

文章编号: 1000-5862(2018)04-0366-08

基于不同 CDM 视角的 CD-CAT 题库建设

涂冬波, 高旭亮, 汪大勋, 蔡 艳

(江西师范大学心理学院 江西 南昌 330022)

摘要: 在 CD-CAT 题库建设过程中, 选择合适的认知诊断模型(CDM)标定题库参数是一个关键的步骤. 若一个题库均采用一个 CDM 来标定题库所有参数, 则可能会造成模型资料失拟. 提出了一种基于混合模型的题库建设思路, 通过 Monte Carlo 模拟研究验证了基于混合模型建立题库的效果, 并与传统的基于单一模型的题库建设进行了比较, 为 CD-CAT 在选用合适的模型方面提供了理论借鉴.

关键词: CD-CAT; 认知诊断模型; 混合模型; 题库

中图分类号: B 841.7 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2018.04.08

0 引言

认知诊断模型(CDM)是认知诊断评估的一个核心技术, CDM 的好坏直接决定了诊断评估的准确性和有效性^[1]. 目前, 对于 CDM 的开发和研究逐渐成为测量学家们关注的重点内容. 基于不同的理论假设, 国内外学者开发了多种 CDM, 这些模型可以分为约束条件非常少、适用面相对较广的饱和模型(Saturated Model)和约束条件较多的简化模型(Reduced Model). 饱和模型有 G-DINA^[2]、GDM^[3]和 LCDM^[4]等. 饱和 CDM 的优点是属性之间的作用机制没有严格的假设, 模型比较灵活, 适用面更广; 缺点是模型复杂, 不容易理解, 需要较大的实测样本来估计题目参数; 常用的简化模型有 DINA^[5]、DINO^[6]、A-CDM^[2]和 R-RUM^[7]等. 与饱和 CDM 相比, 简化模型具有以下优点: (i) 模型简单, 题目参数较少, 参数更易于准确估计; (ii) 简化 CDM 的诊断结果更加直观和易于解释; (iii) 有研究建议, 当简化 CDM 和饱和 CDM 在模型拟合度相当的条件下, 优先选择简化模型^[8-9]. 也有研究发现与 G-DINA 相比, 使用恰当的简化 CDM 会有更高的模式判断率^[10].

与饱和模型相比, 简化模型的缺点是模型的理论假设比较严格, 如 DINA 模型假设考生只有掌握了题目考察的所有属性才有较大的答对概率, 而

DINO 模型则假设只要掌握其中考察的一个属性就有较大的答对概率. A-CDM 和 RRUM 则假设属性之间没有交互效应. 然而在实际应用中, 属性之间的潜在作用机制可能非常复杂, 这在很大程度上会限制简化模型的应用范围.

通过一定的约束条件 G-DINA 可以简化为 DINA、DINO、A-CDM 和 RRUM 等模型. 为了兼顾简化模型与饱和模型的双重优点, de la Torre^[2]提出了在 G-DINA 的框架下, 采用 Wald 统计量为每道题选择一个最适合的模型, 即一个测验中每个题目采用一个最优的模型(可能是饱和模型也可能是简化模型)进行分析, 为了表达方便, 将这类模型简称为混合模型(Mixed Model). 混合模型兼顾了饱和模型和简化模型的优点, 理论上具有更好的应用前景.

认知诊断理论是认知心理学和现代测量学的产物, 它的目的在于对个体内部知识结构进行诊断, 并为后续的补救教学提供指导. 认知诊断计算机化自适应测验(CD-CAT)不仅具有自适应测验的优点而且能够实现诊断的功能, 使得 CD-CAT 逐渐成为国内外研究的热点领域.

CD-CAT 的首要部分是题库的建设, 而建立题库一个关键的步骤是选择合适的 CDM. 目前国内外关于 CD-CAT 的题库建设均是采用一个 CDM 来标定题库所有参数, 但事实上一个 CDM 不太可能适用于所有的题目. 题库相当于由很多题目组成的一个较长的诊断测验.

收稿日期: 2018-03-12

基金项目: 国家自然科学基金(31660278, 31760288), 江西省教育厅高校人文社科(XL1508, XL1507)和江西省社会科学规划(17JY12)资助项目.

作者简介: 涂冬波(1978-) 男, 江西南昌人, 教授, 博士, 博士生导师. 主要从事心理统计与测量的研究. E-mail: tudongbo@aligun.com

de la Torre 等^[8], Ma Wenchao 等^[11], de la Torre 等^[12]以及 Tu Dongbo 等^[13]通过分析实证数据均发现一份测验中不同的题目拟合不同的模型,如有的题目拟合简约模型而有的题目拟合饱和模型。据此,本文利用 de la Torre^[2]的思路而提出了一种基于混合模型的题库建设思路。查阅国内外相关文献,并没有发现在 CD-CAT 题库建设时的认知诊断模型选用与比较的研究。基于此,本研究拟重点比较 3 种 CDM(饱和模型、简化模型和混合模型)视角下题库建设的效果及特点,以期在实际应用者在 CD-CAT 题库建设提供方法学支持和借鉴。

在本研究中 3 种 CDM 视角分别是基于饱和模型的题库,也即选用 G-DINA 来标定题库所有项目参数;另一种是基于简化模型的题库,包括 DINA、DINO、A-CDM 和 RRUM 模型,分别用一个简化模型标定所有题目参数;最后一种是基于混合模型的题库,为题库的每一个题目选择合适的 CDM,每题选用的可能是饱和模型也可能是简化模型。

1 G-DINA 模型

G-DINA 模型的公式为

$$P(X_{ij} = 1 | \alpha_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{l=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} + \cdots + \delta_{j12 \cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}, \quad (1)$$

