

文章编号: 1000-5862(2016)04-0363-06

# 一种提升题库安全性的选题策略

贺翔, 罗芬, 甘登文\*, 丁树良, 汪文义

(江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 题库的安全性是实施计算机化自适应测验的关键问题. 受原有的动态  $a$  分层方法和均值不等式启发, 构造了动态  $a$  分层的选题策略. 模拟实验结果表明: 新的选题策略保持了原有动态  $a$  分层的测验精度, 进一步提高了测验的安全性.

关键词: 项目信息量; 区分度; 选题策略

中图分类号: B 841 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2016.04.07

## 0 引言

计算机化自适应测验 (Computerized Adaptive Test, CAT) [1] 是用项目反应理论建立题库, 并由计算机根据被试能力水平自动选择试题, 最终对被试能力作出估计的一种新型测验, 是心理测量、计算机技术和现代教育技术充分结合的产物 [2]. CAT 与传统的“千人一卷”的考试不同, 它的目的是为每一个考生安排一场最优的测验 [3]. CAT 每次选择适合被试当前估计能力的项目, 真正做到了“量体裁衣, 因人施测”, 因为每份试卷都不相同, 从而在 CAT 实施之初, 许多人认为这可以保证测验的安全性, 最大程度的减少作弊, 同时还能够快速准确地测量出被试的水平. 目前, 国外有较多大规模的测验采用 CAT 的形式进行考试, 如 GRE、全美医护人员资格考试、TOFEL 等. 在我国, CAT 应用于征兵、中学教学、党务工作者专业知识考试自适应系统等 [4].

在 CAT 中, 核心问题是选题策略. 两级评分中最大 Fisher 信息量选题策略 (MFI) 因其测验长度短, 测验效率高而闻名, 但高区分度的项目, 使用次数多, 曝光率高, 对题库的安全性、考试的公平性产生不利影响. Chang Hua-Hua 等 [5] 的  $a$  分层 ( $a$ -STR) 方法针对 MFI 做出了改进, 降低了项目的曝光率.  $a$  分层主观上想保证在能力初估阶段, 选择区分度小的项目, 当能力估计精度逐步提高时, 使用区分度大的项目, 即逐步提升区分度对选题策略的影响, 从而

控制项目的曝光率. 这种选题策略在宏观层面保证了逐步升“ $a$ ”, 但却无法保证在微观层面也逐步升“ $a$ ” [2]. 程小扬在分层的情况下提出将曝光控制因子引入到 CAT 的选题策略中, 其中将曝光控制因子记为  $ecf(j)$ , 其计算方法为  $ecf_j = m_j / \bar{m}$ , 这里  $m_j$  表示项目  $j$  被前  $m-1$  个考生调用的次数,  $\bar{m}$  为前  $m-1$  个考生调用题库中所有项目的平均数, 即  $\bar{m} = \sum_{j=1}^M m_j / M$ ,  $M$  为题库中题目总数, 该方法保证了宏观和微观层面均实现升“ $a$ ”, 有效改善了项目的曝光率. 李萍等在沿用引入曝光控制因子的 CAT 选题策略中, 将定长测验的测验长度以及不定长测验中被试累积信息量加入到新选题策略中, 提出了不需要事先分层而在测验过程中自动控制区分度作用的选题策略. 该方法引入了新的区分度幂函数  $a(j, i)$  来逐步自动调节区分度对信息函数值的影响, 其中文献 [6] 的区分度幂函数  $a(j, i)$  分 2 种情况: (i) 定长测验  $a(j, i) = a_j^{2(I_{\text{test}} - L(i)) / I_{\text{test}}}$ , 其中  $a_j$  为第  $j$  个项目的区分度,  $L(i)$  为第  $i$  个被试当前已作答的项目个数,  $I_{\text{test}}$  为预先确定的测验长度; (ii) 不定长测验  $a(j, i) = a_j^{2(I_{\text{nfor}} - I_{\text{nfor}}(i)) / I_{\text{nfor}}}$ , 其中  $I_{\text{nfor}}$  为测验终止时的信息总量, 即目标信息量,  $I_{\text{nfor}}(i)$  为被试  $i$  当前作答项目的累计信息量. 总之区分度幂函数  $a(j, i)$  是使得区分度的幂随着 CAT 的进程的深入越来越减小, 从剩余题库中选取使得

