

文章编号: 1000-5862(2015)05-0449-04

# 计算机化自适应诊断测验双重约束变长终止规则

艾国金, 甘登文\*, 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:** 提出3种变长CD-CAT终止规则, 在DINA模型、4种属性层级结构下, 采用ED选题策略, 讨论提出的3种方法与常用CD-CAT终止规则在7个评价指标上的优劣. 研究结果显示: 新的3种方法较常用终止规则有其自身优势.

**关键词:** 计算机化自适应测验; 认知诊断; 变长终止规则; Monte Carlo 模拟

**中图分类号:** B 842.1 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.05.02

## 0 引言

近些年来, 随着CD-CAT的受关注度越来越高, 国内外研究者和应用者对它的研究也越来越深入. 通常CD-CAT采用定长形式作为其测验终止规则, 即当被试答题数达到某个预设长度, 测验终止. 然而, 在CD-CAT中未能找出与Fisher信息量类似的指标衡量其测量误差, 因此CD-CAT变长终止规则通常采用其他指标作为其测验终止的条件. 文献[1]提出的3种方法较传统终止规则而言, 在人均用时、人均用题数等方面有较大改善, 但只考虑了当前阶段不同知识状态的后验概率之间的关系, 并没有考虑临近或相邻的知识状态的后验概率之间的关系. 本文在文献[1]的基础上, 参考临近后验概率之差法思想<sup>[2]</sup>, 提出3种CD-CAT双重约束变长终止规则.

## 1 不定长CD-CAT终止规则

### 1.1 Hsu等方法

C. L. Hsu等在Tatsuoka给出的变长终止规则建议<sup>[3]</sup>的基础上, 提出双重约束终止规则<sup>[4]</sup>. 第1重约束沿用了C. Tatsuoka的当被试通过作答得到某一种知识状态的后验概率超过0.8时, 则测验停止的建议并对它作了相应调整, 即当被试属于某个知

识状态的最大后验概率( $P_{1st}$ )大于预先设定值时, 测验完成; C. L. Hsu等在第1重约束的基础上对第2大后验概率( $P_{2nd}$ )作了相应约束得到第2重约束规则, 即当被试属于某个知识状态的 $P_{1st}$ 大于预先设定值且 $P_{2nd}$ 小于某个预定值时, 测验完成. 第2重约束规则中的 $P_{2nd}$ 的计算公式为

$$P_{2nd} = \frac{1 - P_{1st}}{2^K - 1} + \frac{(2^K - 2)(1 - P_{1st})d}{2^K - 1}, \quad (1)$$

其中 $K$ 为考察属性个数,  $P_{1st}$ 为最大后验概率,  $P_{2nd}$ 为第2大后验概率,  $d$ 为随机控制参数( $0 \leq d \leq 1$ ), 可根据需要取值.

### 1.2 临近后验概率之差法

临近后验概率之差法(Difference of the Adjacent Posterior Probability Method, DAPP)<sup>[2]</sup>认为当被试通过作答, 属于某个“真实”的知识状态的后验概率会越来越大. 于是考虑采用前后相邻的并且属于同一个知识状态的最大后验概率之间的绝对距离足够小时, 作为测验终止的条件, 即 $|P_{1st}^{t+1}(\hat{\alpha}_i) - P_{1st}^t(\hat{\alpha}_i)| < \varepsilon$ , 测验终止. 其中 $P_{1st}^{t+1}(\hat{\alpha}_i)$ 为被试作答完 $t+1$ 个项目后知识状态为 $\hat{\alpha}_i$ 的最大后验概率,  $P_{1st}^t$ 表示被试作答完 $t$ 个项目后知识状态为 $\hat{\alpha}_i$ 的最大后验概率,  $\varepsilon$ 为0~1之间足够小的数.

### 1.3 艾国金等的方法

艾国金等在Hsu等方法和DAPP法的启发下, 同样采用知识状态的后验概率作为终止规则的参考依据, 提出3种终止规则<sup>[1]</sup>: (i) 方法1, 最大后验概

收稿日期: 2015-03-19

基金项目: 国家自然科学基金(31500909, 30860084, 31160203, 31100756, 31360237), 国家社会科学基金(12BYY055)和江西省教育厅科技计划(GJJ13207, GJJ13226, GJJ13227, GJJ13208, GJJ13209)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.

