

基于估计类中心和广义距离的认知诊断^①

余 燕, 张自力

西南大学 计算机与信息科学学院, 重庆 400715

摘要: 认知诊断能诊断学生的属性掌握模式, 依据诊断结果能更好地组织个性化教学, 提高教学质量. 针对现有方法诊断准确率不够高的难题, 提出了一种基于估计类中心和广义距离的认知诊断方法. 模拟研究表明: 在大多数情形下, 本文方法的诊断准确率更高. 诊断准确率的提高能促进个性化教学的发展, 对教学质量的提高有着重要的现实意义.

关 键 词: 类中心; 广义距离; 认知诊断; 个性化教学

中图分类号: B841

文献标志码: A

文章编号: 1000-5471(2018)05-0037-05

在传统教学模式下, 教育从业者用测验分数评估学生的学习成果, 但测验分数笼统、不准确^[1]. Tatsuka^[2] 提出规则空间模型(rule space model, RSM), 诊断学生的属性掌握模式, 开启了从微观层面评价学生学习成果的研究. Leighton, Girel 和 Hunka^[3] 提出了属性层级方法(attribute hierarchy method, AHM), 根据判归方式, AHM 分为 A 方法(AHM-A)和 B 方法(AHM-B). 祝玉芳、邓丽萍和周世科等^[4] 证明 AHM-A 比 AHM-B 的诊断准确率更高. Gierl, Cui 和 Hunka 将多层感知器神经网络用于认知诊断^[5], 为认知诊断研究提供了新思路. 涂冬波、蔡艳和丁树良^[6] 将 BP 神经网络用于认知诊断, 并扩大了训练集.

虽然国内外在认知诊断方面的研究已有不少成就, 但诊断准确率方面还有提升的空间. 本文把项目反应理论和广义距离相结合, 提出了一种基于估计类中心和广义距离的认知诊断方法(cognitive diagnosis based on estimated class center and generalized distance, E-GDD). 同时, 展开模拟研究分析 E-GDD 的性能.

1 方法建立

本文利用 Q 矩阵理论推导学生的期望反应模式(expected response pattern, ERP); 若发生猜测或失误, 估计 ERP 对应学生对每个测验项目的正确作答概率, 以 ERP 对应学生对所有测验项目的正确作答概率组成的向量为类中心; 引入项目区分度和难度参数定义广义距离, 把学生的观察反应模式(observed response pattern, ORP)判归到与其广义距离最近的类中心中, 诊断学生的属性掌握模式.

1.1 建立期望反应模式

ERP 是学生对测验项目在理想状态下作答后的得分向量, 1 表示作答正确, 0 表示作答错误. 事实上, 学生实际作答时很难达到理想状态, 我们难以从作答试卷中采集学生的 ERP, 故应用 Q 矩阵理论间接建立. 首先, 从试卷的 J 个测验项目中提取考查的 k 个认知属性, 建立表示项目与认知属性间关联关系的 Q 矩阵和表示认知属性间逻辑关系的属性层级结构; 其次, 从 2^k 种认知属性的组合中剔除不满足层级结构的组合得到 T 类典型的属性掌握模式(1 表示掌握, 0 表示未掌握); 再假设每一类属性掌握模式对应一类学

① 收稿日期: 2017-12-19

作者简介: 余 燕(1992-), 女, 重庆丰都人, 硕士研究生, 主要从事认知诊断、测验数据挖掘研究.

生,让学生在理想状态下对 J 个测验项目作答,建立期望反应模式.

1.2 估计类中心

ORP 是学生对每个测验项目实际作答后的得分向量,1 表示作答正确,0 表示作答错误.若发生猜测或失误,则学生的 ORP 与其自身的 EPR 之间有误差.本文利用软件 BILOG-GM^[7] 估计 ERP 对应学生的能力值、项目的区分度参数和难度参数,应用双参数逻辑斯蒂克模型(2 parameters logistic model, 2PLM)^[8] 进行推导,以 ERP 对应学生对每一个测验项目的正确作答概率组成的向量为类中心.2PLM 如公式(1) 所示:

$$P_{ij} = \frac{\exp(Da_j(\theta_i - b_j))}{1 + \exp(Da_j(\theta_i - b_j))} \quad (1)$$

其中: $D \approx 1.7$; θ_i 为第 t 类 ERP 对应学生的能力值; a_j 为测验项目 j 的区分度参数; b_j 为测验项目 j 的难度参数; P_{ij} 为第 t 类 ERP 对应学生正确作答测验项目 j 的概率.