其中 K_j^* 为第 j 题所考察的属性个数, δ_{j0} 为项目 j 的截距,也称为基线概率; δ_{jk} 为项目 j 上属性 k 的主效应,它表示被试掌握单一认知属性 k 对正确作答的贡献值,其值越大说明掌握该属性对答对项目的贡献越大; $\delta_{jkk'}$ 为项目 j 上属性 k 和属性 k' 的交互效应, $\delta_{j12 \cdots K_j^*}$ 为项目 j 测量所有属性间的交互效应。若只保留基线概率 δ_{j0} 和所有属性交互效应 $\delta_{j12 \cdots K_j^*}$, 则(1)式简化为

$$P(X_{ij} = 1 | \alpha_{ij}^*) = \delta_{j0} + \delta_{j12 \cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk},$$

即此时 G-DINA 等价于 DINA 模型。

设属性的主效应与所有交互效应的绝对值均相等,且符合如下关系 $\delta_{jk} = -\delta_{jkk'} = \delta_{jkk''} = \cdots = (-1)^{K_j^*+1} \delta_{j12 \cdots K_j^*}$, $k = 1, \cdots, K_j^* - 1$, $k' = 1, \cdots, K_j^* - 1$ 和 $k'' > k' = 1, \cdots, K_j^* - 1$, 则(1)式简化为 $P(X_{ij} = 1 | \alpha_{ij}^*) = \delta_{j0} + \delta_{jk} \alpha_{lk}$, 即 G-DINA 等价于 DINO 模型。另外,假设属性之间的所有交互效应为 0, 则(1)式简化为

$$P(X_{ij} = 1 | \alpha_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk},$$

此时 G-DINA 等同于 A-CDM 模型。当 G-DINA 采用 log 链接函数时,并且假设所有属性的交互效应不存在,则(1)式简化为

$$\log(P(X_{ij} = 1 | \alpha_{ij}^*)) = v_{j0} + \sum_{k=1}^{K_j^*} v_{jk} \alpha_{lk},$$

此时 G-DINA 等价于 RRUM 模型。

2 实验

本文主要讨论 3 类认知诊断模型下的题库建设,具体如下:(i) 基于饱和模型的题库建设,使用饱和和认知诊断模型来标定题库参数,本研究使用 G-DINA 模型。(ii) 基于简化模型的题库建设,分别使用一种简化认知诊断模型来标定题库参数,本研究简化模型主要涉及 DINA、DINO、RRUM、A-CDM。(iii) 基于混合模型(Mixed CDM)的题库建设,为每一题选择最合适的 CDM,该方法分为 2 步:第 1 步,所有题目采用 G-DINA 来估计参数;第 2 步,应用 Wald 检验比较简化 CDM(DINA、DINO、A-CDM、RRUM)是否可以取代 G-DINA。若 Wald 检验选出了 2 个以上的简化 CDM,则选择 p 值最大所对应的 CDM。即为题库中每一题选择一个最优的模型来估计参数。

为了比较不同 CDM 下 CD-CAT 题库设计效果,采用 Monte Carlo 方法进行研究,题库包括 300 题,包括 5 个自变量:真模型(generating model)、匹配模型(fitted model)、题库质量、选题策略和测验长度。真模型和匹配模型分别是 G-DINA、Mixed Model、DINA、DINO、A-CDM、RRUM;题库质量包括高等和中等 2 个水平;选题策略:MPWKL、SHE。测验长度:5、10、15、20 和 25 题。每个实验条件重复 30 次,结果取 30 次实验的平均值。所有模拟程序使用 R 语言编写。

2.1 基于 3 类 CDM 视角的模拟

(i) 题库质量的模拟。模拟方法参考了 Ma Wenchao 等^[11]的做法,当题目是高质量时,固定 $P[1] = 0.9$, $P[0] = 0.1$, 其中 $P[1]$ 表示掌握题目考察全部属性答对的概率, $P[0]$ 是未掌握任意属性的猜对概率。例如,某个题目考察了 3 个属性(00111),所有可能的知识状态(Knowledge State, KS)是 $2^5 = 32$ 种,所有 KS 在这题上均有一个答对概率,掌握全部属性的答对概率等于 0.9,一个属性未掌握答对概率等于 0.1,剩余 KS 的答对概率从 $[0.1, 0.9]$ 区间中随机抽取,中等质量固定 $P[1] = 0.8$, $P[0] =$

0.2. (ii) 模拟题库: 题库容量是 300 题, 考察 5 个属性, 每个题目最多考察 3 个属性, 保证每一个属性在题库中被考察的次数相等. 题库的模拟分为 2 步: 第 1 步, 用 G-DINA、Mixed CDMs、DINA、DINO、A-CDM 和 RRUM 分别模拟 2 000 人作答 300 题的得分矩阵, 共 6 种得分矩阵; 第 2 步, 将第 1 步得到的 6 种得分矩阵, 用 6 种匹配模型估计题目参数, 从而得到 36 种题库参数. 需要特别指出的是, 同一真模型产生的得分数据可以形成 6 种题库, 而这同一真模型下的 6 种题库的质量是等价的(因为 6 种题库的项目参数均来自同一真模型产生的同一得分矩阵), 从而保证了比较的公平性. 图 1 展示了当真模型是 G-DINA 时, G-DINA、Mixed Model、DINA、DINO、A-CDM 和 RRUM 共 6 种题库的具体生成过程.

