文章编号:1000-5862(2018)01-0074-08

基于加权距离的一种认知诊断方法

罗　慧,熊建华,王晓庆,谭艳芳,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:马氏距离判别方法(MDD)是一种新的认知诊断分类方法,广义距离判别方法(GDD)和海明距离判别方法(HDD)为它的特例.使用香农熵作为马氏距离的权重矩阵,根据距离最小原则将被试的观察反应模式分类到理想反应模式,再由特殊的测验设计,将理想反应模式一一对应到知识状态上.蒙特卡洛模拟研究表明:在 0-1 评分模型下,选择模式匹配率和平均属性匹配率作为评价分类效果的标准,MDD 的分类效果好于 GDD 和 HDD.

关键词:广义距离判别方法;海明距离判别方法;Q 矩阵;香农熵;马氏距离
中图分类号:B 841.7　　文献标志码:A　　DOI:10.16357/j.cnki.issn1000-5862.2018.01.13

0　引言

认知诊断是多学科理论与技术融合的产物,能评估个体的认知结构,揭示个体内部心理加工过程.由于人的内部心理加工过程通常是不可直接观察的,因此对其测量、诊断、评估不容易,有关学者开发的具有认知诊断功能的心理计量模型(简称为认知诊断模型,也称为认知诊断方法)能够解决这一难题[1].据统计,已存在 100 多种认知诊断方法[2],不同的诊断方法其适用范围不同,性质和优缺点也不一样.研究者关注较多的模型有:规则空间模型(RSM)、属性层次模型(AHM)和确定性输入噪音"与门"模型(DINA)等[3],其中 DINA 是一个简约、易于解释和正确诊断率高的模型,从而对其研究成果也较多[4-6].另外,重新参数化统一模型(RUM)和缩减的重新参数化统一模型(R-RUM)也在国内外广泛流行[7].近年来,还涌现了许多隐式认知诊断方法,比如,基于 0-1 评分模型的有:广义距离判别方法(GDD)[8]和海明距离方法[8-10];基于多级评分模型的有:GDD-P[11]、多级评分的广义距离判别方法(GRM-GDD)[12]、多级评分的规则空间方法[13]和多级评分的聚类分析方法[14].

由于认知诊断方法种类众多,研究者们试图对某些方法进行整合,即给出一般化的认知诊断方法,再把某个方法看作是一般化方法的特例.一般来说,概括化程度越高、越抽象,对其本质认识越深刻.孙佳楠等[8]的 GDD,C. Y. Chiu 等[9]和罗照盛等[10]的海明距离判别方法(HDD)都具备简单易用且分类准确率高等特点.本文将它们的本质抽象出来,提出了马氏距离判别方法(mahalanobis distance discrimination,MDD),MDD 可以看作是 GDD 和 HDD 的一般化认知诊断方法,马氏距离判别方法使用不同的权重矩阵,则对应一种新的方法,当然对应的权重矩阵应该有较好的统计意义和较高的分类准确率,本文引入香农熵作为马氏距离的权重,探讨其分类效果.

1　广义距离判别方法和海明距离判别方法概述

本文的认知诊断测验使用 0-1 评分的项目,被试在一个项目上既不猜测也不失误的反应称为这个被试在这个项目上的理想反应,而在一组项目上的理想反应称为理想反应模式,用 IRP(ideal response pattern)表示;将被试在一个项目上的真实作答称为观察反应,在一组项目上的观察反应称为观察反应模式(observed response pattern,ORP),并假设 ORP 中没有缺失数据.

1.1　广义距离判别方法(GDD)

孙佳楠等[8]的 GDD 将 ORP 和 IRP 之间的相似

收稿日期:2017-08-19
基金项目:国家自然科学基金(31360237,31500909)和江西省教育厅科技计划(GJJ150356,GJJ160282)资助项目.
通信作者:甘登文(1955-),男,江西奉新人,教授,主要从事智能教育软件和应用统计的研究.E-mail:gdw120429@126.com

度用如下公式度量:

$$GD(\boldsymbol{Y}_i,\boldsymbol{I}_t) = \sum_{j=1}^{J} GD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)}), \quad (1)$$

其中 $GD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)}) = |\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)}| P_j(\theta_i)^{Y_{ij}} Q_j(\theta_i)^{1-Y_{ij}}$,公式(1)中 $j = 1,2,\cdots,J$,J 为项目数,$i = 1,2,\cdots,N$,N 为被试人数,$\boldsymbol{Y}_i = (\boldsymbol{Y}_{i1},\cdots,\boldsymbol{Y}_{iJ})$ 为第 i 个被试的 ORP,$\boldsymbol{I}_t = (\boldsymbol{I}_1^{(t)},\cdots,\boldsymbol{I}_J^{(t)})$ 为第 t 种 IRP,$GD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)})$ 表示在项目 j 上被试 i 的观察反应 $\boldsymbol{Y}_{ij}$ 与第 t 种理想反应模式中第 j 个分量 $\boldsymbol{I}_j^{(t)}$ 的广义距离. $P_j(\theta_i)$ 和 $Q_j(\theta_i)$ 分别表示能力水平为 θ_i 的被试在项目 j 上答对和答错的概率,并利用项目反应理论(IRT)中 2 参数逻辑斯蒂克模型(2PLM)[15] 的项目特征函数来将 $P_j(\theta_i)$ 和 $Q_j(\theta_i)$ 具体化. $GD(\boldsymbol{Y}_i,\boldsymbol{I}_t)$ 表示被试 i 在所有项目上的观察反应模式和第 t 个理想反应模式的广义距离之和,再按距离最短准则选择 ORP 对应的 IRP. 若对测验 Q 矩阵进行精心设计,即测验 Q 矩阵包含可达矩阵,在 0-1 评分、属性之间不存在补偿作用条件下,它使得 IRP 和知识状态一一对应[16-17],便可将 ORP 分类到这个 IRP 对应的知识状态上,从而实现对被试诊断分类. 研究表明,GDD 具有良好的诊断分类效果[8,18-19].

