

文章编号: 1000-5862(2021)02-0118-08

双题匹配的 MPI 方法及其在平行试卷生成中的应用

汪文义¹ 熊 建¹ 宋丽红² 郑娟娟¹ 胡海洋¹

(1. 江西师范大学计算机信息工程学院 江西 南昌 330022; 2. 江西师范大学教育学院 江西 南昌 330022)

摘要: 大规模考试的公平性备受关注,探索能够生成多份平行测验并能保证测验质量的组卷方法十分重要。原用于计算机自适应测验(CAT)的基于最大优先级指标已开始用于组卷,但只局限于单题之间的匹配。为了增加求解空间,该文提出了一种基于最大优先级指标的双题匹配的组卷方法,实现题库中双题与种子试卷双题之间匹配,并结合平行试卷质量调整方法,用于生成平行试卷。将已有的组卷方法与双题匹配方法进行实验对比,结果表明:双题匹配方法有效地提升了生成的平行测验质量。

关键词: 平行测验; 组卷方法; 计算机自适应测验; 最大优先级指标; 双题匹配; 单题匹配

中图分类号: B 841 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2021.02.02

0 引言

近年来,人们越来越关注大规模考试的公平性,伴随而来的是人们对测验的要求越来越严格。生成的测验在保证质量和数量的同时,还要保证不生成包含重复题目的试卷。随着计算机技术的快速发展,自动组卷方法已成为研究热点。研究人员对自动组卷方法进行了大量的研究^[1-4],如基于遗传算法^[5-6]和蚁群算法^[7-8]等组卷方法。所谓自动组卷方法,就是基于专家标定的测验蓝图由计算机自动生成满足要求的试卷^[9]。通常,需要使用一些优化方法,在给定题库的计量学指标(如项目参数)和内容规范等约束下,得到与种子试卷“相似”的多份平行测验。因为组卷问题是 NP 难问题,所以一般采用启发式算法进行组卷,并缓解后期生成试卷质量下滑的问题^[10-11]。

在项目反应理论模型下,研究人员基于匹配项目参数的方法从题库中寻找与种子试卷的题目匹配的题目生成试卷^[12-14]。例如,原用于计算机自适应测验(CAT)的最大优先级指标(Maximum Priority Index Method, MPI)^[15-17]已开始用于生成平行试卷^[18]。原 MPI 方法主要用于计算机化自适应测验,

因此基本上是序贯选择(sequential selection)单题。用于生成平行试卷的 MPI 方法,也是通过计算种子试卷中的单题和题库中单题的距离来生成试卷。组卷问题是 NP 问题,虽然单题匹配的 MPI 方法可提高组卷速度,但是为了兼顾较大解空间搜索和组卷速度,本文提出双题匹配的最大优先级指标方法(Maximum Priority Index Method of Two Items Matching, TIM-MPI),用来生成多份平行试卷。

本文的研究目标是开发出一种最大优先级指标的改进方法,期望它能提升生成的平行测验的质量。

1 平行试卷生成方法

1.1 单题匹配的 MPI 方法

在 CAT 中, MPI 方法是在严格约束条件下选题,用来控制测验的统计和非统计约束,比如试卷内容的平衡(试卷要求在不同内容区域内选题),试卷答题键的平衡(试卷要求答案选项的个数要均匀),试卷曝光控制(防止题目由于多次出现而导致曝光)等。

在 CAT 中, Cheng Ying 等^[15]提出了 MPI 方法:

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

收稿日期: 2020-10-05

基金项目: 国家自然科学基金(62067005, 61967009, 31500909), 汉考国际科研基金(CTI2019B10)和江西师范大学教学改革研究(JXSDJG1848)资助项目。

作者简介: 汪文义(1983—),男,湖南衡山人,副教授,博士,主要从事教育测量与信息处理的研究。E-mail: wenyiwang@jxnu.edu.cn

其中 I_j 为项目 j 在当前被试能力为 θ 时评估得到的费希尔 (Fisher) 信息量 f_k 是在约束 k 下的缺额比例 c_{jk} 表示项目 j 和约束 k 是否相关, 若相关, 则 $c_{jk} = 1$, 否则为 0, w_k 为约束 k 的权重.

