

文章编号: 1000-5862(2016)04-0377-05

带区分度约束的选题策略研究

丁加林 熊建华 罗 芬 甘登文* 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: 最大优先级指标(MPI)选题策略可以较好地满足非统计性约束,按 a 分层的选题策略可以有效提高高低区分度项目的利用率,结合两者的优势,构造了附加区分度约束的两阶段MPI选题策略。Monte Carlo模拟研究表明:新选题策略在题库的未使用率方面有明显改进,在测量精度和约束条件控制等评价指标上较现有方法差异不大。

关键词: 非统计约束; 选题策略; a 分层; 最大优先级指标方法

中图分类号: B 841 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2016.04.09

0 引言

计算机化自适应测验(CAT)因其能够因人施测,为每个被试选择最合适的题目,使得被试能力得到准确估计,并且用比纸笔测验更少的题目就能达到相同测量精度等优势而备受关注^[1]。实施CAT测验,选题策略是重中之重,成为关注的焦点。由于传统的CAT中最大信息量选题策略^[1]使得高分度的项目易被重复选择而影响题库的安全性,并且一般CAT注重统计学约束,而一个面向实用的CAT除了满足统计性约束外,还应该考虑满足各种非统计性约束^[2]。这些非统计性约束一般包括:内容平衡(即从每一个内容范围中选择确定数量的题目)、答案平衡(即正确答案应该近似均匀的分布在A、B、C、D的4个选项当中)和允许在一个测验当中有有限个特殊项目,例如含有消极字词(比如题干中出现“下面哪个选项不成立”的字眼)的项目,还有一种特殊的非统计约束是控制项目的曝光。

目前,CAT中控制非统计性约束的常用选题策略是加权离差模型(WDM)和最大优先级指标(MPI)^[3-6]。这2种选题策略,它们可以在一定程度上控制CAT中非统计性约束。Cheng Ying^[7]提出的MPI方法在其实验条件下相较于加权离差方法、最大信息量方法和随机选择方法在若干评价指标下有明显改善,但题库利用率较低。李佳等^[8]和潘奕娆

等^[9]改进的MPI方法与程莹的MPI方法的实验条件相同,所得的题库未使用率依然较高(只使用了题库中不到50%的题目),这就不可避免的造成了资源和财力的浪费。而按 a 分层的选题策略^[2]能有效提高高低区分度项目的使用,即提高了题库的使用率。因此,在改进的MPI方法的基础上,探讨将区分度 a 当作一个非统计性约束以提高题库利用率是有意义的。

1 选题策略研究

1.1 几种非统计性约束选题策略

最大优先级指标方法(MPI)^[3]、基于算术平方根平均数形式的MPI方法(AMPI)^[8]和改进的最大优先级指标方法(MMPI)^[9],每种方法都采用2阶段选题策略^[4]。

1.1.1 MPI方法 MPI方法选取的是对当前能力估计值所提供的信息量与约束条件乘积最大的项目,其目标函数为

$$PI_j = I_j^t \prod_{k=1}^K (w_k f_k)^{c_{jk}},$$

其中 I_j^t 为项目 j 对当前能力估计值所提供的信息量, K 为约束条件的总数, w_k 为约束条件 k 所对应的权重, $f_k = (u_k - x_k) / u_k$ 为约束 k 的缺额比例,在第1阶段 $u_k = l_k$, l_k 为约束 k 的下界, x_k 为在符合约束的项目集合里已经被选择的项目个数;在第2阶段 $u_k =$

收稿日期: 2016-03-29

基金项目: 国家自然科学基金(31500909, 31360237, 31300876, 31160203, 31100756, 30860084), 教育部人文社会科学研究青年基金(13YJC880060)和江西省教育科学2013年度一般课题(13YB032)资助项目。

通信作者: 甘登文(1956-),男,江西奉新人,教授,主要从事计算机辅助教学和统计应用方面的研究。

μ_k μ_k 为约束 k 的上界. $C = (C_{jk})_{J \times K}$ 是项目与约束条件的相关矩阵, 当 $C_{jk} = 1$ 时, 表示项目 j 受到约束条件 k 的约束; 当 $C_{jk} = 0$ 时, 表示项目 j 并未受到 k 的约束. 矩阵 C 通常在选题之前由专家定义. 在选题过程中, 优先选择使目标函数最大的项目.