1.3 诊断属性掌握模式

若 X_{ij} 表示学生 i 在测验项目 j 上的观察得分, α_{ij} 表示第 t 类 ERP 对应学生在测验项目 j 上的期望得分.定义 X_{ij} 到 α_{ij} 的广义距离如公式(2) 所示:

$$\text{dist}(X_{ij}, \alpha_{ij}) = \begin{cases} X_{ij} - P_{ij}, & X_{ij} = 1 \text{ and } \alpha_{ij} = 0 \\ P_{ij} - X_{ij}, & X_{ij} = 0 \text{ and } \alpha_{ij} = 1 \\ 0, & X_{ij} = \alpha_{ij} \end{cases} \quad (2)$$

a_j 和 b_j 会在一定程度上影响学生作答时的猜测或失误概率.本文将标准化后的 a_j 和 b_j 融入到广义距离中,度量 ORP 与 ERP 的相似性.其中,采用 Min-max 标准化 a_j 和 b_j ,同时对标准化后的 a_j 和 b_j 加 1 以防止标准化后的 $b_j = 0$.在假定 ORP 不含缺失值的情形下,定义 X_i 到 α_i 的广义距离如公式(3) 所示:

$$\text{dist}(X_i, \alpha_i) = \sum_{j=1}^J \left(\frac{a_j}{b_j} + \text{dist}(X_{ij}, \alpha_{ij}) \right) \quad (3)$$

将第 i 个 ORP 与每个 ERP 进行比较,计算广义距离.基于距离越小越相似的思想,把第 i 个 ORP 判归到与其广义距离最近的 ERP 中,根据期望反应模式与典型属性掌握模式间的一一对应关系,诊断学生的属性掌握模式为与其广义距离最近的期望反应模式对应的典型属性掌握模式.

2 实 验

模拟研究可将诊断得到的学生的属性掌握模式与其真实的属性掌握模式进行对比,便于比较不同方法的诊断准确率.本实验用 R 语言实现 E-GDD,RSM^[2],AHM-A^[3],用 Matlab 实现 BP 神经网络的认知诊断方法^[5-6],诊断学生的属性掌握模式,选择模式判准确率(pattern match ratio, PMR)和边际判准确率(marginal match ratio, MMR)评价各方法的诊断准确率,对比分析 E-GDD 的性能.其中,对于 BP 神经网络的方法^[6],考虑 3 种训练集下的诊断准确率,从中选择准确率最高情形与 E-GDD 进行对比.3 种训练集分别为:期望反应模式及对应的典型属性掌握模式(CD-BP);期望反应模式加带有 0.1% 项目滑动概率的非期望反应模式及对应的典型属性掌握模式(CD-BP(0.1%));期望反应模式加带有 50% 项目滑动概率的非期望反应模式及对应的典型属性掌握模式(CD-BP(50%)).

2.1 数据模拟

属性层级结构和项目滑动概率会在一定程度上影响诊断准确率. Leighton,Girel 和 Hunka^[3] 提出了 4 种属性层级结构(直线型、收敛型、发散型和无结构型).实际测验中的属性层级结构都由这 4 种基本结构组合而来^[9].因此,本文参考 Leighton,Girel 和 Hunka^[3] 的方法,考虑 4 种属性层级结构模拟 4 组期望反应模式,每组期望反应模式考虑 4 种项目滑动概率(2%,5%,10%和 15%)^[4,6] 模拟 2000 名学生的实际观察反应模式,得到 16 组模拟数据.其中,每种搭配实验重复 30 次以减少误差.