率与第 2 大后验概率之差法; (ii) 方法 2, 最大后验概率与最小后验概率之差法; (iii) 方法 3, 方法 1 与方法 2 差的绝对值法. 计算公式分别为

$$P_{1st} - P_{2nd} = M, \quad (2)$$

$$P_{1st} - P_{min} = N, \quad (3)$$

$$|M - N| < \xi, \quad (4)$$

其中  $P_{1st}$ 、 $P_{2nd}$  与 (1) 式定义相同,  $P_{min}$  为最小后验概率,  $M$ 、 $N$  均为 0 ~ 1 之间的数且越大越好,  $\xi$  为 0 ~ 1 之间足够小的数.

#### 1.4 双重约束变长终止规则

与 C. L. Hsu 等方法不同的是文献 [1] 除考虑了  $P_{1st}$  与  $P_{2nd}$  之外还比较了  $P_{1st}$  与  $P_{2nd}$ ,  $P_{1st}$  与  $P_{min}$  及  $P_{2nd}$  与  $P_{min}$ , 与 DAPP 不同的是文献 [1] 虽然对  $P_{1st}$ 、 $P_{2nd}$  和  $P_{min}$  进行了比较, 却只对当前部分知识状态的后验概率进行比较. 众所周知, 随着被试作答项目数的增多, 被试属于某个“真实”的知识状态的后验概率会越来越大, 如果将相邻的从属于同一个知识状态的后验概率之间的变化与文献 [1] 的 3 种方法综合起来考虑是否会对测验结果起到积极地作用, 即既要比较  $P_{1st}$ 、 $P_{2nd}$ 、 $P_{min}$ , 又要考虑  $PPLS_{1st}^{t+1}$  与  $PPLS_{1st}^t$  之间的变化. 按照此思路, 提出 3 种混合方法也可以称为 3 种双重约束变长终止规则 (Dual

Restrictions Variable-Length Termination Rule, DRVL).

1) DRVL 1: 在考虑  $P_{1st} - P_{2nd} = M$  的同时, 还要考虑相邻 2 次作答后, 从属于同一个知识状态的最大后验概率值之间的变化, 即当  $P_{1st} - P_{2nd} = M$  与  $|P_{1st}^{t+1}(\hat{\alpha}_i) - P_{1st}^t(\hat{\alpha}_i)| < \varepsilon$  共同满足时, 测验终止.

2) DRVL 2: 在考虑  $P_{1st} - P_{min} = N$  的同时, 还要考虑相邻 2 次作答后, 从属于同一个知识状态的最大后验概率值之间的变化, 即当  $P_{1st} - P_{min} = N$  与  $|P_{1st}^{t+1}(\hat{\alpha}_i) - P_{1st}^t(\hat{\alpha}_i)| < \varepsilon$  共同满足时, 测验终止.

3) DRVL 3: 在考虑  $|M - N| < \xi$  的同时, 还要考虑相邻 2 次作答后, 从属于同一个知识状态的最大后验概率值之间的变化, 即当  $|M - N| < \xi$  与  $|P_{1st}^{t+1}(\hat{\alpha}_i) - P_{1st}^t(\hat{\alpha}_i)| < \varepsilon$  共同满足时, 测验终止.

## 2 CD-CAT 模拟研究

### 2.1 被试及题库的生成

模拟实验中, 本文共考察了 6 个属性 ( $K = 6$ ), 分为以下 4 种属性层级结构: 线型、收敛型、发散型、无结构型<sup>[5]</sup>, 如图 1 所示.

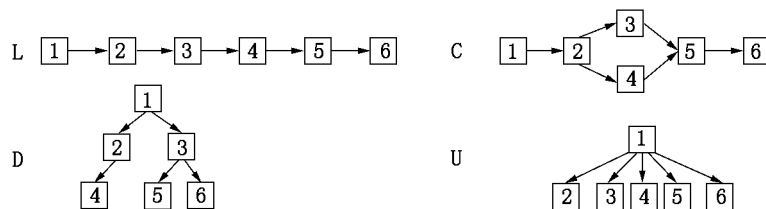


图 1 4 种属性层级结构图

根据属性间的层级关系可以导出每种属性层级关系下所有可能的测验考核项目模式, 它们分别为 6, 7, 15 和 32 种. 每一个属性层级关系下的所有可能的测验考核项目模式可以组成一个矩阵即潜在  $Q$  阵 ( $Q_p$ ), 在  $Q_p$  的尾部添加全 0 列构成学生  $Q$  阵 ( $Q_s$ ).  $Q_s$  的每一列代表一种知识状态. 被试人数  $N = 1000$ , 将  $Q_s$  的列平均分配给这 1000 人, 对于不能均分的, 采取随机指派.