1.1.2 改进的 MPI 方法 针对 Cheng Ying 提出的最大优先级指标方法 (MPI) 中会出现违反约束条件及偏向选择相对约束条件少的项目的不足, 文献 [9] 提出的改进的 MPI 方法的目标函数为

$$PI_j = I_j^t \prod_{k=1}^K (g_k(w_k f_k + 1))^{C_{jk}},$$

引入 g_k 用于判断约束条件 k 是否达到上下界, 当 $g_k = 0$ 时, 表示约束条件 k 达到了其上下界; 当 $g_k = 1$ 时, 表示约束条件 k 还没有达到其上下界.

1.1.3 算术平方根平均数形式的 MPI 选题策略 文献 [8] 将 MMPI 方法与各类平均数形式相结合, 根据其实验所得结论, 可知采用算术平方根平均数形式的 MPI 选题策略在测量精度、项目曝光均匀性、平均违规次数、题库利用率等方面表现最好, 其目标函数为

$$PI_j = (\lambda (I_j^t)^2 + (1 - \lambda) \cdot ((\prod_{k=1}^K (g_k(w_k f_k + 1))^{C_{jk}})^2))^{1/2}, \quad (1)$$

其中 $\lambda = L_{(j)} / l_{\text{test}}$, λ 是一个从小到大动态变化的变量, 在测验初期先满足内容平衡等非统计约束, 而在测验中后期尽量满足测验信息量的要求. $L_{(j)}$ 为当前测验长度, l_{test} 为测验总长度, 在本文中测验长度 $L = 12$ [7], K 、 I_j^t 、 f_k 以及 w_k 的含义同前所述.

1.2 附加区分度约束的 MPI 新选题策略

按 a 分层的选题策略 [10-12] 是先将题库中的项目按照区分度大小进行排序后, 再根据事先约定将题库分成若干层, 在测试过程中, 根据被试作答情况逐层调用适当数目的难度与当前能力估计值相匹配的项目, 直至满足测验结束条件.

受到计算机化自适应测验选题策略的启发, 上述选题策略题库中未曝光项目数过高的原因是 Fisher 信息量和区分度的平方成正比, 在 MPI 中区分度的影响也较大, 参照文献 [7], 本文探讨增加区分度 a 作为一个约束条件的方法, 结合文献 [8] 和文献 [12] 的研究成果, 令目标函数为以下的形式时可以提高题库的利用率, 即

$$PI_j = (\lambda (I_j^t / a_j^{(2-2\sqrt{I_j^t/16}))^2 + (1 - \lambda) \cdot ((\prod_{k=1}^K (g_k(w_k f_k + 1))^{C_{jk}})^2))^{1/2}. \quad (2)$$

此外, 文献 [7] 指出, 可以通过将 λ 变换成不同

的形式以满足不同的测验要求, 当测验要求较高的测量精度时, 可以让 λ 变化的更快一些; 当测验要求要控制曝光率时, 就可以让 λ 变化的缓慢一些. 经过进行多次实验, 发现当 λ 变换成正弦函数的形式时, 题库的曝光率是最均匀的. 因此, 为了更好地控制曝光率, 令 $\lambda_1 = \sin(\lambda)$, $\lambda \in [0, 1]$ 其中 $\lambda = L_{(j)} / l_{\text{test}}$, 则 (1) 式变为如下目标函数

$$PI_j = (\lambda_1 (I_j^t)^2 + (1 - \lambda_1) \cdot ((\prod_{k=1}^K (g_k(w_k f_k + 1))^{C_{jk}})^2))^{1/2}. \quad (3)$$

再结合 (2) ~ (3) 式, 可将目标函数变为

$$PI_j = (\lambda_1 (I_j^t / a_j^{(2-2\sqrt{I_j^t/16}))^2 + (1 - \lambda_1) \cdot ((\prod_{k=1}^K (g_k(w_k f_k + 1))^{C_{jk}})^2))^{1/2}. \quad (4)$$

以 (2) ~ (4) 式为目标函数的选题策略分别称为新方法 1、新方法 2 和新方法 3.

