

HO - GDINA 模型的 EM 算法参数估计^{*}

易 芹¹, 田 伟², 杨 涛², 辛 涛², 刘彦楼³

(1. 北京第五中学分校, 北京 100009; 2. 中国基础教育质量监测协同创新中心, 北京 100875;
3. 中国教育大数据研究院, 济宁 273165)

摘 要: Generalized DINA Model (G - DINA) 为认知诊断模型提供了一个一般性的理论框架, 而高阶诊断模型不仅能描述被试的总体水平, 还能描述被试对属性的掌握情况(微观的认知状态)以及被试掌握属性与能力的关系, 提供更丰富的信息。如果能把这两者结合起来, 可能对实际诊断工作的操作有较大帮助。文章首先对考虑高阶结构的整合性模型——HO - GDINA 模型的形式进行讨论, 探讨其参数估计 EM 算法的实现, 并用模拟过程对模型的估计精度进行研究, 结果验证了 HO - GDINA 的 EM 算法的正确性, 并且说明该算法对该模型有较高估计精确度。然后用饱和模型在约束条件下的特殊形式 HO - DINA 模型对“分数减法”这一经典数据进行 EM 算法参数估计和具体分析, 展示了 HO - GDINA 在实际情况中的具体使用, 并与 de la Torre 之前用 MCMC 估计算法得到的研究结果做比较, 基本一致, 进一步表明 HO - GDINA 模型的参数估计 EM 算法在实际情境中的特殊形式下仍然适用。

关键词: 认知诊断模型; HO - GDINA 模型; EM 算法; 估计精度

中图分类号: B841.2

文献标识码: A

文章编号: 1003 - 5184(2017)05 - 0441 - 08

1 引言

认知诊断模型 (Cognitive Diagnosis Model, CDM) 是一种统计工具, 它可以根据被试的作答数据推断被试潜在的知识掌握状态 (Haertel, 1989; Templin & Henson, 2006) 提供认知诊断信息, 报告学生的知识结构、加工技能等 (陈秋梅, 张敏强, 2010), 进行针对性的补救教学以有效促进教育发展 (关丹丹, 2009; 陈慧麟, 陈劲松, 2013)。针对实际情况的多样性和特殊性, 研究者们开发出不同的 CDM (甘媛源, 余嘉元, 2001; 刘声涛, 戴海琦, 周骏, 2006), 已有研究表明, 到目前为止, 已开发出的 CDM 达到百种以上 (辛涛, 乐美玲, 张佳慧, 2012)。

这些模型都是基于一定的假设基础, 有其适用的范围, 且各有各的优缺点 (Tatsuoka, 1983, 1995; Muraki, 1992; Henson, Templin, & Willse, 2009)。但是不同的数学公式是否代表着本质不同的模型或者这些模型之间到底有多大程度的相关却还不够被大家所详细讨论。de la Torre (2011) 提出了一般化模型——Generalized DINA Model (G - DINA), 作为整合性的广义一般性模型, G - DINA 模型为 CDM 提供了一个一般性的理论框架, 有较大应用前景。另一方面, de la Torre 和 Douglas (2004) 认为在一个相对较窄的测验领域, 可以假定测验属性与某种通用

能力是相关的, 并基于此提出了 High - Order DINA (HO - DINA) 这一高阶模型。它不仅能描述被试的总体水平——宏观的高阶 (higher - order) 能力, 还能描述被试对属性的掌握情况 (微观的认知状态) 以及被试掌握属性与能力的关系, 从而能够提供更丰富的信息。国内外学者 (涂冬波, 蔡艳, 戴海琦, 丁树良, 2010; Bolt & Fu, 2004; Hartz, 2002) 均对开发高阶结构的认知诊断测量模型进行过尝试和研究。

参数估计对于模型的发展与实现有着非常重要的意义 (涂冬波, 蔡艳, 戴海琦, 漆书青, 2008, 2013; Maris, 1999)。不同的估计方法也都有其优缺点。MCMC 方法虽然在操作上相对简单, 但它是一种计算密集型的算法, 当模型参数量、数据量或者缺失数据量增加时, 模型及抽样的复杂程度必然增加, 由此计算耗时量必然增加。EM 算法在参数估计的方法中是一个相对稳定的方法 (Yuling, Brian, & Alan, 2013), 而且有较好的准确率和效率 (Muraki, 1992; 张淑梅, 包钰, 保佳, 李姗姗, 2014), 如果能克服其推导公式较复杂的过程, 当模型所涉及的参数数量较多时, 不失为一个更好的选择。而在以往的研究中, 并没有文献对高阶结构下整合性模型的 EM 算法参数估计作系统地研究。