$$f_j = I_j(\hat{\theta}) / (ecf(j) a(j, i)) \quad (1)$$

收稿日期: 2015-12-27

基金项目: 国家自然科学基金 (31360237, 31160203, 31500909), 江西省教育厅科学技术研究 (GJJ13207, GJJ13208, GJJ13209, GJJ13226, GJJ13227) 资助项目.

通信作者: 甘登文 (1956-), 男, 江西奉新人, 教授, 主要从事智能教学软件和应用统计的研究.

达到最大的项目,其中  $I_j(\hat{\theta})$  为项目  $j$  在当前被试能力估计值处的信息量.戴懿等<sup>[7]</sup>把影子题库<sup>[8]</sup>与曝光因子  $ecf(j)$  相结合构成新的选题策略.李佳等<sup>[9]</sup>将改进的最大优先级指标(MMPI)和各类平均数形式相结合得到4种新的选题策略,在提高测验精度、控制项目曝光均匀性、降低平均违规次数、提高题库利用率等方面均表现更好.经定长测验和不定长测验的蒙特卡洛模拟,MMPI下算术平方根平均数形式的选题策略表现也比较优秀.

以上所述对CAT项目安全性控制的方法中,部分项目曝光率高的问题仍然存在.本文在文献[6]的基础上,将其与李佳<sup>[9]</sup>策略中平均数的表示形式相结合,提出了一种新的选题策略,以期获得更加均匀的题库利用率.

## 1 新的选题策略

### 1.1 新方法1

针对文献[2]和文献[6]在使用曝光因子的基础上又使用区分度的幂函数来平衡项目的曝光率.文献[6]考虑的是最大信息量选题策略的变化,故题库利用率仍受区分度的约束,新方法1希望削弱区分度对选题策略的影响.

令  $a(j) = a_j^2$  在剩余题库中,选取使得  $f$  最大的项目作为下一题  $f = I_j(\hat{\theta}) / (ecf(j) a(j))$ ,探索新策略对项目精度和曝光率的影响.

### 1.2 新方法2

假设  $\lambda$  在  $[0, 1]$  之间取值,平均数主要包括几何平均数  $x^\lambda y^{1-\lambda}$ ,算术平均数  $\lambda x + (1-\lambda)y$ ,算术平方根平均数  $(\lambda x^2 + (1-\lambda)y^2)^{1/2}$  以及调和平均数<sup>[9]</sup>  $1/(\lambda/x + (1-\lambda)/y)$ .

当  $x, y$  为非负时,根据均值不等式有:  $1/(\lambda/x + (1-\lambda)/y) \leq \sqrt{xy} \leq (x+y)$ ,进一步得出  $1/(\lambda/x + (1-\lambda)/y) \leq x^\lambda y^{1-\lambda} \leq \lambda x + (1-\lambda)y \leq \sqrt{\lambda x^2 + (1-\lambda)y^2}$ .

依据不同均值不等式的大小关系,选取算术平方根不等式,将文献[6]进行拆分,由于算术平方不等式与其平方同时达到最大,为计算简便取其平方,同时沿用新方法1削弱区分度对选题的影响,进一步降低项目曝光率.

记  $x = I_j(\hat{\theta}) / a(j)$ ,  $y = 1/ecf(j)$ , 令

$$f = \max_{j \in R_i} (\lambda x^2 + (1-\lambda)y^2), \quad (2)$$

(2)式称为新方法2,选取使得  $f$  最大的项目作为下

一题,其中  $\lambda$  为权重,在定长情况下  $\lambda$  的取值为  $\sin(L(i)/L_{\text{test}})$ ; 在不定长情况下  $\lambda = \sin(I_{\text{info}}(i)/I_{\text{info}}) R_i$  为当前被试  $i$  的剩余项题库.

