文章编号: 1000-5862(2012)05-0452-04

一阶段选题的最大优先级指标的修正

汤　楠, 丁树良*

(江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 为了平衡测验中限制条件而提出最大优先级指标(MPI)一阶段选题方法可能出现越界, 对该情况进行分析, 并给出修正方案, 修正后基本解决了限制条件越界的问题.

关键词: 最大优先级指标; 一阶段; 选题策略; 计算机化自适应测验

中图分类号: O 626.4　　文献标志码: A

1　问题的提出

计算机自适应测验(CAT)是将计算机技术运用于教育测量的一种新方法, 能够更智能地针对被试选题[1]. CAT 主要包括以下几个方面: 题库、选题策略、终止规则以及能力估计程序. 这几个组件当中, 最关键的是选题策略, 它能够根据被试作答的反应从题库中选取合适的项目施测, 对于较高水平的考生, 调用的题库中一部分难度相对较高的项目, 而对于较低水平的考生则是选取一部分难度较低的项目. 此外它需要兼顾的条件较多,如曝光率、测验精度、限制条件等, 其中带限制条件相关方法的选题策略是针对测验要求来设计的, Chang Huahua 等[2]于 2003 年提出按内容分层(c-STR)的选题策略, 这种选题策略主要是针对内容平衡, 如考试的内容有可能涉及到多个章节, 这样就不能用传统的选题策略从题库中选取项目, 而应在选题的过程中从各章中抽取项目, 从各章中选取多少题可以看成是一种约束条件, 其主要思想是按内容分层, 之后从各层中选题. Cheng Ying 提出了最大优先级指标(MPI)[3-4], 针对限制条件(内容也可以是限制条件的一种)提出了上下界的概念, 并先后提出了两阶段选题和一阶段选题 2 种方法. 但是, 在实验过程中发现 MPI 一阶段选题可能存在越界, 本文对该问题进行分析并给出修正方案, 修正后的方法基本上避免了越界情况的发生.

1.1　MPI 一阶段简介

假定 $\boldsymbol{C}=(c_{jk})$是 1 个 J 行 K 列的限制关系矩阵, J 是项目总数, K 是限制条件数, 其中 $c_{jk}=1$ 表示限制条件 k 与项目 j 相关, $c_{jk}=0$ 表示不相关. 矩阵 $\boldsymbol{C}$ 通常预先由该领域的专家给出.

记μ_k 为内容域 k(k=1, 2, ⋯, K)选择的项目数, 其中 K 为限制条件总数,且满足如下 2 个条件

$$l_k \leqslant \mu_k \leqslant u_k, \sum_{k=1}^{K}\mu_k = L,$$

l_k 为限制条件 k 的下界, u_k 为限制条件的上界.

MPI 两阶段选题给出了 2 个最大优先级指标, 一个是关于约束的下界, 另一个是关于约束的上界. 先处理所有下界以后, 且作答的项目尚未达到测验规定长度, 这时再考虑各个约束的上界. 而 MPI 一阶段选题只给出一个优先级指标, MPI 一阶段选题指标为

$$PI_j^t = I_j^t \prod_{k=1}^{K}(f_{1k} f_{2k})^{c_{jk}},$$

其中 K 为限制条件的个数, I_j^t 代表项目 j 在被试当前能力值下第 t 个项目的 Fisher 信息量. f_{1k} 为限制条件 k 与上界的距离, 具体形式为 $f_{1k}=(u_k - x_k - 1)/u_k$, 而 f_{2k} 的计算公式为 $f_{2k}=((L-l_k)-(T-x_k))/(L-l_k)=1-(T-x_k)/(L-l_k)$, 其中$T$ 为已经选择的项目数, L 为测验长度, x_k 为已选择满足限制条件 k 的项目数.

收稿日期: 2012-05-26

基金项目: 国家自然科学基金(30860084, 31160203, 311007561, 60263005), 国家社会科学基金(12BYY055), 国家教育科学规划(CCA110109), 教育部人文社科(09JJCXLX012, 10YJCXLX049, 11YJC190002), 江西省教育厅科技计划(GJJ11385, GJJ10238), 全国教育考试科研规划课题(2009JKS2009)和高等学校博士学科点专项科研基金(20113604110001)资助项目.

作者简介: 丁树良(1949-), 男, 江西樟树人, 教授, 博士生导师, 主要从事计算机辅助教学、应用及教育和心理测量方面的研究.