^{*} 基金项目: 国家自然科学基金面上项目 (31371047)。

通讯作者: 辛涛, E-mail: xintao@bnu.edu.cn。

综上,文章希望能在同一个模型中,不仅描述被试的总体水平——宏观的高阶(higher-order)能力,同时描述被试对属性的掌握情况(微观的认知状态)以及被试掌握属性与能力的关系。所以,提出了高阶结构下的G-DINA模型(HO-GDINA模型)给出了其相应的形式。从某一层面上来说,HO-GDINA模型并非一个全新的模型,它是对于HO-DINA以及G-DINA模型研究的进一步拓展,使之适用于更一般化的情景,这是一个模型整合与关系梳理的过程,其提出更大的意义在于从一个更大更广与整合的角度来认识已有的认知诊断模型和其与IRT之间的联系。因此,这一工作是认知诊断模型发展的必然过程,其对于实际情况也更有包容性。提出新模型后,文章将对高阶结构下整合性模型的EM算法进行研究,探讨对其参数估计的实现方法。然后用其特定的模型HO-DINA对“分数减法经典数据”(de la Torre, 2004; De Carlo, 2010)进行相应的参数估计与具体分析,与先前的研究结果做比较,展示HO-GDINA模型在实际情况中的一种使用方式,并进一步辅助说明参数估计的正确性,得出结论。

2 HO-DINA模型和HO-GDINA模型

文章把考虑高阶结构(High-Order)的Generalized DINA Model(G-DINA)模型简记为HO-GDINA模型。

2.1 HO-DINA模型

HO-DINA模型是建立在传统DINA模型基础之上的。De la Torre和Douglas(2004)指出,一个能评估属性掌握情况的模型应该与一种或几种更高阶的一般智力或一般能力相联系,那些具备了更高水平能力(θ)的被试更有可能具备测验项目要求的认知属性。包含了高阶结构的DINA模型就称为higher-order DINA模型(HO-DINA),如下函数关系来表征高阶结构:

$$P(Y_{ij} = 1 | \vec{\alpha}_i) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}} \quad (1)$$

$$P(\vec{\alpha} | \theta) = \prod_{k=1}^K P(\alpha_k | \theta) \quad P(\alpha_k | \theta) = \frac{\exp(\lambda_k + \lambda_{0k}\theta)}{1 + \exp(\lambda_k + \lambda_{0k}\theta)} \quad (2)$$

其中(1)式中的参数含义与DINA模型(张潇,沙如雪, 2013; Von Davier, 2014)中的含义一致,(2)式中为 λ_k 属性 k 的截距参数;表示 λ_{0k} 属性 k 在能力维度上的负荷(可以是多维的,但大多数情况下是一维的)。

2.2 HO-GDINA模型

通常,不同的属性掌握模式向量在正确作答概

率上可能有不同程度上的差异,即项目的各认知属性对试题答对概率有着不同的贡献比例,掌握部分认知属性的被试也有一定的答对概率,也就是说可以把被试不同的组别分得更细。沿着这样的思路,在DINA模型的基础上,de la Torre(2011)提出了G-DINA这一模型。与DINA模型相比,G-DINA放松了一些对条件的限制,它的饱和模型和基于其它连接函数的认知诊断的一般模型是可以进行等价的。

G-DINA模型的公式表达如下:

$$P(\vec{\alpha}_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{ijk}^* + \sum_{k' > k}^{K_j^*} \sum_{k=1}^{K_j^*} \delta_{jkk'} \alpha_{ijk}^* \alpha_{ijk'}^* + \dots + \delta_{j1 \dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{ijk}^* \quad (3)$$

其中, δ_{j0} 是项目 j 的截距参数, δ_{jk} 是 α_{ijk}^* 的主要影响因子, $\delta_{jkk'}$ 是 α_{ijk}^* 和 $\alpha_{ijk'}^*$ 的交互影响因子, $\delta_{j1 \dots K_j^*}$ 是 $\alpha_{ij1}^*, \dots, \alpha_{ijK_j^*}^*$ 的交互影响因子。所有项目的参数个数为 $\sum_{j=1}^J 2^{K_j^*}$ 。