2.2 结果分析

首先分析 BP 神经网络在 3 种训练集情形下的诊断准确率.表 1 记录了 3 种训练集对应 16 种条件下 BP 神经网络的 PMR 和 MMR.

表 1 BP 神经网络在不同训练集下的 PMR 和 MMR

属性层级结构	项目滑动概率/%	PMR			MMR		
		CD-BP(0.1%)	CD-BP	CD-BP(50%)	CD-BP(0.1%)	CD-BP	CD-BP(50%)
直线型	2	0.936	0.913	0.908	0.987	0.982	0.981
	5	0.851	0.795	0.805	0.969	0.956	0.959
	10	0.710	0.618	0.654	0.934	0.908	0.926
	15	0.574	0.479	0.538	0.897	0.866	0.900
收敛型	2	0.919	0.908	0.502	0.984	0.983	0.901
	5	0.816	0.791	0.513	0.962	0.957	0.891
	10	0.657	0.620	0.365	0.923	0.915	0.854
	15	0.529	0.480	0.305	0.889	0.875	0.835
发散型	2	0.915	0.902	0.367	0.984	0.983	0.838
	5	0.801	0.763	0.276	0.960	0.954	0.790
	10	0.619	0.568	0.183	0.916	0.906	0.740
	15	0.474	0.416	0.134	0.874	0.857	0.717
无结构型	2	0.902	0.901	0.557	0.983	0.983	0.868
	5	0.769	0.761	0.281	0.957	0.955	0.766
	10	0.569	0.555	0.118	0.907	0.905	0.685
	15	0.405	0.390	0.067	0.855	0.852	0.641

分析表 1, 属性层级结构和实际观察反应模式的项目滑动概率一定时, CD-BP(0.1%) 的 PMR 和 MMR 都是最高的, 说明 CD-BP(0.1%) 在扩大训练集的同时较好保证了样本的质量, PMR 和 MMR 相对于 CD-BP 有所提高, 因此, 本文选择 CD-BP(0.1%) 作为 E-GDD 的对照方法.

图 1 表示了 4 种方法在相同属性层级结构、不同项目滑动概率条件下的 PMR.