对于每种类型的属性层级结构, 均有相应的项目模式  $q_j$  ( $q_j$  为  $Q_s$  的某一列), 每个项目模式的属性相同但参数不同, 在模拟实验中每类模式的项目个数设置为 100, 项目的失误参数 ( $s$ ) 和猜测参数 ( $g$ ) 均服从均匀分布  $U(0.05, 0.25)$ , 以此建立题库<sup>[6]</sup>.

### 2.2 测试的构建

模拟实验采用 DINA 模型<sup>[7]</sup> 作为认知诊断模

型, 利用 ED 选题策略<sup>[8]</sup> 进行选题并采用最大后验概率方法<sup>[9]</sup> 对被试的知识状态进行估计. 实验中将定长终止规则、文献 [4] 方法 (以下称为 Tatsuoka 法)、Hus 等方法中的第 2 重约束 (以下简称 Hus 方法 2)、艾国金等<sup>[1]</sup> 的 3 种方法 (以下称作方法 1、方法 2 和方法 3) 作为对比终止规则, 其中定长终止规则长度  $L$  设置为 30, Hus 方法 2 做如下设置:  $P_{1st} > 0.95$ ,  $\mu = 0.25$ , 文献 [1] 中 3 种方法做如下设置:  $M > 0.99$ ,  $N > 0.99$ ,  $\xi = 0.001$ . 3 种双重约束变长终止规则中的  $\varepsilon$  取值范围参考文献 [1] 中  $\xi$  的值, 即  $\varepsilon = 0.001$ .

### 2.3 模拟得分阵及评价指标

在 DINA 模型下, 被试  $i$  在模拟的真实知识状态下对项目  $j$  的得分概率为  $P_j(\alpha_i)$ , 随机产生一个服从  $(0, 1)$  均匀分布的随机数  $r$ , 如果  $r < P_j(\alpha_i)$ , 那么

被试在项目  $j$  上得 1 分, 否则得 0 分, 由此得到被试的得分矩阵.

采用模式判准率、单个被试最大用题数、人均测验用时、测试重叠率、人均测验用题数、单个被试最小用题数、 $\chi^2$  统计量等 7 个评价指标, 具体计算公式参考文献 [1]. 7 个指标中, 除模式判准率越高越好, 其他 6 个指标均是越小越好.

### 3 结果及其分析

Monte Carlo 模拟实验采用重复 100 次求平均值的方法, 得到 4 种属性层级结构下不同终止规则在 7 个评价指标上的表现, 具体表现如表 1 ~ 表 5 所示.

表 1 4 种结构下不同终止规则的模式判准率

属性层级结构	定长 $L=30$	Tatsuoka 法	Hsu 方法 2	方法 1	方法 2	方法 3	DRVL1	DRVL2	DRVL3
L	0.999 6	0.868 3	0.989 2	0.999 2	0.998 9	0.989 9	0.999 1	0.999 3	0.999 3
C	0.999 3	0.867 8	0.987 7	0.999 0	0.998 8	0.990 2	0.999 2	0.999 1	0.999 0
D	0.997 3	0.858 5	0.989 5	0.997 1	0.997 3	0.986 1	0.996 9	0.997 2	0.996 8
U	0.991 7	0.843 8	0.985 3	0.990 1	0.990 6	0.990 3	0.991 1	0.991 2	0.990 4

表 2 4 种结构下不同终止规则模拟实验人均测验用时 单位: s

属性层级结构	定长 $L=30$	Tatsuoka 法	Hsu 方法 2	方法 1	方法 2	方法 3	DRVL1	DRVL2	DRVL3
L	0.002 4	0.000 5	0.002 0	0.001 4	0.001 1	0.001 1	0.001 5	0.001 3	0.001 3
C	0.002 6	0.000 6	0.002 1	0.001 6	0.001 3	0.001 3	0.001 7	0.001 5	0.001 5
D	0.006 1	0.001 8	0.005 1	0.004 5	0.004 1	0.003 8	0.004 7	0.004 2	0.004 2
U	0.029 6	0.011 3	0.025 8	0.0250	0.024 1	0.022 7	0.025 2	0.024 3	0.023 8

表 3 4 种结构下不同终止规则模拟实验人均测验用题数 单位: 个

属性层级结构	定长 $L=30$	Tatsuoka 法	Hsu 方法 2	方法 1	方法 2	方法 3	DRVL1	DRVL2	DRVL3
L	30	5.594	19.913	14.281	13.771	12.386	14.724	14.556	14.452
C	30	6.030	18.304	15.324	14.727	13.483	15.717	15.459	15.343
D	30	8.556	22.483	20.328	19.690	18.039	20.575	20.141	19.693
U	30	11.427	25.649	24.829	24.276	22.871	24.958	24.537	23.928