### 1.3 新方法3

通过对文献[6]进行分析,发现其方法平衡项目曝光率的同时没有损失太大的精度,由于测验精度和题库安全性是一对相互冲突的指标,所以新方法1和新方法2中  $a(j) = a_j^2$  减少了区分度对信息量的影响,平衡了项目的选取,必然带来精度的下降.新方法2是对新方法1的改进,故将文献[6]与新方法2结合设计新方法3,在降低项目曝光率的同时,又维持了测量精度.

对于定长:当  $L(i) \leq \lambda L_{\text{test}}$  选题用文献[6];当  $L(i) > \lambda L_{\text{test}}$  时,用新方法2.对于不定长:当  $I_{\text{info}}(i) \leq \lambda I_{\text{info}}$  时,选题用文献[6],否则用新方法2.其中  $\lambda$  取黄金分割点<sup>[11]</sup>  $\lambda = 0.618$ .

## 2 实验设计

本文的测验为0-1评分测验,采用3参数Logistic模型(3PLM),其答对概率为

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

其中  $D$  取1.7.

### 2.1 被试及题库的模拟

采用Monte Carlo模拟CAT,模拟生成1000个被试,被试能力水平服从标准正态分布  $\theta \sim N(0, 1)$ .模拟生成4种题库,每个题库均为1000个项目;题库1:  $\ln \alpha \sim N(0, 1)$ ,  $b \sim N(0, 1)$ ;题库2:  $\ln \alpha \sim N(0, 1)$ ,  $b \sim U(-3, 3)$ ;题库3:  $a \sim U(0.2, 2.5)$ ,  $b \sim N(0, 1)$ ;题库4:  $a \sim U(0.2, 2.5)$ ,  $b \sim U(-3, 3)$ .以上4个题库中限定  $a$  的取值范围为  $[0.2, 2.5]$ ;限定  $b$  的取值范围为  $[-3, 3]$ ;猜测参数  $c$  服从贝塔分布,记为  $c \sim \text{Beta}(5, 17)$ .

### 2.2 能力初估阶段CAT模拟

在测验的探测阶段,采用随机的从题库里抽取3道不同的试题给被试作答.被试能力的初始估计值为得分与失分之比的自然对数.若被试全部答对,则被试初估能力为3;若被试全部答错,则被试初估能力为-3.得到被试的初估能力之后,则进入精确探测阶段.被试在此阶段所做题目不计入定长测验时被试的测验长度;所作答的信息量不计入不定长测验时被试的累计信息量;与精确探测阶段的剩余题库无关.

2.3 测验的结束条件和能力估计方法

本阶段分定长和不定长 2 种测验: (i) 定长测验 除文献 [2] 将题库分成 4 层, 每层选 6 题外, 其它选题策略均不分层, 测验长度为 24. (ii) 不定长测验 文献 [2] 题库分成 4 层, 各层终止的累计信息量满足  $I_k = I(k/T)^2$ , 其中  $I_{\text{总}}$  为目标信息量  $k$  为当前所在层数,  $T$  为题库所分层数, 本文层数  $T$  为 4. 其他方法不分层, 所有选题策略的信息量达到 16 时终止, 最大答题数为 32. 在此阶段, 使用 EAP 方法进行能力估计, 直到达到测验终止条件时为止.

2.4 评价指标

本文采用以下 7 个评价指标: 能力估计准确性 ( $A_{BS}$ )、能力估计标准差 ( $S_e$ )、测验效率 ( $E_{ff}$ )、项目调用均匀性 ( $S_e$ )、卡方统计量 ( $\chi^2$ )、测验重叠率 ( $R_l$ )、人均用题数. 以上这些指标, 测验效率是越大

越好, 而其它都是越小越好.

3 实验及结果分析

实验分定长和不定长 2 个实验, 每个实验重复 30 次, 在 Matlab2010 下进行.