1.2　海明距离判别方法(HDD)

罗照盛等[10] 考察了 C. Y. Chiu 等[9] 的非参数认知诊断方法,结合可达阵在认知诊断测验编制中的重要作用[16-17],也提出了应用海明距离作判别的方法,即 HDD. 该方法使用海明距离定义被试的 ORP 与每种 IRP 之间的距离,依据距离最短的准则对被试进行诊断分类,当某个被试的 ORP 与多个 IRP 间的海明距离相同且均为最小值时,R 方法和 B 方法是有效的辅助手段. 被试 i 的 ORP 与第 t 种 IRP 之间的海明距离定义为

$$HD(\boldsymbol{Y}_i,\boldsymbol{I}_t) = \sum_{j=1}^{J} HD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)}), \quad (2)$$

其中 $HD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)}) = |\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)}|$,$J$、$\boldsymbol{Y}_i$ 和 $\boldsymbol{I}_t$ 的含义与 GDD 中一致,测验 Q 矩阵的设计也和 GDD 相同,在项目 j 上被试 i 的观察反应 $\boldsymbol{Y}_{ij}$ 与第 t 种理想反应 $\boldsymbol{I}_j^{(t)}$ 的第 j 个分量的海明距离用 $HD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)})$ 表示,$HD(\boldsymbol{Y}_i,\boldsymbol{I}_t)$ 表示被试 i 在所有 J 个项目上的海明距离之和. 因为 HDD 的定义中没有额外的参数,故不需要进行参数估计,操作步骤非常简单,在相同实验条件下与 GDD 相比,具有较高的分类准确率[10].

2　马氏距离判别方法(MDD)

2.1　马氏距离的定义

马氏距离是一种加权距离,它可以有效估算 2 个不同样本之间的相似度[20]. 在本研究中,被试 i 的观察反应模式 ORP 用向量 $\boldsymbol{Y}_i$ 表示,理想反应模式 IRP 用向量 $\boldsymbol{I}_t$ 表示,$\boldsymbol{W}_i$ 表示其权重矩阵(对角矩阵),只要 $\boldsymbol{W}_i$ 是一个正定阵,则 ORP 和 IRP 之间的距离就是马氏距离,具体公式如下:

$$d^2(\boldsymbol{Y}_i,\boldsymbol{I}_t) = (\boldsymbol{Y}_i - \boldsymbol{I}_t)^{\mathrm{T}} \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{I}_t),$$

若 $\boldsymbol{W}_i$ 表示方差-协方差矩阵的时候,$\boldsymbol{W}_i$ 写成逆矩阵的形式,而且这时候它不仅会考虑各种特性之间的关联,而且与测量单位无关,这是和欧氏距离的主要区别.

2.2　香农熵

在信息论中,熵表示某一随机事件不确定性的量度. 假设某事件可能有 n 种结果,每种结果发生的概率分布 $X = \{p_1,p_2,\cdots,p_n\}$ $(0 \leqslant p_i \leqslant 1, i = 1,2,\cdots,n)$ 且 $\sum_{i=1}^{n} p_i = 1$,则第 i 种结果自身提供的信息量为 $I_i = -\log p_i$,则该事件所有结果的平均信息量为

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i, \quad (3)$$

这里的平均信息量为香农熵. 假如 26 个字母中每个字母在文章中出现的次数相同,则每个字母提供的信息量为 $I_i = -\log \frac{1}{26}$,而所有字母的平均信息量为 $H(X) = -\sum_{i=1}^{26} \frac{1}{26} \log \frac{1}{26}$.

当概率分布 X 的不确定性程度越大时,其对应的香农熵 $H(X)$ 的值越大; 反之,熵的值越小,其典型性质有[21]:

(i) 对于概率分布 $X = \{1/n,1/n,\cdots,1/n\}$,有 $H(X) = \log n$;

(ii) 对于概率分布 $X = \{0,\cdots,0,1,0,\cdots,0\}$,有 $H(X) = 0$;

(iii) 对于任意概率分布 X,有 $0 \leqslant H(X) \leqslant \log(n)$.

由性质可知,当该事件的 n 种可能的结果以相同的概率(即 $1/n$)出现时,香农熵的值最大,不确定性也最大; 当该事件的其中一种结果以 1 的概率出现,其他 $n-1$ 种结果出现的概率为 0 时,香农熵的值最小,等于 0,唯一确定. 公式(3)中对数函数的底可以取不同的值,通过预研究发现,对于本研究采用以 e 为底的自然对数表示的香农熵计算结果更好,且与 3 个性质也仅仅相差一个常数因子,不影响诊断判别的结果,故下文使用自然对数表示的香农熵.

