

文章编号: 1000-5862(2017)04-0394-07

海明距离判别法分类准确率的影响因素

康春花 杨亚坤 曾平飞*

(浙江师范大学教师教育学院 浙江 金华 321004)

摘要: 为探讨海明距离判别法(HDD)的非参数优势,通过一个5因素混合实验,考察了4个因素(属性层级、测验长度、样本容量、知识状态分布)对HDD的3种判别方法(R方法、B方法、W方法)分类准确率的影响。结果表明:1)属性层级和测验长度均会影响HDD判准率,属性层级越紧密、测验长度越长,HDD判准率越高;2)HDD对样本容量无依赖,可适于小样本评估;3)HDD的R方法、B方法、W方法的分类准确率无差异;4)HDD无需被试知识状态分布的正态性假设,更适于均匀分布。

关键词: 海明距离判别法; 属性层级; 测验长度; 样本容量; 知识状态分布

中图分类号: B 841.7 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2017.04.11

0 引言

认知诊断评估(cognitive diagnostic assessment, CDA)为教师了解学生的认知结构和心理过程提供了新思路和新方法。尽管CDA的思想易于理解,但目前的诊断分类方法多数基于复杂的参数模型,对于无统计学或计算机背景的研究者及一线教师来说,应用颇为困难,这一定程度上限制了CDA的应用和推广。相比参数方法,非参数方法具有无需前提假设、对数据分布形态无要求、不受样本容量限制、无需参数估计、计算简便等特点^[1]。为此,研究者开始将非参数方法用于认知诊断分类,如V. Vapnik^[2]依据风险最小化原则提出了机器学习方法—支持向量机(support vector machines, SVM); Chiu Chia-yi等^[3]提出了聚类诊断分析法,康春花等^[4]将其扩展为多级评分的聚类诊断法(grade response cluster diagnostic method, GRCDM)。聚类方法虽简单明了,但它是一种非监督的分类方法。Chiu Chia-yi等^[5]在此基础上进一步提出了邻近理想反应模式(ideal response pattern, IRP)的非参数诊断方法,通过计算观察反应模式(observed response pattern, ORP)与IRP间的海明距离实现对被试的诊断分类,罗照盛等^[6]把该方法明确界定为海明距离判别法(hamming distance discrimination, HDD)。模拟研究表明HDD简

单有效,和广义距离判别法(generalized distance discrimination, GDD)^[7]判准率不相上下。然而,模型或方法一经提出,研究者需从多角度多因素论证该方法的有效性和稳健性,判准率的影响因素一直备受关注。

认知诊断模型或方法的影响因素可概括为3个方面:(i)与测验有关的因素,如属性层级、 Q 矩阵、属性个数、测验长度;(ii)与被试有关的因素,如被试知识状态分布、样本容量、失误率;(iii)诊断方法因素,如不同诊断模型或方法、模型与题目特征是否吻合(问题解决时属性之间的补偿性特征)。影响因素的研究,可为方法的有效性验证及方法使用的选择提供信息和依据。关于参数模型的影响因素已有较多的研究,而在非参数方面,研究者仅探讨了0-1和多级计分聚类分析法的影响因素^[4,8-9]。结果表明:聚类诊断法的判准率具有不受被试的知识状态分布的影响、对样本容量无依赖等非参数特性。那么,同属非参数诊断方法,不同因素对HDD判准率的影响如何? HDD作为一种非常简洁的非参数方法,能否体现非参数的特点和优势? 本文拟通过一个多因素的模拟研究,从测验、被试和方法这3个角度(属性层级、测验长度、样本容量、被试知识状态分布、HDD 3种判别方法)探讨HDD判准率影响因素,以期更深入地了解非参数诊断法的特点与优势,丰富非参数方法的研究内容,为CDA的应用和推广提供有用信息。

收稿日期: 2016-12-22

基金项目: 教育部人文社会科学研究一般课题(16YJA190002)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