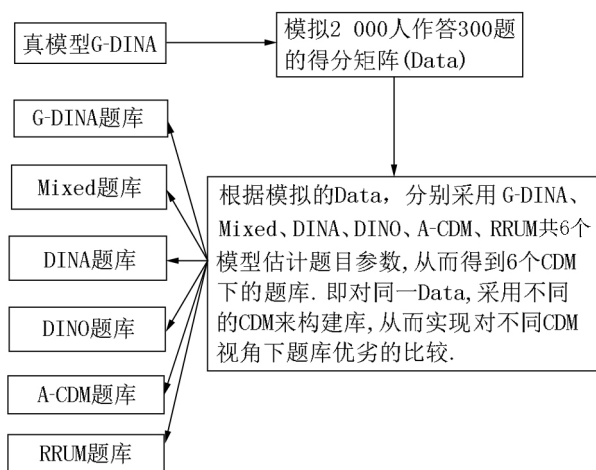


图 1 真模型为 G-DINA 下各题库的生成流程

真模型 Mixed Model(即混合模型)是指分别从 300 题的 Q 矩阵里随机抽取 60 题, 随机分配给 G-DINA、DINA、DINO、A-CDM 和 RRUM 分别模拟得分, 组成 300 题的得分矩阵.

当真模型为 DINA 时, 则只需将图 1 中的真模型改为 DINA 即可; 当真模型为混合模型时, 则也是将图 1 中真模型变为 Mixed 模型, 其余类推.

2.2 选题策略

CD-CAT 常用的选题策略主要有 2 类: 一类是基于 KL 信息量选题; 另一类是基于香农熵选题(Shannon Entropy SHE). 其中基于 KL 信息的选题策略包括: KL、PWKL、HKL 和 MPWKL. 有研究发现 MPWKL 在众多基于 KL 信息中表现最好的一种方法^[14-15], 而 SHE 是基于香农熵信息的代表, 因此, 本研究选用了 2 类具有代表性且表现较好的选题策略: MPWKL 和 SHE.

$$M_{PWKLj} = \sum_{d=1}^{2^k} \left(\sum_{L=1}^{2^k} \sum_{x=0}^1 \left(P(x_j = x | \hat{\alpha}_L^t) \cdot \right. \right.$$

$$\left. \log \left(\frac{P(x_j = x | \hat{\alpha}_L^t)}{P(x_j = x | \alpha_L)} \right) \pi(\alpha_L | X_i) \right) \right).$$

MPWKL 和 PWKL 的区别在于后者选择了最能区分当前估计的 KS 和剩余 KS 的题目作为下一题, 也就是 PWKL 默认把当前估计的 KS 当作被试真实的 KS, 而当前估计的 KS 有可能并不准确, 尤其在测验开始阶段. 而 MPWKL 将所有可能的 2^k 个 KS 当作被试潜在 KS, PWKL 选择了一个点的信息量, MPWKL 选择了一个面的信息量, 与 PWKL 相比, MPWKL 利用了更多的信息从而有利于选出更准确的题目. SHE 选题公式为

$$S_{HEj} = \sum_{y=0}^1 \left(\sum_{c=1}^{2^k} \left(\pi(a_c | x_i^{t+1}) \log \frac{1}{\pi(a_c | x_i^{t+1})} \right) \cdot \left(\sum_{c=1}^{2^k} (P(x_{ij} = y | x_i^t)) \right) \right),$$

其中 $\pi(a_c | x_i^{t+1})$ 为知识状态 a_c 的后验概率, y 为得分, 它的取值可以是 0 或 1, $P(x_{ij} = y | x_i^t) = \pi(a_c | x_i^t) P(x_{ij} = y | a_c)$ 表示被试 i 在第 j 题得 y 分的期望概率. SHE 方法从剩余题库中选择最小 S_{HEj} 的题目作为下一题.

2.3 CD-CAT 的知识状态估计方法

采用极大似然估计方法(Maximum Likelihood Estimation, MLE)估计被试知识状态.

2.4 评价指标

评价指标包括 2 个方面: (i) 模式判准率, 评价估计的 KS 是否准确; (ii) 题库安全性, 包括 2 个指标: 题库使用均匀指标卡方值和测验重叠率(Test Overlap Rate, TOR).

模式判准率指估计 KS 和真实 KS 一致的人数占总人数的比例, 其计算公式为 $P_{MR} = N_p / N$, N_p 为被试知识状态判准的人数, 即被试的估计的 KS 和真实 KS 完全相等的人数, N 为参加测验的总人数.

题库使用均匀性指标(卡方值)计算公式为

$$\chi^2 = \sum_{j=1}^J (e_{ij} - L/J)^2 / (L/J),$$

L 为测验的长度, J 为题库的容量, e_{ij} 为第 j 个题的曝光率, 等于作答第 j 题的被试数除以总人数 N , 卡方值 χ^2 越小越好, 越小代表题库使用越均匀.

测验重叠率指标为

$$TOR = \left(\sum_{j=1}^J T_j (T_j - 1) \right) / (LN(N - 1)),$$

其中 T_j 表示第 j 题被使用的次数, L 和 N 同上, 重叠率指标也是越小越好, 越小表示不同被试之间作答相同题目的比例越小.

