数据时采用二级计分或多级计分方法来估计考生的知识状态,上述计分标准阻碍了其自身更加充分使用选项信息^[9].在二级计分中,考生的答案只有与正确选项完全一致才能得分.而在多级计分的条件下,少选可得部分分数.较二级计分来说,这虽然能够进一步区分考生的知识状态,但上述2种计分方法的计分仍停留在题目层面上,并未深入到具体的选项层面.具体表现是:在考生得部分分数的情况下,考生可能存在不同选项组合,未充分利用选项的诊断信息,同时少选正确选项的学生均会被赋予相同的分数.

以上2种计分方法均停留在题目层面上,对测验信息的使用效率较低,忽略了在选项中包含的大量诊断信息^[10-12].这不仅是对测验编制的极大浪费,还会降低估计精度.因此,若要提高多项选择题更加高效的分析认知诊断功能,则主要有2个方面的技术问题:1)如何选用合适的计分方法,记录更多的测验信息;2)如何采用更加合理的诊断方法以更加充分使用选项信息,提高估计考生的知识状态准确性.这些都是值得深究的问题.

1 多项选择题的认知诊断分析方法

本文结合认知诊断测验的特征提出了一种基于选项信息,利用不同学生在选项上的反应,分析其背后代表的属性掌握模式的多项选择题认知诊断分析方法.在计分方法方面,提出了选项向量计分(vector option scoring method, VOSM)方法.在判别方法方面,结合了距离判别方法和认知诊断多项选择题的特点,提出了空间向量距离判别法(space vector distance discriminant, SVDD).

1.1 选项向量计分

问题1 把8平均分成3份,每份是多少?列式为().

A. $8 \div 3$ B. $8 \times 1/3$ C. 8×3 D. $8 \div 1/3$

选项向量计分以学生在选项上的反应进行计分,从以往将学生的作答反应处理为得分的思路中跳出来,保留学生的原始信息,不必局限于少选的

情况.若学生选择一个选项则记为1,不选为0.从表1可以看出:这种按选项计分的方法能直接清晰地看到个体在每个项目上的作答,得到不同作答反应背后的知识状态,提供更多的诊断信息.

多项选择题包含题干部分、正确选项和错误选项部分.在编制多项选择题认知诊断测验时,不仅可以对题目所考察的指数属性进行编码,还可以对正确选项所涉及的知识属性进行编码.如表1中呈现的4选2编码例子,该题目考察了4个属性(A_1 问题表征、 A_2 整数的认识与意义、 A_3 整数的运算规则、 A_4 分数与小数的认识与意义).考生全部掌握4个属性才能选择正确选项AB,若只掌握了 A_1 、 A_2 、 A_3 属性,则会选择选项A,若没有掌握正确选项的属性,则会选择错误答案组合中的一种.

表1 例1多项选择题4选2编码例子

选项	A_1	A_2	A_3	A_4
A. $8 \div 3$	1	1	1	0
B. $8 \times 1/3$	1	1	1	0
C. 8×3				
D. $8 \div 1/3$				

在二级计分中,考生的答案只有与正确选项完全一致,这道题目才算做对,这种采取“全或无”的数据提取方式忽视了未掌握一个属性的考生与未掌握多个属性的考生之间的差异.如考生单独选择了A或B和选择其他错误答案组合的考生同样得0分,这显然无法区分选择A或B的考生和选择其他错误组合的考生.

在多级计分中,考生的答案与正确选项完全一致,则记为满分,少答记一半的分数,错答和多答不得分.如问题1正确选项为AB,考生只有选择AB才能得满分,选择A或B得一半分,其他情况不得分.这种对少选计一半分的方法比二级计分更科学,能更进一步区分不同考生.不过同是问题1少选的情况,在无法区分学生选择的是A还是B时,其背后可能是怎样的知识掌握状态呢?