1 海明距离判别法

在信息论中,海明距离指2个等长字符串对应位置的不同字符个数。如字符串A(which)和字符串B(whore)的对应位置,有3个不同的字符,为ich、ere,则A与B的海明距离为3。在0-1计分的CDA中,假如被试的ORP和IRP均用含0、1元素的字符串来表示,如ORP为(0,0,1),IRP为(0,1,1),则ORP与IRP间的海明距离为1。据此,在CDA中,研究者将海明距离定义为ORP与IRP对应元素不同的个数^[5-6],公式为

$$d_h(y, \eta) = \sum_{j=1}^J |y_j - \eta_j|, \quad (1)$$

其中 y 和 η 分别表示ORP和IRP, y_j 指被试ORP的第 j 个元素, η_j 指IRP的第 j 个元素。由此,HDD通过计算ORP与所有IRP的海明距离,将被试判为与其ORP距离最小的IRP所对应的属性掌握模式(Attribute Mastery Pattern,AMP)。

按照HDD的思想,有可能出现某ORP与多个IRP间的距离最小且相等,即一对多的情形。为此,研究者提出了多种处理方法,分别有R方法、B方法和W方法^[5-6]。R方法为随机抽取方法,即从多个距离最小的IRP中随机选择一个IRP作为该ORP对应的IRP,进而实现对被试掌握模式的估计。B方法为贝叶斯方法:首先,将被试分为2类,第1类为ORP与IRP之间有唯一的最小海明距离,此类被试可以直接归类;第2类为一个ORP对应多个IRP的被试。对于第2类被试,B方法的处理如下:(i)计算第1类被试中各AMP人数的比例;(ii)将第1类被试中各IRP所占的比例作为第2类被试ORP所对应的IRP的后验概率进行Bayes判别。如果此时出现ORP对应的IRP的AMP比例均为0或相等,再采用随机方法选取被试的AMP。W方法为加权海明距离(weighted Hamming distance)法^[5],它以被试在每个项目上的正确作答比例为权重,计算公式为

$$d_{wh}(y, \eta) = \sum_{j=1}^J \frac{1}{\bar{P}_j(1 - \bar{P}_j)} |y_j - \eta_j|, \quad (2)$$

其中 \bar{P}_j 为项目 j 的正确作答率。

2 研究设计

2.1 实验设计

研究固定属性数目为7个,考察4个因素:属性层级(直线型、发散型、收敛型、无结构型和独立型,

如图1)、测验长度($J=21, 42$)、样本容量($N=100, 500, 1000$)、被试知识状态分布(正态分布、均匀分布)对海明距离3种判别方法(R方法、B方法和W方法)分类准确率的影响。实验为 $5 \times 2 \times 3 \times 2 \times 3$ 的5因素混合设计,其中判别方法为被试内变量,其它4个变量为被试间变量。由此,需模拟被试间处理水平结合 $5 \times 2 \times 3 \times 2$ 共60种条件下的实验数据,然后分别用3种HDD判别方法对每种条件下的数据进行分析,共有180种条件下的判准率结果,为减少误差,每种条件重复30次,最后的判准率结果为各条件在30次重复实验中的均值。

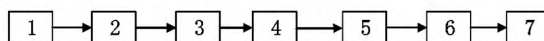
2.2 数据模拟

采用J. P. Leighton^[10]的方法,在IRP的基础上,模拟生成60种条件下,被试作答失误率为10%的观察反应矩阵,每种条件重复30次。具体为:先给被试在每个项目的IRP生成一个服从(0,1)均匀分布的随机数 r ;然后将 r 与0.05和0.95进行比较,获得被试的观察反应,如果 $r < 0.05$ 或 $r > 0.95$,则被试得分发生改变,即得1分变为0分、得0分变为1;否则,该项目得分保持不变。

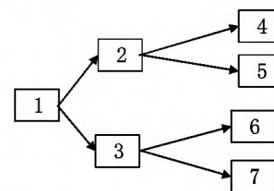
2.3 数据分析与评价指标

自编R语言程序,实现HDD的3种判别法的判别。采用边际判准率(marginal match ratio,MMR)和模式判准率(pattern match ratio,PMR)作为评价指标。

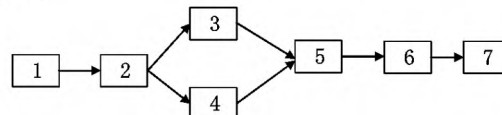
直线型



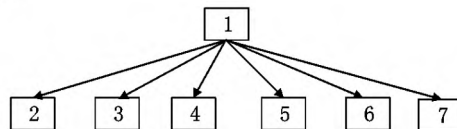
分支型



收敛型



无结构型



独立型

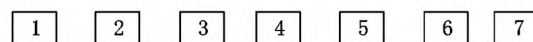


图1 7个属性5种不同的属性层级

3 研究结果

本研究为 $5 \times 2 \times 3 \times 2 \times 3$ 的 5 因素混合设计,其中属性层级(5)、测验长度(2)、样本容量(3)、知识状态分布(2)为被试间变量,HDD 的 3 种判别方法为被