表 1 ~ 表 4 为实验 1, 定长测验的结果, 方法 1、方法 2 和方法 3 在均匀性、卡方、重叠率方面比文献 [2] 和文献 [6] 有较大提高. 尤其是方法 2 在卡方上的提高特别明显, 在表 1 和表 3 中, 有关评价项目曝光率的指标甚至优于随机选题. 但新方法 1 和新方法 2 在能力估计准确性、能力估计标准差方面较文献 [2] 和文献 [6] 有所降低. 在测验效率方面, 3 种新方法均不如文献 [6]. 方法 3 在损失少许测验精度的同时, 极大地降低了项目的曝光率, 提高了测验的安全性.

表 1  $\ln \alpha \sim N(0, 1)$   $b \sim N(0, 1)$   $\rho \sim \text{Beta}(5, 17)$  定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.336 4	0.335 3	24.000 0	0.233 6	4.835 9	0.974 7	0.024 0
MIC 选题	0.132 6	0.147 9	24.000 0	1.750 1	73.415 9	224.580 0	0.247 8
文献 [2]	0.225 5	0.247 3	24.000 0	0.602 8	7.303 4	2.222 6	0.025 2
文献 [6]	0.173 6	0.189 8	24.000 0	1.046 5	20.136 7	16.895 4	0.039 9
新方法 1	0.244 6	0.267 3	24.000 0	0.505 7	4.896 4	0.999 1	0.024 0
新方法 2	0.252 9	0.271 9	24.000 0	0.472 3	3.669 8	0.561 2	0.023 6
新方法 3	0.215 7	0.236 9	24.000 0	0.634 8	6.323 9	1.666 5	0.024 7

表 2  $\ln \alpha \sim N(0, 1)$   $b \sim U(-3, 3)$   $\rho \sim \text{Beta}(5, 17)$  定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.382 0	0.352 2	24.000 0	0.151 7	4.816 9	0.967 3	0.024 0
MIC 选题	0.131 4	0.137 0	24.000 0	1.586 7	78.346 1	255.755 7	0.279 0
文献 [2]	0.230 4	0.232 1	24.000 0	0.564 0	14.722 5	9.032 3	0.032 1
文献 [6]	0.182 0	0.186 6	24.000 0	0.932 3	25.927 7	28.010 4	0.051 1
新方法 1	0.266 0	0.262 9	24.000 0	0.427 8	11.338 8	5.357 8	0.028 4
新方法 2	0.305 1	0.296 6	24.000 0	0.346 8	5.142 7	1.102 0	0.024 1
新方法 3	0.234 8	0.237 2	24.000 0	0.521 5	9.414 8	3.693 5	0.026 7

表 3  $a \sim U(0.2, 2.5)$   $b \sim N(0, 1)$   $\rho \sim \text{Beta}(5, 17)$  定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.280 0	0.293 7	24.000 0	0.347 6	4.810 6	0.964 6	0.024 0
MIC 选题	0.120 9	0.135 6	24.000 0	2.105 8	70.285 3	205.836 4	0.229 1
文献 [2]	0.177 8	0.198 6	24.000 0	0.949 7	5.641 2	1.326 1	0.024 4
文献 [6]	0.149 6	0.169 0	24.000 0	1.339 3	14.741 8	9.055 0	0.032 1
新方法 1	0.187 5	0.209 2	24.000 0	0.851 1	4.448 4	0.824 6	0.023 8
新方法 2	0.196 4	0.219 4	24.000 0	0.790 3	3.554 8	0.526 6	0.023 6
新方法 3	0.177 7	0.198 1	24.000 0	0.936 5	4.584 9	0.875 9	0.023 9

表4  $a \sim U(0.2, 2.5)$   $b \sim U(-3, 3)$   $\rho \sim \text{Beta}(5, 17)$  定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.339 1	0.345 8	24.000 0	0.211 4	4.838 3	0.975 5	0.024 0
MIC 选题	0.124 1	0.139 7	24.000 0	1.911 0	70.910 0	209.512 7	0.232 7
文献[2]	0.185 2	0.200 7	24.000 0	0.906 2	15.980 6	10.642 4	0.033 7
文献[6]	0.166 5	0.181 6	24.000 0	1.222 9	22.771 2	21.605 6	0.044 7
新方法1	0.208 8	0.228 2	24.000 0	0.725 9	13.252 6	7.318 7	0.030 3
新方法2	0.247 8	0.267 6	24.000 0	0.537 6	4.960 7	1.025 5	0.024 0
新方法3	0.209 9	0.227 1	24.000 0	0.744 8	8.230 8	2.822 9	0.025 8