2.3　用马氏距离统一表示 GDD 和 HDD

将(1)式变换成 $MD(\boldsymbol{Y}_{ij},\boldsymbol{I}_j^{(t)}) = (\boldsymbol{Y}_{ij} -$

$\boldsymbol{I}_j^{(t)})^2 P_j(\theta_i)^{Y_{ij}} Q_j(\theta_i)^{1-Y_{ij}} = (\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)})^{\mathrm{T}} P_j(\theta_i)^{Y_{ij}} Q_j(\theta_i)^{1-Y_{ij}} \cdot (\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)})$,便可将 GDD 表示成马氏距离的表达形式:

$$GD(\boldsymbol{Y}_i, \boldsymbol{I}_t) = (\boldsymbol{Y}_i - \boldsymbol{I}_t)^{\mathrm{T}} \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{I}_t). \quad (4)$$

(4) 式中的权重矩阵为

$$\boldsymbol{W}_i = \begin{pmatrix} P_1(\theta_i)^{Y_{i1}} Q_1(\theta_i)^{1-Y_{i1}} & & 0 \\ & \ddots & \\ 0 & & P_J(\theta_i)^{Y_{iJ}} Q_J(\theta_i)^{1-Y_{iJ}} \end{pmatrix}.$$

同样,(2) 式进行简单变换,即可得 $HD(\boldsymbol{Y}_{ij}, \boldsymbol{I}_j^{(t)}) = (\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)})^2 = (\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)})^{\mathrm{T}} (\boldsymbol{Y}_{ij} - \boldsymbol{I}_j^{(t)})$,则 HDD 的马氏距离表达形式为

$$HD(\boldsymbol{Y}_i, \boldsymbol{I}_t) = (\boldsymbol{Y}_i - \boldsymbol{I}_t)^{\mathrm{T}} \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{I}_t), \quad (5)$$

(5) 式中的权重矩阵 $\boldsymbol{W}_i = \mathbf{E}$(单位矩阵).

由(4) 式和(5) 式可见,GDD 和 HDD 的本质都是马氏距离,只是对马氏距离中权重矩阵进行了不同的定义而得到的,因此马氏距离判别方法是一种

$$\boldsymbol{W}_i = \begin{pmatrix} -P_1(\theta_i)\ln P_1(\theta_i) - Q_1(\theta_i)\ln Q_1(\theta_i) & & 0 \\ & \ddots & \\ 0 & & -P_J(\theta_i)\ln P_J(\theta_i) - Q_J(\theta_i)\ln Q_J(\theta_i) \end{pmatrix}.$$

在表达式 $H(X_{ij}) = -P_j(\theta_i)\ln P_j(\theta_i) - Q_j(\theta_i) \cdot \ln Q_j(\theta_i)$ 中,当 $P_j(\theta_i)$ 取值为 1 或 0 时,即肯定答对或答错时,香农熵的值为 0,达到最小,唯一确定;而当 $P_j(\theta_i) = 1/2$ 时,即答对和答错的概率相等时,香农熵的值达到最大,最不确定;其他的情况介于两者之间. 将香农熵作为权重,这点和(4) 式中的 GDD 有很大不同,因为要将 ORP 判别到使马氏距离最小的 IRP 上时,不仅要考虑 ORP 与 IRP 之间的接近程度,还要考察 ORP 的确定性程度,即使 ORP 与某个 IRP 接近,但 ORP 的不确定性大,也认为它们之间的"接近"是不可靠的. 由于采用和 GDD 中相同的认知诊断测验设计,IRP 能够一一对应到知识状态,就可以达到对被试诊断分类的目的.

另外,这里 $P_j(\theta_i)$ 和 $Q_j(\theta_i)$ 考虑 2 种定义,一种是与 GDD 中的定义一致,即以 IRT 下 2PLM 为例,$P_j(\theta_i)$ 和 $Q_j(\theta_i)$ 分别表示能力水平为 θ_i 的被试答对和答错项目 j 的概率;另一种定义是使用经典测量理论(CTT) 下的通过率来表示,这种定义计算简单,用 P_j 和 Q_j 分别表示在项目 j 上的通过率和未通过率.

2.5 应用马氏距离做诊断

计算某个被试 i 的 ORP 到所有 IRP 的马氏距离,根据距离最短的原则找到使马氏距离最小的那个 IRP,在 IRP 与知识状态一一对应的前提下,便可更一般化的认知诊断方法.

2.4 定义反应模式间的马氏距离

从马氏距离定义的角度,导出更一般化的 ORP 和 IRP 之间的马氏距离. 定义被试 i 的 ORP 与第 t 种 IRP 之间的马氏距离,并将被试 i 分类到使马氏距离最小的 IRP 上,公式如下:

$$MD(\boldsymbol{Y}_i, \boldsymbol{I}_t) = (\boldsymbol{Y}_i - \boldsymbol{I}_t)^{\mathrm{T}} \boldsymbol{W}_i (\boldsymbol{Y}_i - \boldsymbol{I}_t).$$

$MD(\boldsymbol{Y}_i, \boldsymbol{I}_t)$ 定义了 $\boldsymbol{Y}_i$ 到 $\boldsymbol{I}_t$ 的马氏距离,$\boldsymbol{Y}_i$ 和 $\boldsymbol{I}_t$ 的含义与 GDD 中一致,测验 Q 矩阵的设计也与 GDD 相同. 并将香农熵作为权重矩阵(对角矩阵) 的对角元,由于本文中的项目都采用 0-1 评分,故香农熵中的可能结果只有 2 种,即答对概率(P) 和答错概率($Q = 1 - P$),因此某个被试 i 在某个项目 j 上使用自然对数的香农熵表示为 $H(X_{ij}) = -P_j(\theta_i)\ln P_j(\theta_i) - Q_j(\theta_i)\ln Q_j(\theta_i)$,而被试 i 在所有项目 J 上的香农熵可以用对角矩阵表示为

将被试 i 分类到这个 IRP 对应的知识状态上.

3 蒙特卡洛模拟研究

本研究主要考查认知诊断模型(3 水平)(MDD 方法、GDD 和 HDD 方法)、属性层级关系(5 水平)(直线型、收敛型、发散型、无结构型和独立型) 和失误概率(5 水平)(0.02,0.05,0.10,0.15 和 0.2) 的影响,总共有 $3 \times 5 \times 5 = 75$ 种处理水平进行实验. 为了结论的稳定性和减少实验误差,每种实验条件下的模拟次数为 50 次,具体实验设计如下.

