Stage Adaptive and Timed Item Selection Methods for Computerized Classification Test

Huang Yingshi, Ren He, Chen Ping (Corresponding Author)
Collaborative Innovation Center of Assessment for Basic Education Quality, Beijing Normal University, Beijing, 100875, pchen@bnu.edu.cn

Abstract：Computerized classification test is designed to divide examinees into different proficiency categories. A corollary problem is how to select items for the test to satisfy all stakeholders. One of the challenges faced by test developers is item exposure control. As CCT trends to select items with high information at the cut scores to make decisions reliable, the top informatic items are under high exposure rate while the less informatic ones with little chance to be used. Another issue that is pertinent for CCT is time control. It is crucial to reduce the differences between examinees' test-taking time and the number of examinees who exceed the time limitation. The objects of this study are to (1) propose the stage adaptive item selection method (SAI) for CCT, i.e., matching the current need for decision making with the rank of item information; (2) optimize the GMFIT method to CCT (M-GMFCT) and combine with the idea of stage adaptive method to introduce the timed-SAI method.

Study 1 investigated the nature of the SAI and the performance in item usage as well as classification accuracy compared with the traditional method. Specifically, 29 levels of examinees' ability (from -3.5 to 3.5 by 0.25) and two levels of test length (10 and 20 items) were simulated. The item bank contained 500 items and the cut score was set at 0. The weighting parameter was set at 1. Each condition was repeated 100 times and the selected priority index, item exposure rate, and percentage of correct classifications (PCC) were computed. Results show that: (1) the new method picks the informatic items when the current ability estimate are close to the cut score and recommends the less informatic items otherwise; (2) the item exposure rates of high-quality items under the traditional method are close to 1 while most of the items have lower exposure rate with the SAI method; (3) the PCCs of the SAI method, for the most part, are higher than the traditional method.

Study 2 determined whether the M-GMFCT and the timed-SAI can shrink the mean and variance of the test-taking time. 1000 examinees were generated from $N(0, 1)$. The hierarchical model was used to model the response and response time. And the correlation between ability and speed was 0.59. The correlation between item difficulty and time density was 0.65. The sensitivity parameter was 0.5. The remaining conditions were the same as in study 1. Each condition was repeated 100 times and the test time and PCC were computed under length 10 when the χ^2 statistic was calculated under three levels of length (10, 15, and 20 items). Results show that: (1) the timed-SAI method holds the smallest mean and variance of test time and the M-GMFCT takes second place while the traditional one has the longest and most fluctuant test time; (2) the PCCs of the M-GMFCT and the traditional method are similar; (3) the χ^2 statistics of the timed-SAI method are at the lowest level among the others under all test lengths.

Future directions may seek to further progress in extending the stage adaptive and timed item selection methods in the multidimensional and polytomous scenario and using the conditional dependent model to be more in line with the practical need and make full use of the information provided by response time.

Keywords：computerized classification test, item selection method, response time, exposure control

阶段适应与融入反应时的计算机化分类测验选题策略

黄颖诗　任赫　陈平（通讯作者）

北京师范大学中国基础教育质量监测协同创新中心，北京，100875，pchen@bnu.edu.cn

摘 要　计算机化分类测验（Computerized Classification Testing, CCT）的目的在于将考生划分到不同类别。如何选择题目以构建满足各方面需求的试卷一直是 CCT 的研究热点。题目曝光控制是首要面临的问题，现有基于分界分数的选题策略固定选择在分界分数处具有高信息量的题目，导致题库中大量题目浪费而高信息量题目曝光率极高。测验时间控制是亟待解决的第二个问题，目前不同考生完成测验的时间往往差异较大，且部分考生可能超时。因此本研究：（1）提出阶段适应的 CCT 选题策略（SAI），即根据当前决策需求匹配相应信息量的题目；（2）优化融入反应时的 CCT 选题策略（M-GMFCT），并结合阶段适应思想提出 timed-SAI 策略。

研究一采用模拟方法考察 SAI 能否“按需分配”题目，并比较它与传统方法在题目均衡使用和分类准确性方面的表现。因此，从区间[-3.5, 3.5]以 0.25 为步长生成 29 个考生能力点，并考虑两种测验长度（10 和 20 题）。题库大小为 500 题，分界分数取 0，权重参数取 1。每种条件重复 100 次，评价指标为：选择优先指数、题目曝光率与正确分类百分比（PCC）。结果表明：（1）能力距分界分数越近，SAI 越倾向于选择高信息量题目，反之则推荐低信息量题目；（2）传统方法中高信息量题目的曝光率接近 1，而 SAI 中大部分题目的曝光率低于 0.2；（3）在大多数情况下，SAI 的 PCC 都高于传统方法。

研究二考察 M-GMFCT 与 timed-SAI 能否减小测验时长的均值与方差。为贴合实际，从 $N(0, 1)$中随机抽取 1000 名考生的能力。采用层次模型对作答和反应时联合建模，并设置考生能力与速度的相关 $\rho_{\theta\tau} = 0.59$、题目难度与时间强度的相关$\rho_{b\beta} = 0.65$。敏感参数取 0.5，其余条件同研究一。每种条件重复 100 次，评价指标为：测验长度为 10 时每名考生的测验时间和 PCC 以及测验长度为 10、15 和 20 题时的题目曝光χ^2统计量。结果表明：（1）timed-SAI 的测验时长均值与方差最小，M-GMFCT 次之，传统方法最大；（2）M-GMFCT 与传统方法的 PCC 无显著差异；（3）timed-SAI 的χ^2在所有题目水平下均最低。

未来研究可拓广至多维多分类测验情景、考虑条件依赖的反应时模型、分析不同类型反应时对决策判断的帮助。

关键词　计算机化分类测验，选题策略，反应时，曝光控制