表 4 4 种结构下不同终止规则模拟实验单个被试最小用题数 单位: 个

属性层级结构	定长 $L=30$	Tatsuoka 法	Hsu 方法 2	方法 1	方法 2	方法 3	DRVL1	DRVL2	DRVL3
L	30	3	4	5	5	2	6	6	6
C	30	3	5	5	5	2	6	6	6
D	30	4	5	7	6	2	7	7	7
U	30	6	6	8	8	7	9	9	9

表 5 不同终止规则模拟实验  $\chi^2$  指标和测试重叠率指标

指标	定长 $L=30$	Tatsuoka 法	Hsu 方法 2	方法 1	方法 2	方法 3	DRVL1	DRVL2	DRVL3
$\chi^2$ 指标	22.121 5	11.396 5	19.263 7	20.738 8	20.407 6	19.044 4	20.657 3	20.280 6	19.655 7
$R_t$ 指标	0.011 6	0.003 0	0.008 5	0.007 0	0.006 8	0.006 2	0.006 8	0.006 6	0.006 5

### 4 讨论

在不同属性层级关系下, DRVL1、DRVL2 和 DRVL3 在 2.3 节中的 7 个指标上的表现不尽相同, 在这 3 种方法中 DRVL3 的表现较为突出. 新方法只在 DINA 模型、ED 选题策略及 Leighton 等提出的 4

种属性层级关系的情况下进行了讨论, 如果改成其他认知诊断模型、选题策略或丁树良等提出的属性层级关系<sup>[10-11]</sup>, 新方法是否适用, 如果不适用该如何修改. 另外由于研究是在模拟数据的情况下进行的, 如果将模拟数据改成真实数据, 新方法能否表现出其在模拟数据下的优势, 这些都将是下一步值得研究的内容.

## 5 参考文献

- [1] 艾国金,甘登文,丁树良,等. 不定长认知诊断计算机化自适应测验终止规则研究 [J]. 江西师范大学学报:自然科学版 2014 38(5):442-443.
- [2] 郭磊 郑蝉金 边玉芳. 变长 CD-CAT 中的曝光控制欲终止规则 [J]. 心理学报 2015 47(1):130-134.
- [3] Tatsuoka C. Data analytic methods for latent partially ordered classification models [J]. Applied Statistics 2002 , 51(3):337-350.
- [4] Hsu C L ,Wang Wenchung ,Chen Shuying. Variable-length computerized adaptive testing based on cognitive diagnosis models [J]. Applied Psychological Measurement 2014 4: 6-7.
- [5] Leighton J P ,Gierl M ,Hunka S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka's rule-space approach [J]. Journal of Educational Measurement 2004 41(3):205-236.
- [6] 唐小娟,丁树良,毛萌萌,等. 基于属性层级结构的认知诊断测验的组卷 [J]. 心理学探新 2013 33(3):252-259.
- [7] dela Torre J. DINA model and parameter estimation: a didactic [J]. Journal of Educational and Behavioral Statistics 2009 34:115-130.
- [8] 尚志勇,丁树良. 认知诊断自适应测验选题策略探新 [J]. 江西师范大学学报:自然科学版 2011 35(4):418-421.
- [9] 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 2002:85-91,97-142.
- [10] 丁树良,汪文义,罗芬. 多级评分认知诊断测验蓝图的设计:根树型结构 [J]. 江西师范大学学报:自然科学版 2014 35(2):265-269.
- [11] 丁树良,罗芬,汪文义. 多级评分认知诊断测验蓝图的设计:独立型和收敛型结构 [J]. 江西师范大学学报:自然科学版 2014 35(3):111-118.

## The Dual Restrictions Variable-Length Termination Rule in Cognitive Diagnosis Computerized Adaptive Testing

AI Guojin ,GAN Dengwen\* ,DING Shuliang

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

**Abstract:** Three new variable-length termination rules in cognitive diagnosis computerized adaptive testing is proposed. Results of the new methods with pre-existing methods were compared under seven evaluation indexes using the deterministic inputs ,noisy “and” gate ( DINA) model ,four kinds of attribute hierarchies and the expected discrimination method ( ED) . Data showed that the new three variable-length termination rules have their own advantages.

**Key words:** cognitive diagnosis; computerized adaptive testing; variable-length; termination rules; monte carlo simulation

( 责任编辑: 冉小晓)