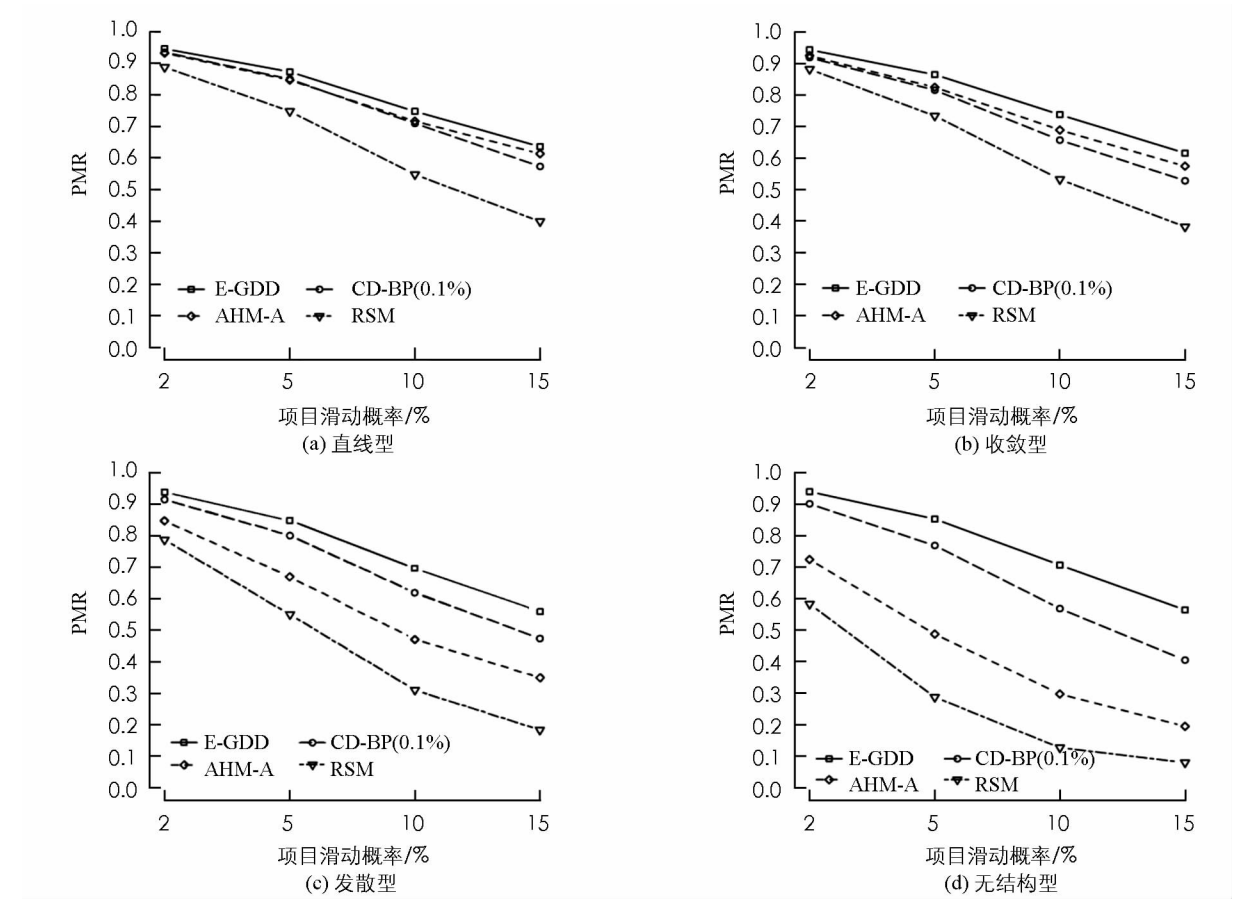


图 1 4 种认知诊断方法的模式判准率

在同一属性层级结构下,4 种方法的 PMR 都随着项目滑动概率的增加而降低,说明项目滑动概率越大,学生的实际观察反应模式与其真实的期望反应模式偏离得越远,准确诊断的难度越大.

当项目滑动概率一定时,对于直线型和收敛型层级结构,AHM-A 比 CD-BP(0.1%)和 RSM 的 PMR 高;对于发散型和无结构型层级结构,CD-BP(0.1%)比 AHM-A 和 RSM 的 PMR 高.但总体来看,在同一属性层级结构和项目滑动概率下,E-GDD 的 PMR 最高,若属性层级结构为无结构型且项目滑动概率为 15%,E-GDD 比 CD-BP(0.1%)、AHM-A 和 RSM 中准确率最高的 CD-BP(0.1%)的 PMR 高出最多,约为 15%.

在同一属性层级结构下,实际观察反应模式的项目滑动概率在 2%,5%,10%和 15%之间变化时,4 种认知诊断方法 PMR 的变化范围不一样.对于直线型和收敛型层级结构,RSM,AHM-A,CD-BP(0.1%)和 E-GDD 的 PMR 的范围非常接近,分别为(0.4, 0.9),(0.6, 1.0),(0.5, 1.0)和(0.6, 1.0);对于发散型层级结构,RSM,AHM-A,CD-BP(0.1%)和 E-GDD 的 PMR 范围分别为(0.1, 0.8),(0.3, 0.9),(0.4, 1.0)和(0.5, 1.0);对于无结构型层级结构,RSM,AHM-A,CD-BP(0.1%)和 E-GDD 的 PMR 的范围分别为(0.0, 0.6),(0.2, 0.8),(0.4, 0.9)和(0.5, 1.0).通过以上分析发现:属性层级结构越发散,4 种方法 PMR 的范围越低,但 E-GDD 始终比 RSM,AHM-A 和 CD-BP(0.1%)的范围高.这说明 E-GDD 的诊断准确率受属性层级结构的影响更小.

图 2 表示了 4 种方法在相同属性层级结构、不同项目滑动概率条件下的 MMR.