从表2可以发现:选项向量计分是采用向量的形式对考生的题目选项的情况进行记录,一方面,采用选项向量计分可以通过考生选项向量对不同考生进行区分,另一方面可以明确知道在多级评分中得部分分的考生究竟选择的是A还是B,保留了考生在选项层面的信息.同时可以发现:无论是二级计分还是多级计分,只要采用一定的分数转化规则,选项向量计分就均可以转化成二级计分或者多

级计分,而二级计分或多级计分却没有此功能.可以说,二级计分或多级计分是向量计分经过处理后的另一种分数表现形式,但同时也丢失了选项层面的信息.

表 2 3 种计分方法的实例

题目	选项	二级计分	多级计分	选项向量计分
问题 1	AB	4	4	(1 1 0 0)
	A	0	2	(1 0 0 0)
	B	0	2	(0 1 0 0)

1.2 空间向量距离判别法

认知诊断判别方法根据是否需要参数估计,可将认知诊断方法分为参数方法和非参数方法^[13].现在有许多研究者提出了非参数诊断方法用于考生知识状态的判别^[14-17].非参数方法的具有操作方便、受样本量的影响较小、前提假设弱、条件假设少和不需要进行复杂的参数估计等优点,这极大地节省了时间和精力,十分适用于课堂评估^[18].

研究结合具有认知诊断功能多项选择题的特点,同时考虑非参数距离判别法的优点和课堂实践的需求,进一步在原有欧氏距离判别法的基础上拓展出适合多项选择题选项向量计分的判别方法.欧氏距离判别法在使用过程中,理想反应和观察反应采用的均是分数.在认知诊断多项选择题中,若计分采用选项向量的方式,为了能够使距离判别法适用于选项向量计分,多项选择题的距离判别法可将理想反应和观察反应同样采用向量.具体来说,首先计算出考生作答的选项向量和所有属性掌握模式理想作答反应选项向量之间的空间向量距离;随后选取距离最小的理想反应模式所代表的知识状态(KS)作为考生的真实知识状态.值得注意的是,在多项选择题的距离判别法计算中,若考生的观察选项向量不是正确选项组合的子集,则作答选项向量替换成元素为 0 的向量.

$$d(x_i, \eta_m) = \sqrt{\sum_{j=1}^J (x_{ij} - \eta_{mj})^2}. \quad (1)$$

1) m 为所有理想反应模式, J 为所有的项目, x_i 为考生 i 的观察反应模式的选项向量, η_m 为第 m 个的理想反应模式的选项向量, $d(x_i, \eta_m)$ 为所有的项目上关于 x_i 与 η_m 的空间向量距离.

2) 选取空间向量距离最小的理想作答选项向量所代表的知识状态(KS)作为考生 i 的真实知识状态,即

$$C = \arg \min_{m \in \{1, 2, \dots, M\}} d(x_i, \eta_m). \quad (2)$$

3) 某一作答反应模式在存在多个且一致最小距离的理想作答选项向量时采用随机判别的方法,随机选择一个理想作答选项向量.

认知诊断测验将可观察的作答选项向量对应到不可观察的知识状态,并找到与知识状态一一对应的理想选项向量测验理论.空间向量距离判别法理想采用如下方式计算理想作答选项向量. η_{mj} 是第 m 个的属性掌握模式的理想选项向量:

$$\eta_{mj} = (\eta_{mj1}, \eta_{mj2}, \dots, \eta_{mjo}) \quad (3)$$

其中 o 为选项个数,在理想情况下,当选项 q_o 为正确选项,且属性掌握模式 m 考生具备该选项考察的所有属性时,考生会选择 q_o ,即 $\eta_{mjo} = 1$;反之,当 q_o 不是正确选项或者属性掌握模式 m 考生不具备该选项考察的所有属性时,属性掌握模式 m 考生不选择 q_o ,即 $\eta_{mjo} = 0$. η_{mjo} 计算公式为

$$\eta_{mjo} = \prod_{k=1}^K \alpha_k^{q_{jok}}, \quad (4)$$

其中 η_{mjo} 为在项目 j 正确选项 o 上的理想反应, q_{jok} 为正确选项 o 的 Q 向量的元素, q_o 为正确选项 o 所涉及的知识属性 Q 向量,即若选项 o 测量了属性 k 表示为 1,未测量表示为 0,即

$$q_{ok} = \begin{cases} 0, & \text{选项 } o \text{ 未测量属性 } k, \\ 1, & \text{选项 } o \text{ 测量了属性 } k. \end{cases} \quad (5)$$