在组卷方法中, 由于求费希尔信息量的过程比较复杂, Wang Shiyu 等^[18] 对程莹的 MPI 方法进行了简化, 用距离函数来代替费希尔信息量:

$$M_{PI_j} = \frac{1}{d_{ij}} \prod_{k=1}^K (w_k f_k)^{c_{jk}}.$$

距离函数的思想是在组卷过程中需要让生成试卷和种子试卷之间“相似”, 即要让生成试卷题目和种子试卷题目之间的距离最小. 在项目反应理论中, 可以通过项目难度和区分度来计算距离, 从而选出最“相似”的题目^[13]. 设 a_i 为项目 i 的难度参数, b_i 为项目 i 的区分度参数. 对于项目 i 和项目 j , 它们之间的距离平方可写为

$$d_{ij}^2 = \lambda_1 (a_i - a_j)^2 + \lambda_2 (b_i - b_j)^2,$$

其中 λ_1, λ_2 是难度和区分度统计量的可调整权重, 且 $\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1$. 在组卷的第 1 阶段中, 主要依据 M_{PI_j} 指标依次从题库中寻找与种子试卷中各题最匹配的试题, 并将选中的试题从题库中删除, 依次生成多份试卷. 根据试题的项目参数难度和区分度, 可以得到种子试卷的试题 i_k 和某份试卷的对应试题 j_k 的距离平方, 再对所有试题的距离平方求和:

$$d = \sum_{k=1}^N d_{i_k j_k}^2.$$

在组卷的第 2 阶段中, 主要采用 Wang Shiyu 等^[18] 提出的调整平行试卷之间质量的启发式算法, 在一定的标准和限制下生成多份质量相当的平行测验. Chen Peihua^[19] 应用类似的算法生成平行试卷. 该算法的基本思路是: 首先, 每份平行测验与种子试卷的总距离设置为 0, 要在组装的测验中的 N 个题目的顺序随机确定, 每次根据该顺序从题库中选择具有最大 M_{PI} 值的 T 个题目, 将第 1 组 T 项随机分配给 T 份测验, 并更新每份测验的总测验距离; 然后, 按照 N 个题目顺序依次取出后面的试题, 选择与之匹配的 T 个最佳题目, 按 M_{PI} 值升序排序, 按顺序分配给依总测验距离的降序排列的 T 份测验; 最后, 在分配之后, 更新来自种子试卷的每份平行测验的总距离, 该过程直到为 T 份平行测验的每份测验都选择了 N 个题目时终止.

1.2 双题匹配的 MPI 方法

无论是 Wang Shiyu 等^[18] 的方法, 还是其他的

MPI 方法^[20], 对于 MPI 方法的题目处理还停留在单题匹配. 已有的单题匹配 MPI 方法在匹配题目时, 由于初始选题具有随机性, 所以其选题结果并不是最优解. 针对这一个情况, 本文提出了一种针对 MPI 的双题匹配的组卷方法. 新方法借鉴了混合整数线性规划的思想, 对题目匹配进行组合, 扩大了求解空间. 基于 Wang Shiyu 等^[18] 提出的 MPI 方法, 对距离函数进行了处理, 设给定种子试卷有 1, 2, \dots , N 道题目, 从中随机选择 2 道题 i_1 和 i_2 , 再从题库中随机抽取 2 道题 j_1 和 j_2 , 可以计算种子试卷随机抽取得到 2 道题与题库随机抽取的 2 道题之间的距离平方和. 种子题目和题库题目之间的关系如图 1 所示.

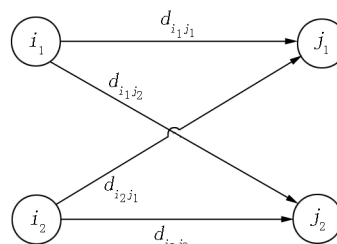


图1 种子题目和题库题目的关系

由于不知种子题目 i_1, i_2 与题库题目 j_1, j_2 之间哪 2 个题目距离更近, 当求双题的距离时, 需要考虑 2 种情况, 第 1 种情况是

$$d_1^2 = \lambda_1 ((a_{i1} - a_{j1})^2 + (a_{i2} - a_{j2})^2) + \lambda_2 ((b_{i1} - b_{j1})^2 + (b_{i2} - b_{j2})^2);$$

第 2 种情况是

$$d_2^2 = \lambda_1 ((a_{i1} - a_{j2})^2 + (a_{i2} - a_{j1})^2) + \lambda_2 ((b_{i1} - b_{j2})^2 + (b_{i2} - b_{j1})^2).$$

当 $d_1^2 < d_2^2$ 时, 选择 d_1^2 作为 TIM-MPI 的距离函数 ($d_{i_1 i_2 j_1 j_2}^2$). 反之就选 d_2^2 作为距离函数 ($d_{i_1 i_2 j_2 j_1}^2$). 因此, 经过计算可得由种子试卷的题目和题库的题目之间的距离组成的距离矩阵. 利用排列组合的方式, 可以穷举种子试卷的双题与题库的双题的组合. 根据距离矩阵, 可以得到种子试卷的题目 i_1, i_2 和题库的题目 j_1, j_2 之间的距离, 并算出各题的 M_{PI}^{TIM} 值:

$$M_{PI_{i_1 i_2 j_1 j_2}}^{\text{TIM}} = \frac{1}{d_{i_1 i_2 j_1 j_2}^2} \prod_{k=1}^K (w_k f_k)^{c_{j_1 k}} \prod_{k=1}^K (w_k f_k)^{c_{j_2 k}},$$

其中 f_k 是在约束 k 下的缺额比例, w_k 是约束 k 的权重, $c_{j_1 k}$ 和 $c_{j_2 k}$ 分别表示项目 j_1 、项目 j_2 和约束 k 是否相关, 与 MPI 方法中的含义相同.