与HO-DINA类似,在G-DINA的基础上,加入高阶结构就可称为HO-GDINA模型。其中各参数含义与HO-DINA中一样。HO-GDINA模型的

参数个数一共为 $\sum_{j=1}^J 2^{K_j^*} + 2K_j^*$

$$P(\vec{\alpha} | \theta) = \prod_{k=1}^{K_j^*} P(\alpha_k | \theta) \quad P(\vec{\alpha}_k | \theta) = \frac{\exp[\lambda_{0k}(\theta - \lambda_k)]}{1 + \exp[\lambda_{0k}(\theta - \lambda_k)]} \quad (4)$$

3 模型参数估计的EM算法

文章考虑采用de la Torre(2011)对G-DINA模型参数估计的处理思路对项目参数进行估计,再推导出属性参数的估计公式。 λ 参数估计分为两部分来进行。对基本模型G-DINA中的各 δ 采用边际极大似然估计方法(marginalized maximum likelihood estimation, MML)与设计矩阵形式下最小二乘法结合的方式进行估计,对高阶结构中的 λ 参数直接采用MML进行估计。具体如下:

3.1 项目参数 δ 的估计过程

先将似然函数对概率值 P 求偏导,令其为0,解出 P 的估计值,再在所有掌握模式下用矩阵形式下的最小二乘法获得参数估计值。

在用MML估计方法计算可得 $\hat{P}(\vec{\alpha}_{ij}^*)$ 后,利用G-DINA模型在设计矩阵 D_j 下的形式

$$P(\alpha_j) = D_j \delta_j \quad (5)$$

(模型公式的矩阵书写,不改变本质)来估计项目参数值可得到参数向量的最小二乘估计值。

$$\hat{\delta}_j = f[\hat{P}(\vec{\alpha}_{ij}^*)] = (D_j' D_j)^{-1} D_j' W_j \hat{P}(\vec{\alpha}_{ij}^*) \quad (6)$$

即参数向量 δ_j 是关于 $\hat{P}(\vec{\alpha}_{ij}^*)$ 的函数。

3.2 高阶结构中属性参数 λ 的估计过程

模型对数似然函数为 $l(\vec{\lambda}) = \log [L(\vec{\lambda})] =$

$$\log \prod_{i=1}^N L(\vec{\lambda}_i) = \sum_{i=1}^N \log L(\vec{\lambda}_i)$$

对参数 λ 求偏导(de la Torre 2009a 2009b) 整理可得:

当 $\lambda = \lambda_{0k}$ 时,

$$\frac{\partial l(\vec{\lambda})}{\partial \lambda_{0k}} = \int (\theta - \lambda_{0k}) [R_{lkq} - I_q P(\alpha_{lk}^* | \theta)] p(\theta) d\theta \quad (7)$$

当 $\lambda = \lambda_k$ 时,

$$\frac{\partial l(\vec{\lambda})}{\partial \lambda_k} = \int -\lambda_{0k} [R_{lkq} - I_q P(\alpha_{lk}^* | \theta)] p(\theta) d\theta \quad (8)$$

其中,

$$I_{lq} = \sum_{i=1}^N \frac{L(\vec{\lambda}_i | l^*) L(l^* | \theta)}{L(\vec{\lambda}_i)} I_q = \sum_{i=1}^L \sum_{i=1}^N \frac{L(\vec{\lambda}_i | l^*) L(l^* | \theta)}{L(\vec{\lambda}_i)} R_{lkq} = \sum_{i=1}^L \alpha_{lk}^* \sum_{i=1}^N \frac{L(\vec{\lambda}_i | l^*) L(l^* | \theta)}{L(\vec{\lambda}_i)}$$

令偏导都为 0, 则在程序中可用迭代法求解出 λ 的值。

3.3 被试属性掌握模式 α_i 参数

文章采用最大化后验估计(Maximum a posteriori estimation, MAP) 方法来得到每个被试的属性掌握模式。经过计算可得每个被试掌握模式 α_i 的后验概率为

$$p(\alpha_i | \vec{\lambda}_i) = \frac{p(\alpha_i) P(\vec{\lambda}_i | \alpha_i)}{\sum_{i=1}^L P(\vec{\lambda}_i | \alpha_i) p(\alpha_i)} \quad (9)$$