1.3 多项选择题的认知诊断分析过程

多项选择题的认知诊断分析过程具体如下:

1) 确定测验 Q 矩阵,正确选项的 Q 向量; 2) 属性层级结构获得所有的知识状态; 3) 采用选项向量计分记录考生的作答反应; 4) 依据知识状态和选项 Q 向量,计算所有知识状态的理想选项向量; 5) 依据考生观察选项向量和理想选项向量,通过空间向量距离判别法估计考生的属性掌握模式.具体流程如图 1 所示.

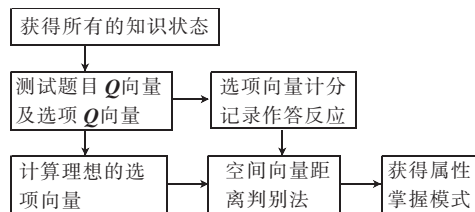


图 1 多项选择题的认知诊断分析流程图

2 模拟研究

模拟认知诊断多项选择题分析过程,探究在不同因素条件下选项向量计分的效果及其稳健性和空间向量距离判别法的性能.

2.1 研究设计

本研究在固定属性层级结构为独立型的情况下, 研究采用 3 种计分方法(二级、多级、选项向量计分) \times 2 种属性个数($K=4, 6$) \times 3 种知识状态分布(unif、norm-0. 5、norm-0. 8) \times 2 种测验长度($L=10, 15$) \times 4 种滑动概率(0. 05、0. 10、0. 15、0. 20) = 144 种条件下进行模拟研究. 其中计分方法、测验长度、滑动概率为组内变量, 属性个数和知识状态为组件变量. 结果如表 3 和表 4 所示.

15) \times 4 种滑动概率(0. 05、0. 10、0. 15、0. 20) = 144 种条件下进行模拟研究. 其中计分方法、测验长度、滑动概率为组内变量, 属性个数和知识状态为组件变量. 结果如表 3 和表 4 所示.

表 3 不同条件下的属性分类准确性($K=4$)

滑动概率	测验长度	知识状态分别	二级计分	多级计分	选项向量计分
0. 05	10	unif	0. 713	0. 945	0. 989
		norm-0. 5	0. 792	0. 934	0. 993
		norm-0. 8	0. 775	0. 925	0. 995
	15	unif	0. 721	0. 902	0. 999
		norm-0. 5	0. 797	0. 917	0. 999
		norm-0. 8	0. 766	0. 886	1. 000
0. 10	10	unif	0. 686	0. 879	0. 964
		norm-0. 5	0. 778	0. 868	0. 980
		norm-0. 8	0. 735	0. 867	0. 983
	15	unif	0. 682	0. 915	0. 994
		norm-0. 5	0. 800	0. 906	0. 997
		norm-0. 8	0. 766	0. 891	0. 997
0. 15	10	unif	0. 643	0. 806	0. 966
		norm-0. 5	0. 731	0. 777	0. 966
		norm-0. 8	0. 699	0. 764	0. 974
	15	unif	0. 682	0. 876	0. 980
		norm-0. 5	0. 777	0. 832	0. 983
		norm-0. 8	0. 746	0. 822	0. 987
0. 20	10	unif	0. 617	0. 766	0. 928
		norm-0. 5	0. 665	0. 688	0. 926
		norm-0. 8	0. 675	0. 681	0. 943
	15	unif	0. 651	0. 802	0. 945
		norm-0. 5	0. 729	0. 754	0. 946
		norm-0. 8	0. 698	0. 767	0. 961

采用 Monte Carlo 模拟方法生成数据, 过程如下:

1) 在模拟考生方面, 研究中模拟考生 2 000 人, 其中 unif 表示考生的知识状态由均匀分布生成; norm-0. 5 考生的知识状态由多元正态分布且协方差为 0. 5 生成; norm-0. 8 考生的知识状态由多元正态分布且协方差为 0. 8 生成.

