

文章编号: 1000-5862(2019)01-0059-09

多级计分的 CD-CAT 曝光控制方法

高椿雷, 罗照盛, 丁树良

(江西师范大学初等教育学院, 江西 南昌 330022)

摘要: 相对于 0-1 计分, 多级计分考察的认知能力水平更高, 能够提供更丰富的诊断信息, 同教学实践更相符, 而现有研究中鲜有基于多级计分的 CD-CAT. 该文提出了 5 种新的曝光控制的选题策略, 分别是 BSCDI_A、BSCDTID_A、BSCDI_D、BSCDTID_D 和 RP_PWCID. 将这 5 种方法同已有的 PWKL 和 RP_PWKL 进行比较, 并尝试将这 7 种选题策略从 0-1 计分拓展到多级计分. 研究结果表明: 不管是 0-1 计分还是多级计分, 采用统一量纲后的综合指标进行评价时 RP_PWCID($\beta = 0.5$) 表现最好, 题库使用均匀性各项指标均是最优. 区分度分层方法(BSCDI_A、BSCDTID_A、BSCDI_D、BSCDTID_D) 控制曝光的效果略差于 RP 法. 在多级计分下 β 值仍然具有调节判断率和题库使用的作用.

关键词: P-DINA 模型; CD-CAT; 0-1 计分; 多级计分; 曝光控制选题方法

中图分类号: B 841 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2019.01.11

0 引言

认知诊断评估(cognitive diagnosis assessment, CDA)作为形成性评价(formative assessment)改变了传统考试的目的和意义. 传统的总结性评价(summary assessment)试图通过考试达到排序、晋升等目的, 将学生和考试放在对立的位置. 随着测验的发展, 人们不满足于单纯提供单一的成绩, 希望精细地了解学生的知识掌握情况, 获得更丰富的诊断信息. CDA 满足了该功能, 它是教师和学生用于改进教学、自我学习的工具, 服务于当前的知识学习. 当然, 现阶段研究者也尝试开发传统的项目反应理论模型的诊断功能, 并取得了良好的效果^[1].

随着 CDA 的迅速发展, 其发展趋势已经从理论的研究转向了具有实践的可实施性的方向. 认知诊断同计算机化自适应测验的结合(cognitive diagnosis computerized adaptive testing, CD-CAT)是 CDA 最成功的发展和应用^[2-8]. CD-CAT 同时具有认知诊断和自适应的特点, 可以使用较少的题目在较短的时间内评价出考生知识状态的掌握情况, 这种即时的评价适用于课堂练习, 为教师的补救教学、学生的自我学习提供依据.

由于形成性评价的多样性要求, 近些年认知诊断模型(cognitive diagnostic model, CDM)发展迅速^[9-12]. 依据项目的计分方式不同, 将 CDM 划分为 0-1 计分模型和多级计分模型, 如 DINA(the deterministic input noisy output “AND” gate model)模型^[13-14]、DINO(the deterministic input noisy output “OR” gate model)模型^[15]、rRUM(reduced reparameterized unified model)模型^[16]等是 0-1 计分模型. GDM(generalized diagnosis model)^[17]、P-DINA^[1]、PC-DINA(partial credit DINA)^[18]和 sequential cognitive diagnosis model^[19]等是现有的为数不多的多级计分模型. 在教育考试实践中多级计分应用广泛, 如综合分析题、作文题和应用题等. 多级计分题相较于 0-1 计分题目能够提供更多的信息, 更贴合教育实践. 在心理测评领域, 多级计分的利用率更高, 如李克特量表就是典型的多级计分方法. 涂冬波等^[1]将多级计分模型和 DINA 模型相结合提出的 P-DINA 模型适用于各种评分的数据资料, 且判断率较高, 具有较好的借鉴和应用价值.

现有的 CD-CAT 选题策略着重于提高测验的判断率, 如 PWKL(posterior-weighted Kullback-Leibler index)^[20]、MPWKL(the modified PWKL index)、GDI(the generalized deterministic inputs noisy “and” gate

收稿日期: 2018-09-08

基金项目: 国家自然科学基金(31660279)和江西省社会科学规划(16JY36)资助项目.

作者简介: 高椿雷(1983-), 女, 山东烟台人, 讲师, 博士, 主要从事心理与教育测量评价. E-mail: gaochunlei51@qq.com

model discrimination index)^[3]、SHE(Shannon entropy method)^[21]、互信息(mutual information)^[22]等,判准率提高的同时,测验曝光率较高,题库使用均匀性差,部分项目反复使用,出现项目漂移。CD-CAT 题库开发者在编写项目时必须同复杂的认知诊断测验蓝图密切联系^[23],这增加了项目编写的难度,其题库的建设相对于其他测验方式需要花费更多的人力物力。因此提出效果良好的项目曝光控制选题策略对于 CD-CAT 有较大价值。

针对该问题, Wang Chun 等^[23]提出了 RP(restrictive progressive)法和 RT(restrictive threshold)法,其主要思想是在选择项目的过程中,在原有的信息量基础上增加随机元素,用于平衡项目的曝光率并保证对判准率损失最小。陈平^[24]提出了最大修正优先指标(Maximum Modified Priority Index, MMPI)用于控制 CD-CAT 的曝光。MMPI 法和 RP 法有相似之处,这 2 种方法都是在 MPI(maximum priority index)法^[25]的基础上进行改进和拓展,其中 RP 法增加了进度控制(progressive control)部分。研究结果表明,RP 法在较小损失判准率的情况下能够有效改善题库使用的均匀性^[23]。毛秀珍等^[26]将传统 CAT 中使用区分度分层来控制曝光的思想应用于 CD-CAT,通过 CDM 的区分度指标 CDI(cognitive diagnosis index)进行分层后使用现有选题策略如 PWKL 进行选题。

Wang Chun 等^[23]只将 RP 法用于 PWKL。为了考察 RP 法同其它选题策略结合后的效果,本文将 RP 法同 Zheng Chanjin^[27]提出的 PWCDI(PW cognitive diagnosis index)法相结合,并比较 RP_PWCDI 和 RP_PWKL 的曝光控制效果。毛秀珍等^[26]使用区分度分层时,只采用了 CDI 一种区分度指标,分层后采用 PWKL 选题效果一般。现有 CDM 的区分度计算方法有 CDI 和 CTTID(classical test theory item discrimination),因此本文使用了 CDI 和 CTTID 这 2 种区分度方法分层,然后采用二分搜索(binary searching, BS)选题。现有曝光控制方法是应用于 0-1 计分的 CD-CAT,多级计分下的 CD-CAT 如何控制曝光尚未见到相关研究。