3 研究结果

3.1 不同 CDM 视角下模式判准率比较

CDM 视角下模式判准率比较结果如图 2 ~ 图 7 所示。

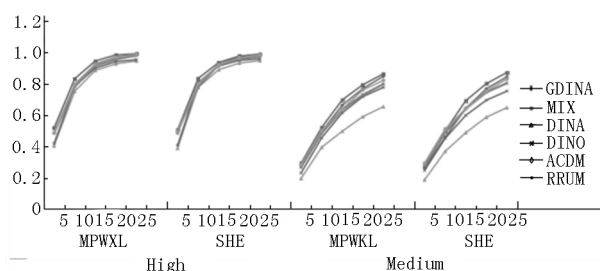


图2 G-DINA 是真模型下的各 CDM 的模式判准率

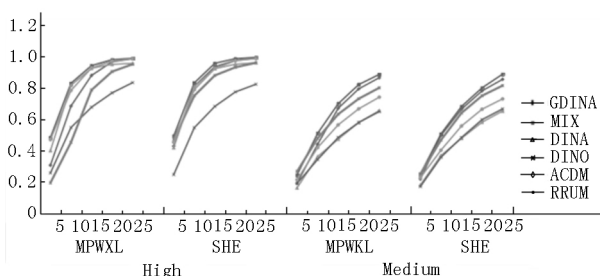


图3 Mixed 是真模型下的各 CDM 的模式判准率

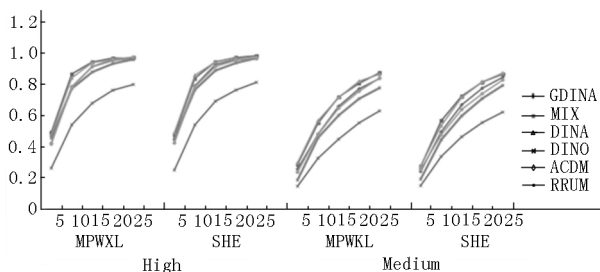


图4 DINA 是真模型下的各 CDM 的模式判准率

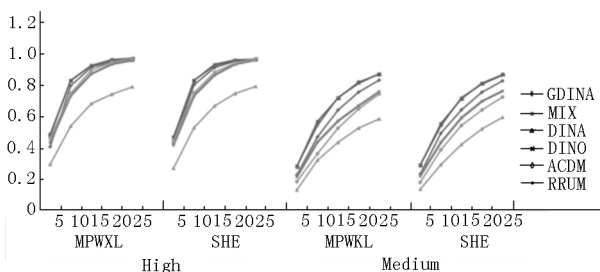


图5 DINO 是真模型下的各 CDM 的模式判准率

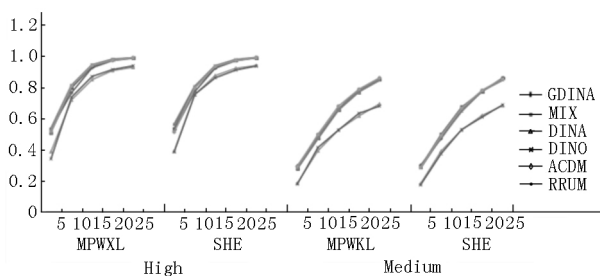


图6 A-CDM 是真模型下的各 CDM 的模式判准率

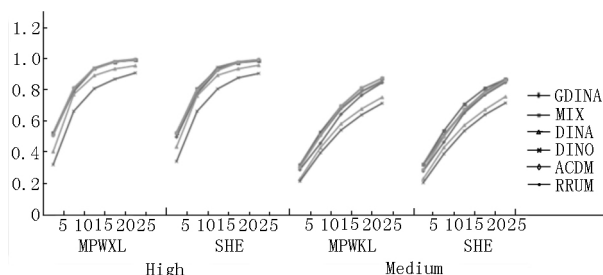


图7 RRUM 是真模型下的各 CDM 的模式判准率

当真模型是 G-DINA 时,高质量题库下 Mixed 选择的 CDM 包括:66% 是 G-DINA,16% 是 A-CDM,17% 是 RRUM,1% 是 DINO,Mixed 的模式判准率略优于 G-DINA.中等质量题库下 Mixed 选择的 CDM 包括:51% 是 G-DINA,26% 是 A-CDM,17% 是 RRUM,6% 是 DINA 和 DINO,各模型之间的效果差异有所增加.这与 G. Rojas 等^[10]的研究结论相符,饱和 CDM 和简化 CDM 的拟合度在统计学意义上无显著差异时,使用简化 CDM 更有利.当 Mixed 是真模型时,高质量题库下 Mixed 选择的 CDM 包括:18% 是 G-DINA,18% 是 DINA,20% 是 DINO,20% 是 A-CDM 和 24% 的 RRUM.中等质量题库下 Mixed 选择的 CDM 包括:17% 是 G-DINA,19% 是 DINA,20% 是 DINO,23% 是 A-CDM 和 23% 的 RRUM,Mixed 的判准率始终最高的,DINO 的结果始终是最差的,G-DINA 的效果和 Mixed 是最接近的,这是由于 G-DINA 作为饱和的 CDM,适用于不同属性作用机制的题目,但是其他简化模型 DINA 和 DINO 是完全不同的 2 类模型,A-CDM 和 RRUM 是 2 个相似的 CDM,若分别应用单模型来估计 Mixed 的题目,必定有较大部分题目参数是估不准的.

而当真模型是 DINA 时,高质量题库下 Mixed 选择的 CDM 包括:27% 是 G-DINA,2% 是 A-CDM,12% 是 RRUM,9% 是 DINO,50% 是 DINA.中质量题库下 Mixed 选择的 CDM 包括:26% 是 G-DINA,15% 是 RRUM,59% 是 DINA.Mixed 和真模型 DINA 的结果几乎是一样的,其次是 G-DINA、RRUM、A-CDM 和 DINO.

当真模型是 DINO 时,高质量题库下 Mixed 选择的 CDM 包括:32% 是 G-DINA,62% 是 DINO,其他模型占 6%.中质量题库下 Mixed 选择的 CDM 包括:36% 是 G-DINA,61% 是 DINO,其他模型占 3%.Mixed 和 DINO 的结果几乎是一样的,其次分别是 G-DINA、A-CDM、RRUM 和 DINA.