2 模拟研究

2.1 被试及题库模拟

为了试验结果可以与文献 [7] 中的研究结果做对比, 采用文献 [7] 题库需求, 即整个题库的题目要覆盖初等代数的 15 个考核内容. 每个测验要覆盖所有考核内容 ($C_1 \sim C_{15}$) 要求最多考核一个不等式 (INQ) 和 2 个敌对题 (NEG), 答案为 A、B、C、D 的项目 (Key A ~ Key D) 要分别有 2 ~ 8 个, 在第 1 层区分度 a_1 的项目有 2 ~ 6 个, 在其它 3 个区分度层的项目区分度 ($a_2 \sim a_4$) 要分别有 2 ~ 8 个, 每个约束的权重选择同文献 [7]. 同文献 [7], 采用 3PLM 模型, 使用的题库由模拟产生的 642 道题目组成, 项目参数分别服从 $a \sim U(0.5, 2.5)$, $b \sim N(0, 1)$, $c \sim \text{Beta}(5, 17)$. 共模拟生成 39 000 名被试, 且被试的能力真值服从标准正态分布, 即 $\theta \sim N(0, 1)$. 依据题库的项目参数情况以及为了方便研究分析本文将 a 分成 4 层, 第 1 层 a_1 的取值范围是 (0.5, 1.0), 第 2 层 a_2 的取值范围是 (1.0, 1.5), 第 3 层 a_3 的取值范围是 (1.5, 2.0), 第 4 层 a_4 的取值范围是 (2.0, 2.5), 再将 a 作为一个约束加入到题库当中, 即可得到如表 1 所示的新题库.

2.2 模拟 CAT 的施测过程

本文讨论 3PLM 模型下的满足非统计学指标的选题策略, 测验长度为 12 的定长测验, 假设被试的初始能力估计值为 0, 计算题库中所有题目的信息量, 然后在 10 个信息量最大的题目当中随机选择一

个作为第1道题, 接下来的项目根据不同的选题策略在使目标函数最大的2个项目当中随机选择一个, 采用EAP方法对能力进行估计. 进行以下的CAT测验: (i) 仅在原始题库上, 使用MPI、MMPI、AMPI、新方法1、新方法2和新方法3共6种选题策略进行CAT测验; (ii) 将1中提到的6种选题策略分别应用于原始题库和添加了约束 a 的新题库中进行CAT测验.

表1 题库组成

约束条件	权重	下界	上界	可用项目数
C_1	5.0	0	1	28
C_2	15.0	1	1	58
C_3	10.0	0	1	31
C_4	10.0	0	1	25
C_5	20.0	1	1	73
C_6	15.0	1	1	50
C_7	15.0	1	1	63
C_8	5.0	0	1	22
C_9	15.0	1	1	36
C_{10}	15.0	1	1	26
C_{11}	15.0	1	2	71
C_{12}	5.0	0	1	18
C_{13}	5.0	0	1	23
C_{14}	10.0	1	1	92
C_{15}	5.0	0	1	26
NEG	5.0	0	2	7
INQ	5.0	0	1	19
Key A	0.5	2	8	164
Key B	0.5	2	8	156
Key C	0.5	2	8	158
Key D	0.5	2	8	164
a_1	1.0	2	6	129
a_2	1.0	2	8	160
a_3	1.0	2	8	183
a_4	1.0	2	8	190

2.3 评价指标

本文用能力估计准确性、平均违背率、曝光控制(项目从未曝光率、项目过度曝光率)等评价指标来评价比较选题策略的优劣, 除了相关系数越大越好外, 其它的指标都是越小则表示选题策略越好.

1) 能力估计准确性. 能力总体偏差 $B_{\text{ias}} = \sum_{i=1}^N (\hat{\theta}_i - \theta_i) / N$; 能力均方误差 $M_{\text{SE}} = \sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2 / N$; 相关系数

$$\rho_{\theta, \hat{\theta}} = \left(N \sum_{i=1}^N (\hat{\theta}_i \theta_i) - \sum_{i=1}^N \theta_i \sum_{i=1}^N \hat{\theta}_i \right) /$$

$$\left(\sqrt{N \sum_{i=1}^N \theta_i^2 - \left(\sum_{i=1}^N \theta_i \right)^2} \sqrt{N \sum_{i=1}^N \hat{\theta}_i^2 - \left(\sum_{i=1}^N \hat{\theta}_i \right)^2} \right),$$

其中 N 为被试总人数, θ_i 为第 i 个被试的能力真值, $\hat{\theta}_i$ 为第 i 个被试的能力估计值. 能力误差均值和均方差越小说明能力估计越准确, 相关系数 $\rho_{\theta, \hat{\theta}}$ 反映了被试能力真值与其能力估计值的相关性, 若其值越大, 则说明能力估计越准确.

2) 曝光控制. 主要包括项目从未曝光率(即在整个测验当中都没有被被试选中的项目与题库项目总数的比值)、项目过度曝光率(一般指曝光率超过0.2的项目所占题库总数的比例). 项目从未曝光率 $N_{\text{everExpo}} = \text{sum}(I_{\text{temUseTime}} = 0) / M$, 项目过度曝光率 $O_{\text{verExpo}} = \text{sum}(I_{\text{temUseTime}} > 0.2) / M$, 其中 M 为题库的总题数, $I_{\text{temUseTime}}$ 是题库中每个项目在整个测验当中的使用次数.