2) 在测验 Q 矩阵方面, 测验 Q 矩阵的生成采用蔡艳等^[19]的做法. 在保证测验 Q 矩阵至少包含 1 个 R 矩阵的情况下, 使 Q 矩阵中其他元素是否考核的概率为 0. 5, 对不符合属性层级的考核模式重新模

拟直至符合属性层级; 在研究中多项选择题的正确选项 Q 向量为题目 Q 向量的子集, 每个题目包含 2~4 个正确选项.

3) 在作答反应生成方面, 首先分别计算 3 种计分方法的理想作答反应, 然后进行滑动概率生成 3 种计分方法的作答反应. 其具体滑动方式如下. 首先, 对考生在每个项目上的作答概率生成一个服从(0, 1) 均匀分布的随机数, 然后根据设定的滑动概率和随机数作比较, 若随机数小于滑动概率, 则从错误选项的随机组合中挑选一个选项作为考生的观察反应; 若随机数大于滑动概率, 则将考生的理

想反应作为观察反应.

4) 在判别方法方面 ,对二级计分、多级计分方法得到的数据使用距离判别法分析 ,选项计分方法得到的数据使用空间向量距离判别法分析.

为减少随机误差 ,每种实验条件模拟 30 批数据.

2.2 评价指标

评价指标分别为模式判准确率 (pattern correct

claooification rate ,PCCR) ,计算公式为

$$P_{CCR} = \sum_{i=1}^N n_i / N. \tag{6}$$

P_{CCR} 考查考生属性掌握模式的返真性. 假设测验共考查了 K 个属性 ,有 N 个考生参加了测验 ,考生 i 真实的属性掌握向量元素记为 X_i ,但把该考生归类为 Z_i ,若有 $X_i = Z_i$,则记 $n_i = 1$;否则记 $n_i = 0$. P_{CCR} 的值处于 0~1 之间 ,值越高越好.

表 4 不同条件下的属性分类准确性($K=6$)

滑动概率	测验长度	知识状态分别	二级计分	多级计分	选项向量计分
0.05	10	unif	0.638	0.775	0.824
		norm-0.5	0.701	0.781	0.872
		norm-0.8	0.650	0.745	0.847
	15	unif	0.847	0.920	0.990
		norm-0.5	0.843	0.910	0.990
		norm-0.8	0.779	0.892	0.995
0.10	10	unif	0.628	0.792	0.928
		norm-0.5	0.679	0.707	0.928
		norm-0.8	0.666	0.704	0.935
	15	unif	0.784	0.833	0.964
		norm-0.5	0.801	0.813	0.968
		norm-0.8	0.752	0.817	0.977
0.15	10	unif	0.536	0.668	0.830
		norm-0.5	0.597	0.574	0.859
		norm-0.8	0.562	0.550	0.871
	15	unif	0.696	0.731	0.917
		norm-0.5	0.730	0.686	0.924
		norm-0.8	0.681	0.686	0.929
0.20	10	unif	0.466	0.571	0.729
		norm-0.5	0.482	0.426	0.741
		norm-0.8	0.433	0.388	0.723
	15	unif	0.630	0.635	0.892
		norm-0.5	0.626	0.536	0.902
		norm-0.8	0.578	0.545	0.915

2.3 研究结果

从不同滑动概率来看 ,随着题目滑动概率的增大 ,3 种计分方法的模式判准确率均在下降. 在 4 种题目滑动概率的条件下 ,选项向量计分和空间向量距离判别法受题目滑动概率影响相对较小 ,而二级计分和多级计分受题目滑动概率的影响较大. 同时 ,相比于其他方法 ,选项向量计分和空间向量距离判

别法在不同滑动概率条件下均具有更高的模式判准确率 ,特别是在滑动概率较大的条件下 ,选项向量计分和空间向量距离判别法的模式判准确率的优势更加明显.