当题库中的某双题与种子试卷的双题的 M_{PI}^{TIM} 值最大时, 将题库中匹配双题的题号按照种子试卷双题的题号顺序放入测验矩阵中, 并标记种子试卷中已匹配的双题, 同时删除题库中已经选择过的双题. 再在题库中继续寻找与种子试卷未标记的双题

匹配且有最大 M_{PI}^{TIM} 值的双题. 重复这个过程, 直到所有的种子试卷的题目在题库中都找到了对应的双题组合为止, 并将匹配的题号都按照种子试卷对应的题号顺序放入测验矩阵中. 此外, 在生成试卷的过程中需考虑到一个特殊情况, 当种子试卷的测验长度为奇数时, 最后一次种子试卷的双题与题库的双题匹配之后, 种子试卷只剩下 1 道题目. 因此, 当种子试卷只有 1 道题没有被标记时, 采用 Wang Shiyu 等^[18] 的方法进行单题匹配, 从题库中匹配到最后一个题目, 并更新题库. 使用以上方法得到种子试卷与题库匹配的测验矩阵后并生成多份试卷, 再通过调整试卷质量的算法生成多份平行测验, 其步骤流程同 MPI 的启发式算法相同.

2 实验模拟研究

2.1 研究设计

基于已有研究设计^[8], 本文主要考虑的实验因素有: 题库大小为 540、内容约束、测验长度为 30、平行测验份数为 5. 各个题目均服从 3PLM(在计算距离时不考虑猜测参数), 区分度参数服从正态分布, 其均值为 1、标准差为 0.3(约束区分度 ≥ 0.2), 难度参数服从标准正态分布, 猜测参数服从 $[0, 0.3]$ 上的均匀分布. 题库中题目涉及 3 方面内容, 其中内容 A、B、C 的题目数量比例分别为 40%、30%、30%. 种子试卷由 30 个题目组成, 其中各内容的题目数量比例也分别为 40%、30%、30%. 重复试验 $R(1\ 000)$ 次, 每次随机生成符合上述要求的题库参数和种子试卷, 分别使用单题匹配 MPI 组卷方法和双题匹配 MPI 组卷方法, 生成 5 份平行试卷. 为直接满足测验的内容约束, 仅直接按单题或双题距离大小从各内容子题库中寻找与种子试卷单题或双题(内容相同)最匹配的试题, 没有采用缺额和权重因子.

2.2 评价标准

主要应用如下评价标准, 对单题匹配的 MPI 组卷方法和双题匹配的 MPI 组卷方法的表现进行比较:

(i) 生成试卷和种子试卷的测验特征曲线、测验信息函数图像接近程度. van der Linden^[5] 认为若 2 份测验的信息函数的曲线十分接近, 即生成试卷和种子试卷在不同能力处的信息量差异较小时, 则可以认为 2 份测验在统计学意义上是等效的. 测验特征曲线的接近程度也可作为组卷质量的评价标

准. 绘制生成试卷和种子试卷的测验信息函数、测验特征曲线图像, 可直观判断这 2 种组卷方法的优劣.

(ii) 测验信息函数、测验特征曲线的均方差(mean square deviation, MSD) 指标. 该指标用来评估生成试卷和种子试卷在测验信息量或测验特征曲线中的差异量, 其计算公式为

$$M_{SD_i} = \left(\sum_{i=1}^n (T_{i_0}(\theta_i) - T_{i_t}(\theta_i))^2 \right) / n,$$

$$M_{SD_c} = \left(\sum_{i=1}^n (T_{c_0}(\theta_i) - T_{c_t}(\theta_i))^2 \right) / n,$$

其中 $T_{i_0}(\theta_i)$ 和 $T_{i_t}(\theta_i)$ 分别表示能力水平在 $\theta_i (i = 1, 2, \dots, n)$ 处的生成试卷和种子试卷的信息量; $T_{c_0}(\theta_i)$ 和 $T_{c_t}(\theta_i)$ 分别表示能力水平在 $\theta_i (i = 1, 2, \dots, n)$ 处生成试卷和种子试卷的测验特征函数值. 能力节点数 n 设置为 61, 能力范围是 $-3 \sim 3$, 其步长为 0.1.