取该值最大所对应的掌握模式为对该被试估计的属性掌握模式。有研究表明, MAP 方法可以得到更高的属性边际判断率(Huebner & Wang 2011)。

3.4 高阶结构中能力 θ 参数

其估计过程与高阶结构中的参数 λ 的估计过程类似, 让似然函数对 θ 求偏导, 整理可得:

$$\frac{\partial l(\vec{\lambda})}{\partial \theta} = \int \lambda_{0k} [R_{lkq} - I_q P(\alpha_{lk}^* | \theta)] p(\theta) d\theta + \int I_q d\theta \cdot p(\theta) \quad (10)$$

令偏导都为 0, 可用迭代法求解出各被试的 θ 值。

4 模拟研究—HO - GDINA 模型参数估计 EM 算法的实现及估计精度

用 MonteCarlo 模拟方法得到的数据来验证研究设计的 EM 算法的可行性, 并考察估计结果与模拟数据间的偏差(BIAS)、均方根误差(RMSE) 和平均绝对离差指标(ABSE), 用以体现算法的估计精度。

4.1 数据模拟过程

假设测验项目长度为 30, 所有项目共涉及属性个数为 5, 被试人数为 5000。

(1) 首先给定 HO - GDINA 模型能力值的分布和各参数分布, 并从相应分布中随机生成参数真值 θ 、 λ 和 δ 。这里项目参数 δ 和属性参数 λ 真值分布均采用 delaTorre&Douglas(2004) 所采用的分布。 θ 设置为(-6, 6) 上的均匀分布, 假设所有属性的斜率参数相同, 即 $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$, 设 $\lambda = (1, -0.71, -1.16, 0.45, -0.5)$, 项目的 Q 矩阵表 1 所示。

表 1 模拟数据 Q 矩阵

项目号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
q1	1	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	0	0	0	0	1	1	0	0	0	1	1	1	0	0
q2	0	1	0	0	0	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	0	1	1	0	1	0	0	1	1
q3	0	0	1	0	0	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	1	1	1	0
q4	0	0	0	1	0	0	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	1	1	0	0	1	1	0	0	1
q5	0	0	0	0	1	0	1	0	0	1	0	1	1	0	1	0	0	0	0	1	0	0	0	1	1	0	0	1	1	1

(2) 根据参数真值及 HO - GDINA 模型高阶函数生成被试的属性掌握模式概率, 从而随机生成属性掌握模式的真值, 进而得到正确作答概率, 再随机生成反应作答矩阵。

(3) 根据生成的作答矩阵采用 HO - GDINA 模型的自编的 R 算法估计所有参数(每个项目都选择饱和模型)。

(4) 重复(1) 到(3) 步 1000 次, 即重复实验 1000 次, 减少随机误差。

(5) 将估计出来的参数与第(1) 步生成的参数真值进行比较, 以考察参数估计的精度。估计值与真值之间的差异越小说明该估计方法越有效, 且能考察参数估计的返真性。

4.2 评价指标

(1) 偏差 (BIAS)、均方根误差 (RMSE) 和平均绝对离差指标 (ABSE)

如果用 π_j 表示参数真值, $\hat{\pi}_j$ 表示参数估计值, r 表示重复次数, 那么对应的公式分别为:

$$BIAS(\pi_j) = \frac{1}{1000} \sum_{r=1}^{1000} \frac{(\hat{\pi}_j - \pi_j)}{\pi_j} \quad (11)$$

$$RMSE(\pi_j) = \sqrt{\frac{1}{1000} \sum_{r=1}^{1000} (\hat{\pi}_j - \pi_j)^2} \quad (12)$$

$$ABSE(\pi_j) = \frac{1}{1000} \sum_{r=1}^{1000} |\hat{\pi}_j - \pi_j| \quad (13)$$

(2) 边际判准率 (Marginal MatchRate, MMR; de la Torre, 2010) 和模式判准率 (Pattern MatchRate, PMR)。公式分别为:

$$MMR(\alpha_k) = \sum_{i=1}^N \frac{I(\hat{\alpha}_{ik} = \alpha_{ik})}{N} \quad (14)$$

$$PMR(\vec{\alpha}_1) = \sum_{i=1}^N \frac{\prod_{k=1}^K I(\hat{\alpha}_{ik} = \alpha_{ik})}{N} \quad (15)$$

4.3 分析工具

采用 R 语言自编程序

4.4 研究结果

表 2 和表 3 分别是项目参数和属性参数的返真性情况。表 2 中每个项目的参数返真均值不论是偏差 (bias)、均方根误差 (RMSE) 还是平均绝对离差指标 (ABSE) 基本都小于 0.02, 比较理想。表 3 中对属性参数的估计也比较理想, 各参数的返真值基本都达到较高精度。