从不同的测验长度来看 ,当测验长度 $10 \leq L \leq 15$ 时 ,3 种计分方法在不同的测验长度中的判准确率有所差异 ,同时均表现出随着题目数量增多 ,其模式判准确率变大. 但相比于其他 2 种方法 ,选项向量计

分和空间向量距离判别法在不同测验长度条件下均具有更高的模式判准率。

从不同的知识状态分布来看,选项向量计分和空间向量距离判别法分析的多项选择题的模式判准率的差异较小。通过比较在不同分布状态下判准率的均值也发现,均匀分布和多元正态分布的判准率均值近似相等。这说明选项向量计分和空间向量距离判别法在进行多项选择题分析时受知识状态分布的影响很小,且具有很高的属性分类准确性。

从不同的属性个数来看,随着属性个数的增加,3种计分方法的模式判准率均在下降。具体来说,随着 $K=4$ 增加到 $K=6$,二级计分、多级计分和选项向量计分的模式判准率均有较大幅度的下降。值得注意的是,在不同属性个数条件下,采用选项向量计分和空间向量距离判别法对多项选择题进行分析都具有更高的模式判准率。这可能是因为选项向量计分和空间向量距离判别法更加充分地使用选项信息,每道题目得到的可用于诊断的信息更多,因此能够更准确地估计考生属性掌握模式。

3 实证研究

通过实证数据验证选项向量计分和空间向量距离判别法在实际测验中分析多项选择题认知诊断的有效性。

3.1 研究设计

测验内容来源于张淑君^[20]开发的小学数学与代数认知诊断测验题库。本研究选取题库中的题目改编为认知诊断多项选择题测验。

测验对象为浙江省金华市某小学六年级上学期的学生,一共3个班,120人,发放认知诊断测验试卷120份,回收试卷120份,剔除无效试卷2份,最终获得有效数据118份。所考察属性如表5所示。

表5 测验考查的认知属性

认知属性	内容
A_1	问题表征-数学情景、关系和条件
A_2	整数的认识与意义
A_3	分数和小数的认识与意义
A_4	整数、小数的四则混合运算法则
A_5	量之间的单位换算
A_6	找数量关系和用字母表示数

本研究分别使用选项向量计分、二级计分、多

级计分3种方法对实证数据进行计分,随后采用距离判别法估计属性掌握模式,比较不同计分方法在多项选择题的评分估计中的表现。

3.2 评价指标

属性通过率(attribute mastery rate, AMR)是计算在不同类别下每个属性的通过概率,即优秀类别学生的属性通过率较高,不足类别学生的属性通过率较低,计算公式为

$$A_{MR} = N_{mastery} / N, \quad (7)$$

其中 $N_{mastery}$ 为掌握某个属性的人数, N 为考生总人数。

3.3 研究结果

借助六年级学生上学期期中数学科目的考试成绩,取总分排名前27%及后27%的作为高分组和低分组,其余学生为中等组,计算不同分类下的属性通过率以及诊断结果与学生真实情况的一致性。

通常一种分析方法的效度越好越能区分不同类型的学生,其平均属性通过率相差越大。根据二级计分、多级计分和选项向量计分3种计分方法的结果计算不同类型学生的平均属性通过率,具体结果如表6所示。由表6可以看到,不同类型学生的属性通过率确实存在着差异,高分组的学生属性通过率最高,其次是中等组的学生,低分组的学生属性通过率最低。