(iii) 生成试卷与种子试卷的距离平方和. 若种子试卷的第 i_k 题与在第 t 份平行试卷中第 $j_k^{(t)}$ 题匹配, 则可计算第 t 份平行试卷与种子试卷的距离平方和为

$$d_t^2 = \sum_{k=1}^N d_{i_k j_k^{(t)}}^2.$$

(iv) TIM-MPI 方法的占优率. 其计算公式为

$$\xi_{TIM-MPI} = \frac{1}{R} \sum_{r=1}^R V_r,$$

其中 $V_r = 1$ 表示在第 r 次试验时双题匹配方法生成的所有平行试卷与种子试卷的距离平方和小于单题匹配方法, 即在相同题库参数和种子试卷条件下, 从距离指标来看双题匹配法所生成的平行试卷优于单题匹配方法; 否则, $V_r = 0$. 占优率 $\xi_{TIM-MPI}$ 较大表示双题匹配法所生成的平行试卷多数优于单题匹配方法.

(v) 这 2 种方法在第 r 次试验时的平均运行时间.

3 实验结果及分析

从总体来看, TIM-MPI 方法在 1 000 次组卷过程中的占优率 $\xi_{TIM-MPI}$ 为 0.839. 这表明在组卷过程中, TIM-MPI 方法生成的测验优于 MPI 方法生成的测验的次数, 远超过 MPI 方法优于 TIM-MPI 方法的次数. MPI 方法和 TIM-MPI 方法在 1 000 次组卷过程中的平均运行时间分别为 0.003 s 和 1.872 s. 由此可见, 在组卷速度方面, MPI 方法优于 TIM-MPI 方法. 表 1 显示了在 1 000 次组卷过程中由 MPI 和

TIM-MPI 这 2 种方法所生成的 5 份测验的均方差的描述性统计分析结果,他们分别为测验信息函数均方差和测验特征曲线均方差的平均值、标准偏差、最大值和最小值.由表 1 可以发现 MPI 方法产生的测验信息函数均方差的平均值、标准偏差、最大值和最

小值均大于 TIM-MPI 方法产生的相对应的值.而 MPI 方法产生的测验特征曲线均方差只有标准偏差(0.195)比 TIM-MPI 方法产生的相对应值(0.196)小,MPI 方法产生的其他均方差的统计量值均大于 TIM-MPI 方法产生的相对应值.

表 1 2 种方法的均方差

测验内容	方法	平均值	标准偏差	最大值	最小值
测验信息函数	MPI	0.446	0.251	0.801	0.189
	TIM-MPI	0.435	0.247	0.782	0.183
测验特征曲线	MPI	0.263	0.195	0.547	0.078
	TIM-MPI	0.258	0.196	0.546	0.074

表 2 给出了 MPI 方法和 TIM-MPI 方法所生成的 5 份测验与种子试卷之间距离平方和在 1 000 次试验下的平均值.由表 2 可以发现 TIM-MPI 方法生成的 5 份测验与种子试卷之间距离平方和均值均比 MPI 方法生成的 5 份测验与种子试卷之间距离平方和均值要小.

表 2 2 种方法的距离平方和的平均值

方法	测验 1	测验 2	测验 3	测验 4	测验 5
MPI	1.562	1.633	1.681	1.727	1.771
TIM-MPI	1.502	1.562	1.610	1.668	1.733

下面统计 MPI 优于 TIM-MPI 的结果,即考虑 $V_r=0$ 的所有测验.表 3 给出了这 2 种方法的测验信息函数均方差的平均值、标准偏差、最大值和最小值.由表 3 可以发现 MPI 方法产生的平均值(0.533)略大于 TIM-MPI 方法产生的平均值(0.528),这 2 种方法产生的标准偏差相等,MPI 方法产生的最大值和最小值均小于 TIM-MPI 方法产生的最大值和最小值.表 4 给出了这 2 种方法的测验特征曲线均方差的平均值、标准偏差、最大值和最小值.由表 4 可以发现 MPI 方法产生的标准偏差,最大值和最小值均小于 TIM-MPI 方法产生的值,并且它们差距较小.MPI 方法产生的平均值和 TIM-MPI 方法产生的平均值相等.表 5 给出这 2 种方法的距离平方和的平均值.由表 5 可以发现 MPI 方法生成的每份测验与种子试卷之间距离的平均值均比 TIM-MPI 方法生成的每份测验与种子试卷之间距离的平均值更小,差异量约为 0.1.

表 3 2 种方法的测验信息函数的均方差($V_r=0$)

方法	测验信息函数			
	平均值	标准偏差	最大值	最小值
MPI	0.533	0.272	0.908	0.244
TIM-MPI	0.528	0.272	0.911	0.248

表 4 2 种方法的测验特征曲线的均方差($V_r=0$)

方法	测验特征曲线			
	平均值	标准偏差	最大值	最小值
MPI	0.300	0.210	0.607	0.096
TIM-MPI	0.300	0.212	0.614	0.101

表 5 2 种方法的距离平方和的平均值($V_r=0$)

方法	测验 1	测验 2	测验 3	测验 4	测验 5
MPI	1.817	1.894	1.950	2.000	2.062
TIM-MPI	1.881	1.967	2.041	2.132	2.242

下面统计 TIM-MPI 优于 MPI 的结果.对于 $V_r=1$ 的所有测验,表 6 给出了这 2 种方法的测验信息函数均方差.由表 6 可以发现 TIM-MPI 方法产生的均方差的相关统计量均比 MPI 方法更小.表 7 给出了这 2 种方法的测验特征曲线均方差.由表 7 可以发现 TIM-MPI 方法仅标准偏差(0.193)比 MPI 方法产生的相对应值(0.192)更大,其余值均更小.表 8 给出了这 2 种方法的距离平方和的平均值.由表 8 可以发现 TIM-MPI 方法生成的 5 份测验和种子试卷之间距离平方和的平均值,比 MPI 方法的更小.