表5 ~ 表8为实验2 不定长测验的结果. 从表5 ~ 表7 中可以看出3 种新方法在均匀性和卡方上与文献[2] 和文献[6] 相比均有优势. 在测验长度方面, 人均用题数均高于文献[6]; 表8 中新方法2 和新方法3 在均匀性、卡方、重叠率方面均好于文献[2] 和文献[6]. 新方法2 在4 种题库下有关项目曝光率方面

的指标, 甚至要优于随机选题. 其能力估计准确性和能力估计标准差在表5 和表7 上得到了一定保持, 但在表6 和表8 上表现较差. 新方法1 与文献[2] 和文献[6] 相比, 并无多少优势. 新方法3 在降低项目曝光率的同时, 又较好地保持了测量精度.

表5  $\ln \alpha \sim N(0, 1)$   $b \sim N(0, 1)$   $\rho \sim \text{Beta}(5, 17)$  不定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.292 9	0.311 7	31.993 9	0.223 8	5.603 6	0.981 9	0.032 0
MIC 选题	0.200 3	0.226 3	10.552 5	1.621 4	53.565 2	271.907 2	0.281 7
文献[2]	0.205 4	0.227 8	30.914 5	0.530 7	8.664 1	2.428 4	0.032 4
文献[6]	0.205 0	0.227 2	26.187 1	0.640 3	7.605 4	2.209 0	0.027 4
新方法1	0.213 7	0.238 8	30.047 9	0.469 8	5.540 2	1.021 9	0.030 1
新方法2	0.235 6	0.258 0	30.882 8	0.416 9	2.810 1	0.255 8	0.030 2
新方法3	0.211 9	0.235 8	30.421 4	0.496 9	4.924 4	0.797 2	0.030 2

表6  $\ln \alpha \sim N(0, 1)$   $b \sim U(-3, 3)$   $\rho \sim \text{Beta}(5, 17)$  不定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.338 3	0.348 9	32.000 0	0.156 7	5.604 6	0.982 3	0.032 0
MIC 选题	0.200 9	0.223 9	10.486 9	1.631 1	56.933 1	309.091 1	0.318 9
文献[2]	0.204 6	0.226 7	30.518 5	0.539 0	18.575 6	11.307 6	0.040 9
文献[6]	0.214 5	0.235 4	27.717 0	0.571 1	16.468 2	9.784 9	0.036 5
新方法1	0.233 9	0.254 3	29.471 7	0.437 3	14.988 1	7.623 2	0.036 1
新方法2	0.315 5	0.302 4	30.625 3	0.313 2	1.948 0	0.124 0	0.029 8
新方法3	0.227 8	0.247 4	30.653 5	0.428 5	8.281 7	2.237 6	0.031 9

表7  $a \sim U(0.2, 2.5)$   $b \sim N(0, 1)$   $\rho \sim \text{Beta}(5, 17)$  不定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.252 7	0.267 3	31.614 1	0.337 8	5.525 8	0.966 4	0.031 6
MIC 选题	0.203 5	0.229 2	8.926 0	1.954 3	49.193 3	271.125 2	0.279 3
文献[2]	0.197 7	0.219 2	22.754 8	0.755 8	8.116 4	2.895 2	0.024 7
文献[6]	0.203 4	0.226 7	17.228 0	1.003 2	4.357 7	1.102 3	0.017 3
新方法1	0.206 0	0.229 6	20.088 2	0.834 6	3.937 4	0.771 8	0.019 9
新方法2	0.206 7	0.229 4	21.528 4	0.757 1	2.222 2	0.229 4	0.020 8
新方法3	0.199 2	0.222 6	20.005 8	0.826 6	3.159 9	0.499 1	0.019 5