表 2 HO - GDINA 模型下项目参数的返真性情况

项目编号	bias 平均值	RMSE 平均值	ABSE 平均值
1	-0.001	0.005	0.082
2	0.013	0.005	0.062
3	-0.011	0.005	0.083
4	0.039	0.004	0.060
5	0.014	0.004	0.053
6	-0.053	0.004	0.063
7	0.098	0.007	0.011
8	-0.090	0.012	0.016
9	-0.014	0.010	0.018
10	0.034	0.006	0.087
11	-0.031	0.021	0.032
12	0.014	0.018	0.028
13	0.005	0.010	0.016
14	-0.008	0.007	0.011
15	-0.002	0.009	0.013
16	0.050	0.002	0.040
17	-0.113	0.006	0.095
18	-0.061	0.002	0.031
19	0.034	0.004	0.062
20	-0.011	0.003	0.054
21	0.067	0.009	0.014
22	0.017	0.005	0.068
23	-0.015	0.013	0.024
24	0.021	0.004	0.067
25	0.032	0.005	0.081
26	0.032	0.022	0.035
27	0.007	0.009	0.014
28	-0.005	0.006	0.093
29	-0.003	0.008	0.012
30	0.004	0.010	0.017

表 3 HO - GDINA 模型下属性参数的返真性情况

	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
bias 值	-0.0110	0.0119	0.0411	-0.0322	-0.0540	0.0982
RMSE 值	0.0008	0.0041	0.0037	0.0009	0.0014	0.0017
ABSE 值	0.0135	0.0718	0.0605	0.0143	0.0251	0.0242

表 4 中显示, 模型对每个属性的边际判准率都高达 0.95 以上, 模式判准率也有较高的准确度, 被试能力偏差 (bias)、均方根误差 (RMSE) 和平均绝对离差指标 (ABSE) 的平均值都比较理想, 但基本都比属性参数大, 这可能是由于能力参数的估计值在

一定程度上依赖于项目参数 (见估计程序迭代公式)。而且被试参数 (被试属性参数和被试能力参数) 1000 次实验结果的平均标准差都很小, 说明程序估计有良好的稳健性。

表 4 HO - GDINA 模型下被试参数掌握属性判准率与能力估计情况

	单个属性边际判准率 (MMR)					模式判准率 (PMR)	被试能力值		
	α_1	α_2	α_3	α_4	α_5		bias	RMSE	ABSE
平均值	0.953	0.954	0.950	0.952	0.951	0.794	0.072	0.023	0.068
标准差	0.0000003	0.0000142	0.0000083	0.0000086	0.0000031	0.0000106	0.006	0.002	0.007

以上这些结果表明研究对项目参数、属性参数和被试参数估计返真性较好, 估计的精度较高, 稳健性

较强 (赵顶位, 戴海琦, 2012; de la Torre, 2008)。HO - GDINA 模型的参数估计 EM 算法具有可行性, 同时也

说明 HO – GDINA 具有较高的判准率。

5 实证研究—分数减法经典数据分析

文章对 de la Torre 等人曾经在研究中使用的经典“分数减法数据”进行分析,结合数据特点,由于要正确回答相关项目必须把项目所涉及到的所有属性全部掌握,所以在考虑到基础模型时,文章具体使用 HO – DINA 模型进行相关分析,这一连接模型比其它的模型更适合本数据。分析过程采用作者编写的 EM

算法进行参数估计,通过与前人研究中 MCMC 方法的结果进行比较来辅助说明算法的正确性。

5.1 实证数据

经典“分数减法数据”是由 20 个分数减法项目构成,一共涉及到 8 个属性,并有 2144 名中学生参与了测试。最开始使用该数据的是 K. Tatsuoka (1990),文章用该公开数据作了相应的分析,其 Q 矩阵如表 5 所示:

表 5 分数减法数据 Q 矩阵

项目号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
q_1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0
q_2	0	0	0	1	1	0	1	0	1	1	1	0	1	1	0	1	1	1	1	1
q_3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
q_4	1	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
q_5	0	0	0	1	0	0	0	0	0	1	1	0	1	0	0	0	1	1	1	1
q_6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
q_7	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
q_8	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0