在高分组的学生中,选项向量计分和空间向量距离判别法的平均属性通过率为0.934,二级计分、多级计分的平均属性通过率分别为0.889、0.980;在中等组的学生中,3种方法的平均属性通过率分别为0.840、0.934、0.862;在低分组的学生中,3种方法的平均属性通过率分别为0.742、0.871、0.737。首先,比较3种方法在高分组与中等组学生中的属性通过率,选项向量计分和空间向量距离判别法分析这2种类别的平均属性通过率的差距为0.072,而二级计分和多级计分的差距分别为0.049、0.046。选项向量计分和空间向量距离判别法区分不同类型的学生的能力最强,其次是二级计分,多级计分较差;选项向量计分和空间向量距离判别法判别中等组学生和低分组学生的平均属性通过率的差距为0.125,而二级计分和多级计分分

别为 0.098、0.063,选项向量计分和空间向量距离判别法区分中等学生与低分学生的能力同样较其他 2 种方法强。故可知,在区分 3 种类型学生的能力上,选项向量计分和空间向量距离判别法均优于二级计分和多级计分法。

表 6 不同类型学生的 AMR

学生类型	属性	二级计分	多级计分	选项向量计分
高分组	A_1	0.970	1.000	1.000
	A_2	0.818	0.909	0.697
	A_3	0.879	1.000	0.909
	A_4	1.000	1.000	1.000
	A_5	0.758	0.970	1.000
	A_6	0.909	1.000	1.000
	均值	0.889	0.980	0.934
中等组	A_1	0.792	0.981	0.925
	A_2	0.774	0.792	0.717
	A_3	0.981	1.000	0.811
	A_4	0.906	0.962	0.906
	A_5	0.736	0.887	0.906
	A_6	0.849	0.981	0.906
	均值	0.840	0.934	0.862
低分组	A_1	0.742	0.968	0.903
	A_2	0.677	0.645	0.645
	A_3	0.871	0.935	0.806
	A_4	0.871	0.968	0.677
	A_5	0.516	0.710	0.645
	A_6	0.774	1.000	0.742
	均值	0.742	0.871	0.737

为了进一步验证选项向量计分和空间向量距离判别法分析的结果是否符合学生的实际情况,分别比较 3 种方法估计学生的属性个数与六年级学生期中考试成绩相关系数,具体计算结果如表 7 所示。

表 7 属性个数与成绩的一致性

方法	属性个数与成绩的一致性
二级计分	0.304
多级计分	0.337
选项向量计分	0.414

由表 7 可知:在这 3 种计分方法中,选项向量计分法和空间向量距离判别法估计学生掌握的属性

个数与六年级学生期中成绩相关系数最高,多级计分次之,二级计分相对最低。这说明采用选项向量计分和空间向量距离判别法估计的结果更加符合学生的实际情况。

4 讨论与结论

4.1 讨论

已有的多项选择题认知诊断分析方法仅仅是在题目分数层面上进行评分,这虽然简便易行,但是容易丢失信息,导致不能准确地估计考生的知识状态^[11]。选项向量计分和空间向量距离判别法的分析方法充分考虑了学生在每个选项上的反应是与该选项所考查的属性相关联的,通过选项考查的属性,得到更多的诊断信息。相对其他的方法,选项向量计分和空间向量距离判别法可利用的信息量更大。从模拟研究的结果来看,选项向量计分和空间向量距离判别法在不同条件下均表现出准确的计分。

选项向量计分和空间向量距离判别法与其他方法一样呈现随属性个数的增多模式判准率会下降^[14-21]的趋势。但相较于其他方法,在不同条件下,选项向量计分和空间向量距离判别法均具有更高的属性分类准确性。

3 种方法随着测验长度的增加,模式判准率呈现不断上升的趋势,这与已有研究的结论相同^[22]。但二级计分和多级计分法表现较不稳定,这可能是由于二级计分和多级计分法在考生少选的条件下计分存在缺陷,笼统地将不同的作答反应记为同样的分数,区分不同属性掌握模式的能力较弱。相较而言,选项向量计分则是详尽记录了考生各种选项信息,因此选项向量计分和空间向量距离判别法在不同测验长度上均具有最好的模式判准率。