针对以上问题和思路,本文提出了 5 种新的曝光控制选题策略,同已有的 PWKL 和 RP_PWKL 进行比较,并将这 7 种方法从 0-1 计分推广到多级计分的 CD-CAT。这 5 种方法分别是: (i) BSCDI_A(binary searching algorithm and stratification with ascend-

ing CDI) 法, (ii) BSCDI_D(BS and stratification with descending CDI) 法, (iii) BSCTTID_A(BS and stratification with ascending CTT item discrimination) 法, (iv) BSCTTID_D(BS and stratification with descending CTTID) 法, (v) 将现有的 RP 法同 PWCDI 相结合,提出了 RP_PWCDI 法。

1 曝光控制方法

1.1 BSCDI_A 方法

1.1.1 BS 选题法 Zheng Chanjin 等^[27-28]将计算机科学中的线性搜索(linear searching)和 BS 算法迁移到 CD-CAT 的选题中,其中 BS 算法本质上与二分法(halving algorithm, HA)^[29]相同。BS 的模式匹配是指理想项目的 Q 矩阵能够将当前的后验分布的概率密度半分。

$$B_{Sij} = \left| \sum_{i=1}^{2^K} e\pi_i^{(t)}(\alpha_c) - 0.5 \right|, \quad (1)$$

K 为测验所考察的属性个数。 e 是一个非 0 即 1 的数组,当 $\alpha_{IMP} - Q_j \geq 0$ 时 $e = 1$, 反之为 0。 α_{IMP} 是理想掌握模式, Q_j 是第 j 题的 Q 矩阵。 $\pi_i^{(t)}(\alpha_c)$ 是被试 i 在完成第 t 题后的后验概率。BS 的第 $(t+1)$ 题从 BS_i 值最小的项目中随机选取一个作为被试的下一题目。从 (1) 式可以看出 BS 直观易于计算,它本身含有随机因素,适合用于曝光控制。其次,BS 由于只需要计算被试的后验概率,在拓展到多级或其他模型时,不需要进行公式推导。

1.1.2 CDI CDA 的目的在于能够准确可靠地将被试分类到潜在类别,每个潜在类别就是一种具体的属性掌握模式。具体来说,一份好的认知诊断测验应该能够区别出属性掌握模式不同的被试。因此, R. Henson 等^[30]提出了 CDI 指标作为认知诊断模型的区分度:

$$C_{DIj} = \frac{1}{\sum_{u \neq v} h(\alpha_h, \alpha_v)^{-1}} \sum_{u \neq v} h(\alpha_h, \alpha_v)^{-1} D_{juv}, \quad (2)$$

$$D_{juv} = \sum_{x=0}^1 \left(\log \frac{P(X_j = x | \alpha_u)}{P(X_j = x | \alpha_v)} \right) P(X_j = x | \alpha_u), \quad (3)$$

$$h(\alpha_h, \alpha_v) = \sum_{k=1}^K |\alpha_u - \alpha_v|, \quad (4)$$

$h(\alpha_h, \alpha_v)$ 为 2 种模式之间的海明距离。从 (2) ~ (4) 式中可以发现 D_{juv} 用于度量 2 种模式之间的差异, CDI 是对 D_{juv} 的距离加权,若 2 种模式之间的差异越大,则权重越大,反之越小。

在 0-1 计分中对 D_{juv} 的分解为

$$\left(\log \frac{P(X_j = 1 | \alpha_u)}{P(X_j = 1 | \alpha_v)} \right) P(X_j = 1 | \alpha_u) + \\ \left(\log \frac{P(X_j = 0 | \alpha_u)}{P(X_j = 0 | \alpha_v)} \right) P(X_j = 0 | \alpha_u).$$

从 0-1 计分推导到多级计分时,这部分不再只是包含正确作答和错误作答 2 种情况,用 m_f 表示该题的满分值,则这部分公式可表示为

$$\left(\log \frac{P(X_j = m_f | \alpha_u)}{P(X_j = m_f | \alpha_v)} \right) P(X_j = m_f | \alpha_u) + \cdots + \\ \left(\log \frac{P(X_j = 0 | \alpha_u)}{P(X_j = 0 | \alpha_v)} \right) P(X_j = 0 | \alpha_u). \quad (5)$$

(5) 式 CDI 中 D_{juv} 的多级计分下的计算公式转换为

$$D_{juv} = \sum_{x=0}^{m_f} \left(\log \frac{P(X_j = x | \alpha_u)}{P(X_j = x | \alpha_v)} \right) P(X_j = x | \alpha_u). \quad (6)$$

1.1.3 两者结合的 BSCDI_A 方法 Chang Huahua 等^[31]提出采用区分度分层的方法控制 CAT 中的项目曝光.该方法的实施如下:将题库根据区分度参数分成 K 层;将测验分成 K 段;从第 K 层中根据难度 b 和能力估计值 $\hat{\theta}$ 接近的原则选取出第 n_k 题施测(其中 $n_1 + \cdots + n_k =$ 测验长度);重复上一步直到测验结束.在确定 K 时需要考虑题库中区分度的差异、题库的质量和大小、测验的长度等因素.为了使项目曝光一致每层中选取的项目个数应相同.本研究也采用该思想进行 CD-CAT 的曝光控制.

BSCDI_A 方法的思想是在测验实施之前计算所有项目的 CDI 指标,将该指标按照升序的顺序排序,确定整个题库的层数 n ,层数可以依据测验长度、题库大小和曝光率的控制调整,测验长度/ n 层是每层中选题的个数.这同 a 分层^[31]的思想相一致,强制被试在区分度较低的题库中选择最初的项目,随着项目数量的增加,选择的项目质量越好.依据本研究中题库的大小和测验的长度将题库分成 4 层,测验长度为 16 题,每个被试在每层选择 4 题.将题库分层后采用 BS 选题.从思想上来看 BSCDI 作为曝光控制方法计算简单、易于实施.

1.2 BSCDI_D 方法

Chang Huahua 等^[31]提出了 a 分层用于基于 IRT 的 CAT 的曝光控制方法,该方法简单有效,得到业内的认同.在初始阶段被试能力不明确时应采用区分度较低的题目,区分度高的项目应用于测验

后期,这样可提高测验效率和精度. BSCDI_A 是基于该思想分层的,然而,基于 IRT 和 CDA 的 CAT 是有区别的. CDA 中的知识状态是离散的偏序集合, CDM 是离散偏序的分类模型.不管是基于 IRT 的 CAT 还是基于 CDA 的 CAT 的选题策略,均要找到一批合适的项目,实现最快的从估计值到真值的收敛.离散偏序分类模型的最优收敛速度是指数收敛^[21],而 IRT 中 $\hat{\theta}$ 的最优收敛为 \sqrt{T} (T 为测验项目数)^[32],从 $\hat{\alpha}$ 到 α 的速度要快于从 $\hat{\theta}$ 到 θ ^[33].因此,在基于 IRT 的 CAT 初期使用质量较差的项目对其收敛的影响较小,在基于 CDA 的 CAT 初期使用质量较好的项目可加速其收敛.