3) 满足约束条件情况. 主要使用平均违背率(\bar{V})来表示约束条件被满足的情况, 即为每个被试在测验当中违背的约束条件的平均值, 它的表达式为 $\bar{V} = \sum_{i=1}^N V_i / (NK)$, 其中 K 为约束条件的总数, 当选择原始题库(没有加入 a 分层的题库)时, $K = 21$; 当选择将区分度作为约束的题库时, $K = 25$, V_i 为被试 i 在测验中违背约束条件的次数.

3 结果分析

每个实验重复10次, 得到图1. 图1描述的是各方法在未融入 a 分层前, 在题库使用率和平均违背率指标上的情况; 表2为各方法在融入 a 分层前后的各评价指标值.

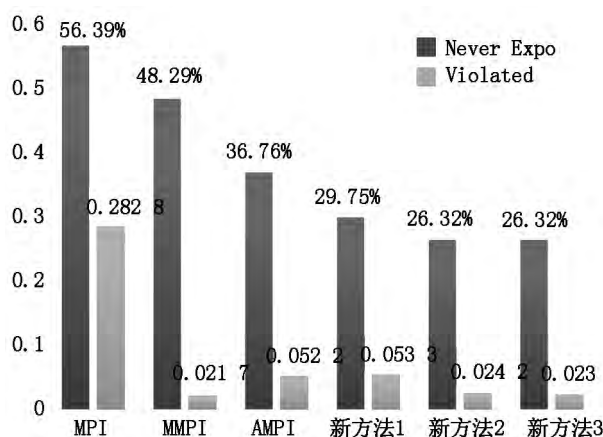
图1 未加 a 前各方法的题库利用率和平均违背率

图1表明, 未融入 a 分层的各选题策略在 Never

Expo 和 \bar{V} 指标上,新方法均好于或相当于已有的 3 种选题策略.新方法 2 和新方法 3 在各项指标上略好于新方法 1.

从表 2 中,通过分别比较 MPI 方法、MMPI 方法以及 AMPI 和它们各自融入了 a 分层思想后构造的方法.可以看出融入 a 分层的思想之后,在题库的未曝光率方面均有了明显的改善,其他的评价指标上差别不大.

综上所述,各方法在融入 a 分层思想后,在题库的未曝光率这一评价指标上都有了明显的改善;本

文中新提出来的附加区分度约束的 MPI 选题策略与 MPI 方法、MMPI 选题策略相比较,可能在约束条件平均违背率上会有所增加,这是由于约束条件的增加,违背的可能也就更大(约束条件增加了 19%,违背率大约增加了 10.8%,表 2 显示,增加约束条件后,过度曝光增加的百分比较大,这表明新方法恶化了过度曝光率且减小了未曝光率,即改善了题库的使用率),但在题库的未曝光率这一评价指标上都有了明显的改善.

表 2 各种方法加 a 与不加 a 的评价指标值

	B_{ias}	M_{SE}	$\rho_{(\theta, \hat{\theta})}$	Over Expo	Never Expo	\bar{V}
MPI	0.001 5	0.061 2	0.968 0	0.018 7	0.563 9	0.294
MPIa	0.001 4	0.084 0	0.955 6	0.033 9	0.182 2	0.322
MMPI	0.006 8	0.074 1	0.960 9	0.051 4	0.482 9	0.022
MMPIa	0.001 3	0.089 4	0.952 7	0.020 9	0.160 4	0.041
AMPI	0.001 8	0.075 4	0.958 5	0.014 0	0.365 6	0.052
AMPIa	0.002 1	0.114 0	0.961 7	0.029 8	0.160 4	0.062
方法 1	0.004 1	0.107 1	0.947 1	0.095 0	0.297 5	0.053
方法 1a	-0.005 5	0.117 3	0.941 9	0.190 0	0.141 7	0.058
方法 2	0.001 4	0.070 6	0.965 2	0.109 0	0.263 2	0.024
方法 2a	0.009 9	0.087 5	0.956 7	0.218 1	0.129 3	0.035
方法 3	0.005 8	0.073 5	0.964 0	0.119 9	0.263 2	0.023
方法 3a	0.002 7	0.093 7	0.954 0	0.244 5	0.126 2	0.030

注:后面没有添加 a 表示没有使用附加区分度约束,添加 a 表示使用了附加区分度约束.