3.1 测验 Q 阵的设计

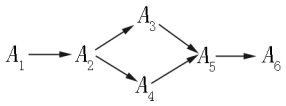
为了将 MDD 方法与 GDD 和 HDD 方法在同等条件下进行比较,本研究的实验条件与文献[8,10]相同. 主要考查的基本属性层级结构包括:直线型(linear)、收敛型(convergent)、发散型(divergent)、无结构型(unstructured) 和独立型(independent) 共 5 种(见图 1),其他更加复杂的属性层级结构可以由这 5 种基本层级结构复合而成. 当属性之间不存在补偿作用、测验 Q 矩阵包含可达矩阵时,则知识状态与 IRP 能够一一对应. 依照属性之间的层级结构,得到 5 种属性层级结构下的典型项目考核形式,它们分别为直线型 6 个项目、收敛型 7 个项目、发散型 15 个项目、无结构型 32 个项目和独立型 63 个项

目. 为了保证参数估计的精度和各种层级结构下的测验长度基本一致,将 5 种典型项目考核形式分别重复出现在测验中的次数设为 5 次、5 次、2 次、1 次和 1 次,其中独立型层级关系下,由于测验长度的限制,将其典型项目考核形式按属性个数从小到大排序并取前30 个项目. 因此,本研究中5 种属性层级关系的测验项目个数分别是 30、35、30、32 和 30 个(见图 2).

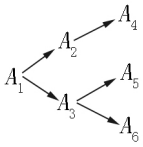
直线型(Linear)

$$A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow A_4 \rightarrow A_5 \rightarrow A_6$$

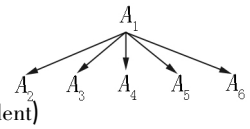
收敛型(Convergent)



发散型(Divergent)



无结构型(Unstructured)



独立型(Independent)

$$A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \quad A_6$$

图 1　5 种基本的属性层级关系

直线型(Linear)

$$Q = \begin{pmatrix} 111111111111111111111111111111 \\ 011111011111011111011111011111 \\ 001111001111001111001111001111 \\ 000111000111000111000111000111 \\ 000011000011000011000011000011 \\ 000001000001000001000001000001 \end{pmatrix}$$

发散型(Divergent)

$$Q = \begin{pmatrix} 111111111111111111111111111111 \\ 010100111111011010100111111011 \\ 001011111111111001011111111111 \\ 000100000111001000100000111001 \\ 000010010010111000010010010111 \\ 000001001001111000001001001111 \end{pmatrix}$$

独立型(Independent)

$$Q = \begin{pmatrix} 100000111110000000000111111111 \\ 010000100001111000000111100000 \\ 001000010001000111000100011100 \\ 000100001000100100110010010011 \\ 000010000100010010101001001010 \\ 000001000010001001011000100101 \end{pmatrix}$$

收敛型(Convergent)

$$Q = \begin{pmatrix} 111111111111111111111111111111 \\ 011111011111011111011111011111 \\ 001011100101110010111001011100 \\ 000111100011110001111000111100 \\ 000011000001100000110000011000 \\ 000001000000100000010000001000 \end{pmatrix}$$

无结构型(Unstructured)

$$Q = \begin{pmatrix} 111111111111111111111111111111 \\ 010000111100011100110011010101 \\ 001000100011111100001111001100 \\ 000100010010010011111111100001 \\ 000010001001001010101010111111 \\ 000001000100100101010101111111 \end{pmatrix}$$

图 2　5 种基本属性层级关系的测验 Q 矩阵

3.2　被试观察反应模式的模拟

本研究中,在各种实验条件下均采用 1 000 人的被试样本容量,模拟 ORP 包括以下步骤:(i) 分别计算 5 种属性层级关系下被试的理想掌握模式(ideal master pattern, IMP)(又称为知识状态),分别为直线型 7 种、收敛型 8 种、发散型 16 种、无结构型 33 种和独立型 64 种. (ii) 结合对测验 Q 阵的设计,通过 IMP 和测验 Q 矩阵求取相应的 IRP,对每种 IRP 求取其对应的测验总分,对总分从小到大排序,为了使这些对应总分的人能符合标准正态分布,模拟 1 000 人并按比例分配到各 IRP 上,并给测验总分相同的 IRP 分配相同的被试人数,便可得到各种属性层级关系下 1 000 个被试的理想掌握模式 IMP,在没有任何失误的前提下,计算被试的 IRP. (iii) 根据计算得到的 IRP,在不同失误概率($slip$) 下(即0.02, 0.05,0.10,0.15 和 0.2),模拟产生被试的 ORP[22].

3.3　判别准确性的衡量

采用 2 种指标来衡量判别准确性,即模式匹配率(PMR) 和平均属性匹配率(AAMR)[8,10,19]. 二者公式为

$$PMR = \sum_{i=1}^{N} N_{i_correct} / N,$$

$$AAMR = \sum_{i=1}^{N} \sum_{k=1}^{K} N_{ik_correct} / (K \times N),$$

其中 N 为被试总人数,K 为属性个数,$N_{i_correct}$ 表示被试 i 估计的知识状态与模拟产生的真实的知识状态是否完全一致,完全一致为 1,否则为 0. $N_{ik_correct}$ 表示被试 i 估计的知识状态的某个分量与模拟产生的真实的知识状态的对应分量是否一致,一致为 1,否则为 0.