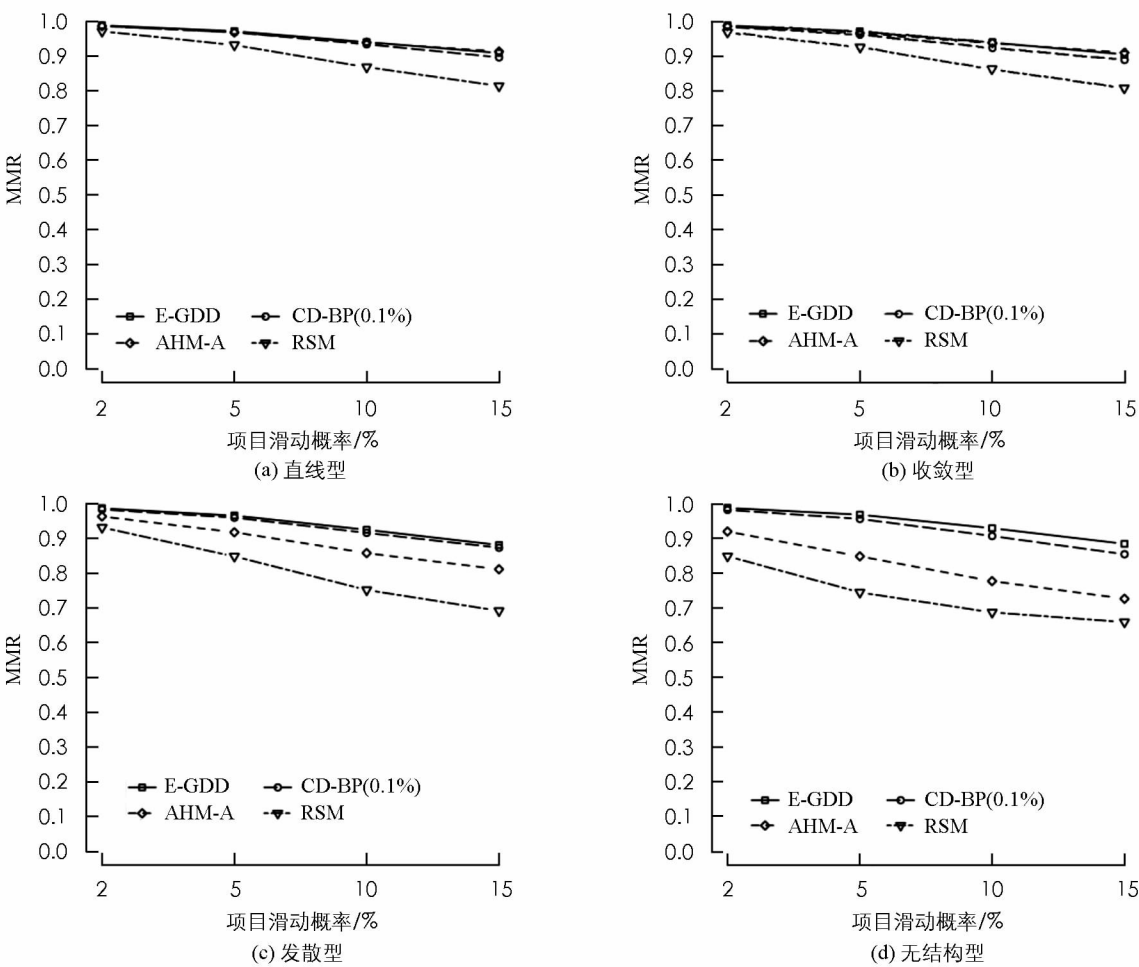


图 2 4 种认知诊断方法的边际判准率

若属性层级结构和项目滑动概率一定,对同一种方法的诊断准确率进行评价,PMR 小于 MMR.因为 PMR 和 MMR 的评价角度不一样,PMR 计算学生的属性掌握模式被正确诊断的比例,MMR 计算学生对认知属性的掌握情况被正确诊断的比例.