试内变量.3 种判别方法分别在 60 种数据模拟条件下重复 30 次,判准率均值见表 1 和表 2.总体而言,HDD 具有较高的判准率,各条件下 MMR 均值均在 0.90 以上,且与 PMR 的变化趋势一致,接下来以 PMR 为例,阐述各因素与 HDD 判准率之间的关系.

表 1 测验长度 21 时的判准率

层级关系类型	人数	R 方法		B 方法		W 方法		
		PMR	MMR	PMR	MMR	PMR	MMR	
均匀分布	直线型	100	0.853 3	0.973 0	0.844 0	0.970 6	0.848 7	0.971 8
		500	0.844 7	0.971 0	0.844 1	0.970 7	0.844 9	0.971 0
		1 000	0.846 0	0.971 3	0.843 1	0.970 0	0.844 1	0.970 4
	收敛型	100	0.822 7	0.966 3	0.825 7	0.967 3	0.828 0	0.968 2
		500	0.825 1	0.967 6	0.826 1	0.967 7	0.827 3	0.968 2
		1 000	0.831 1	0.968 2	0.834 4	0.969 2	0.835 6	0.969 5
	分支型	100	0.708 0	0.938 7	0.719 0	0.942 8	0.707 7	0.938 6
		500	0.711 0	0.941 8	0.728 5	0.947 7	0.708 2	0.940 0
		1 000	0.709 7	0.939 0	0.730 4	0.946 3	0.700 1	0.935 1
	无结构型	100	0.616 0	0.922 3	0.627 0	0.925 3	0.632 3	0.926 9
		500	0.629 9	0.927 7	0.644 9	0.931 5	0.646 3	0.931 9
		1 000	0.622 0	0.924 9	0.636 5	0.928 8	0.638 3	0.929 4
独立型	100	0.562 7	0.916 5	0.560 0	0.915 9	0.561 3	0.916 1	
	500	0.562 7	0.916 0	0.562 2	0.915 8	0.563 8	0.916 2	
	1 000	0.568 3	0.917 4	0.567 4	0.917 0	0.567 5	0.917 1	
正态分布	直线型	100	0.813 0	0.963 5	0.834 0	0.970 6	0.790 7	0.955 7
		500	0.823 4	0.965 4	0.843 6	0.972 1	0.801 2	0.957 8
		1 000	0.819 1	0.965 7	0.837 9	0.971 7	0.800 2	0.959 5
	收敛型	100	0.806 7	0.963 6	0.822 3	0.967 9	0.791 0	0.958 2
		500	0.812 5	0.965 3	0.832 5	0.971 2	0.802 3	0.961 5
		1 000	0.810 5	0.964 3	0.830 5	0.970 2	0.798 2	0.960 4
	分支型	100	0.670 7	0.933 5	0.670 3	0.934 7	0.658 3	0.927 4
		500	0.658 2	0.930 3	0.669 9	0.934 4	0.641 5	0.921 3
		1 000	0.667 2	0.932 7	0.677 5	0.936 5	0.653 0	0.925 7
	无结构型	100	0.571 3	0.921 2	0.565 7	0.920 1	0.555 7	0.911 0
		500	0.566 7	0.918 6	0.567 1	0.919 0	0.555 4	0.910 2
		1 000	0.572 2	0.919 1	0.573 8	0.920 0	0.559 8	0.910 6
独立型	100	0.543 0	0.913 3	0.538 3	0.912 8	0.543 7	0.914 0	
	500	0.531 0	0.909 0	0.528 1	0.908 1	0.530 7	0.908 7	
	1 000	0.522 3	0.906 8	0.521 3	0.906 6	0.522 7	0.906 9	

3.1 海明距离 3 种方法的判准率差异极小

纵观表 1 和表 2,无论是测验长度为 21 题还是 42 题,被试知识状态分布为均匀分布还是正态分布,PMR 和 MMR 均值在 R、B、W 方法上的差异均较小,且变化趋势一致.为更直观地看出 HDD 的 3 种判别方法对判准率的影响,以测验长度 42 题为例,呈现均匀和正态分布时,其余各条件下 3 种方法在判准率上的差异(见图 2).从图 2 可以看出,均匀分布和正态分布时,A、B、W 方法在各属性层级下的判准率变化趋势几乎完全一致,且除了分支型时,B

方法稍好(但最多也不超过 0.02,分别为 0.792 9、0.806 4、0.788 8),其它属性层级下 3 种方法的判准率几乎无差异.