基于该原因,本研究尝试计算项目的 CDI 值后采用降序排序,确定题库的层数,采用 BS 选题.这样可实现被试首先选择和完成的是区分度最好的项目.

1.3 BSCTTID_A 方法

在 CTT 中,项目的区分度计算包括鉴别指数法和题总相关法.其中鉴别指数法是将被试根据校标成绩或测验总分排序,取 27% 的高分端被试为高分组,27% 的低分端被试为低分组,分别计算出每道题的通过率,二者之差是鉴别指数,该思想同样可运用到 CDM 中.与 CTT 不同,CDM 的项目通常考察多个属性,不能像 CTT 中单维的总分分布中划分出高分组和低分组,因此, A. A. Rupp 等^[34]提出在 CDM 中计算区分度的基本问题在于“项目是否能够较好地地区分出掌握更多属性的被试和掌握更少属性的被试”,即 P_{ah} 是掌握了项目所考察属性的被试正确作答该项目的概率, P_{al} 是没有掌握该项目所考察属性的被试正确作答该项目的概率.有了这 2 种概率,结合 CTT 的鉴别指数法,适用于所有 CDM 的对于项目 i 的一般的区分度计算方法公式为 $D_i = P_{ah} - P_{al}$,该方法可用于计算项目水平的项目区分度,也可用于计算属性水平的项目区分度.

对于 DINA 模型来说

$$P_{ah} = (1 - s_i)^1 g_i^{1-1} = (1 - s_i), P_{al} = (1 - s_i)^0 g_i^{1-0} = g_i, D_{i,DINA} = P_{ah} - P_{al} = (1 - s_i) - g_i.$$

由于 DINA 模型的作答以被试掌握了所有属性为基础,因此其项目水平和属性水平的项目区分度计算公式相同.

BSCTTID_A 选题方法与 BSCDI_A 思想相同,计算项目的 CTID 值,采用升序排序,确定层数后

采用 BS 选题,该方法计算方式简单快速.

在 CTT 中该方法的思想是用于 0-1 计分项目的区分度的计算,将它运用到 DINA 模型时可以发现其主要依据是 $(1 - s_i) - g_i$,即 $s_i + g_i$ 越小越好.在从 0-1 拓展到多级计时,采用下式作为多级计分中的

CTTID 计算方法 $D_{i, P-DINA} = \sum_{j=1}^{m_j} (s_{ji} + g_{ji})$. 0-1 计时 $D_{i, DINA}$ 是越大越好,多级计时 $D_{i, P-DINA}$ 则是越小越好.当题库中项目的等级数不同时需对等级数较少的项目加权.由于本研究的题库只包含了等级数为 4 级的题目,故在此对于权重就不另外讨论.

1.4 BSCTTID_D 方法

与 BSCDI_D 的思想相同,BSCTTID_D 首先计算题库中所有项目的 CTTID 值,将所有项目的 CTTID 值按照降序排列,根据相关指标确定测验分层个数 n ,被试在每层完成的题目个数是测验长度/层数 n ,同样采用 BS 选题.被试首先做的是区分度高的题目,在测验初期的判断率较高.从题库使用的角度可知,升序时被试选择项目从低区分度到高区分度,降序时则正好相反.由于每层中选择的项目个数相同,其题库使用情况应该同升序比较接近.

1.5 RP_PWKL 方法

Wang Chun 等^[23]提出了用于 CDM 的曝光控制方法 RP (restrictive progressive method),其选题指标公式为 $R_{P-PWKL_j} = (1 - \exp_j/r) \cdot ((1 - x/L) R_j + P_{WKL_j} \beta x/L)$, x 为已经施测的项目, L 为测验长度, r 为测验控制的最大曝光率, $H^* = \max(P_{WKL}(x_j))$, R_j 服从于 $U(0, H^*)$, β 为一个任意数,根据测验的具体需要调整. Wang Chun 等^[23]通过预实验模拟表明 β 越大,估计精度提高,同时 x^2 值增大.即测验的判断率同题库使用均匀性成反比,判断率越高则均匀性越差,自适应测验需要在这两者之间进行权衡.因此,测验精度是关注的重点,可选择较大的 β 值,若更重视测验的安全性, β 可调整得更低.在 Wang Chun 等^[23]的研究中 β 取 0.5、1、2、3 共 4 个值,本研究同样采用这 4 个值作为 β 的取值点.

PWKL 的公式为 $P_{WKL_j}^{(i)}(\alpha_i^{(i)}) = \sum_{c=1}^{2^K} (\pi_i^{(i)}(\alpha_c) \cdot \sum_{x=0}^1 (\log(\frac{P(X_j = x | \alpha_i)}{P(X_j = x | \alpha_c)})) P(X_j = x | \alpha_i))$.

该方法在从 0-1 拓展到多级计时,将

$$\sum_{x=0}^1 (\log(\frac{P(X_j = x | \alpha_i)}{P(X_j = x | \alpha_c)})) P(X_j = x | \alpha_i)$$

拓展到 (6) 式,其他部分不需要变动. Wang Chun 等^[23] β 取 3、2、1 和 0.5 这 4 个值,为了考察在多级计分中 RP 方法的表现,本研究中依然采用原文的方法,分别研究 β 取值 3、2、1 和 0.5 时,判断率和曝光率的变化,最大曝光率固定在 0.2.

1.6 RP_PWCDI 方法

CDI 是在 KL 信息量的基础上增加海明距离作为权重获得的,从广义上来讲,它也是一种选题方法. PWCDI^[27-28] 是在 CDI 的 D_{juv} 部分增加后验概率作为权重获得的,其公式表示为

$$P_{WCDI_j} = \frac{1}{\sum_{u \neq v} d(\alpha_u, \alpha_v)^{-1}} \sum_{u \neq v} d(\alpha_u, \alpha_v)^{-1} P_{WD_{juv}},$$

$$P_{WD_{juv}} = \sum_{x=0}^1 \pi(\alpha_u) \pi(\alpha_v) P(X_j = x | \alpha_u) \cdot (\log(P(X_j = x | \alpha_u) / P(X_j = x | \alpha_v))). \quad (7)$$

通过公式可发现, PWCDI 方法同 MPWKL (modified PWKL)^[3] 非常接近,二者的区别在于 PWCDI 比 MPWKL 增加了距离加权.