表 6 2 种方法的测验信息函数的均方差($V_r=1$)

方法	测验信息函数			
	平均值	标准偏差	最大值	最小值
MPI	0.430	0.247	0.780	0.178
TIM-MPI	0.417	0.242	0.757	0.170

表 7 2 种方法的测验特征曲线的均方差($V_r=1$)

方法	测验特征曲线			
	平均值	标准偏差	最大值	最小值
MPI	0.256	0.192	0.535	0.075
TIM-MPI	0.250	0.193	0.533	0.068

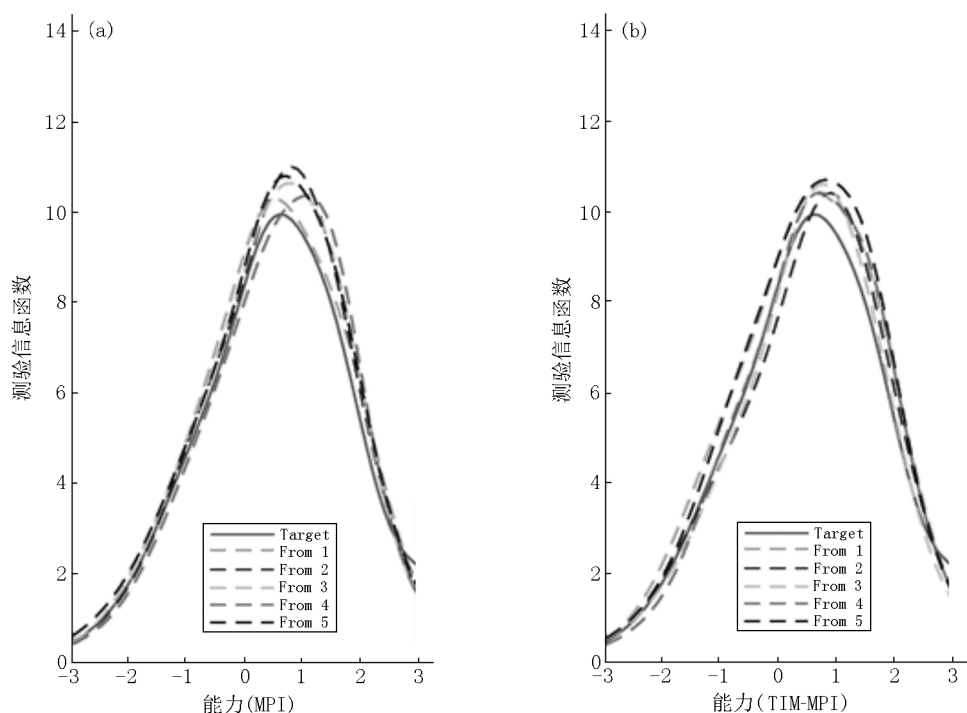
表 8 2 种方法的距离平方和的平均值($V_r=1$)

方法	测验 1	测验 2	测验 3	测验 4	测验 5
MPI	1.513	1.583	1.630	1.675	1.715
TIM-MPI	1.429	1.484	1.527	1.578	1.635

图 2 描绘了在 MPI 方法优于 TIM-MPI 方法时,这 2 种方法的测验信息函数曲线.如图 2(a)所示,

能力水平在 $-3 \sim 0$ 之间,所有生成试卷的测验信息函数都接近种子试卷的测验信息函数.能力水平在 $0 \sim 3$ 之间,所有生成试卷的测验信息函数都是从正

方向上接近种子试卷的测验信息函数.在图 2(b)中相对于图 2(a)而言,可以看到生成试卷的测验信息函数和种子试卷的测验信息函数之间偏差稍大.