5.2 数据分析

为了进一步验证 EM 算法的准确性,文章将其项目参数估计结果与 de la Torre 和 Douglas(2004)所报告的 MCMC 算法结果作对照比较。

5.3 研究结果

从表 6 可以看出 g 参数的估计值范围在 0.003 ~ 0.42 之间,且有 17 个项目的 g 值小于 0.2,属于一般可接受范围; s 参数的估计值范围在 0.03 ~ 0.32,且有 16 个项目的 s 值小于 0.2,属于一般可接

受范围。使用文章中的 EM 算法所得到的项目参数值基本与 de la Torre 和 Douglas(2004)所报告的 MCMC 算法参数结果相吻合,从一定程度上验证了模型算法的正确性。总体来看,第 8 个项目的 g 参数和 s 参数都不太理想,可能与其 Q 矩阵的设定(Rupp & Templin,2008,2010;涂冬波,蔡艳,戴海琦,2012a,2012b)不够准确有关系(只涉及到一个简单属性)。

表 6 文章 EM 算法估计结果与 MCMC 算法估计结果比较

项目号	EM 算法				MCMC 算法			
	g	SE(g)	$1-s$	SE($1-s$)	g	SE(g)	$1-s$	SE($1-s$)
1	0.04	0.02	0.89	0.02	0.04	0.01	0.90	0.01
2	0.03	0.01	0.97	0.02	0.03	0.01	0.96	0.01
3	0.003	0.01	0.87	0.01	0.00	0.00	0.88	0.01
4	0.22	0.02	0.83	0.02	0.22	0.01	0.89	0.01
5	0.30	0.02	0.94	0.03	0.30	0.02	0.96	0.01
6	0.04	0.03	0.95	0.01	0.03	0.02	0.96	0.01
7	0.03	0.01	0.83	0.02	0.03	0.01	0.81	0.01
8	0.42	0.01	0.82	0.02	0.44	0.03	0.81	0.01
9	0.19	0.02	0.76	0.02	0.18	0.03	0.75	0.01
10	0.03	0.01	0.78	0.01	0.03	0.01	0.79	0.01
11	0.07	0.01	0.89	0.02	0.07	0.01	0.93	0.01
12	0.13	0.02	0.92	0.01	0.13	0.02	0.96	0.01
13	0.02	0.01	0.68	0.02	0.02	0.00	0.67	0.02
14	0.04	0.01	0.95	0.01	0.05	0.01	0.93	0.01
15	0.03	0.01	0.86	0.02	0.04	0.01	0.86	0.01
16	0.10	0.01	0.88	0.01	0.10	0.02	0.88	0.01
17	0.04	0.01	0.82	0.01	0.04	0.01	0.86	0.01
18	0.12	0.01	0.83	0.02	0.13	0.01	0.85	0.01
19	0.02	0.01	0.74	0.02	0.02	0.00	0.76	0.02
20	0.01	0.01	0.83	0.02	0.01	0.00	0.84	0.01

表 7 是属性参数的结果,从中可以看出属性 1 和属性 5 的截距参数相对大一些,属性 6 和属性 7 的截距参数相对小一些,这从一定程度上可以说明相对而言,掌握属性 1 和属性 5 困难一些,而掌握属性 6 和属性 7 容易一些。

表 7 分数减法数据属性参数的估计情况

属性	λ_0	λ_k
α_1	3.02	0.02
α_2	3.02	-0.92
α_3	3.02	-0.48
α_4	3.02	-0.31
α_5	3.02	0.06
α_6	3.02	-2.02
α_7	3.02	-1.21
α_8	3.02	-0.87

根据 MAP 的估计方法对被试的属性掌握模式进行分类,可得到表 8 的结果。较大部分的被试 (39.5%) 掌握了所有的属性;所有被试都掌握属性 6,其中有 11.2% 的被试只掌握了属性 6;掌握属性 5 的被试最少 (49.4%),掌握属性 1 的被试比例也很低,仅占总体的 51%。这一结论所抽取的 536 名被试中没有被试一个属性都没有掌握。

表 8 分数减法数据被试掌握模式分布

掌握模式	频数	百分比 (%)
00000100	240	11.2
00000110	50	2.3
00000111	38	1.8
00010110	30	1.4

表 9 分数减法数据项目拟合结果

项目号	1	2	3	4	5	6	7	8	9	10
RMSEA	0.056	0.076	0.078	0.134	0.097	0.069	0.078	0.065	0.087	0.067
项目号	11	12	13	14	15	16	17	18	19	20
RMSEA	0.102	0.078	0.089	0.052	0.04	0.067	0.078	0.054	0.064	0.043

文章采用 EM 算法分别在 HO-DINA 和 DINA 模型下对数据进行分析,以此可检验 EM 算法下 HO-DINA 的模型拟合情况。从表 10 中可以看出,模

表 10 分数减法数据模型拟合结果

模型	AIC	BIC	-2LL	CAIC	NP
HO-DINA(EM)	8756.466	8981.253	8643.527	9023.433	49
DINA(EM)	8874.567	9134.598	8621.476	9217.634	40

注: -2LL = -2 log likelihood; AIC = Akaike's information criterion; BIC = Bayesian information criterion; CAIC = Computer - Assisted Indexing and Categorizing; NP = number of parameters.