$$M_{PWKL_{ij}}^{(i)} = \sum_{d=1}^{2^K} (\sum_{c=1}^{2^K} (\sum_{x=0}^1 \log(\frac{P(X_j = x | \alpha_d)}{P(X_j = x | \alpha_c)})) P(X_j = x | \alpha_d) \pi_i^{(i)}(\alpha_c) \pi_i^{(i)}(\alpha_d)).$$

在从 0-1 推广到多级时,将 (7) 式转换成

$$P_{WD_{juv}} = \pi(\alpha_u) \pi(\alpha_v) \sum_{x=0}^{m_j} (\log(P(X_j = x | \alpha_u) / P(X_j = x | \alpha_v))) P(X_j = x | \alpha_u).$$

从 PWCDI^[27] 的研究结果可以发现 PWCDI 是一种判断率较高的测验方式,同时其曝光率也非常高.本研究尝试将 RP 方法推广到 PWCDI,观察它是否能够在控制曝光率的基础上仍然保持较好的判断率.由于 PWCDI 与 PWKL 相同,都是取最大值,因此 RP_PWCDI 的公式为 $R_{P-PWCDI_j} = (1 - e^j/r) \cdot ((1 - x/L) R_j + P_{WCDI_j} \beta x/L)$.

在 RP_PWCDI 下同样关注 β 的取值对多级计分下的 CD-CAT 的曝光控制的效果.

本研究使用的 2 级积分模型是 DINA 模型^[29],多级计分模型是 P-DINA 模型^[1],模型的算法请参见相关文献,这里不再累述.

2 0-1 计分下的曝光控制选题方法

本研究采用 Matlab R2013b 自编程序进行模拟研究。

2.1 实验设计

现有的 CD-CAT 考察属性个数多为 4 ~ 8 个^[25, 733], 因此本研究将属性个数定为 5 个, 属性之间相互独立。题库中的项目数固定为 400 题, 根据文献[20]的研究结果, 每个项目考查每个属性的概率为 0.2。项目参数服从均匀分布, 其取值区间为 [0.05, 0.25]。Chang Huahua 等^[31]指出, 为了保证题库中项目曝光的均匀性, 每层选取的项目个数应相同。本研究中测验长度是 16 题, 共 4 层, 每层选取 4 题。由于分层选题, 每层选取的项目量是固定的, 只能采用定长的测验方式。

被试人数为 2 000 人, 被试掌握每个属性的概率服从参数为 0.5 的 Bernoulli 分布, 随机生成 2 000 个被试的真值。使用 DINA 模型计算被试正确作答概率, 生成 [0, 1] 之间的随机数, 正确作答概率大于该随机数则作答为 1, 反之则为 0。使用 MAP (Maximum a Posterior) 计算被试的属性掌握模式。

采用之前所介绍的 BSCDI_A、BSCDI_D、BSCT-TID_A、BSCTTID_D、RP_PWCDI、RP_PWKL 6 种方法选题, 为了观测曝光控制的效果, 将没有任何曝光控制的 PWKL 作为基准。实验重复 30 次, 取平均值。

为了验证 β 的取值对于判准率和曝光控制是否存在平衡的作用, 采用文献[23]中 β 的取值, 分别是 0.5、1、2 和 3。

2.2 评价指标

采用模式判准率 (Pattern Classification Correct Rate, PCCR) 作为判准率的评价指标。PCCR 为 $P_{CCR} = \sum_{i=1}^N t_i / N$, t_i 代表第 i 个被试知识状态的估计值和真值一致。PCCR 越大表示对于整个知识状态的估计越准确。

题库使用均匀性的评价指标采用 χ^2 值、测验重叠率 (T_{OR})、最大曝光率 ($E_{R_{max}}$)、最小曝光率 ($E_{R_{min}}$)、曝光率 > 20% 的题目数 ($N_{20\%}$)、曝光率 < 2% 的题目数 ($N_{2\%}$)。其中 χ^2 值反映了曝光率分布的观测值与平均值之间的差异。 $\chi^2 = \sum_{j=1}^N (e_{r_j} - L/N)^2 / (L/N)$, 其中 N 为题库容量, L 为测验长度, e_{r_j} 为项目 j 的曝光率。 χ^2 越小, 项目 j 的曝光率与期望越接近, 项目的

使用越均衡。

测验重叠率是指随机选择的 2 个被试之间期望重叠的项目个数与测验长度之比。 $\hat{T} = (N/L) S_{er}^2 + L/N$, 其中 N 为题库容量, L 为测验长度, S_{er}^2 为项目曝光率的方差。题库中项目使用的越均衡, 测验重叠率越低, 测验的安全性越好。

为了能够更加全面地比较不同方法之间的差异, 采用统一量纲再加权求和的方法综合考虑各评价指标下不同方法的表现。具体做法如下: 对值越大越好的指标, 将该指标上的最大值做分母, 把各选题方法在该指标上的值做分子, 求二者的比值; 对值越小越好的指标, 则将指标上的最小值做分子, 把各选题方法在该指标上的值做分母, 求二者的比值。统一量纲后, 对某一选题方法的所有评价指标比值分别赋加权系数, 加权求和值最大的, 则该选题方法在各个方面的综合效果最好, 反之则最差^[35-37]。为了便于计算, 所有指标设置相同权重。如 BSCDI_A 统一量纲后的指标计算是: $0.8915/0.9865 + 1.3286/11.6898 + 0.0428/0.0688 + 0.0875/0.164 + 0/0.028 + 1 + 0/86 = 3.1730$ 。由于 χ^2 本身波动范围较大, $E_{R_{min}}$ 、 $N_{20\%}$ 、 $N_{2\%}$ 这 3 个值存在 0 的结果, 因此各选题方法统一量纲后的综合指标差异较大, 有不合理之处, 此处仅作参考。

2.3 实验结果

表 1 是 0-1 计分下 6 种曝光控制方法的各项评价指标结果。

从统一量纲后的综合指标可以看出, RP_PWCDI ($\beta = 0.5$) 表现最好, 题库使用均匀性的各项指标均是最优。如前文所述, β 是调节判准率和题库使用的一个指标, 随着 β 值的减少, 题库使用均匀性各个指标表现越来越好。因此, RP_PWCDI ($\beta = 0.5$) 统一量纲后综合指标最高, 达到 6.7965, 随着 β 值的增加, RP_PWCDI 的综合指标值逐渐减小, 当 $\beta = 3$ 时降到最低。RP_PWKL 有着同样趋势。BSCDI_A、BSCDI_D、BSCTTID_A、BSCTTID_D 较为接近, 其中, BSCTTID 要略优于 BSCDI。PWKL 的判准率高, 同时题库使用非常不均衡, 远远高于其他方法。