当真模型是 A-CDM 时,高质量题库下 Mixed 选择的 CDM 包括:4% 是 G-DINA,89% 是 A-CDM,6% 是 RRUM.中质量题库下 Mixed 选择的 CDM 包括:67% 是 A-CDM,27% 是 RRUM,6% 是 G-DINA.G-DI-

NA、Mixed、RRUM 和 A-CDM 的结果几乎是一样的,而表现最差的是 DINA 和 DINO(2 者结果基本相同)。这是因为 A-CDM 的基本假设是属性之间具有累积补偿效应,掌握属性对于答对概率具有线性加和关系。因而,A-CDM 属于部分补偿模型,RRUM、G-DINA 和 A-CDM 具有相似的计量性能,但这与 DINA 和 DINO 的假设并不相符。

当真模型是 RRUM 时,高质量题库下 Mixed 选择的 CDM 包括:92% 是 RRUM,4% 是 A-CDM,4% 是 G-DINA。中质量题库下 Mixed 选择的 CDM 包括:75% 是 RRUM,21% 是 A-CDM,4% 是 G-DINA。G-DINA、Mixed、A-CDM 和 RRUM 的结果很相似,其次是 DINA,最差的是 DINO。中等质量的题库 SHE 选题策略下,Mixed 的效果略优于其他模型。