$T-x_k$ 是已选项目中与约束条件 k 无关的项目数, $L-l_k$ 是测验中除约束条件 k 外还必须选取的项目数, 故 $(T-x_k)/(L-l_k)$ 是与约束 k 无关的缺额比. 于是 f_{2k} 是约束 k 在满足下界后的缺额比. 在施测过程中, 选取剩余题库中 PI_j 值最大的项目给被试作答.

1.2　MPI 一阶段可能出现的问题及其原因

在 Cheng Ying[3-4] 的一阶段指标中, $f_{2k}=((L-l_k)-(T-x_k))/(L-l_k)=1-(T-x_k)/(L-l_k)$. 对于先选择的那一个限制条件, 通过其下界, 可能会优先达到其上界. 在此过程中, 选择其他限制条件的概率较小, 如果一个测验的各限制条件上下限差距较大, 则易使得后选择的限制条件出现越界的现象. 举个简单例子: 对于限制条件 P、Q 来说, P、Q 的上下界均为(2, 4)、(2, 4), 即 l=2, u=4. 题库中带有限制约束条件的项目分别为(1, 0)、(0, 1)各 4 题, 测验长度固定为 L=4. 即限制关系矩阵 $\boldsymbol{C}^{\mathrm{T}}=\begin{pmatrix}1&1&1&1&0&0&0&0\\0&0&0&0&1&1&1&1\end{pmatrix}$. 本例中本来只有(1, 0)和(0, 1)各选而题才能满足条件. 对于第 1 题的选取, 由于在测验开始阶段进行能力的粗略估计, 通常随机选取一题, 这里选取(1, 0). 在选第 2 题时, 计算 $f_{11}=(u_1-x_1-1)/u_1=(4-1-1)/4=0.5$, $f_{21}=1-(T-x_1)/(L-l_1)=1-(1-1)/(5-3)=1$,接下来计算各项目 PI_j 值, 这里暂时不考虑信息量的影响. 对于(1,0)的项目 $PI_j=I\times(0.5\times1)^1\times(0.75\times0.5)^0=0.5\times I$, 对于(0, 1)的项目有 $PI_j=I\times(0.5\times1)^0\times(0.75\times0.5)^1=0.375\times I$, 结果会选(1, 0)的项目. 在选取第 3 题的过程中, 定义 $0^0=1$, 则同理可求 $f_{11}=(4-2-1)/4=0.25$, $f_{21}=1-(2-2)/(4-2)=1$, $f_{12}=(4-0-1)/4=0.75$, $f_{22}=1-(2-0)/(4-2)=0$ 对于(1, 0)的项目 $PI_j=I\times(0.25\times1)^1\times(0.75\times0)^0=0.25\times I$(由于 0^0 无意义, 这里假设为 1), 对于(0, 1)的项目 $PI_j=I\times(0.25\times1)^1(0.75\times0)^0=0\times I$,结果同样会选取(1, 0)的项目. 在本例中, 已经选取了 3 道(1, 0)的题目, 由于测验长度固定 L=4, 则第 2 个约束条件已经不可能达到下界了(l_2=2), 这必将导致该测验出现越界情况. 而且对第 4 题的选取, $f_{11}=(4-3-1)/4=0$, $f_{21}=1-(3-2)/(4-2)=0.5$, $f_{12}=(4-0-1)/4=0.75$, $f_{22}=1-(3-0)/4-2=-0.5$, 此时竟出现了负值, 必然与 Cheng Ying 的本意不符. 所以按原 MPI 方法选题时, 对于优先选择的限制条件, 如果信息量相差不大, 则选取下一个项目时仍可能会优先考虑含有此限制条件的项目, 这有可能导致其他限制条件越界.

2　模拟实验

因为考虑到处理下界的 f_{2k} 存在问题, 对一阶段 PI_j 中的 f_{2k} 进行修改, 修改后的 f_{2k} 指标有 2 个版本, 它们分别为

$$f_{2k}=(T-x_k)/(L-l_k)\,,\qquad \text{(修正 1)},$$

$$f_{2k}=(T-x_k)/(L-l_k)+1\,,\quad \text{(修正 2)}.$$

本实验通过模拟 3 种方法(原 MPI 一阶段, 修正 1 及修正 2)在以上 3 种评价指标上进行对比, 模拟次数为 3. 此外, 在试验过程中, 会随机抽取 5 名被试最终的违规信息在 3 种方法中进行对比.