3.2 海明距离判别法对样本容量无依赖

因 3 种判别方法无差异,接下来均以 B 方法为例,阐述各因素对 HDD 的影响.图 3 为 42 题时,均匀和正态分布下,各属性层级及各样本容量下 B 方法的判准率.从图 3 可以较清楚地看出,无论是在何种分布的何种属性层级下,100 人、500 人和 1 000 人的判准率几乎都在同一水平线上.由此,样本容量

对海明距离判准率无影响,且这种趋势与分布形态和属性层级均无关.

表 2 测验长度 42 时的判准率

层级关系类型	人数	R 方法		B 方法		W 方法		
		PMR	MMR	PMR	MMR	PMR	MMR	
均匀分布	直线型	100	0.945 3	0.991 3	0.944 3	0.991 2	0.945 7	0.991 5
		500	0.944 2	0.991 6	0.941 3	0.991 0	0.942 7	0.991 2
		1 000	0.944 1	0.991 5	0.942 8	0.991 2	0.943 7	0.991 4
	收敛型	100	0.927 3	0.988 5	0.924 7	0.988 0	0.924 0	0.987 8
		500	0.928 6	0.988 8	0.927 1	0.988 4	0.928 1	0.988 8
		1 000	0.927 1	0.988 4	0.927 2	0.988 5	0.927 8	0.988 6
	分支型	100	0.802 3	0.960 1	0.814 7	0.964 2	0.803 3	0.960 7
		500	0.792 9	0.958 8	0.806 4	0.963 3	0.788 8	0.957 3
		1 000	0.806 2	0.962 2	0.819 8	0.966 8	0.798 8	0.959 3
	无结构型	100	0.693 4	0.939 7	0.701 0	0.941 4	0.707 3	0.943 5
		500	0.691 4	0.938 2	0.708 2	0.942 9	0.710 3	0.943 3
		1 000	0.712 0	0.942 9	0.724 5	0.946 3	0.725 8	0.946 7
	独立型	100	0.646 0	0.933 5	0.642 0	0.932 1	0.646 0	0.932 9
		500	0.643 9	0.932 6	0.641 9	0.931 8	0.644 0	0.932 5
		1 000	0.645 1	0.931 5	0.643 1	0.931 1	0.646 2	0.931 8
正态分布	直线型	100	0.943 3	0.991 2	0.946 0	0.991 9	0.944 3	0.991 3
		500	0.943 3	0.991 3	0.945 5	0.991 8	0.941 5	0.990 7
		1 000	0.929 8	0.989 0	0.933 6	0.990 0	0.927 7	0.988 5
	收敛型	100	0.924 0	0.988 3	0.925 3	0.988 6	0.920 0	0.987 2
		500	0.918 9	0.987 0	0.924 5	0.988 5	0.916 9	0.986 4
		1 000	0.919 7	0.987 5	0.922 5	0.988 2	0.918 9	0.987 3
	分支型	100	0.757 7	0.952 7	0.764 7	0.954 5	0.760 0	0.952 1
		500	0.776 4	0.958 5	0.780 7	0.960 2	0.765 9	0.953 9
		1 000	0.753 6	0.952 5	0.760 1	0.9547	0.739 6	0.945 7
	无结构型	100	0.635 3	0.930 4	0.628 7	0.927 9	0.630 7	0.925 5
		500	0.639 8	0.932 2	0.637 3	0.932 2	0.630 1	0.925 9
		1 000	0.635 3	0.931 3	0.635 7	0.931 7	0.624 7	0.924 6
	独立型	100	0.587 0	0.918 3	0.578 3	0.916 4	0.584 7	0.918 3
		500	0.595 3	0.919 9	0.589 5	0.918 3	0.594 7	0.919 5
		1 000	0.595 1	0.920 9	0.592 1	0.920 2	0.595 1	0.921 0

3.3 增加测验长度可提高判准率

从图 4 可直观地看出各分布形态和各属性层级下,测验长度对判准率的影响.由图 4 的右图可知,测验长度对判准率的影响与分布形态和属性层级均

无关.在各种条件下,均为 42 题的判准率要高于 21 题的判准率,如 21 题时 PMR 均值最大为 0.843 7,而 42 题时却高达 0.942 8,并且这种差异在直线型和收敛型时体现更为明显,接近 0.10.