3.2 不同 CDM 视角下题库安全性比较

3.2.1 题库使用均匀性(卡方值)比较 题库使用均匀性的比较结果如图 8~图 13 所示。

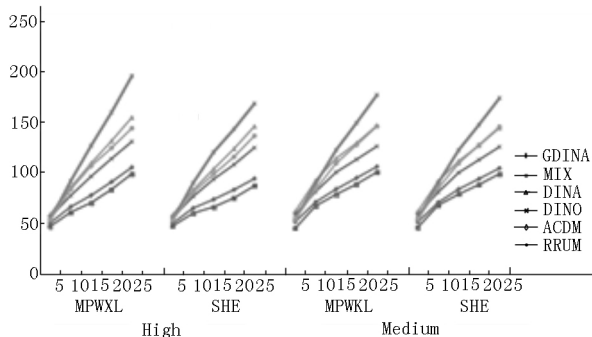


图 8 G-DINA 是真模型下测验卡方值

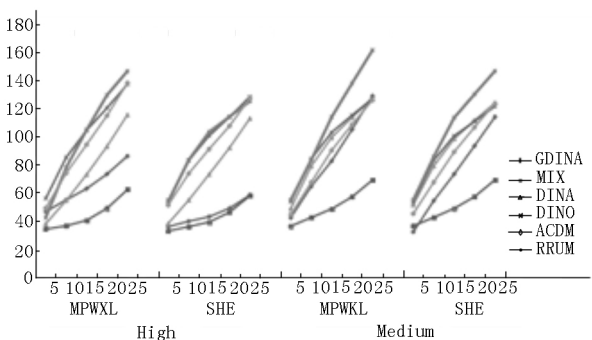


图 9 Mixed 是真模型下测验卡方值

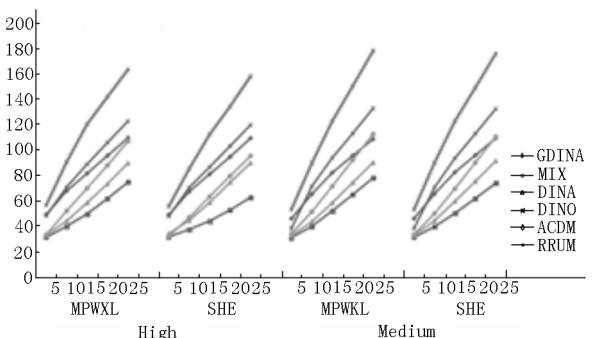


图 10 DINA 是真模型下测验卡方值

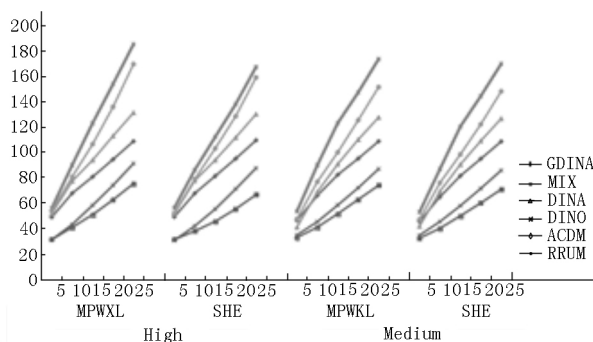


图 11 DINO 是真模型下测验卡方值

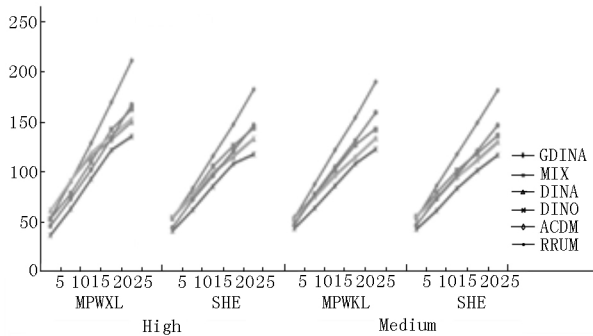


图 12 A-CDM 是真模型下测验卡方值

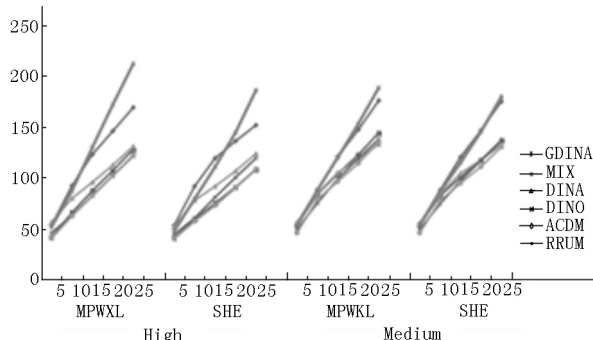


图 13 RRUM 是真模型下测验卡方值

对于题目调用均匀指标卡方值,当真模型是 G-DINA 或 Mixed 时,Mixed 的效果最优,其次是 G-DINA。当真模型是 DINA 时,表现从高到低的排序分别是: Mixed、DINA、RRUM、G-DINA、DINO、A-CDM。当真模型是 DINO 时,表现从高到低的排序分别是: Mixed、DINO、G-DINA、DINA、RRUM、A-CDM。当真模型是 A-CDM 或 RRUM 时,真模型的效果最好, Mixed 的表现居中,而此时 DINA 和 DINO 也有较好的效果,这可能是因为与 A-CDM 和 RRUM 这 2 个真模型相比,DINA 和 DINO 的模型假设、计量性能都与之有较大的不同,这可能会造成 DINA 和 DINO 对大部分题目的参数有较大的估计误差,因而 DINA 或 DINO 模型在选题的时候比较均匀调用了题库的题目。