2.1　实验数据的模拟

本试验采用 GRM 模型, 由于包含限制条件的题库, 故采用定长模拟实验, 测验长度固定为 40. 模拟生成 1 000 个项目, 区分度参数 a 服从对数正态分布, 即 $\ln a\sim N(0,1)$, 等级难度参数 $b_{jt}\sim N(0,1)$. 生成约束条件个数 K=3, 其上下界分别为(10, 14)、(15, 19)、(12, 16), 考虑到 Cheng Ying 原方法中对于题库的限制关系矩阵 $\boldsymbol{C}$ 种类不明显, 本实验生成以下 2 种 $\boldsymbol{C}$ 矩阵: (1)限制关系矩阵 $\boldsymbol{C}$ 为每个项目只对应 1 个限制条件, 如对于总共 3 个限制条件来说, 只能生成(1, 0, 0)、(0, 1, 0)、(0, 0, 1) 共 3 种项目; (2)限制关系矩阵 $\boldsymbol{C}$ 为每个项目对应 1 个或多个限制条件, 如可以生成(1, 1, 1)、(1, 0, 1)、和(1, 0, 0)的项目. 同时模拟 1 000 个被试, 能力真值 $\theta\sim N(0,1)$, 对能力估计采用 EAP 方法.

2.2　评价指标

本文分别从能力估计准确性, 卡方检验统计量和平均违规次数[5-7]对 2 种题库下的 3 种方法进行评价, 具体如下:

(1)能力估计准确性:

$$Recovery=\frac{1}{C}\sum_{j=1}^{C}\left(\sum_{i=1}^{N}\left|\theta_i-\hat{\theta}_{ij}\right|\Big/N\right),$$

其中 θ_i 为被试 i 的能力真实值, $\hat{\theta}_{ij}$ 代表被试 i 在进行第 j 次模拟实验得到的能力估计值. *Recovery* 指标反映了能力真值与能力估计值的平均绝对偏差. *Recovery* 越小, 能力估计越准确.

(2)卡方检验统计量:

$$\chi^2=\sum_{j=1}^{M}\left\{\left[A_j-\left(\sum_{j=1}^{M}A_j/M\right)\right]^2\Bigg/\left(\sum_{j=1}^{M}A_j/M\right)\right\},$$

χ^2 检验统计量值越小, 表明题库曝光率控制更均匀,

安全性则更高.

(3)平均违规次数:

$$\bar{V}=\sum_{i=1}^{M}V_i/N,$$

其中 V_i 表示被试 i 在测验中违背约束条件的总数, N 为被试总数. $\bar{V}$ 指标综合反应了所有被试的违规情况, 值越小则所有被试的违规次数越少.

2.3 实验结果及分析

实验结果如表 1~表 8 所示. 在表 1 和表 2 中, 各方法的能力估计准确性与曝光率控制差别不大, 在此不做深入研究. 从表 1 的平均违规次数 1 列可以看出, 在第 1 种题库下, 原 MPI 方法越界率很高. 对于表 3 中的限制条件 1 这列, 基本上都已经达到甚至超过上界, 分析其原因, 可能当某个限制条件达到上界之后, $f_{2k}=1-(T-x_k)/(L-l_k)$ 可能出现小于 0 的情况, 即 $T-x_k<L-l_k$. 举个例子, 如当含有限制条件 1 和 3 的项目一直被选择, 到了上界之后, $f_{22}=1-(T-x_2)/(L-l_2)=1-(14+16-0)/(40-15)<0$. 再如 $f_{11}>0, f_{21}\geqslant 0, f_{12}>0, f_{22}<0$, 此时说明限制条件 1 已经达到上界, 限制条件 2 还未及下界. 对于项目 $p(0, 1)$ 和 $q(1, 0)$ 来说, $PI_p=I_j^t\prod_{k=1}^{K}(f_{1k},f_{2k})^{c_{jk}}<0$, $PI_q>0$, 从而优先选择项目 q, 可是限制条件 1 已经达到了上界, 本不应该继续选择. 这也就是说某个限制条件达到了上界之后依然会选取包含这个限制条件的项目, 这样就产生了溢出的现象. 然而基于限制条件平衡方法的实验终止规则都是通过固定测验长度来实现, 在第 1 种题库的条件下, 某些限制条件超出上界范围过多之后, 则该被试在测验结束时, 其他限制条件没达到下界的情况是理所应当的.