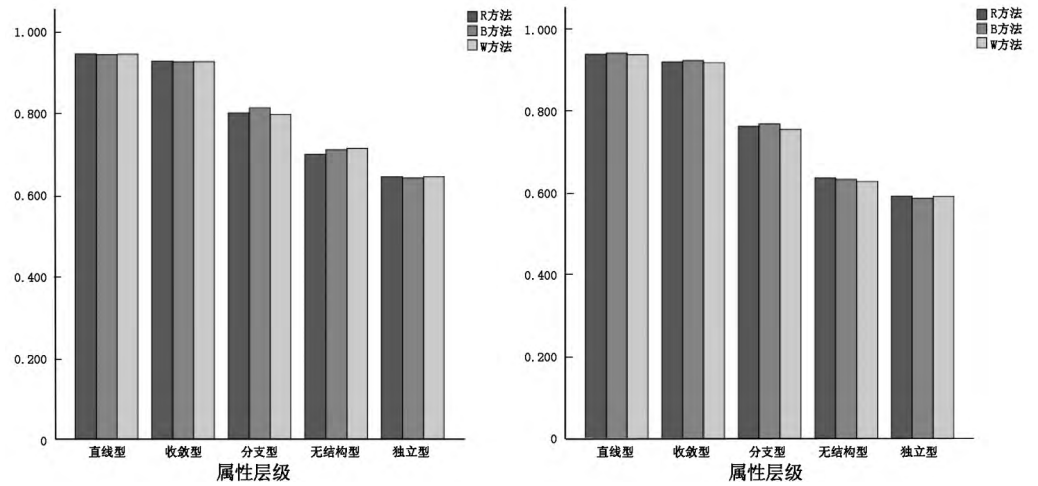


图 2 均匀(左)和正态分布(右)时,各属性层级下 3 种方法的判准率均值

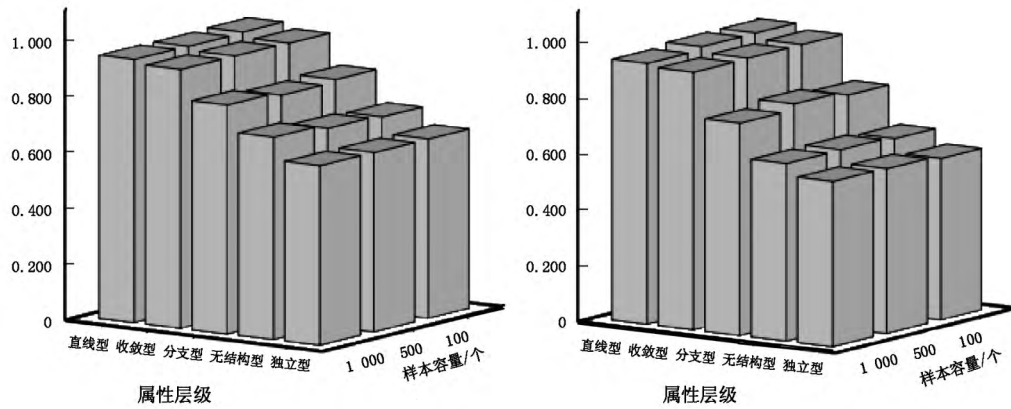


图3 均匀(左)和正态分布(右)时,各属性层级及各样本容量下的判准率均值

3.4 属性层级和知识状态对判准率的影响

综合上述图2~图4可知,无论从哪个角度,HDD判准率受属性层级影响较大,由高到底依次为:直线型、收敛型、发散型、无结构型、独立型,而且这种趋势与样本容量、判别方法、分布形态、测验长度均无关,即无论在何种条件下,属性层级越紧密,

其判准率越高.同样,知识状态对HDD判准率的影响也与样本容量、判别方法和测验长度无关,但与属性层级有关,如图5所示.属性层级为直线型和收敛型时,均匀分布和正态分布的判准率一致;属性层级为分支型、无结构型和独立型时,2种测验长度下,皆为均匀分布判准率要稍高于正态分布.