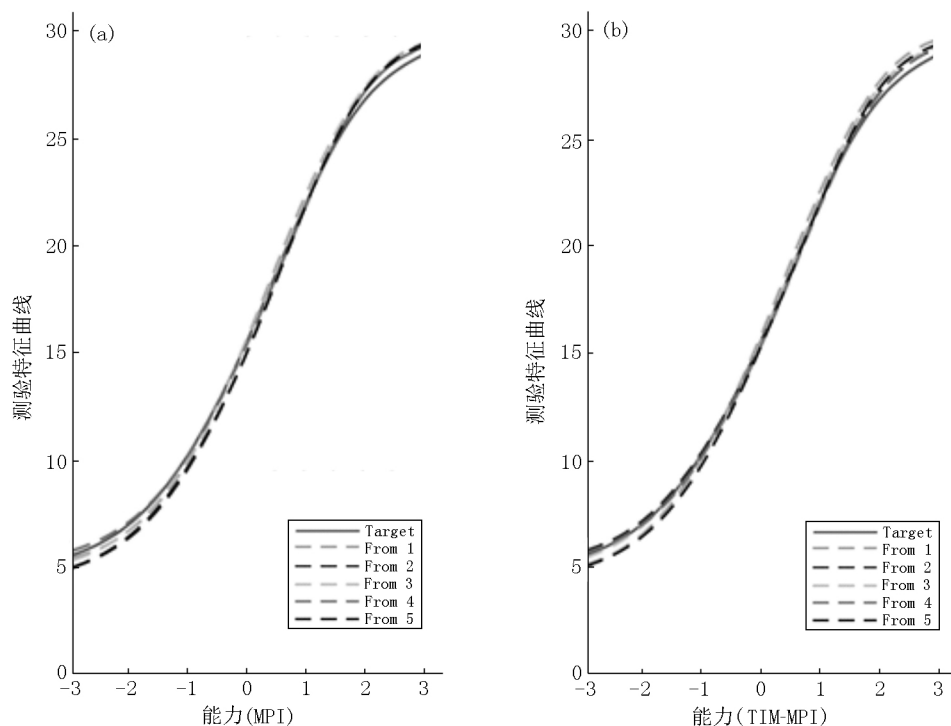


注: $V_r = 0$ 的一次实验结果.

图 2 MPI 和 TIM-MPI 的测验信息函数

图 3 描绘了在 MPI 方法优于 TIM-MPI 方法时这 2 种方法的测验特征曲线.如图 3(a)所示,能力水平在 $-3 \sim 3$ 之间,生成试卷和种子试卷的测验特

征曲线十分接近,在图 3(a)中两端生成试卷和种子试卷的测验特征曲线有偏差.图 3(b)结果基本上与图 3(a)类似.

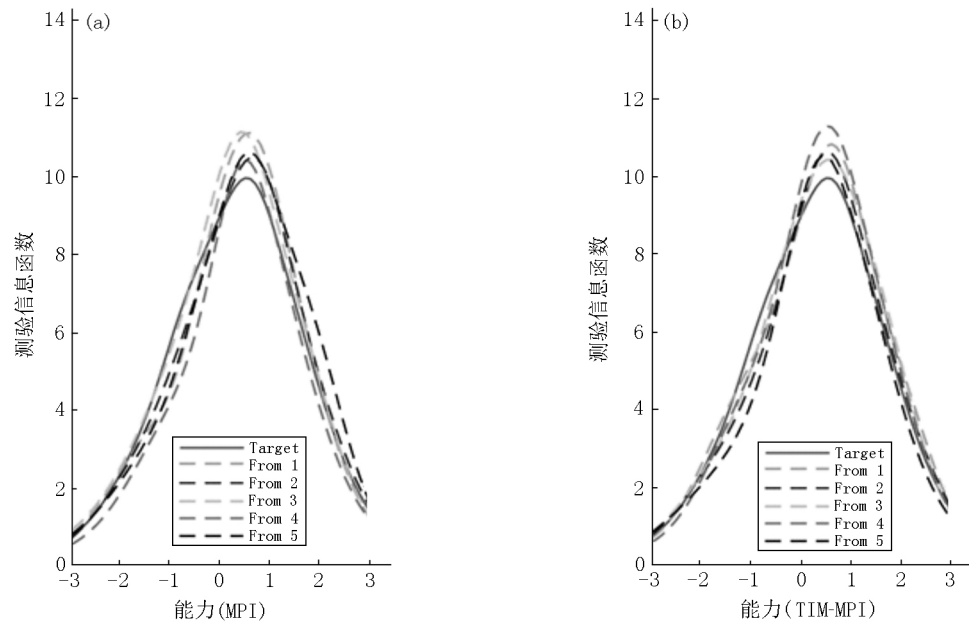


注: $V_r = 0$ 的一次实验结果.

图 3 MPI 和 TIM-MPI 的测验特征曲线

图 4 描绘了在 TIM-MPI 方法优于 MPI 方法时, 这 2 种方法的测验信息函数. 通过图 4(a) 可以看到, 能力水平在 $-3 \sim 0$ 之间, 生成的大部分试卷测验信息函数低于种子试卷的测验信息函数. 能力水平在 $0 \sim 3$ 之间, 大部分测验的测验信息函数高于种子试卷的测验信息函数. 当能力水平较高时, 除只有

一份生成试卷的测验信息函数和种子试卷的测验信息函数有一些差距外, 其余生成试卷的测验信息函数接近种子试卷的测验信息函数. 相对于图 4(a) , 通过图 4(b) 可以看到, 能力水平在 $-3 \sim 3$ 之间, 生成试卷的测验信息函数与种子试卷的测验信息函数之间的偏差较小.