从判准率上来说, 区分度方法的判准率略低于 RP 方法。随着 β 值的逐渐减少, RP_PWCDI 和 RP_PWKL 的判准率略有下降。如 RP_PWCDI 当 $\beta = 3$ 时 $P_{CCR} = 0.9360$, $\beta = 0.5$ 时 $P_{CCR} = 0.9140$ 。

值得注意的是, 当 CDI 和 CTTID 降序排列时, 其判准率的结果与升序没有差异。

从 χ^2 的结果来看, PWKL 的 χ^2 达到 196.727 1, 远远高于其他方法, 由此可见本文中提到的曝光控制方法的效果较好. 曝光控制方法中 BSCDI_A 的 χ^2 最大为 11.689 8, BSCDI_D 也略高于其他方法. BSCDI 和 BSCTTID 两者在判准率上没有显著性差异, 题库使用上 BSCTTID 略优于 BSCDI, 这同 2 种区分度计算方式有关. 使用 BSCDI 进行分层时区分度最差的一层大部分是考察属性个数较多的项目,

小部分项目是项目参数较大但考察属性个数较少的, 当需要在这一层进行选题时, BS 法倾向于选择考察属性个数较少的项目, 这时某些项目被重复选择, 其它项目的曝光不足, 导致题库使用不均衡. RP_PWCDI 和 RP_PWKL 2 种方法随着 β 值的逐渐减少 χ^2 也在减少, 这个结果与 Wang Chun 等^[23]一致. 由于 χ^2 本身浮动范围较大, 不能单纯依据 χ^2 来确定曝光率情况, 还需要观察其他题库使用指标.

表 1 0-1 计分下曝光控制方法的各项指标结果

	判准率	题库使用均匀性						统一量纲后
	P_{CCR}	χ^2	T_{OR}	$E_{R_{max}}$	$E_{R_{min}}$	$N_{20\%}$	$N_{2\%}$	综合指标
PWKL	0.986 5 ^[1]	196.727 1	0.531 6	0.990 0	0	25	329	1.175 6
BSCDI_A	0.891 5	11.689 8	0.068 8	0.164 0	0	0	86	3.173 0
BSCDI_D	0.894 0	9.350 0	0.062 9	0.197 5	0	0	84	3.171 8
BSCTTID_A	0.901 0	5.994 5	0.054 5	0.121 5	0	0	80	3.640 5
BSCTTID_D	0.897 5	6.389 0	0.055 5	0.135 5	0	0	85	3.534 7
RP_PWCDI($\beta=3$)	0.936 0	4.594 1	0.051 0	0.100 5	0.012 0	0	75	4.376 4
RP_PWCDI($\beta=2$)	0.933 5	3.463 6	0.048 18	0.099 5	0.017 0	0	30	4.704 7
RP_PWCDI($\beta=1$)	0.925 0	2.254 6	0.045 2	0.093 0	0.023 0	0	0	6.236 1
RP_PWCDI($\beta=0.5$)	0.914 0	1.328 6 ^[1]	0.042 8 ^[1]	0.087 5 ^[1]	0.028 0 ^[1]	0	0 ^[1]	6.926 5 ^[1]
RP_PWKL($\beta=3$)	0.927 0	4.678 3	0.051 2	0.103 5	0.011 5	0	79	4.315 7
RP_PWKL($\beta=2$)	0.919 5	3.967 7	0.049 4	0.097 0	0	0	50	4.035 4
RP_PWKL($\beta=1$)	0.914 0	2.235 3	0.045 1	0.095 0	0.022 0	0	0	6.176 6
RP_PWKL($\beta=0.5$)	0.915 0	1.355 0	0.042 9	0.091 5	0.027 5	0	0	6.844 1

注: P_{CCR} 为模式判准率; χ^2 为卡方统计量; T_{OR} 为测验重叠率; $E_{R_{max}}$ 为测验最大曝光率; $E_{R_{min}}$ 为测验最小曝光率; $N_{20\%}$ 为题库中曝光率 > 20% 的项目个数; $N_{2\%}$ 为题库中曝光率 < 2% 的项目个数.

TOR 的变化趋势同 χ^2 值相同. 各种曝光方法的测验重叠率均低于 0.07, 这意味着题库使用较为均匀. 其中, SBACDI 最大, 为 0.068 8, RP_PWCDI($\beta=0.5$) 和 RP_PWKL($\beta=0.5$) 最小, 分别为 0.042 8 和 0.042 9.

$E_{R_{max}}$ 作为题库中项目的最大曝光率的指标意味着题库中某一个项目被重复使用的概率, 该指标越大意味着该项目被重复使用的次数越多, 泄露的可能性越大. 在 RP 方法中, 通常对最大曝光率进行限制, 如本文中 RP 方法的最大曝光率限制在 0.2. 这 4 种区分度方法虽然没有对最大曝光率进行限制, 其结果仍然低于 0.2. RP 方法最大曝光率均在 0.1 左右.

对于曝光控制的方法, 判断其优劣的重要指标是题库中曝光率低于 2% 的项目个数. 低于 2% 的项目越多, 说明题库中有更多项目曝光不足, 被使用次数较少. BSCDI_A 曝光率低于 2% 的项目个数最多, 达到了 86 个, RP_PWKL($\beta=3$) 时, 有 79 题曝光率不足 2%. 当 $\beta=1$ 和 $\beta=0.5$ 时 2 种 RP 的曝光率低

于 2% 的项目数均为 0.

3 多级计分下曝光控制选题方法

3.1 实验设计

多级计分中有多对项目参数, 本研究将等级数固定为 4, 则每个项目有 3 对项目参数 S_{ji} 和 G_{ji} 服从于 $U(0.05, 0.25)$. 根据涂冬波等^[1]的设定, S_j 服从于单调递增的原则, G_j 服从于单调递减的原则, 将随机生成的 3 个 S_{ji} 按升序排列, 3 个 G_{ji} 按降序排列. 被试的模拟、题库 Q 矩阵的模拟、测验长度、属性掌握模式的估计方法、评价指标等同测验 1.

为了验证在多级计分下 β 的取值对于判准率和曝光控制是否也存在着平衡的作用, 采用 Wang Chun 等^[23]文章中 β 的取值, 分别是 0.5、1、2 和 3.

3.2 实验结果

表 2 是多级计分下 6 种曝光控制方法的在各项指标上的表现.