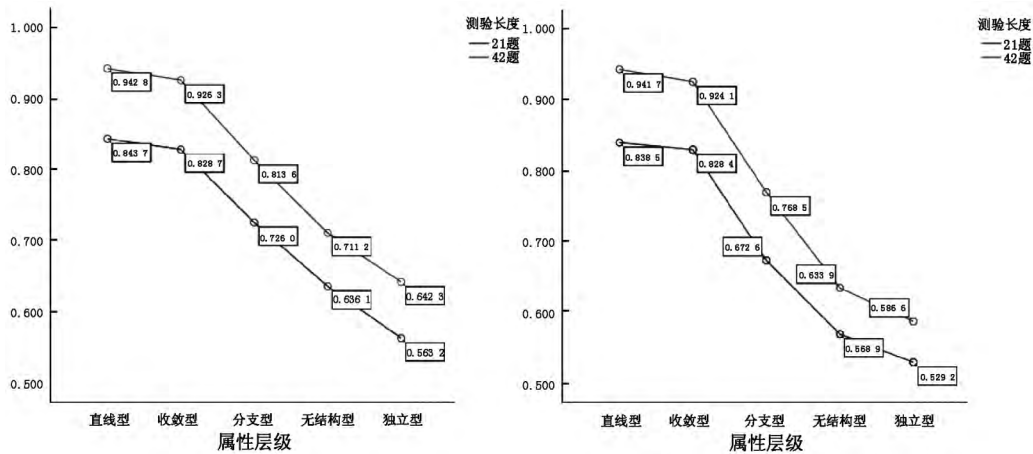


图4 均匀(左)和正态分布(右)时,各属性层级及各测验长度下的判准率均值

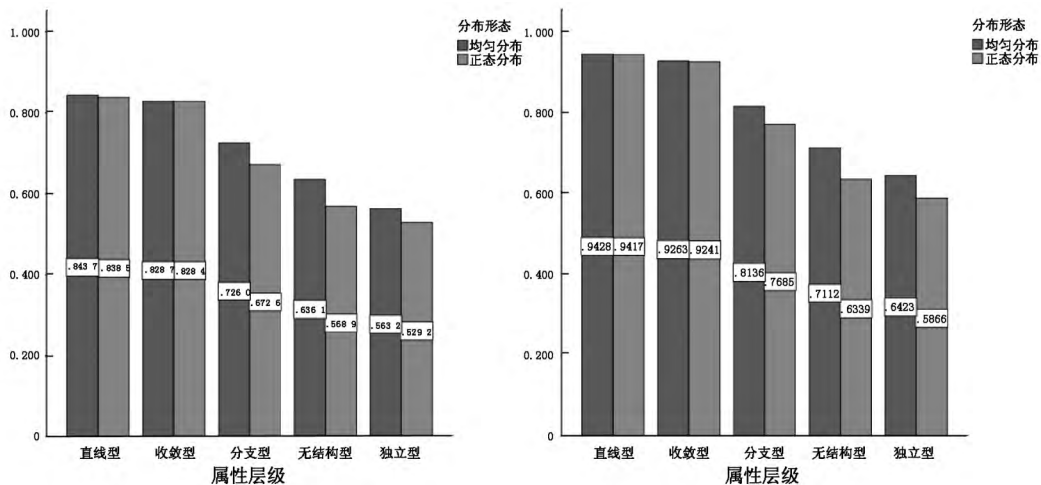


图5 21题(左)和42题(右)时,各属性层级及各知识状态下的判准率均值

4 讨论

本研究从测验(属性层级、测验长度)、被试(知识分布、样本容量)、方法(R、B、W)3个角度的5个因素较全面地考察了HDD判准率的影响因素,为进一步为非参数诊断法的优越性提供更多理论证据,丰富非参数诊断研究,为认知诊断实践的方法选择提供有效信息。