3.2.2 测验重叠率的比较 测验重叠率的比较结果如图 14~图 19 所示。

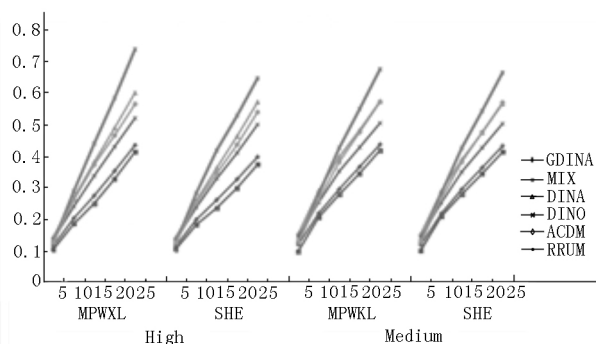


图 14 G-DINA 是真模型下测验 TOR

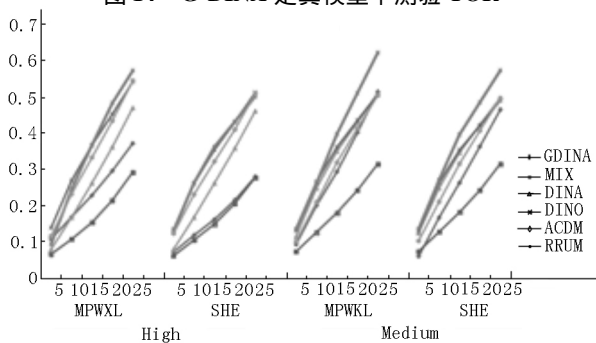


图 15 Mixed 是真模型下测验 TOR

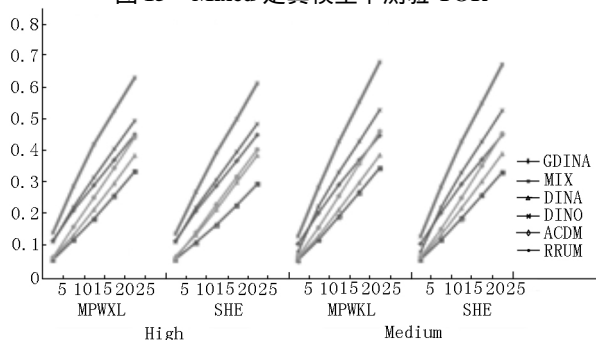


图 16 DINA 是真模型下测验 TOR

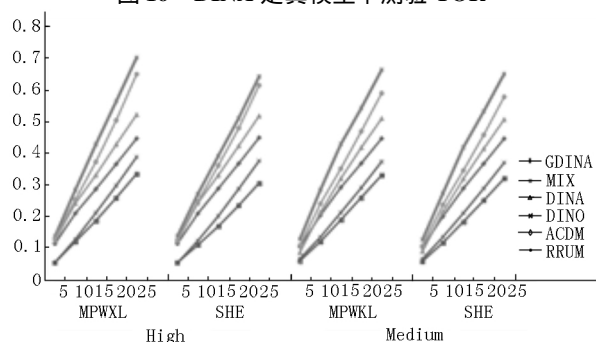


图 17 DINO 是真模型下测验 TOR

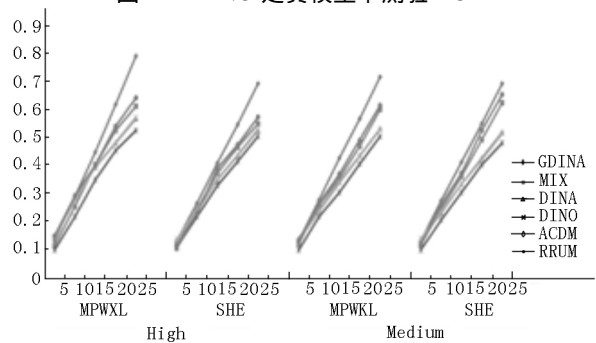


图 18 A-CDM 是真模型下测验 TOR

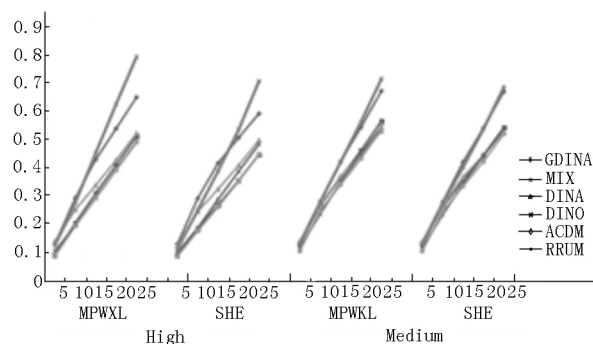


图 19 RRUM 是真模型下测验 TOR

题库重叠率指标和卡方值反映的是题库中题目的使用情况,从图 8 ~ 图 19 的结果可以看出,卡方值和重叠率指标的整体趋势也是一致的。在题库安全性方面,当真模型是 G-DINA、Mixed、DINA、DINO 时, Mixed 的效果是最好的,其次是真模型。而当真模型是 A-CDM 和 RRUM 时,真模型的效果最佳, Mixed 的效果居中。综合来看,当真模型是 Mixed 时, Mixed 的模式判准率和题库安全指标都是最好的,其次是 G-DINA。而当真模型是 G-DINA 时,在中等质量题库时, Mixed 的模式判准率略高于 G-DINA;但在题库安全指标中, Mixed 在高质量和中质量题库中的表现要略优于 G-DINA。当真模型是 DINA 和 DINO 时, Mixed 的模式判准率非常接近于真模型,但在题库安全方面, Mixed 略优于 DINA 和 DINO。当真模型是 A-CDM 和 RRUM 时,在模式判准率方面, Mixed 的表现接近于真模型,在题库安全指标中, Mixed 的表现略低于真模型。

4 总结与讨论

建立一个好的题库是 CD-CAT 成功的首要环节,当前 CD-CAT 题库建设均是基于单一认知诊断模型下构建的。在实际应用中,一份测验中不同题目可能会拟合不同 CDM,若所有题目采用同一个 CDM 来标定题目参数,则有可能造成题目与 CDM 之间失拟。

基于此,本文探讨了基于混合模型视角来建立题库的可行性,以期未来 CD-CAT 在实际应用中选用合适的 CDM 提供一种新的思路和视角。通过模拟研究发现,当题库是由一个模型(可以是饱和模型或简化模型)生成时,综合模式判准率和题库安全性来看,使用混合模型标定的题库参数的效果和真模型是非常接近的。而当真模型是混合模型时,也就是说当题库中的题目拟合不同 CDM 时,与使用混合模型相比,若使用单一模型(包括饱和模型和简

化模型)来构建 CD-CAT,不仅会降低模式判准率,还会提高题库曝光率。

尽管通过模拟实验表明混合模型具有较好的效果,需要强调的是本研究只是模拟研究,未来还需要通过更多的实证数据来多方面验证该思路的真实效果。

另外,需要注意的是在教育评估活动中,测验编制者在编制题目时,可能完全按照某一个模型来编制题目,一种非常理想的情形是通过实测发现,所编制的题目完全拟合预想的模型,如题库中所有题目完全拟合 DINA 或者 RRUM,这 2 个模型是 CD-CAT 的相关模拟和实证研究使用较多的模型。此时,意味着真模型是 DINA 或者 RRUM,这种情况下,通过本文的模拟研究发现,使用混合模型估计参数并建立 CD-CAT 和使用真模型(DINA 和 RRUM)建立产生的效果是非常接近的,也即意味着,与使用真模型相比,使用混合模型标定题库参数并不会明显降低 CD-CAT 的效果。但在实际应用中,上述理想的情形(测验完全拟合预想的模型)可能并不容易满足,即使测验编制者完全按照预想的模型编制题目,也很难保证所有的题目完全拟合预计的单一模型。因为,考生对题目的潜在加工过程是非常复杂的,对于有些题目考生的反应可能并未按照编题者预想的路径,换句话说,即有些题目可能拟合另外的模型,对于这类测验(题库),使用混合模型来估计参数可能效果会更好。

值得强调的是,与传统单一模型相比,混合模型的缺点是使用略微复杂一点,混合模型思路可以分为 2 个步骤:(i)通过 Wald 检验为每个题目选择模型,(ii)根据选择的模型估计题目参数。因此,在建立题库时,使用混合模型来估计题库参数时,要比单一模型多花一点时间,建立题库大量的精力和时间投入在前期的编制测验,以及进行实际测验等,与这些步骤相比,进行参数估计所花费的一点点时间可能是微不足道的。

为了验证与单一模型相比,使用混合模型建立 CD-CAT 会不会降低 CD-CAT 的实际测验效率。在相同的实验条件下,记录了 30 次分别基于混合模型和单一模型 CD-CAT 的运行时间,并对所有模型的运行时间进行了方差分析。结果发现模型之间的运行时间并没有显著差异,即基于混合模型建立 CD-CAT 并不会降低测验的运行效率。

限于时间及精力,本研究还有诸多值得进一步探讨的问题,本研究只通过 Wald 检验比较了 G-

DINA 是否可以被 DINA、DINO、A-CDM 和 RRUM 替代。但事实上题库中很可能有部分题目并不能通过 Wald 检验选择一个简化模型(默认选用 G-DINA),这部分题目可能拟合其他的简化模型。假如,在选择简化模型时有更多的选择,就会有更多的简化模型取代 G-DINA,简化模型对于诊断测验更有利,这时候 Mixed 标定的 CD-CAT 效果可能会更好。另外,本研究模拟的题库都是二级计分,多级的好处是可以指示部分掌握的情况,并提供更多的诊断信息,未来研究可以考虑在多级计分的测验中展开。