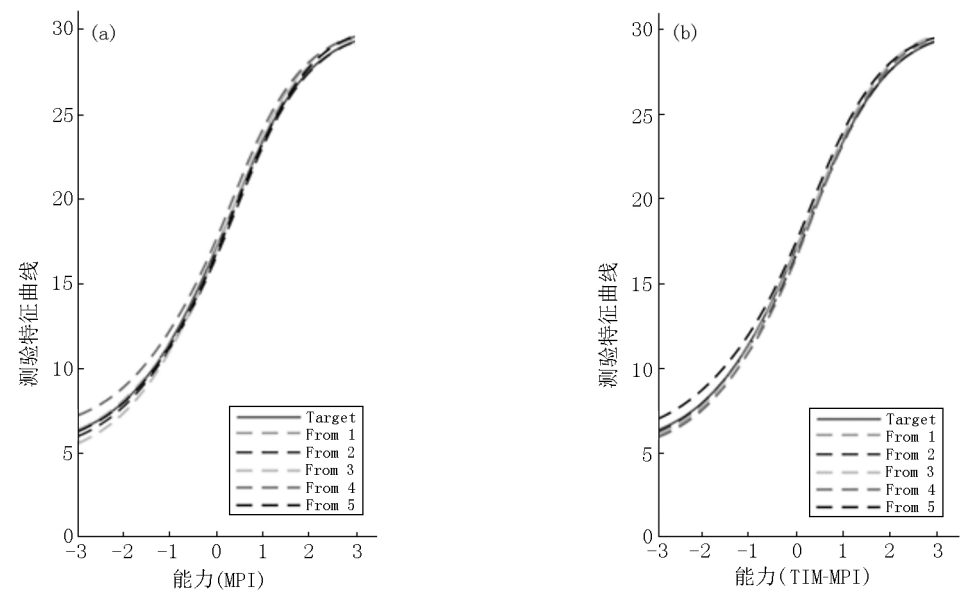


注: $V_r = 1$ 的一次实验结果.

图 4 MPI 和 TIM-MPI 的测验信息函数

图 5 描绘了在 TIM-MPI 方法优于 MPI 方法时 这 2 种方法的测验特征曲线. 由图 5(a) 所示, 分别有一份生成试卷的测验特征曲线低于或高于种子试

卷的测验特征曲线, 其余生成试卷和种子试卷的测验特征曲线接近. 由图 5(b) 所示, 与图 5(a) 相比, 生成试卷与种子试卷的测验特征曲线之间的偏差较小.



注: $V_r = 1$ 的一次实验结果.

图 5 MPI 和 TIM-MPI 的测验特征曲线

4 讨论

从实验数据来看,在 1 000 次组卷中,TIM-MPI 方法以 0.839 的概率优于 MPI 方法;TIM-MPI 方法生成试卷的均方差基本优于 MPI 方法生成试卷的均方差;TIM-MPI 方法生成的 5 份测验与种子试卷之间距离平方和的平均值也基本小于生成的 5 份测验与种子试卷之间距离平方和的平均值;只有在运行时间上,MPI 方法的运行时间要远小于 TIM-MPI 方法的运行时间。

从实验图像来看,将这 2 种方法的生成试卷和种子试卷的测验特征曲线以及测验信息函数的图像进行对比,能直观地看到这 2 种方法在不同情况下的图像表现情况。当 MPI 方法好过 TIM-MPI 方法时,其测验特征曲线和测验信息函数的表现也会好过 TIM-MPI 方法,反之亦然。

在 MPI 方法优于 TIM-MPI 方法的情况下,这 2 种方法在除运行时间外的其他评价标准上差距并不明显;而在 TIM-MPI 方法优于 MPI 方法的情况下,TIM-MPI 方法优势更明显一些。由此可以得出结论:与 MPI 方法相比,新提出的 TIM-MPI 方法可有效地提升平行测验质量。

从本文模拟实验结果来看,为何出现单题匹配优于双题匹配,其可能原因是题库中与种子试题匹配十分好的题目相对较少,该推断仍有待进一步研究。是否可以通过继续增加匹配试题数量来进一步优化组卷方法是值得尝试的。此外,题库难度的模拟只考虑到正态分布,若题库难度模拟符合混合正态分布或均匀分布,则效果如何,这也是值得探究的问题。并且,还可以去探究在生成多少份平行测验的情况下,双题匹配的 MPI 方法表现是否最佳。

5 结论

本文在最大优先级指标(MPI)的基础之上,对其匹配试题数量进行了改进,由原来的单题匹配改为双题匹配。在生成平行试卷的组卷过程中,TIM-MPI 方法生成的多数测验可更好地匹配种子试卷,但也有极少数情况,MPI 方法生成的测验会优于 TIM-MPI 方法生成的测验,所以还需要进一步去改进组卷方法。但总的来讲,采用 TIM-MPI 方法生成