由表 2 可以看出, 多级计分下依然是 RP_PWCDI($\beta=0.5$) 表现最好, 统一量纲后综合指标是

6.830 1, 其次是 RP_PWKL ($\beta = 0.5$). 在多级计分下 β 有着同样的调节判准率和题库使用的作用. BSCDI_A、BSCTTID_A 同 RP_PWKL ($\beta = 3$) 基本持平. 2 种区分度方法题库使用情况升序效果要略优于降序, 这同 0-1 计分有差异. α 分层的思想是在测验的早期阶段整体上了解被试的能力水平, 由于不需要精准的估计, 项目的质量可以稍差. 随着测验的进行, 需要像聚光灯一样精准地估计被试的能力, 项目质量的要求逐渐提高^[38]. 因此, 在基于 IRT 的 CAT 中, α 分层均采用升序的方式. 如前文所说, Wang Chun 等^[33] 理论推导认为 CD-CAT 的收敛速度要快于基于 IRT 的 CAT, 基于该思想本研究提出

采用区分度降序的方式分层. 在 0-1 计时, 升序或降序的被试题库使用差异不大. 原因可能是整个题库中项目质量的区间范围较小. 如 CTTID 中 s 和 g 的取值范围是 $[0.05, 0.25]$, 项目质量差异不够大. 而在多级计分下项目参数的数量由等级数量确定, 如 4 级计分则有 3 对项目参数. 相对于 0-1 计分项目质量差异区间增大了很多. 区分度采用降序时, 后期精准估计时使用区分度不好的项目, 多级计分下项目质量差异较大, 导致后期项目使用不均衡. 收敛速度对于升序和降序的影响没有达到预期效果, 在多级计分的 CD-CAT 中, 区分度升序的效果要优于降序.

表 2 多级计分下 6 种曝光率控制方法各项指标结果

	判准率	题库使用均匀性						统一量纲后
	P_{CCR}	χ^2	T_{OR}	$E_{R_{max}}$	$E_{R_{min}}$	$N_{20\%}$	$N_{2\%}$	综合指标
PWKL	0.971 5 ^[1]	172.033 2	0.469 8	0.977 5	0	22	320	1.192 0
BSCDI_A	0.943 0	4.846 3	0.051 6	0.103 5	0	0	57	3.987 6
BSCDI_D	0.949 5	7.109 9	0.057 3	0.134 5	0	0	83	3.611 4
BSCTTID_A	0.945 5	4.283 4	0.050 2	0.12	0	0	44	3.938 3
BSCTTID_D	0.938 0	7.663 5	0.058 7	0.148	0	0	89	3.505 7
RP_PWCDI ($\beta = 3$)	0.943 5	3.442 3	0.048 1	0.102 5	0.006 5	0	59	4.438 6
RP_PWCDI ($\beta = 2$)	0.940 0	3.431 4	0.048 1	0.099 0	0.009 5	0	46	4.582 3
RP_PWCDI ($\beta = 1$)	0.941 0	2.749 9	0.046 4	0.097 0	0.017 5	0	17	5.054 8
RP_PWCDI ($\beta = 0.5$)	0.921 0	1.554 8 ^[1]	0.043 4 ^[1]	0.103 5	0.026 0 ^[1]	0	0 ^[1]	6.803 1 ^[1]
RP_PWKL ($\beta = 3$)	0.949 0	3.737 7	0.048 9	0.103 5	0.000 5	0	79	4.154 6
RP_PWKL ($\beta = 2$)	0.949 0	3.392 4	0.047 8	0.088 5	0.008 0	0	58	4.650 8
RP_PWKL ($\beta = 1$)	0.944 0	2.591 4	0.046 0	0.092 5	0.015 0	0	22	5.048 8
RP_PWKL ($\beta = 0.5$)	0.936 0	1.866 5	0.044 2	0.088 5 ^[1]	0.022 0	0	0	6.624 5

注: P_{CCR} 为模式判准率; χ^2 为卡方统计量; T_{OR} 为测验重叠率; $E_{R_{max}}$ 为测验最大曝光率; $E_{R_{min}}$ 为测验最小曝光率; $N_{20\%}$ 为题库中曝光率 $> 20\%$ 的项目个数; $N_{2\%}$ 为题库中曝光率 $< 2\%$ 的项目个数.

从判准率来说, 各种方法的判准率结果非常接近. 值得注意的是 β 的变化对于判准率的影响不是特别明显. 如 RP_PWCDI 法, 当 β 的取值是 3、2 和 1 时, 其判准率都在 0.94. RP_PWKL 法 β 取值的最大值和最小值之间只差 0.013. 从 χ^2 的结果来看, BSCDI_D 和 BSCTTID_D 略高于其他方法, 分别是 7.109 9 和 7.663 5.

随着 β 值的减少, RP 方法的 χ^2 逐渐减少, 这同 0-1 计分的趋势是一致的, 说明 β 值的取值越小题库使用越均衡.

从 T_{OR} 、 $E_{R_{max}}$ 2 个指标上可发现 6 种方法的题库使用较为均衡. 如 BSCDI、BSCDIC、BSCTTID 和 BSCTTIDC 的重叠率在 0.05 ~ 0.06, RP 的重叠率均低于 0.05; 最大曝光率低于 0.15.

曝光率低于 2% 的指标上, BSCDI 和 BSCTTID 较大, 分别是 83 和 89. $\beta = 0.5$ 时 2 种 RP 方法

低于 2% 的题目都为 0.

从实验的结果可发现, RP_PWCDI 和 RP_PWKL 在判准率和题库使用上几乎没有差异. 在题库使用上, BSCDI_D 和 BSCTTID_D 要略低于 BSCDI_A 和 BSCTTID_A.

4 讨论

曝光问题是 CAT 测验的重要内容^[34-37], 项目的过度曝光意味着需要花费更多的人力物力维护题库. 本研究将 Chang Huahua 等^[31] 的 α 分层方法推广到 CDM, 将 PWCDI 和 RP 方法相结合, 用于 CD-CAT 的曝光控制. 当前的 CD-CAT 研究都是基于 0-1 计分的模型, 在实践考试中, 0-1 计分的题目, 如选择题、填空题等形式认知能力层次较低 (识记、再认、记忆等), 所能提供的信息十分有限. 多级计分

的题目,形式多样,认知能力层次较高(分析、综合、应用、评价等),能够提供更加丰富的信息^[38-42]。多级计分在心理测评中的应用更加广泛,李克特量表属于典型的多级计分,多级计分 CD-CAT 的提出和应用有利于认知诊断模型在心理测评领域的推广。本研究尝试采用 P-DINA 模型将包含曝光控制的 CD-CAT 从 0-1 计分推广到多级计分。