4.1 HDD受属性层级和测验长度影响明显

首先,在属性层级方面,HDD所得结论与已有研究一致,即属性层级对判准率有一定的影响,属性层级越紧密判准率越高^[11-13]。HDD通过比较ORP与IRP之间的海明距离实现对ORP的分类,属性层级越松散,典型项目考核模式也越多,IRP相似的可能性越大,由此ORP可能与错误的IRP更邻近,使得发生误判的概率增大。

其次,在测验长度方面,已有研究表明:在发散型、收敛型、无结构型中,测验长度应越长越好,但在线型结构中,测验长度并非越长越好^[11];此外,在Q矩阵包含一个R矩阵的情况下^[14],测验长度也非越长越好,而Q矩阵包含R矩阵的个数越多则判准率越高^[11]。HDD结果表明无论在何种属性层级下,均为42题判准率要高于21题时的判准率,由于本研究中42题是在21题的考核模式上直接翻倍,则此时判准率的提高是由于测验长度还是由于R矩阵个数增多导致,值得进一步考察。

4.2 HDD 3种方法无差异

对于HDD可能出现一种ORP对应多个IRP的情况(打结),Chiu Chia-yi等^[5]指出由于被试的作答反应在某些项目上的变异要比在其他项目上大,可以根据项目内变异的大小加以不同的权重,提出了加权海明距离,即W方法。W方法是一种直接利用被试作答信息的判别方式。模拟研究表明,相比R方法,W方法“打结”情况更少发生。此外,罗照盛等^[6]提出了利用被试先验信息的B方法,指出B方法较R方法更多地利用了样本中被试的分布信息,因此更为合理,其研究表明:在5种属性层级下,判准率由高到低依次为HDD的B方法、HDD的R方法、GDD;就HDD而言,B方法要优于R方法,但B方法容易受分布信息准确性的影响,因此当样本量较小时,还是宜采用R方法。本研究为进一步探究HDD的3种方法的优劣信息,同时考察了R方法、B方法和W方法在各种条件下的判准率,结果表明3种方法除了在个别情况下相差0.02,B方法稍好外,其它情况下都几乎无差异。罗照盛等^[6]的研究,也只是在发散型15%失误率的情况下相差不到0.06(0.876 8,0.802 5),其它条件下均不超过

0.02,多数为0~0.001间。因此,综合而言,可以认为HDD的3种方法几乎无差异。

4.3 HDD对样本容量无依赖,适宜小样本评估

在认知诊断方法对样本容量的要求方面,参数模型所需样本容量一般至少在1 000~2 000区间内^[5,12]。而对于非参数方法,由于无需参数估计,即便只有一个被试也可以进行判别分类^[5],相对而言,即使小样本也很适宜^[4,8]。本研究中,为进一步探讨非参数方法对样本容量是否无依赖,以及是否适宜小样本评估,将样本容量设置为100、500和1 000。图3的3维柱状图表明,HDD在3种样本容量下的判准率几乎处于同一水平线,且这种趋势并不受其它变量影响。由此,HDD的结果进一步验证了非参数方法对样本容量无要求的优势,可用于小型诊断评估,为CDA走向实践、走向课堂提供了可能。

4.4 HDD无需正态性假设,较适宜均匀分布

蔡艳等^[15]的研究表明知识状态的分布形态会影响5个认知诊断模型的诊断正确率。对于规则空间方法而言,由高到低分别是:负偏态、正偏态、正态、均匀分布;对于属性层级方法和DINA而言,由高到低分别是:负偏态、正态、正偏态、均匀分布;对于GDD和DINA_HC而言,由高到低分别是:负偏态、均匀、正态、正偏态分布。可见,参数方法对被试知识状态分布确实有要求,正态、正偏态和负偏态的判准率普遍高于均匀分布。而对于非参数方法,如GRCDM的判准率在正态分布和均匀分布时就无差异^[4]。本研究中HDD的判准率在直线型和收敛型时,正态分布和均匀分布条件下的判准率几乎相等,在其它属性层级下,均匀分布的判准率反而高于正态分布。由此,HDD和GRCDM结果一致,体现了非参数方法无需正态性假设的优势,较适宜均匀分布。