已有研究发现题目的滑动概率(题目滑动概率越小题目质量越高,题目滑动概率越大题目质量越差)对诊断分类结果影响最大,题目的滑动概率越小,则其分类准确率也越高,且 2 者间的关系不受属性个数、属性层级关系、测验长度和样本量的影响^[23]。这一点不仅在参数认知诊断方法中适用,在非参数诊断方法中同样得到了证明。不同滑动概率

的题目对选项向量计分和空间向量距离判别法分析结果存在不同程度的影响。当滑动概率小时,选项向量计分和空间向量距离判别法可以根据考生的作答较好地区分不同属性掌握模式的考生,即使独立型属性层级结构拥有 2^K 种属性掌握模式,存在较多的差异较小的属性掌握模式,选项向量计分依然能够区分;当滑动概率较大时,选项向量计分和空间向量距离判别法很难区分差异较小的理想属性掌握模式,可能会发生错判。但相比之下,选项向量计分和空间向量距离判别法的属性掌握模式判断仍显著高于同等条件下的二级计分和多级计分方法。

通过模拟研究结果可知:选项向量计分和空间向量距离判别法较其他2种方法表现更好,且在不同条件下的性能稳定性也更好。实证研究通过比较二级计分、多级计分和选项向量计分方法的属性通过率进一步表明:3种方法均能较为有效地利用属性通过率区分不同类型的学生,即高分组学生属性通过率最高,中等组学生次之,低分组学生属性通过率较低;同时也发现选项向量计分和空间向量距离判别法能更有效地区分不同类型的学生。并且,选项向量计分和空间向量距离判别法得到的数据与学生期中考试成绩的一致性最高,判别结果与学生的真实情况更为接近。以上结果均说明:在获得选项 Q 向量,利用学生在选项上的作答反应之后,有效地提高了计分方法的计分效果,提高了对不同类型学生的区分能力。

4.2 结论

本研究提出用于更贴合多项选择题认知诊断测验实际情况的选项向量计分和空间向量距离判别法分析法,并通过模拟研究和实证研究对此种分析方法的性能和适用性进行验证和比较,得到以下结论:

1) 选项向量计分和空间向量距离判别法充分考虑了学生在每个选项上的反应是与该选项所考查的属性相关联的,通过选项考查的属性,得到更多的诊断信息;

2) 选项向量计分和空间向量距离判别法在不同条件下均表现出最高的模式判准率;

3) 选项向量计分和空间向量距离判别法在实证研究中同样表现良好,得到的结果与学生的真实情况更为接近。

5 参考文献

- [1] LEIGHTON J P, GIERL M J, HUNKA S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuoaka's rule-space approach [J]. Journal of Educational Measurement, 2004, 41: 41.
- [2] YANG Xiaodong, EMBRETSON S E. Construct validity and cognitive diagnostic assessment [M]. Cambridge: Cambridge University Press, 2007.
- [3] LEIGHTON J P, GIERL M J. Cognitive diagnostic assessment for education: theory and applications [M]. Cambridge: Cambridge University Press, 2007.
- [4] ROEDIGER H L, MARSH E J. The positive and negative consequences of multiple-choice testing [J]. Journal of Experimental Psychology: Learning, Memory, and Cognition, 2005, 31(5): 1155-1159.
- [5] JIMMY DE LA TORRE. A cognitive diagnosis model for cognitively based multiple-choice options [J]. Applied Psychological Measurement, 2009, 33(3): 163-183.
- [6] TEMPLIN J L, HENSON R A. Cognitive diagnosis models for nominal response data [EB/OL]. [2022-06-17]. https://www.researchgate.net/profile/Robert-Henson/publication/228894528_Cognitive_diagnosis_models_for_nominal_response_data/links/0a85e5332fadc2ef60000000/Cognitive-diagnosis-models-for-nominal-response-data.pdf.
- [7] OZAKI K. DINA models for multiple-choice items with few parameters: considering incorrect answers [J]. Applied Psychological Measurement, 2015, 39(6): 431-447.
- [8] 郭磊,周文杰.基于选项层面的认知诊断非参数方法[J].心理学报,2021,53(9):1032-1043.
- [9] 张晓涛,马秀谊.选择题的测试功能及其适用性分析[J].中国考试(研究版),2008(2):38-43.
- [10] Thissen D, Steinberg L, Fitzpatrick A R. Multiple-choice models: the distractors are also part of the OVS [J]. Journal of Educational Measurement, 1989, 26(2): 161-176.
- [11] 李瑜.多项选择题认知诊断测验编制及多策略的多项选择题认知诊断模型的开发[D].南昌:江西师范大学,2014.
- [12] 刘拓.干扰项信息在计算机化自适应测验中的利用