表 8  $a \sim U(0.2, 2.5)$   $b \sim U(-3, 3)$   $c \sim Beta(5, 17)$  不定长

策略	$A_{BS}$	$S_e$	人均用题	测验效率	均匀性	卡方	重叠率
随机选题	0.305 1	0.321 8	31.998 1	0.203 1	5.503 7	0.947 2	0.032 0
MIC 选题	0.207 7	0.228 1	9.213 5	1.892 1	51.315 9	285.816 8	0.294 3
文献 [2]	0.204 8	0.224 4	25.539 4	0.667 5	17.274 3	11.684 6	0.036 3
文献 [6]	0.206 5	0.226 9	20.884 1	0.823 6	13.295 9	8.465 2	0.028 4
新方法 1	0.209 3	0.228 1	24.551 3	0.659 3	14.582 1	8.661 6	0.032 2
新方法 2	0.296 2	0.285 9	26.875 4	0.438 5	2.148 9	0.172 1	0.026 1
新方法 3	0.220 0	0.238 1	26.820 6	0.544 2	6.076 1	1.377 0	0.027 2

总之,在定长和不定长测验方面,若只考虑测验的安全性,新方法 2 可认为是一种不错的选题策略;若要兼顾被试能力估计的准确性和题库的安全性,新方法 3 可作为一种供选择的安全、高效的选题策略。

4 讨论

文章中的实验是一个模拟研究,项目参数和被试能力是沿用前人的基础上预先设定为某个理论分布<sup>[2,6]</sup>。在实际的 CAT 中,希望题库题量充分多而且各种难度的题目数量也比较多,但是由于时间或者经费或者其他原因,这不一定能够得到保障;被试的能力分布也不一定这么理想,模拟条件不一定能够满足。这时候采用本文的选题策略实施的 CAT 的效果,就不一定如人意。但是模拟条件和现实的差异总是存在,一方面必须严格进行模拟研究,如果模拟结果都不能令人满意,那么这种方法(或者理论)的价值就值得怀疑;所以新方法在实际应用中具有一定的借鉴意义。另一方面,对于真实的 CAT 的结果,还应该仔细分析,不能够完全硬搬模拟结果,还应该从实际工作发现新的问题,进行新的更加深入的研究。

测量精度与项目的测验信息量相关,测验效率是单个项目平均贡献的信息量。从 Monte Carlo 实验结果中发现,在定长测验中,由于测验长度相同,测验效率高的则测量精度就高,测验效率低的测量精度就低。两者具有正相关性。在不定长测验中,测验的终止条件,除受目标信息量的控制外,还与测验最大答题数相关;当测验效率与人均用题数相乘得到人均信息量时,人均测验信息量大的测验精度会升高,此时测量精度与人均用题数和测验效率有关。通过模拟实验发现,被试在能力初估阶段所做 3 道题目是否计入定长测验时的测验长度、不定长时的被试累积信息量,结果相似。同时鉴于初始 3 题为随机选题,是对被试能力的初估,故本文采取不计入方式处理。

本文在沿用程小扬等<sup>[2]</sup>提出的曝光因子的基础上,提出了关于 3PLM 下的 0-1 评分的 3 种新选题策略,通过新方法 1 和新方法 2 探讨了在选题策略中减少区分度对信息量的影响是否可以降低曝光率。通过实验对比,新选题策略在定长测验和不定长测验中,在有关项目曝光率指标上,与文献 [2] 和文献 [6] 相比,均有不同程度提高。在测量精度、测验效率方面稍逊于文献 [2] 和文献 [6]。新方法主要探索了在减少区分度对项目信息函数值影响的情况下有关项目曝光率的问题。其中在新方法 2 中,虽然不定长测验中人均用题数增加了,但有关项目曝光率的指标较低。新方法 3 是将测验精度比较好的文献 [6] 迁移过来,借助其测量精度高的优点,与项目曝光率比较低的新方法 2 结合,关于这 2 种方法在新方法 3 上占的比例经过试验取黄金分割点  $\lambda = 0.618$ ,是否还有更合适的比例,使效果达到最佳,还有待一步研究;同时发现  $\lambda \rightarrow 0$ ,测量的精度和曝光率越接近新方法 2;  $\lambda \rightarrow 1$  时,其测量的精度和曝光率越接近文献 [6],可以根据测验的需要进行改变。鉴于新方法 2 项目曝光率比较低的优点是否可以迁移,与其他测量精度高的选题策略相结合;是否可以应用到多级评分中,是否可以利用平均数的其他形式对文献 [6] 和文献 [2] 进行改进,都有待于进一步的研究。

5 参考文献

[1] 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.