4.5 研究展望

以7个属性为例,本研究探讨了5个因素对HDD分类准确率的影响,获得了一些有意义的结果,然而,仍有一些问题值得进一步探究。比如,在参数方法中,属性数目不能太多,一般而言,如果属性数目超过7个,会造成模型判准率急速下降^[5,12,16]。然而,研究者关于GRCDM的研究结果却显示:当属性个数为4、7、9个时,20次重复实验下的PMR均值分别为92.5%、97.51%、98.78%,即GRCDM对属性数目无依赖,随属性数目增加判准率反而升高^[8],这是GRCDM特有的,还是非参数方法的特点,值得进一步明确。此外,未来研究应进一步关注非参数方法之间,以及非参数方法与参数方法之间的直接比较,以更进一步验证非参数方法的优势。

5 结论

本研究通过一个 5 因素混合实验,探讨了 4 个因素(属性层级、测验长度、样本容量、知识状态分布)对 HDD 的 3 种判别方法(R 方法、B 方法、W 方法)分类准确率的影响。结果表明:1) HDD 会受属性层级和测验长度影响。不同属性层级下的判准率由高到低依次为:直线型、收敛型、分支型、无结构型、独立型;测验长度越长, HDD 判准率越高,且与其它因素无关。2) HDD 对样本容量无依赖。100、500、1 000 人时, HDD 判准率几乎无差异,表明 HDD 无需大样本数据,可适用于小样本评估。3) HDD 的 R 方法、B 方法、W 方法的分类准确率无差异,实际应用中可任选一种进行。4) HDD 无需被试知识状态分布的正态性假设,更适于均匀分布。

6 参考文献

- [1] 胡竹菁. 心理统计学 [M]. 北京: 高等教育出版社, 2010.
- [2] Vapnik V. The nature of statistical learning theory [M]. New York: Springer Science & Business Media 2000.
- [3] Chiu Chia-yi, Douglas J A, Li X. Cluster analysis for cognitive diagnosis: theory and applications [J]. Psychometrika 2009, 74(4): 633-665.
- [4] 康春花, 任平, 曾平飞. 非参数认知诊断方法: 多级评分的聚类分析 [J]. 心理学报 2015, 47(8): 1077-1088.
- [5] Chiu Chia-yi, Douglas J. A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns [J]. Journal of Classification 2013, 30(2): 225-250.
- [6] 罗照盛, 李喻骏, 喻晓锋, 等. 一种基于 Q 矩阵理论朴素素的认知诊断方法 [J]. 心理学报 2015, 47(2): 264-272.
- [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] 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.
- [9] 康春花, 任平. 聚类诊断分析法诊断正确率的影响因素 [J]. 中国考试 2015(2): 25-32.
- [10] 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 Education Measurement 2004, 41(3): 205-237.
- [11] 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(3): 87-91.
- [12] 涂冬波, 蔡艳, 戴海琦, 等. 一种多级评分的认知诊断模型: P-DINA 模型的开发 [J]. 心理学报 2010, 42(10): 1 011-1 020.
- [13] 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(2): 249-262.
- [14] 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.
- [15] 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): 1 295-1 304.
- [16] 涂冬波, 蔡艳, 戴海琦. 基于 HO-DINA 模型的多级评分认知诊断模型的开发 [J]. 心理科学 2013, 36(4): 984-988.

The Influencing Factors of Diagnostic Accuracy Based on Hamming Distance Discrimination Method

KANG Chunhua, YANG Yakun, ZENG Pingfei*

(College of Teacher Education Zhejiang Normal University, Jinhua Zhejiang 321004, China)

Abstract: In order to explore the advantages of non-parametric methods in cognitive diagnostic assessment and better understand the properties of HDD, Monte Carlo simulation method is used to examine a variety of factors that influence diagnostic accuracy. Fixed the test measured attributes as five, a mixed-factor design is conducted to investigate the effect of attribute hierarchy, test length, sample size and knowledge state distribution on HDD three different discriminating methods (i. e., Random method, Bayes method and Weighted method) classification accuracy. And the results showed that both attribute hierarchy and test length have an impact on the classification accuracy of HDD method, the closer of the test's attributes relate and the longer of test length, the higher of the HDD method classification accuracy. And the sample size has little effect on HDD, so it is suitable for small scale implementation. There is no difference between three HDD methods of classification accuracy. Moreover, the normal distribution of knowledge state is not necessary when applying HDD to test D_{ata} , it performs a little better when the knowledge state is uniform.

Key words: Hamming distance discrimination method; attribute hierarchy; test length; sample size; knowledge state distribution

(责任编辑: 冉小晓)