测验是不错的选择。

6 参考文献

- [1] 舒清录. 一种实用的随机组卷策略研究与实现 [J]. 电脑与信息技术, 2020, 165(3): 34-36, 45.
- [2] 韩啸, 毕波, 唐锦萍. 基于基因表达式编程的计算机组卷算法研究 [J]. 计算机技术与发展, 2020, 30(5): 154-159.
- [3] 王友仁, 张砦, 施玉霞, 等. 题库系统智能成卷理论和组卷方法研究 [J]. 电子科技大学学报, 2006, 35(3): 77-80.
- [4] 李佳, 丁树良. 计算机化自适应测验中能力估计新方法 [J]. 江西师范大学学报: 自然科学版, 2019, 43(2): 36-40.
- [5] 王宣策, 欧阳鑫玉. 基于改进遗传算法的智能组卷系统设计 [J]. 电脑知识与技术, 2020, 16(2): 65-67.
- [6] Wang Lei, Tang Dunbing. An improved adaptive genetic algorithm based on hormone modulation mechanism for job-shop scheduling problem [J]. Expert Systems With Applications, 2011, 38(6): 7243-7250.
- [7] Lin Ying, Gong Yuejiao, Zhang Jun. An adaptive ant colony optimization algorithm for constructing cognitive diagnosis tests [J]. Applied Soft Computing, 2017, 52(1): 1-13.
- [8] Lin Ying, Jiang Yesi, Gong Yuejiao, et al. A discrete multiobjective particle swarm optimizer for automated assembly of parallel cognitive diagnosis tests [J]. IEEE Transactions on Cybernetics, 2019, 49(7): 2792-2805.
- [9] Armstrong R, Jones D, Wu I L. An automated test development of parallel tests from a seed test [J]. Psychometrika, 1992, 57(2): 271-288.
- [10] Swanson L, Stocking M L. A model and heuristic for solving very large item selection problems [J]. Applied Psychological Measurement, 1993, 17(2): 151-166.
- [11] van der Linden W J. Linear models of optimal test design [M]. New York: Springer-Verlag, 2005.
- [12] Adema J J, van der Linden W J. Algorithms for computerized test construction using classical item parameters [J]. Journal of Educational Statistics, 1989, 14(3): 279-290.
- [13] Armstrong R D, Jones D H. Polynomial algorithms for item matching [J]. Applied Psychological Measurement, 1992, 16(4): 365-371.
- [14] Chen Peihua, Chang Huahua, Wu Haiyan. Item selection for the development of parallel forms from an IRT-based

- seed test using a sampling and classification approach [J]. Educational and Psychological Measurement, 2012, 72(6): 933-953.
- [15] 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.
- [16] 潘奕尧. 改进的最大优先指标及在计算机化自适应诊断测验中的应用 [D]. 南昌: 江西师范大学, 2011.
- [17] 汤楠, 丁树良. 一阶段选题的最大优先级指标的修正 [J]. 江西师范大学学报: 自然科学版, 2012, 36(5): 452-455.
- [18] Wang Shiyu, Zheng Yi, Zheng Chanjin, et al. An automated test assembly design for a large-scale Chinese proficiency test [J]. Applied Psychological Measurement, 2016, 40(3): 233-237.
- [19] Chen Peihua. Three-element item selection procedures for multiple forms assembly: an item matching approach [J]. Applied Psychological Measurement, 2016, 40(2): 114-127.
- [20] Chang Huahua. Psychometrics behind computerized adaptive testing [J]. Psychometrika, 2015, 80(1): 1-20.
- [21] 漆书青. 现代教育与心理测量学原理 [M]. 北京: 高等教育出版社, 2002.

The Maximum Priority Index Method of Two Items Matching and Its Application on Constructing Parallel Tests

WANG Wenyi¹, XIONG Jian¹, SONG Lihong², ZHENG Juanjuan¹, HU Haiyang¹

(1. College of Computer and Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

2. College of Education, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: The fairness of large-scale examinations has attracted much attention. It is very important to explore a method for generating multiple parallel tests and ensuring the quality of the tests. The maximum priority index (MPI) method originally proposed in the computer adaptive test (CAT) has been applied to constructing multiple parallel tests, but it is only limited to matching between single item and single item. In order to increase the solution space, a two items matching method based on the maximum priority index is proposed to realize the matching between the two items in the item bank and the two items in the seed test, and the parallel test quality adjustment method is combined to generate parallel tests. The existing test construction method is compared with the two items matching method. The results show that the two items matching method effectively improves the quality of the parallel test.

Key words: parallel tests; test assembly method; computerized adaptive test; maximum priority index; two items matching; single item matching

(责任编辑: 冉小晓)