3.4　认知诊断方法的比较

本文比较了新提出的 MDD 与 GDD 和 HDD 在

模式匹配率和平均属性匹配率上的差异. 对于 MDD 和 GDD 这 2 种方法,权重矩阵的对角元中均涉及到概率参数 P,对于 P 的获得,设计了 2 种方案,一种是基于 IRT 的方法,将 1 000 名被试的 ORP 和所有 IRP 组合起来估计项目参数,使用被试的 ORP 估计能力参数,再根据估计出的参数使用 IRT 模型中 2PLM 的项目特征函数计算正确作答概率 P 和错误作答概率 Q[8];另一种是基于经典理论的方法,即直接使用 1 000 名被试的观察反应模式 ORP 来计算通过率 P 和未通过率 Q. 为了探查不同的参数计算方法对判别结果的影响,故表示为如下 4 种组合: MDD-CTT,MDD-IRT,GDD-CTT,GDD-IRT. 另外,对于 HDD 方法,在文献[10]的结论中指出,HDD 的 B 方法分类效果更好,故本研究只将 HDD 的 B 方法拿来比较,简称为 HDDB,在每种实验条件下均使用以上认知诊断方法对其进行判别分类.

4 研究结果

从表 1 和表 2 的数值可知,和 GDD 或者 HDD 相比,本研究提出的 MDD 方法的表现均为最优的(*PMR* 和 *AAMR* 值较大),而 GDD 和 HDDB 的方法则各有千秋,在直线、收敛和无结构型层级关系下,HDDB 方法优于 GDD 方法,在发散、独立型层级关系下,HDDB 方法介于 GDD-IRT 和 GDD-CTT 之间或比两者都差. MDD-CTT 和 MDD-IRT 是 MDD 方法中因采用不同参数估计方法得到的 2 种不同形式,但它们的 *PMR* 和 *AAMR* 都比较接近,在直线、收敛型层级关系下,MDD-IRT 略优于 MDD-CTT,而在独立型层级关系下 MDD-CTT 略优于 MDD-IRT,在其他层级关系下不分上下. 对于 GDD-CTT 和 GDD-IRT 之间也存在相同的结果.

图 3 和图 4 的折线显示,在同一种层级关系下,当 *slip* 增大时,每种方法的 *PMR* 和 *AAMR* 都呈下降趋势,其中 MDD 方法下降最慢,而其他 2 种方法则各有短长,在直线、收敛型层级关系下,GDD 方法下降较快,在发散、独立型层级关系下,HDD 方法下降较快,无结构层级关系下两者相当.

表 1 各认知诊断方法的模式匹配率的比较

属性层级关系类型	失误概率(*slip*)	模式匹配率(*PMR*)				
		MDD-CTT	MDD-IRT	GDD-CTT	GDD-IRT	HDDB
直线型	0.02	0.999 9	0.999 9	0.980 8	0.998 2	0.999 9
	0.05	0.998 0	0.998 0	0.969 1	0.988 7	0.998 0
	0.10	0.985 6	0.985 7	0.924 2	0.967 4	0.985 2
	0.15	0.952 1	0.953 4	0.879 4	0.928 4	0.950 2
	0.20	0.893 6	0.895 9	0.812 6	0.875 6	0.886 9
收敛型	0.02	0.999 8	0.999 8	0.980 8	0.997 3	0.999 8
	0.05	0.998 0	0.998 1	0.966 7	0.987 7	0.998 0
	0.10	0.984 4	0.984 7	0.916 7	0.970 3	0.984 3
	0.15	0.952 4	0.953 8	0.870 2	0.930 7	0.950 9
	0.20	0.893 7	0.898 2	0.817 9	0.883 6	0.888 7
发散型	0.02	0.986 3	0.986 7	0.985 4	0.970 7	0.981 3
	0.05	0.963 3	0.961 8	0.960 5	0.928 2	0.945 6
	0.10	0.908 3	0.907 6	0.904 7	0.870 0	0.869 5
	0.15	0.841 6	0.842 6	0.828 9	0.809 6	0.783 6
	0.20	0.749 4	0.753 7	0.720 5	0.734 8	0.673 7
无结构型	0.02	0.955 3	0.955 0	0.947 4	0.944 9	0.945 7
	0.05	0.892 7	0.900 7	0.878 1	0.874 9	0.880 3
	0.10	0.803 6	0.822 3	0.782 7	0.773 3	0.792 8
	0.15	0.717 6	0.736 7	0.686 4	0.688 7	0.710 9
	0.20	0.640 3	0.658 2	0.599 3	0.629 8	0.635 3
独立型	0.02	0.988 7	0.978 6	0.988 8	0.959 5	0.984 4
	0.05	0.968 4	0.956 4	0.967 0	0.899 5	0.956 6
	0.10	0.907 4	0.902 3	0.904 5	0.806 5	0.880 3
	0.15	0.824 8	0.819 3	0.814 7	0.723 3	0.777 7
	0.20	0.704 3	0.702 5	0.691 7	0.642 7	0.643 1