和 PMR 一样, 在同一属性层级结构下, 4 种方法的 MMR 都随着项目滑动概率的增加而降低. 实际观察反应模式的项目滑动概率为 15%、属性层级结构为直线型或收敛型时, AHM-A 的 MMR 略微高于 E-GDD 的 MMR. 说明在被 E-GDD 正确诊断的属性掌握模式数量多于被 AHM-A 正确诊断的数量但不明显的情况下, 可能出现被 E-GDD 诊断错误的认知属性掌握情况比被 AHM-A 诊断错误的认知属性掌握情况稍多的现象, 进而得到 AHM-A 的 MMR 比 E-GDD 的 MMR 高的结果.

除以上两种情形以外, 在属性层级结构和项目滑动概率都一定的情形下, E-GDD 的 MMR 始终高于 RSM, AHM-A 和 CD-BP(0.1%) 的 MMR, 若属性层级结构为无结构型且项目滑动概率为 15% 时, E-GDD 比 CD-BP(0.1%), AHM-A 和 RSM 中准确率最高的 CD-BP(0.1%) 的 MMR 高出最多, 约为 3%.

3 总 结

为提高学生认知属性掌握情况的诊断准确率, 更好地服务于个性化教学, 本文将项目区分度参数和难度参数融入到广义距离中, 与项目反应理论相结合, 提出了 E-GDD. 与 AHM-A, RSM 和 BP 神经网络的认知诊断方法对 16 组模拟数据的诊断结果相比, E-GDD 在大多数情形下具有更高的诊断准确率. 诊断准确率提高后, 教师可省去分析学生试卷的时间, 将更多精力投入到个性化教学课堂中, 进而促进个性化教学的发展.

参考文献:

- [1] 鲍孟颖. 运用 DINA 模型对 5 年级学生数学应用题问题解决进行认知诊断 [D]. 上海: 华东师范大学, 2014.
- [2] TATSUOKA K K. Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory [J]. Journal of Educational Measurement, 1983, 20(4): 345—354.
- [3] LEIGHTON J P, GIERL M J, HUNKA S M. The Attribute Hierarchy Method for Cognitive Assessment: A Variation on Tatsuoka's Rule-Space Approach [J]. Journal of Educational Measurement, 2004, 41(3): 205—237.
- [4] 祝玉芳, 邓丽萍, 周世科, 等. 用规则空间模型和属性层次方法进行认知诊断 [J]. 计算机时代, 2008, (6): 1—3.
- [5] GIERL M J, CUI Y, HUNKA S. Using Connectionist Models to Evaluate Examinees' Response Patterns to Achievement Tests [J]. Journal of Modern Applied Statistical Methods, 2008, 7(1): 234—245.
- [6] 涂冬波, 蔡 艳, 丁树良. 认知诊断理论、方法与应用 [M]. 北京: 北京师范大学出版社, 2012.
- [7] KIM S H. BILOG 3 for Windows: Item Analysis and Test Scoring with Binary Logistic Models [J]. Applied Psychological Measurement, 1997, 21(4): 371—376.
- [8] FISCHER G H. The Linear Logistic Test Model as An Instrument in Educational Research [J]. Acta Psychologica, 1973, 37(6): 359—374.
- [9] 丁树良, 毛萌萌, 汪文义, 等. 教育认知诊断测验与认知模型一致性的评估 [J]. 心理学报, 2012, 44(11): 1535—1546.

Cognitive Diagnosis Based on Estimated Class Centers and Generalized Distance

YU-Yan, ZHANG Zi-li

School of Computer and Information Science, Southwest University, Chongqing 400715, China

Abstract: Cognitive diagnosis could diagnose whether students master or not for each attribute, and organize individual teaching according to diagnosed results to improve the quality of teaching. Combining item response theory and similarity measure, a cognitive diagnosis method based on estimated class centers and generalized distance has been proposed to improve diagnostic accuracy. Simulation studies show that in most cases, the accuracy of this method is higher. The improvement of diagnostic accuracy can promote the development of individual teaching and has important practical significance for the improvement of teaching quality.

Key words: class centers; generalized distance; cognitive diagnosis; individual teaching

责任编辑 张 杓