6 结果与讨论

相对于简单模型来说,广义一般性模型分配到每个项目上的参数增加较多,以使模型能将属性掌

续表 8

掌握模式	频数	百分比 (%)
00110111	10	0.5
01000100	120	5.6
01000110	80	3.7
01001111	8	0.4
01100111	210	9.8
01101111	45	2.1
01110110	16	0.7
01110111	120	5.6
01111110	5	0.2
01111111	80	3.7
10100111	4	0.2
10111111	5	0.2
11011111	20	0.9
11100111	28	1.3
11101110	4	0.2
11101111	38	1.8
11110110	8	0.4
11110111	130	6.1
11111110	8	0.4
11111111	847	39.5
总计	2144	100

5.4 项目拟合与模型拟合

在 HO-DINA 模型的 EM 算法下,得到各项目拟合 RMSEA 值如下表 9 所示,均值为 0.074。从表中可知,项目 4 和 11 拟合度相对差一些,也提醒研究者可以进一步检查相应的项目质量或者 Q 矩阵的设置。

型 HO-DINA 模型相对来说比 DINA 模型拟合统计量基本都较小,对数据拟合得更好。

握模式空间划分得更加细致。而高阶结构的引入可以更好地较准被试属性掌握情况与其一般能力之间的关系,模型的复杂度从一定程度上增加了与灵活

性。因此文章以此为着手点进行了相应的研究。研究使用 EM 算法来估计 HO - GDINA 这一高阶广义一般性模型框架下的项目参数。文章通过模拟研究和一个对经典实证数据的分析来验证了算法的准确性和精度,在同一具体模型下,EM 算法对实证数据给出的参数估计结果与前人使用 MCMC 算法得到的结果是基本一致的。

EM 算法参数估计只是对该模型进行的初步研究,在今后的研究中,还可以考虑对 EM 算法和 MC-MC 算法对不同模型在不同条件下的估计效果进行更为细致与系统地比较(如属性个数、项目数量、被试数量等因素对模型估计的影响),以便研究者们在实际使用模型时能更好更快地选择估计方法,得到优质的结果。

当然,在实际应用模型时,研究者应该对 Q 矩阵的设定及属性层级关系的界定做好相应的工作,以确定测验所涉及的认知属性及属性间的逻辑关系。从文章对实证数据的分析中可看出这一工作的重要性,因为它可能会直接影响模型的参数估计及其它相关结果。因此,可以对“如何更准确地设定 Q 矩阵及给定属性层级关系、相关影响如何”等做出相应研究。

此外,“如何对广义一般性模型进行参数检验”、“在实际问题中,如何选择广义一般性模型的具体形式”、“对于多级评分的高阶模型如何实现 EM 算法的估计”和“考虑到实际情况,涉及属性掌握程度的模型如何进行分析”等相关基础和应用领域的问题也都有待进一步探索。