研究结果表明,在多级计分下,同没有曝光控制的 PWKL 相比,本文所提出的这些方法的曝光控制效果较好,其中 RP 方法的曝光控制效果要优于区分度法,尤其是 RP_PWCDI 表现最好。Wang Chun 等^[23]在提出 RP 方法一文中将 β 设定为 3、2、1、0.5 这 4 个值, β 值越小,判准率越低,题库使用越均匀。本文通过模拟研究发现,多级计分下 β 值同样有着调节判准率与题库使用均匀性的作用,随着 β 值的减少,题库使用均匀性越好。

在多级计分下,基于 CDI 和 CTTID 分层后采用 BS 选题的曝光控制方法,题库使用均匀性相对较好,在没有任何约束的情况下,最大项目的曝光率均在 20% 以内,判准率与 RP 方法基本持平。基于 CDI 和 CTTID 的方法分层用于曝光控制计算简单,容易理解。CDI 和 CTTID 作为 CDM 的 2 种区分度指标,在从 0-1 计分拓展到多级计分时,CDI 只需要将正确作答概率和错误作答概率推导到所有级数上,它适用于任何模型。而 CTTID 在推广到不同的模型时计算公式没有这么简便,需要另外推导。因此,在实际应用中,CDI 比 CTTID 更易于推广。

总体来说,在多级计分下,区分度方法的判准率同 RP 法持平,由于没有增加约束,在题库使用均匀性上不如 RP 法。在实际测验中,如果对测验的判准和题库的使用要求较严格时推荐使用 RP 法,其中 RP_PWCDI 无论在判准率和题库使用上均优于 RP_PWKL。当然 RP 法也有缺点,如 β 的取值影响题库使用的均匀性,需要提前调节并确定。相比之下,区分度方法计算相对简单,易于理解。尤其是 CDI 法容易从 0-1 计分推广到多级计分,更加直观形象。在实际测验中,如果测验长度较长,区分度方法也是不错的选择。

本研究所提出的这些多级计分的曝光控制方法是独立于模型的。通过公式可以看出,RP_PWCDI 和 RP_PWKL 是基于作答反应概率来计算信息量的,因此不受模型约束。CDI 作为一种区分度计算方法也是对 2 种模式之间的距离进行海明距离加权,对

2 种模式的计算也是基于作答反应概率的,同样不受模型约束。对于 CTTID 法本文只列出了在 DINA 模型下的计算方法,A. A. Rupp 等^[34]对于其他常用的 0-1 计分模型的 CTTID 公式也进行了推导。CTTID 用于多级计分的其他模型的公式需要根据具体模型进行具体推导。当然,这些曝光控制方法在其他模型中表现如何还需要进一步研究。同样的,在不同属性个数和属性层级结构下这些方法的效果也是今后研究的一个方向。

5 参考文献

- [1] 涂冬波,蔡艳,戴海琦,等.一种多级计分的认知诊断模型:P-DINA 模型的开发[J].心理学报,2010,42(10):1011-1020.
- [2] Hsu C L, Wang Wenchung. Viable-length computerized adaptive testing using the higher order DINA model [J]. Journal of Educational Measurement, 2015, 52(2): 125-143.
- [3] Kaplan M, de la Torre J, Barrada J R. New item selection methods for cognitive diagnosis computerized adaptive testing [J]. Applied Psychological Measurement, 2015, 39(3): 167-188.
- [4] Liu Jingcheng, Ying Zhiliang, Zhang Stephanie. A rate function approach to computerized adaptive testing for cognitive diagnosis [J]. Psychometrika, 2015, 80(2): 468-490.
- [5] 涂冬波,蔡艳.基于属性多极化的认知诊断计算机化自适应测验设计与实现[J].心理学报,2015,47(11): 1405-1414.
- [6] 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.
- [7] 郭磊,郑蝉金,边玉芳.变长 CD-CAT 中的曝光控制与终止规则[J].心理学报,2015,47(1): 129-140.
- [8] 高椿雷,罗照盛,郑蝉金,等.CD-CAT 初始阶段项目选取方法[J].心理科学,2017,40(2): 485-491.
- [9] de la Torre J. The generalized DINA model framework [J]. Psychometrika, 2011, 76(2): 179-199.
- [10] Li Feiming, Cohen A, Bottge B, et al. A latent transition analysis model for assessing change in cognitive skills [J]. Educational and Psychological Measurement, 2016, 76(2): 76-82.
- [11] Zhan Peida, Wang Wenchung, Jiao Hong, et al. The probabilistic-inputs noisy conjunctive models for cognitive diagnosis [EB/OL]. [2018-03-13]. <https://doi.org/10.3389/fpsyg.2018.00997>.