- [D]. 北京: 北京师范大学, 2016.
- [13] 高旭亮, 涂冬波. 参数化认知诊断模型: 心理计量特征、比较及其转换 [J]. 江西师范大学学报(哲学社会科学版), 2017, 50(1): 88-104.
- [14] CHIU C Y, DOUGLAS J. A Nonparametric approach to cognitive diagnosis by proximity to ideal response patterns [J]. Journal of Classification, 2013, 30(2): 225-250.
- [15] 康春花, 任平, 曾平飞. 多级评分聚类诊断法的影响因素 [J]. 心理学报, 2016, 48(7): 891-902.
- [16] 罗欢, 丁树良, 汪文义, 等. 属性不等权重的多级评分属性层级方法 [J]. 心理学报, 2010, 42(4): 528-538.
- [17] 康春花, 杨亚坤, 曾平飞. 一种混合计分的非参数认知诊断方法: 曼哈顿距离判别法 [J]. 心理科学, 2019, 42(2): 455-462.
- [18] 胡竹菁. 心理统计学 [M]. 北京: 高等教育出版社, 2010.
- [19] 蔡艳, 涂冬波, 丁树良. 五大认知诊断模型的诊断正确率比较及其影响因素: 基于分布形态、属性数及样本容量的比较 [J]. 心理学报, 2013, 45(11): 1295-1304.
- [20] 张淑君. 基于 MDD 的 npCD-CAT 研究及其在数与代数领域的应用 [D]. 金华: 浙江师范大学, 2019.
- [21] HENSON R, DOUGLAS J. Test construction for cognitive diagnosis [J]. Applied Psychological Measurement, 2005, 29(4): 262-277.
- [22] 颜远海, 丁树良, 汪文义. 影响 AHM 与 DINA 诊断准确率的因素研究 [J]. 江西师范大学学报(自然科学版), 2011, 35(6): 640-645.
- [23] 聂畅, 孙小坚, 辛涛. 基于 BP 神经网络的认知诊断评估分类准确率影响因素分析 [J]. 中国考试, 2021(3): 28-35.

The Analytical Method of Multiple Choice and Multiple Answer Cognitive Diagnostic Tests with Option Information

LI Junjie^{1,2}, ZHENG Huijing¹, NIU Yizhe¹, KANG Chunhua¹, ZENG Pingfei^{1*}

(1. Assessment of Education Quality, Zhejiang Normal University, Jinhua Zhejiang 321004, China;

2. Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing 100875, China)

Abstract: As an important type of cognitive diagnostic test, multiple choice and multiple answer questions have many advantages and uniqueness, especially in the aspect of correctness and non-uniqueness. At present, most cognitive diagnostic models use second-level scoring or multilevel scoring methods to estimate the knowledge state of the examinees when processing the response data of multiple choice and multiple answer questions. However, the scoring of the above two scoring methods is still at the item level, not in-depth into the specific option level, and make full use of the option information in multiple choice and multiple answer questions. Therefore, the method of vector option scoring method and space vector distance discriminant method is developed to analyze the cognitive diagnostic test of multiple choice questions, and the performance of vector option scoring method and space vector distance discriminant Method is verified to analyze the cognitive diagnostic test of multiple choice and multiple answer questions through simulation study and empirical study.

Key words: multiple choice and multiple answer questions; cognitive diagnostic tests; the scoring method; discriminant method

(责任编辑: 冉小晓)