参考文献

- 陈秋梅,张敏强.(2010).认知诊断模型发展及其应用方法述评.心理科学进展,18,522-529.
- 陈慧麟,陈劲松.(2013).G-DINA 认知诊断模型在语言测验中的验证.心理科学,36(2),1470-1475.
- 关丹丹.(2009).认知诊断理论与考试评价.中国考试(研究版),3(3),8-12.
- 甘媛源,余嘉元.(2001).认知诊断模型研究新进展.湖北大学学报(哲学社会科学版),37,121-124.
- 刘声涛,戴海崎,周骏.(2006).新一代测验理论——认知诊断理论的源起与特征.心理学探新,26(3),73-77.
- 辛涛,乐美玲,张佳慧.(2012).教育测量理论新进展及发展趋势.中国考试,5(5),3-11.
- 涂冬波,蔡艳,戴海崎,漆书青.(2008).现代测量理论下四大认知诊断模型述评.心理学探新,28(3),64-68.
- 涂冬波,蔡艳,戴海崎,丁树良.(2010).小学儿童数学问题解决的认知诊断.心理科学,33(6),1461-1466.
- 涂冬波,蔡艳,戴海崎,丁树良.(2011).HO-DINA 模型的 MCMC 参数估计及模型性能研究.心理科学,34(6),1476-1481.
- 涂冬波,蔡艳,戴海崎.(2012a).基于 DINA 模型的 Q 矩阵修正方法.心理学报,44(4),558-568.
- 涂冬波,蔡艳,丁树良.(2012b).认知诊断理论、方法与应用.北京:北京师范大学出版社.
- 涂冬波,蔡艳,戴海崎.(2013).基于 HO-DINA 模型的多级评分认知诊断模型的开发.心理科学,36(4),984-988.
- 赵顶位,戴海崎.(2012).4~8 年级学生几何类比推理问题解决认知诊断.心理学探新,32(3),77-81.
- 张潇,沙如雪.(2013).认知诊断 DINA 模型研究进展.中国考试,1(1),32-37.
- 张淑梅,包钰,保佳,李姗姗.(2014).项目之间具有相关性的认知诊断模型.北京师范大学学报(自然科学版),50,331-336.
- de la Torre J., & Douglas. (2004). Higher-Order Latent trait Models for Cognitive Diagnosis. Psychometric, 69, 333-353.
- de la Torre J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. Journal of Educational Measurement, 45, 346-362.
- de la Torre J. (2009a). DINA model and parameter estimation: A didactic. Journal of Educational and Behavioral Statistics, 34, 115-130.
- de la Torre J. (2009b). A cognitive diagnosis model for cognitively based multiple-choice options. Applied Psychological Measurement, 33, 163-183.
- de la Torre J., Hong Y., & Deng W. L. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. Journal of Educational Measurement, 47, 227-249.
- De Carlo L. T. (2010). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. Applied Psychological Measurement, 35, 8-24.
- de la Torre J. (2011). The generalized DINA model framework. Psychometric, 76, 179-199.
- Haertel E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. Journal of Educational Measurement, 26, 301-321.
- Henson R., Templin J., & Willse J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. Psychometrika, 74, 191-210.
- Leighton J. P., Gierl M., & Hunka S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsukawa's rule-space approach. Journal of Educational Measurement, 41, 205-237.
- Muraki E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159-176.
- Maris E. (1999). Estimating multiple classification latent class models. Psychometrika, 64, 187-212.

- Rupp, A. A. & Templin, J. (2008). Effects of Q - matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement* 68, 78 - 98.
- Rupp, A. A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford Press.
- Tatsuoka, K. K. (1983). Rulespace: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement* 20, 345 - 354.
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessments* (pp. 327 - 359). New York, NY: Guilford Press.
- Templin, J., & Henson, R. (2006). *Measurement of psychological disorders using cognitive diagnosis models*. Psychological Methods.
- Von Davier, M. (2014). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67, 49 - 71.
- Yuling, F., Brian, T. H., & Alan, H. (2013). Parameter Estimation of the Reduced RUM Using the EM Algorithm. *Applied Psychological Measurement* 38, 137 - 150.

EM Algorithm Parameter Estimation of HO - GDINA Model

Yi Qin¹, Tian Wei², Yang Tao², Xin Tao², Liu Yanlou³

(1. Beijing Fifth Middle School, Beijing 100009; 2. Collaborative Innovation Center for Quality Monitoring of Basic Education in China, Beijing 100875; 3. China Institute of Educational Data, Jinan 273165)

Abstract: Generalized DINA Model (G - DINA) is a cognitive diagnostic model providing a general theoretical framework. Higher - order latent traits can provide not only the overall level of participants, but also the cognitive situation and the relationship between general aptitude and specific knowledge, which provides more abundant information. If we combine the two, it may be helpful for the actual diagnosis. First, the HO - GDINA model are discussed with the form of model and the EM algorithm for estimating parameters. In the simulation study, the results validate the correctness of HO - GDINA with EM algorithm. Then, we use the saturated model under the constraint condition HO - DINA model to analyze the classical data "subtract of fraction" with EM algorithm. Comparing with de la Torre's research of MCMC estimation algorithm, the results are almost the same, signifying that the parameter estimation with EM algorithm of the special HO - GDINA model under the actual situation still apply.

Key words: CDMs; HO - GDINA model; EM algorithm; estimation accuracy