[2] 程小扬,丁树良,严深海,等.引入曝光因子的计算机化自适应测验选题策略[J].心理学报,2011,43(2):203-212.

[3] Meijer R R, Nering M L. Computerized adaptive testing: overview and introduction[J]. Applied Psychological Measurement, 1999, 23(3):187-194.

[4] 刘珍,丁树良,林海菁.基于 GPCM 的 CAT 选题策略比较[J].心理学报,2008,40(5):618-625.

- [5] Chang Hua-Hua ,Qian Jiahe ,Ying Zhiliang. Alpha-stratified multistage computerized adaptive testing [J]. Applied Psychological Measurement ,1999 ,23( 23) : 211-222.
- [6] 李萍 ,甘登文 ,丁树良 . 自动控制区分度作用的选题策略研究 [J]. 江西师范大学学报: 自然科学版 ,2013 ,37( 1) : 101-105.
- [7] 戴懿 ,甘登文 ,丁树良 . 结合影子题库的选题策略 [J]. 江西师范大学学报: 自然科学版 ,2013 ,37( 6) : 657-660.
- [8] 陈平 ,丁树良 ,林海菁 . 等级反应模型下计算机化自适应测验选题策略 [J]. 心理学报 ,2006 ,38( 3) : 461-467.
- [9] 李佳 ,丁树良 ,方剑英 . 基于平均数形式的选题策略比较 [J]. 江西师范大学学报: 自然科学版 ,2015 ,39( 1) : 69-72.
- [10] 胡姗 ,丁树良 ,程 艳 . CAT 分层终止规则探究 [J]. 江西师范大学学报: 自然科学版 ,2014 ,38( 5) : 445-448.
- [11] 程起航 . 浅谈黄金分割在数学与生活中的应用 [J]. 数理化学学习 ,2008( 5) : 27-29.
- [12] 罗芬 ,丁树良 ,王晓庆 . 多级评分计算机自适应测验动态综合选题策略 [J]. 心理学报 ,2012 ,44( 3) : 400-412.

## The Item Selection Strategy of Raising Item Pool Security

HE Xiang ,LUO Fen ,GAN Dengwen\* ,DING Shuliang ,WANG Wenyi

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

**Abstract:** The safety of item pool is the key point to the implementation of computerized adaptive test ,inspired by the original dynamic  $\alpha$ -stratified method and mean inequality ,constructing another item selection strategy of the dynamic  $\alpha$ -stratified. The simulation experimental results showed that the new item selection strategy to keep the test precision of the original dynamic  $\alpha$ -stratified method and further improves the security of the test.

**Key words:** item information; discrimination; item selection strategy

( 责任编辑: 冉小晓)

( 上接第 353 页)

## The Positive Solutions of a Class of Third-Order Boundary-Value Problems with Derivative Dependence

LI Fengyan<sup>1</sup> ,SHI Jinchuan<sup>2</sup>

( 1. Department of Business and Trade ,College of Liaoning Administration ,Shenyang Liaoning 110161 ,China;

2. Keya College ,Shenyang University of Chemical Technology ,Shenyang Liaoning 110167 ,China)

**Abstract:** By means of the methods of fixed point index computation ,the existence of positive solutions of a class of third-order boundary-value problems with derivative dependence is proved. The boundary-value problems considered possess different order ,boundary conditions and singularity from those in previous works. An example is given to apply the results obtained.

**Key words:** derivative dependence; third-order boundary-value; positive solution

( 责任编辑: 曾剑锋)