表 2　各认知诊断方法的平均属性匹配率的比较

属性层级关系类型	失误概率(*slip*)	平均属性匹配率(*AAMR*)				
		MDD-CTT	MDD-IRT	GDD-CTT	GDD-IRT	HDDB
直线型	0.02	1.000 0	1.000 0	0.996 7	0.999 7	1.000 0
	0.05	0.999 7	0.999 7	0.994 4	0.998 1	0.999 7
	0.10	0.997 5	0.997 5	0.986 0	0.994 4	0.997 1
	0.15	0.991 1	0.991 6	0.977 6	0.987 2	0.989 6
	0.20	0.979 0	0.979 7	0.962 9	0.976 2	0.974 7
收敛型	0.02	1.000 0	1.000 0	0.996 7	0.999 6	1.000 0
	0.05	0.999 7	0.999 7	0.994 0	0.997 9	0.999 6
	0.10	0.997 3	0.997 4	0.985 0	0.995 0	0.997 1
	0.15	0.991 3	0.991 8	0.976 1	0.987 7	0.989 8
	0.20	0.979 6	0.981 1	0.964 5	0.978 6	0.974 9
发散型	0.02	0.997 6	0.997 7	0.997 2	0.994 8	0.995 6
	0.05	0.993 2	0.993 0	0.992 3	0.987 1	0.986 1
	0.10	0.981 4	0.981 5	0.980 5	0.975 2	0.961 8
	0.15	0.965 5	0.966 2	0.963 8	0.960 1	0.931 1
	0.20	0.941 1	0.942 6	0.939 0	0.941 3	0.889 1
无结构型	0.02	0.990 3	0.991 6	0.989 9	0.988 6	0.990 5
	0.05	0.975 9	0.980 2	0.974 8	0.973 9	0.977 2
	0.10	0.951 9	0.961 4	0.950 1	0.950 6	0.955 0
	0.15	0.925 7	0.937 7	0.922 1	0.927 0	0.924 4
	0.20	0.899 3	0.912 8	0.894 6	0.907 4	0.890 5
独立型	0.02	0.997 6	0.996 2	0.997 9	0.992 7	0.994 6
	0.05	0.993 1	0.991 8	0.993 3	0.981 1	0.984 7
	0.10	0.978 1	0.979 1	0.978 7	0.959 9	0.954 5
	0.15	0.955 8	0.957 7	0.955 5	0.936 1	0.910 1
	0.20	0.921 6	0.925 6	0.921 6	0.908 4	0.850 2