通过表 2 可以看出, 第 2 种题库下的原 MPI 方法越界率很高, 第 3 种方法(修正 2)也会出现小范围越界的情况, 而第 2 种方法越界率为 0. 所以修正后的方法基本能解决原 MPI 方法一阶段中越界的问题.

表 1 第 **1** 种 ***C*** 下各方法的表现

3 种方法	能力估计准确性	卡方检验统计量	平均违规次数
$f_{2k}=1-\dfrac{T-x_k}{L-l_k}$	0.104	428. 0	1.72
$f_{2k}=\dfrac{T-x_k}{L-l_k}$ (修正 1)	0.094	461.2	0.03
$f_{2k}=1+\dfrac{T-x_k}{L-l_k}$ (修正 2)	0.093	404.1	0.01

表 2 第 **2** 种 ***C*** 下各方法的表现

	能力估计准确性	卡方检验统计量	平均违规次数
$f_{2k}=1-\dfrac{T-x_k}{L-l_k}$	0. 100	481.2	1.80
$f_{2k}=\dfrac{T-x_k}{L-l_k}$ (修正 1)	0.099	537.5	0
$f_{2k}=1+\dfrac{T-x_k}{L-l_k}$ (修正 2)	0. 095	468.7	0.16

表 3 第 **1** 种 ***C*** 下原 **MPI** 方法被试作答情况

被试编号	约束 1 (10,14)	约束 2 (15,19)	约束 3 (12,16)
1	14	18	8
2	15	8	17
3	14	9	17
4	14	19	7
5	14	10	16

表 4 第 **2** 种 ***C*** 下原 **MPI** 被试作答情况

被试编号	约束 1	约束 2	约束 3
1	16	10	17
2	16	11	17
3	17	12	18
4	17	12	17
5	18	12	17

表 5 第 **1** 种 ***C*** 下修正 **1** 被试作答情况

被试编号	约束 1	约束 2	约束 3
1	12	15	13
2	12	15	13
3	12	15	13
4	12	15	13
5	12	15	13

表 6 第 **2** 种 ***C*** 下修正 **1** 被试作答情况

被试编号	约束 1	约束 2	约束 3
1	12	16	14
2	12	16	14
3	12	16	14
4	12	16	14
5	12	15	13

表 7 第 **1** 种 ***C*** 下修正 **2** 被试作答情况

被试编号	约束 1	约束 2	约束 3
1	12	15	13
2	12	15	13
3	11	16	13
4	12	16	12
5	12	15	13

表 8　第 2 种 C 下修正 2 被试作答情况

被试编号	约束 1	约束 2	约束 3
1	14	18	16
2	13	18	15
3	13	17	15
4	14	18	15
5	14	18	16

3　讨论

本文基于 GRM 模型基础之上, 修正的方法在越界率方面比原方法有较大的改善, 对 GPCM 模型中此方法是否同样适用值得深入探索. 从本文的实验结果可以看出 3 种方法的曝光率都较高, 这将严重威胁到题库的安全性. 在修正后的方法中怎样添加曝光控制的相关方法值得研究.

本文的修正方法虽然基本解决了越界的情况, 但仍有发生很少的越界现象, 并且不稳定, 可见修正方法还是不完美. 怎样才能“根治”越界的问题值得深入研究.

4　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] Yi Qing, Chang Huahua. *a*-stratified CAT design with content blocking [J]. British Journal of Mathematical & Statistical Psychology, 2003, 56: 359-378.

[3] Cheng Ying, Chang Huahua. 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 Huahua, Douglas Jeffery, et al. Constraint-weighted a-stratification for computerized adaptive testing with nonstatistical constraints [J]. Educational and Psychological Measurement, 2009, 69(1): 35-49.

[5] 陈平, 丁树良, 林海菁, 等. 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.

[6] 戴海琦, 陈德枝, 丁树良. 多级评分题计算机自适应测验选题策略比较 [J]. 心理学报, 2006, 38(5): 778-783.

[7] 潘奕娆, 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.

Amendment on Maximum Priority Index in One Phase Strategy

TANG Nan, DING Shu-liang*

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

Abstract: The maximum priority index in one-stage strategy may appear to cross the border. This situation is analyzed and the rectification for it is given. The amendment basically solves the problem beyond the restrict conditions.

Key words: maximum priority index; one-phase; strategy; computerized adaptive test

(责任编辑: 冉小晓)