4 总结与讨论

本文通过尝试着将 a 分层的思想融入到 MPI 选题策略当中,以及令目标函数当中的 $\lambda_1 = \sin\lambda$,其中 $\lambda \in [0, 1]$,使得题库的使用率得以提高.通过与之前所提出来的满足非统计性约束的选题策略(主要是 MPI 选题策略和在其基础上改进的选题策略)相对比,所得实验结果表明:

1) 在未融入 a 分层思想之前,将目标函数(1)式中的 I_j^t 替换为 $I_j^t / a_j^{(2-2\sqrt{I_j^t/16})}$,能够提高题库的利用率.令目标函数(3)式中的 $\lambda_1 = \sin\lambda$, $\lambda \in [0, 1]$,得到的新方法 2 在题库的未曝光率这一指标上都要优于其他的方法,且相较于以(4)式为目标函数的新方法 3,2 种方法在这一指标上几乎一样.

2) 新提出的 MPI 选题策略能够较好地满足了各种约束条件.所有的新方法在违背率这一指标上都达到了千分之几,且都要小于原来的 MPI 选题策略在这一指标上的数值.

3) 融入 a 分层思想后,各选题策略在题库的使

用率上均得到较大改善,新提出的选题策略在题库的使用率上也得到了进一步的提高,题库的使用率基本上都提高了 10% 左右,甚至达到了 20% 以上,这一点主要表现在从未曝光率(Never Expo)这一指标上.

这样就使得之前提出来的 MPI 选题策略以及其改进的选题策略所出现的弊端即题库的利用率不高得到较好的改善,避免了财力物力资源的浪费.综合以上实验结论,附加区分度 a 后的新方法 3 可能更加适合运用于实际工作当中,但是在最大曝光率等方面比文献[7]中所提出来的 MPI 选题策略可能略差,另外,本文中考虑的是将定长作为终止条件,能否在不定长的测验当中使用也有同样的效果?如何设定新策略的权?能否用于多级评分项目的 CAT?将在以后的研究中作进一步探讨.

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,2002:254-255.
- [2] Cheng Ying, Chang Hua-Hua, Douglas J, et al. Constraint-

- weighted a -stratification for computerized adaptive testing with nonstatistical constraints [J]. Educational and Psychological Measurement 2009 69(1): 35-49.
- [3] Cheng Ying ,Chang Hua-Hua. The maximum priority index method for severely constrained item selection in computerized adaptive testing [J]. British Journal of Mathematical and Statistical Psychology 2009 62(2): 369-383.
- [4] Cheng Ying ,Chang Hua-Hua ,Yi Qing. Two-phase item selection procedure for flexible content balancing in CAT [J]. Applied Psychological Measurement 2007 31: 467-482.
- [5] 汤楠. 最大优先级指标选题策略的相关研究[D]. 南昌: 江西师范大学 2012.
- [6] 汤楠 ,丁树良 ,余丹. 最大优先级指标下选题策略的研究[EB/OL]. [2015-11-10]. <http://cpfd.cnki.com/Article/CPFDTOTAL-ZGXG201110001968.htm>.
- [7] Cheng Ying. Computerized adaptive testing-new developments and applications [D]. Illinois: University of Illinois at Urbana-Champaign 2007.
- [8] 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.
- [9] 潘奕尧 ,丁树良 ,尚志勇. 改进的最大优先级指标方法 [J]. 江西师范大学学报: 自然科学版 2011 35(2): 213-215.
- [10] Chang Hua-Hua ,Ying Zhiliang. A global information approach to computerized adaptive testing [J]. Applied Psychological Measurement 20 213-229.
- [11] 艾国金 ,甘登文 ,丁树良. 计算机化自适应认知诊断测验按模式分层选题策略 [J]. 江西师范大学学报: 自然科学版 2014 38(3): 270-274.
- [12] 程小扬 ,丁树良. 子题库题量不平衡的按 a 分层选题策略 [J]. 江西师范大学学报: 自然科学版 2011 35(1): 5-9.
- [13] 程小扬 ,丁树良. 拓广分部评分模型下计算机自适应测验变加权选题策略 [J]. 心理科学 2011 34(4): 965-969.

The Item Selection Strategies with Discrimination Constraint

DING Jialin ,XIONG Jianhua ,LUO Fen ,GAN Dengwen* ,WANG Wenyi

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

Abstract: MPI can well meet the statistical constraints ,and a -stratified method can effectively improve the utilization rate of low discrimination item. Combining the advantages of MPI and a -stratified method ,a two-phase MPI item selection strategy with additional distinction constraint is constructed. The simulation study of Monte Carlo shows that the new item selection strategy has improved a lot in the inavailability of item bank ,which is about the same as the existing approach in measurement accuracy ,constraint management and other evaluation in dices.

Key words: the statistical constraints; item selection strategy; a -stratified method; maximum priority index method

(责任编辑: 冉小晓)