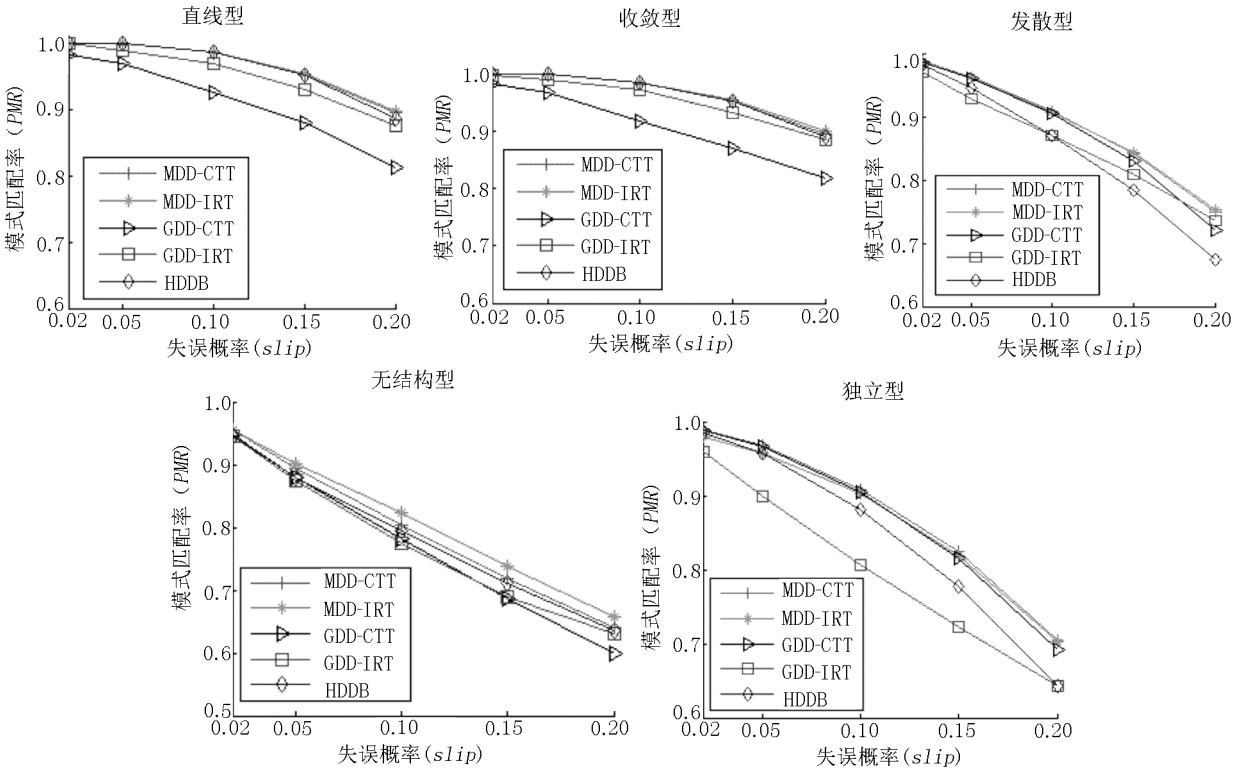


图 3　各认知诊断方法的模式匹配率的比较

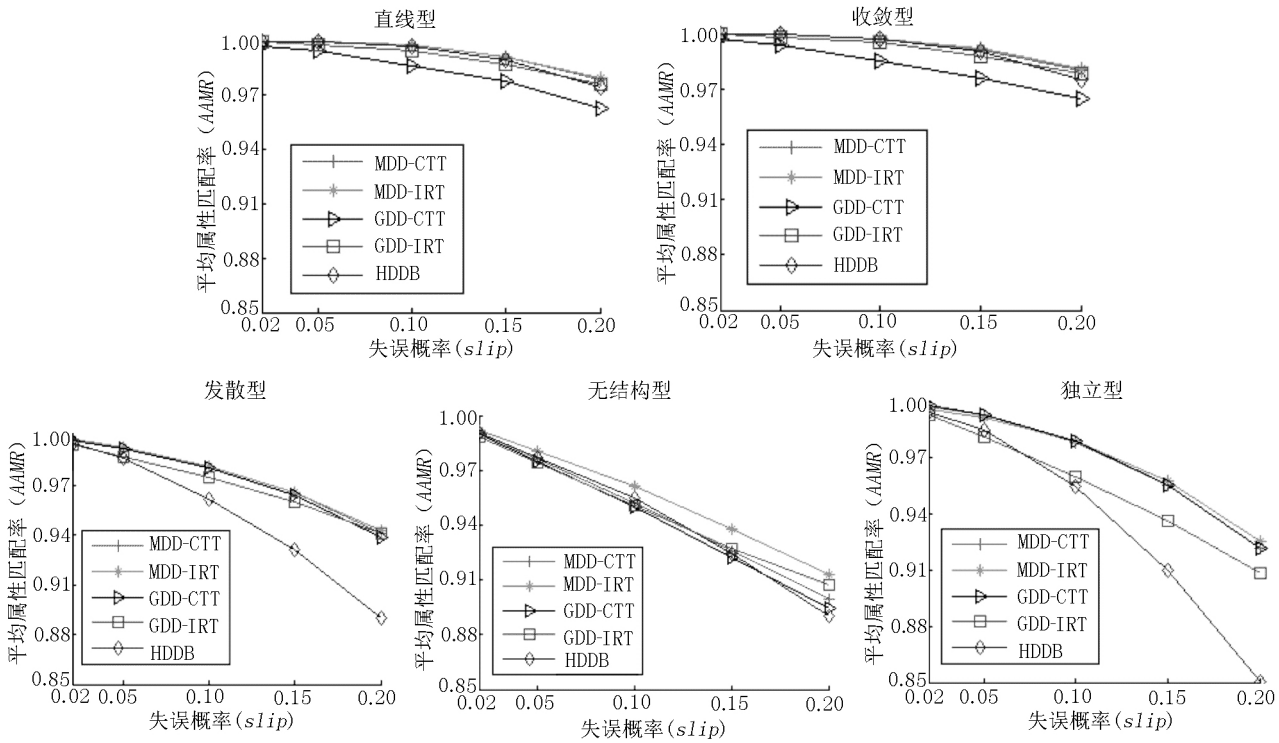


图 4 各认知诊断方法的平均属性匹配率的比较

5 小结与讨论

1) 属性层级结构对各种认知诊断方法的分类准确率均有影响,在 5 种基本结构中,直线型和收敛型准确率最高,其次是发散性和独立型,无结构型最低,因此认知诊断中属性层级关系的认定非常重要.

2) 认知诊断方法的分类准确率还受失误概率(*slip*)的影响,*slip* 越小,分类越准确,*slip* 越大,分类准确率越低. 而本研究讨论的 3 种认知诊断方法中,分类准确率随 *slip* 下降最慢的是 MDD,其次是 GDD,而下降最快的方法是 HDDB. 但由于不同方法中对距离的权重定义不同而导致了不同的分类准确率,比如 HDDB 方法中权重矩阵采用单位阵,虽然计算简单,但也损失了项目参数这部分信息,所以当失误率增大时,ORP 与 IRP 之间的差异就越明显,导致分类准确率下降迅速; 而 MDD 方法和 GDD 方法中权重矩阵的定义均和项目参数有关,从而对 ORP 与 IRP 之间的距离有一定的修正作用,因此下降较慢.

3) 在 3 种认知诊断方法中,分类准确率排在前两位的是 MDD-IRT 和 MDD-CTT,而且 MDD-CTT 和 MDD-IRT 相比,分类准确率只有略微的差别,并且使用 IRT 方法计算项目参数和能力参数有前提条件,比如需要的被试样本量较大,一般 1 000 人以上,当样本量较小时,比如少于 500 人,参数估计精度较差; 其次,估计方法也比较复杂. 而通过 CTT 方法计算通过率直观简单,对被试人数的要求也相对较少,因此即使样本量较小,马氏距离判别法(MDD-CTT)依然可以得到比较好的结果.

4) 本文引入的马氏距离判别方法实际上和认知诊断测验的 Q 矩阵设计方法相结合,特别重视可达阵的作用. 另外,在 Q 矩阵设计中属性平衡也很重要,但是对于某些属性层级关系,很难做到属性的平衡,比如线型、收敛型和发散型,若要考察下层某个节点属性,上层节点属性必须在考察范围内,因此顶点属性和上层节点属性考察的频率较高; 对于无结构型,除了顶点属性频次高之外,其他属性均匀出现 16 次; 对于独立型层级结构,由于典型项目考核模式有 63 种,而测验长度受限制,在预研究时也尝试过不同的 Q 矩阵设计,比如只使用 1 个属性的项目(各属性平衡性最好),按属性个数从小到大选择前 30 个项目(各属性平衡性较好),随机选择 30 个典型项目考核模式(各属性平衡性较差)等,随着属性平衡性的下降,各方法的判准率也呈下降趋势,但在同等条件下,马氏距离判别法在各方法中依然是表现最好的.

5) 由于知识状态是潜在向量,即不可直接观察,所以一般对新开发的认知诊断模型的表现(特别是对知识状态的判准率)的评估,都要进行模拟

实验. 若在能较好地控制实验条件的模拟实验中,新方法的认知诊断判准率都不理想,则对于千变万化的实际测验获得的数据,很难有好的表现. 但是,除模拟之外,应尽量对实测数据进行分析研究,本文受条件限制,没有对实测数据进行分析,这是一个遗憾.

6 展望

1) 在独立和无结构层级关系中,每种认知诊断方法的 *PMR* 都不太高,特别是当失误概率较大时,*PMR* 较低,考虑通过测验 *Q* 矩阵的设计来提高判准率.

2) 本研究用香农熵作为马氏距离的权重矩阵(对角阵)的对角元素,取得了较好的分类效果,可考虑定义不同的权重矩阵,即可得到不同的分类方法. 对于收集到的实际数据集,可以通过几个模型拟合指标,比如相对拟合指标 AIC 和 BIC 或者绝对拟合指标进行拟合,从而遴选最合适的权重矩阵来进行判别分类.