5 参考文献

- [1] 高旭亮,涂冬波.参数化认知诊断模型:心理计量性能特征、比较及其转换[J].江西师范大学学报:哲学社会科学版,2017,50(1):88-104.
- [2] de la Torre J. The generalized DINA model framework[J]. Psychometrika, 2011, 76(2):179-199.
- [3] von Davier M. A general diagnostic model applied to language testing data[J]. British Journal of Mathematical and Statistical Psychology, 2008, 61(2):287-307.
- [4] Henson R A, Templin J L, Willse J T. Defining a family of cognitive diagnosis models using log-linear models with latent variables[J]. Psychometrika, 2009, 74(2):191.
- [5] Haertel E H. Using restricted latent class models to map the skill structure of achievement items[J]. Journal of Educational Measurement, 2006, 26(4):301-321.
- [6] Templin J L, Henson R. Measurement of psychological disorders using cognitive diagnosis models[J]. Psychological Methods, 2006, 11(3):287-305.
- [7] Hartz S M. A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality[J]. American Journal of Gastroenterology, 2002, 95(4):906-909.
- [8] de la Torre J, Lee Y S. Evaluating the Wald test for item level comparison of saturated and reduced models in cognitive diagnosis[J]. Journal of Educational Measurement, 2013, 50(4):355-373.
- [9] Li Hiongli, Hunter C V, Lei Puiwa. The selection of cognitive diagnostic models for a reading comprehension test[EB/OL]. [2017-06-19]. https://www.researchgate.net/publication/281273001_The_selection_of_cognitive_diagnostic_models_for_a_reading_test.
- [10] Rojas G, de la Torre J, Olea J. Choosing between general and specific cognitive diagnosis models when the sample size is small[C]//In Annual Meeting of the National Council of Measurement in Education, Vancouver, British Columbia, Canada, 2012.

- [11] Ma Wenchao ,Jaconangelo C ,de la Torre J. Model similarity ,model selection ,and attribute classification [J]. Applied Psychological Measurement ,2016 ,40(3) :200-207.
- [12] de la Torre J ,van der Ark L A ,Rossi G. Analysis of clinical data from a cognitive diagnosis modeling framework [J]. Measurement and Evaluation in Counseling and Development ,2017 ,9(3) :210-213.
- [13] Tu Dongbo ,Gao Xuliang ,Wang Daxun ,et al. A new measurement of internet addiction using diagnostic classification models [J]. Frontiers in Psychology ,2017 ,8:1768.
- [14] Kaplan M ,de la Torre J ,Barrada J R. New item selection methods for cognitive diagnosis computerized adaptive testing [J]. Applied Psychological Measurement ,2015 ,39(3) :167-188.
- [15] 张龙 ,涂冬波. 多级计分题项目功能差异常用检测方法 及比较 [J]. 江西师范大学大学学报:自然科学版 , 2015 ,39(5) :441-448.

The Study on the CD-CAT Item Bank Based on the Perspective of Different CDMs

TU Dongbo ,GAO Xuliang ,WANG Daxun ,CAI Yan

(College of Psychology ,Jiangxi Normal University ,Nanchang Jiangxi 330022 ,China)

Abstract: The item bank is a very critical part of the cognitive diagnostic computerized adaptive testing (CD-CAT). In regular CD-CAT, an item bank uses the same cognitive diagnostic model to calibrate the item parameters, however, all the items in item bank can't be fully fitted with the same CDM. In the study the advantage of the method proposed by de la Torre to select an appropriate CDM for each item of item bank within a G-DINA framework has been taken for convenience, this method is referred to as a Mixed CDMs, and use a complex simulation study to compare the performance of Mixed CDMs and five special models (G-DINA, DINA, DINO, A-CDM and RRUM) in CD-CAT. The results show that across all conditions, the Mixed CDMs provided the ideal classification accuracy and utilization rate of the item bank.

Key words: CD-CAT; CDM; Mixed CDMs; item bank

(责任编辑: 冉小晓)

(上接第 356 页)

- [9] Mahmood H ,Anwar I ,Zafar M K. A construction of Cohen-Macaulay f -graphs [J]. Journal of Algebra and Its Applications ,2014 ,13(6) :1450012.
- [10] Aharoni R ,Berger E ,Meshulam R. Eigenvalues and homology of flag complexes and vector representations of graphs [J]. Geom Funct Anal ,2005 ,15(3) :555-566.
- [11] Caroline J Klivans. Threshold graphs ,shifted complexes , and graphical complexes [J]. Discrete Math ,2007 ,307(21) :2591-2597.
- [12] Roy Meshulam. Domination numbers and homology [J]. J Combin Theory: Ser A ,2003 ,102(2) :321-330.
- [13] Berger E. Relation between the maximal degree of a graph and the topological connectivity of its independence complex [J]. Banff International Research Station for Mathematical Innovation and Discovery ,2013.
- [14] Earl J ,Vander Meulen K N ,Van Tuyl A. Independence complexes of well-covered circulant graphs [J]. Experimental Mathematics ,2016 ,25(4) :441-451.
- [15] Woodroffe R. Vertex decomposable graphs and obstructions to shellability [J]. Proceedings of the American Mathematical Society ,2009 ,137(10) :3235-3246.

The Study about the Vertex Decomposability of f -Graph and Its Complement

WU Jinping ,GUO Jin*

(College of Information Science and Technology ,Hainan University ,Haikou Hainan 570228 ,China)

Abstract: A class of algebraic graph f -graph is characterized in this paper. It is proved that all f -graphs are vertex decomposable and the vertex decomposability of the complement of f -graph is also characterized.

Key words: vertex decomposable; f -graph; bipartite graph; Cohen-Macaulayness

(责任编辑: 曾剑锋)