- [12] 蔡艳,涂冬波.属性多极化的认知诊断模型拓展及其 Q 矩阵设计[J].心理学报,2015,47(10):1300-1308.
- [13] Haertel E H, Wiley D E. Presentations of ability structures: Implications for testing [M]. Hillsdale: Erlbaum, 1993: 359-384.
- [14] Junker B W, Sijtsma K. Cognitive assessment models with few assumptions and connections with nonparametric item response theory [J]. Applied Psychological Measurement, 2001, 25(3): 258-272.
- [15] Templin J, Henson R. Measurement of psychological disorders using cognitive diagnosis models [J]. Psychological Methods, 2006, 11(3): 287-305.
- [16] Hartz S. A bayesian framework for the unified model for assessing cognitive abilities: blending theory with practice [D]. Urbana: University of Illinois at Urbana-Champaign, 2002.
- [17] von Davier M. A general diagnostic model applied to language testing data [J]. British Journal of Mathematical and Statistical Psychology, 2008, 61(2): 287-307.
- [18] de la Torre J. The partial-credit DINA model [C]. Athens: The International meeting of the Psychometric Society, 2010.
- [19] Ma Wenchao, de la Torre J. A sequential cognitive diagnosis model for polytomous responses [J]. British Journal of Mathematical and Statistical Psychology, 2016, 69(3): 253-275.
- [20] Cheng Ying. When cognitive diagnosis meets computerized adaptive testing: CD-CAT [J]. Psychometrika, 2009, 74(4): 619-632.
- [21] Tatsuoka C. Data analytic methods for latent partially ordered classification models [J]. Journal of the Royal Statistical Society, 2002, 51(3): 337-350.
- [22] Wang Chun. Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length [J]. Educational and Psychological Measurement, 2013, 73(6): 1017-1035.
- [23] Wang Chun, Chuang Huahua, Huebner A. Restrictive stochastic item selection methods in cognitive diagnostic CAT [J]. Journal of Educational Measurement, 2011, 48(3): 255-273.
- [24] 陈平. 认知诊断计算机化自适应测验的项目增补: 以DINA模型为例 [D]. 北京: 北京师范大学, 2011.
- [25] Cheng Ying, Chang Huahua. The maximum priority index method for severely constrained item selection in computerized adaptive testing [J]. The British Psychological Society, 2009, 62(2): 369-383.
- [26] 毛秀珍, 辛涛. 认知诊断CAT中项目曝光控制方法的比较 [J]. 心理学报, 2013, 45(6): 694-703.
- [27] Zheng Chanjin. Some practical item selection algorithms in cognitive diagnostic computerized adaptive testing-smart diagnosis for smart learning [D]. Urbana: University of Illinois at Urbana-Champaign, 2015.
- [28] Zheng Chanjin, Wang Chun. Application of binary searching for item exposure control in cognitive diagnostic computerized adaptive testing [J]. Applied Psychological Measurement, 2017, 41(7): 561-576.
- [29] Tatsuoka C, Ferguson T S. Sequential classification on partially ordered sets [J]. J R Statist Soc B, 2003, 65(1): 143-157.
- [30] Henson R, Douglas J. Test construction for cognitive diagnosis [J]. Applied Psychological Measurement, 2005, 29(4): 62-77.
- [31] Chang Huahua, Ying Zhiliang. A stratified multistage computerized adaptive testing [J]. Applied Psychological Measurement, 1999, 23(3): 211-222.
- [32] Chang Huahua, Stout W. The asymptotic posterior normality of the latent trait in an IRT model [J]. Psychometrika, 1993, 58(1): 37-52.
- [33] Wang Chun, Zheng Chanjin, Chang Huahua. An enhanced approach to combine item response theory with cognitive diagnosis in adaptive testing [J]. Journal of Educational Measurement, 2014, 51(4): 358-380.
- [34] Rupp A A, Templin J, Henson R A. Diagnostic measurement: theory, methods, and application [M]. NY: The Guilford Press, 2010.
- [35] 陈德枝. Samejima 等级反应模型下 CAT 选题策略比较研究 [D]. 南昌: 江西师范大学, 2010.
- [36] 刘珍, 丁树良, 林海菁. 基于GPCM的计算机自适应测验选题策略比较 [J]. 心理学报, 2008, 40(5): 618-625.
- [37] 罗照盛, 喻晓峰, 高椿雷, 等. 基于属性掌握概率的认知诊断计算机化自适应测验选题策略 [J]. 心理学报, 2015, 47(5): 679-688.
- [38] Chang Huahua. Psychometrics behind computerized adaptive testing [J]. Psychometrika, 2015, 80(1): 1-20.
- [39] 詹沛达, 陈平, 边玉芳. 使用验证性补偿多维IRT模型进行认知诊断评估 [J]. 心理学报, 2016, 48(10): 1347-1356.
- [40] Huang Hungyu, Wang Wenchun. The random-effect DINA model [J]. Journal of Educational Measurement, 2014, 51(1): 75-97.
- [41] Maris E. Estimating multiple classification latent class models [J]. Psychometrika, 1999, 64(2): 187-212.
- [42] Van der Linden W J, Hambleton R K. Handbook of modern item response theory [M]. New York: Springer, 1997: 85-100.

- cells in vitro stimulates human umbilical cord vein endothelial cell proliferation [J]. American Journal of Therapeutics 2006 ,13(3) : 205-209.
- [15] Wan Huida ,Li D. Highly efficient biotransformation of ginsenoside Rb1 and Rg3 using β -galactosidase from *Aspergillus* sp [J]. Rsc Advances 2015 5(96) : 78874-78879.
- [16] Zhao Haibo ,Johnathan E Holladay ,Heather Brown ,et al. Metal chlorides in ionic liquid solvents convert sugars to 5-hydroxymethylfurfural [J]. Science ,2007 ,1597: 316-319.

The Catalytic Transformation of Ginsenoside Re by Metal Ions

DING Shan¹ ,LIU Chunying² ,XU Longquan¹ ,SONG Jianguo¹ ,YU Hongshan^{1*}

(1. College of Biology Engineering ,Dalian Polytechnic University ,Dalian Liaoning 116034 ,China;

2. College of Life Science and Technology ,Dalian University ,Dalian Liaoning 116622 ,China)

Abstract: In this paper ,metal ion catalytic conversion of ginsenoside Re1 to generate Rg2 as the target ,investigation of metal ions in the organic water changes as catalytic reaction solvent and the ethanol water system under the catalysis of Fe^{3+} reaction product changes with ethanol concentration ,reaction temperature ,iron ion concentration and reaction time were also discussed. The results show that the ethanol water solvent catalytic reaction products 20(S , R) -Rg2 the most ,the catalytic reaction conditions optimization results were: ethanol concentration 50% , Fe^{3+} concentration was $0.8 \text{ mol} \cdot \text{L}^{-1}$,the reaction at a temperature below 40°C ,12 h ,the product rate was 64.8% ,and the content of 20(S , R) -Rg2 was 74.3% . The paper provided a new method for the catalytic conversion of ginsenosides to rare saponins.

Key words: ginsenoside; catalytic conversion; condition optimization; HPLC-ELSD

(责任编辑: 刘显亮)

(上接第 67 页)

The Exposure Control Methods of Polytomous CD-CAT

GAO Chunlei ,LUO Zhaosheng ,DING Shuliang

(Elementary Education College ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China)

Abstract: Many CD-CAT selection strategies focused on improving the classification correct rate ,and ignoring the exposure rate. Most of them has high exposure rate ,poor uniformity of item bank ,and items drift. It spends more manpower and financial resources on constructing item bank. So reasonable and effective exposure control methods are needed to improve item bank utilization for CD-CAT. In order to improve item bank utilization ,five exposure control methods are presented ,they are binary searching algorithm and stratification with ascending Cognitive Diagnosis Index(BSCDI_A) ,binary searching algorithm and stratification with descending Cognitive Diagnosis Index(BSCDI_D) ,binary searching algorithm and stratification with ascending CTT item discrimination(BSCTTID_A) ,binary searching algorithm and stratification with descending CTT item discrimination(BSCTTID_D) ,combing RP with PWCDI(RP_PWCDI) . Two simulation experiments were made to observe the five exposure control methods under dichotomous and polytomous models respectively. The results showed that under dichotomous model ,the maximum exposure rate of the five methods was under 20% ,the testing overlap rate was under 7% ,BSCDI_A、BSCDI_D、BSCTTID_A、BSCTTID_D had the similar PCCR and item bank utilization. Under polytomous model ,the PCCR of BSCDI_A、BSCDI_D、BSCTTID_A、BSCTTID_D was a little increased ,and almost the same with RP_PWKL and RP_PWCDI. BSCDI_D and BSCTTID_D were a little poorer than BSCDI_A and BSCTTID_A on item bank utilization.

Key words: P-DINA model; CD-CAT; dichotomous; polytomous; exposure control methods

(责任编辑: 冉小晓)