3) 本研究对已有的 GDD 方法和 HDD 方法抽象出它们的本质,并构造了更具一般性的马氏距离判别法,并在 0-1 评分项目上通过模拟手段探索了其效果,下一步应该将其应用于实测数据,进行诊断分析,进一步考察其性质; 同时还可以研究将该方法推广到多级评分测验中.

7 参考文献

[1] 涂冬波,蔡艳,丁树良. 认知诊断理论方法、方法与应用 [M]. 北京: 北京师范大学出版社,2012.

[2] 辛涛,乐美玲,张佳慧. 教育测量理论新进展及发展趋势 [J]. 中国考试,2012(5):3-11.

[3] 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.

[4] 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):1011-1013.

[5] de la Torre J. The generalized DINA model framework [J]. Psychometrika,2011,76(2):179-183.

[6] 张淑梅,包钰,郭文海. 一种多级评分的广义认知诊断模型 [J]. 心理学探新,2013,33(5):444-450.

[7] 汪文义,宋丽红. 教育认知诊断评估理论与技术研究 [M]. 北京: 北京师范大学出版集团,2015.

[8] 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.

[9] 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.

[10] 罗照盛,李喻骏,喻晓锋,等. 一种基于 *Q* 矩阵理论朴素的认知诊断方法 [J]. 心理学报,2015,47(2):264-272.

[11] 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.

[12] 李娟,丁树良,罗芬. 基于等级反应模型的广义距离判别法 [J]. 江西师范大学学报: 自然科学版,2012,36(6):636-639.

[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(1):249-262.

[14] 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.

[15] 漆书青,戴海琦,丁树良. 现代教育与心理测量学原理 [M]. 北京: 高等教育出版社,2002.

[16] 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.

[17] 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.

[18] 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.

[19] 涂冬波,蔡艳,戴海琦. 几种常用非补偿型认知诊断模型的比较与选用: 基于属性层级关系的考量 [J]. 心理学报,2013,45(2):243-252.

[20] 张尧庭. 多元统计分析选讲 [M]. 北京: 中国统计出版社,2002:42-50.

[21] 吴成茂. 基于加权香农熵的图像阈值法 [J]. 计算机工程与应用,2008,44(18):177-180.

[22] 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.

(下转第 88 页)

cal performance of aligned electrospun polyimide nanofiber belt at high temperature [J]. Materials Letters, 2015, 40: 12-15.

[23] Tamai S, Oikawa H, Ohta M, et al. Melt processable copolyimides based on 4,4′-bis (3-aminophenoxy) biphenyl [J]. Polymer, 1998, 39(10): 1945-1949.

The Synthesis and Characterization of Polyimides from A Novel Synthesized 3,3′-(*m*-Phenylene) Dianhydride Monomer

XU Wenhui, YANG Ting, YU Ying, ZHANG Chunwu, DING Yichun, HOU Haoqing*

(College of Chemistry and Chemical Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: 3,3′-(*m*-phenylene) dianhydride (3,3′-RsDPA) monomer is synthesized by using 3-chlorophthalic anhydride and 1,3-benzenediol as raw materials, and the reaction time and temperature are optimized for the synthesis. Subsequently, a series of polyimide (PI) molding powder are prepared by the polymerization of the 3,3′-RsDPA with 4,4′-oxydianiline (4,4′-ODA), 4,4′-(1,3-phenylenedioxy) dianiline (TPEQ), 1,4-diaminobenzene (PDA), and 1,3-diaminobenzene (MPD), respectively, followed with chemical imidization. During the polymerization, phthalic anhydride (PA) is used as the blocking agent. The thermal and mechanical properties of the synthesized polyimides are characterized. Results show that the polyimides exhibit excellent thermal stabilities with the 5% weight loss temperatures ($T_{d5\%}$) being 525 ~ 531 ℃ and 526 ~ 538 ℃, respectively under air and nitrogen atmosphere. With the increase of the rigidity of diamine monomer, the glass transition temperature (T_g) increases from 218 ℃ to 261 ℃. When the diamine monomer is PDA, PI (3,3′-RsDPA-PDA) shows obvious melting crystallization behavior, and the melting temperature (T_m) is 327 ℃. The excellent thermal property and good processability enables the synthesized polyimide powders to have potential application in 3D printing technology.

Key words: 3,3′-RsDPA; synthesis; polyimide; 3D printing; melt crystallization

(责任编辑: 刘显亮)

(上接第 81 页)

The Generalized Cognitive Diagnosis Method Based on Weighted Distance

LUO Hui, XIONG Jianhua, WANG Xiaoqing, TAN Yanfang, GAN Dengwen*

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

Abstract: Cognitive diagnostic model (CDM) is an important part of cognitive diagnosis, the main aim for CDM is to discriminate examinees into different classes. Although there existed a lot of CDMs, researchers proposed many new CDMs still. Among them, the generalized distance discrimination (GDD) and the hamming distance discrimination (HDD) have some advantages, such as, simple and easy to use, high classification accuracy, thus receive more and more attention. Mahalanobis distance discrimination (MDD) is a generalized CDM, GDD and HDD are the special cases. Mahalanobis Distance (MD) is employed for MDD to calculate the distance between an examinee's Observed Response Pattern (ORP) and all kinds of Ideal Response Pattern (IRP), and specifies the Shannon Entropy as weight. According to the principle of minimum distance and special test design, IRP can be mapped one-to-one to the state of knowledge. Under binary scoring model, the pattern match ratio and average attribute match ratio were selected as the criteria for evaluating the classification accuracy, the Monte Carlo simulation study show that the performance of MDD was better than GDD and HDD.

Key words: GDD; HDD; Q matrix; shannon entropy; Mahalanobis distance

(责任编辑: 冉小晓)