

R-DINA 模型参数估计 EM 算法准确性检验

宋丽红¹, 戴海琦², 汪文义³, 丁树良³

(1. 江西师范大学 初等教育学院, 南昌 330027; 2. 江西师范大学 心理学院, 南昌 330022;

3. 江西师范大学 计算机信息工程学院, 南昌 330022)

摘 要: DINA 模型是一个倍受关注和得到广泛应用的认知诊断模型, 但 DINA 模型在每个项目上只将被试分为掌握和未掌握两类。文章提出的改进的 DINA 模型(R-DINA) 可以对被试进行更为细致的分类。文章首先简要介绍 R-DINA 模型和该模型参数估计的 EM 算法, 然后设计模拟实验对 EM 算法的参数估计准确性进行检验。实验结果表明, R-DINA 模型的 EM 算法估计结果稳定, 项目参数估计精度和被试分类准确性较高。

关键词: 认知诊断模型; R-DINA 模型; 参数估计; EM 算法

中图分类号: B841.2

文献标识码: A

文章编号: 1003-5184(2012)05-0410-04

1 引言

认知诊断评估(cognitive diagnostic assessment, CDA) 是一种基于认知心理学并结合现代测量学的方法, 旨在探索人类在特定领域的潜在认知过程和认知结构的测验形式, 是现代测量理论的核心内容。认知诊断评估这一目的的实现需要相应的具有认知诊断功能的测量模型, 即认知诊断模型(cognitive diagnostic model, CDM) 来建立被试的不可观察的属性掌握和可观察的项目反应间的关系, 从而达到对不可观察的属性掌握模式进行推断。

在众多的认知诊断模型中, DINA 模型由于模型简单, 模型参数解释性好在国内外得到了广泛的应用(de la Torre, 2008, 2009, 2011; de la Torre, Hong & Deng, 2010; de la Torre & Karelitz, 2009; de la Torre & Lee, 2010; Huebner & Wang, 2011; Junker & Sijtsma, 2001; Liu, You, Wang, Ding, & Chang, in press; Tatsuoka, 2002; 杨智为, 卓淑瑜, 郭伯臣, 陈亭宇, 2011)。然而, DINA 模型只将被试在项目上的潜在知识状态分为掌握组与未掌握组两类, 忽视了未掌握组的被试在属性掌握多少上的差异。针对 DINA 模型这一不足, 文章对 DINA 模型中的潜在反应变量(latent response variables, Junker & Sijtsma, 2001) 重新进行定义, 提出了改进的 DINA 模型(Revised deterministic inputs, noisy “and” gate, R-DINA)。R-DINA 模型假设被试掌握项目所测属性的数量越多, 正确作答项目的概率越大。因此, R-DINA 模型可以区分完全未掌握组被试和部分掌握组被试在项目作答概率上的差异。也就是说, 掌握了

部分属性的被试的正确作答概率要高于一个属性都没掌握的被试的正确作答概率。R-DINA 模型定义被试 i 在项目 j 的正确作答概率如下:

$$P_j(\alpha_i) = P(x_{ij} = 1 | \alpha_i) = g_j^{1-\eta_{ij}} (1 - g_j)^{\eta_{ij}}$$

其中 α_i 表示被试 i 的知识状态(列向量), x_{ij} 表示被试 i 在项目 j 的得分, q_j 为项目 j 在属性关联 Q 矩阵中的列(列向量), 即项目 j 考察的属性向量, s_j 表示掌握项目 j 所考察的所有属性的被试组的失误概率, g_j 表示没掌握项目 j 所考察的任何属性的被试组的猜测概率, $\eta_{ij} = \alpha'_i q_j / q'_j q_j$ (α'_i 和 q'_j 分别表示相应列向量的转置), 这里的 η_{ij} 与 DINA 模型中 η_{ij} 不一样, 它的取值不是非 0 即 1, 而是推广到 0 到 1 间的某个分数。它的分子是知识状态为 α_i 的被试所具备的属性集合与正确解答项目 j 所需属性的集合的交集中属性数, 分母则是正确回答项目所需属性数, η_{ij} 是这两个数的比例。如果将这里定义的 η_{ij} 向下取整, 则它与 DINA 模型定义中的 η_{ij} 完全相同。

2 R-DINA 模型参数估计的 EM 算法

提出一个模型, 必须解决参数估计问题, 否则只具有理论意义, 在实际工作中将无法使用。这里采用 EM(expectation-maximization) (von Davier, 2005; de la Torre, 2009, 2011) 算法实现 R-DINA 模型的参数估计。

EM 算法是一个迭代算法, 是处理缺失数据统计模型中参数的极大似然估计问题的一种方案。EM 算法每个迭代中都含有两步: 期望步(expectation, E-step) 和极大化步(maximization step, M-

step) (漆书青,戴海崎,丁树良,2002)。记 $f(x|\alpha|\xi)$ 为完全数据 $(x|\alpha)$ 的联合似然函数,其中 $\xi = (\varphi|\beta)$ 表示待估计的知识状态 α 的分布参数和项目参数,EM 第 t 步迭代包括下面两个步骤:

E-Step: 计算 $Q(\varphi|\beta|\varphi^{(t)}|\beta^{(t)}) = E[\ln f(x|\alpha|\varphi|\beta)|x|\varphi^{(t)}|\beta^{(t)}]$;

M-Step: $(\varphi^{(t+1)}|\beta^{(t+1)}) = \arg \max_{\varphi|\beta \in \Omega} Q(\varphi|\beta|\varphi^{(t)}|\beta^{(t)})$ 。

在第 t 步迭代下的循环中,期望步是在给定 $\xi^{(t)}$ (由上一轮极大化 M-step 计算得到或第一轮循环给定的初始值) 条件下计算联合似然函数的期望 $E[\ln f(x|\alpha|\varphi|\beta)|x|\varphi^{(t)}|\beta^{(t)}]$; 极大化步即最大

化 $Q(\varphi|\beta|\varphi^{(t)}|\beta^{(t)})$, 估计得到 $\xi^{(t+1)}$, 重复这个过程直至收敛,可估计得到模型项目参数 $\hat{\beta}$ (在此不设置也不估计 φ 参数)。在估计得到 $\hat{\beta}$ 后,可通过经典的估计方法,如最大似然估计方法(maximum likelihood estimation, MLE)、最大后验估计方法(maximum a posteriori, MAP)和期望后验估计方法(expected a posteriori, EAP)估计得到被试的知识状态。R-DINA 模型 EM 参数估计流程图可见图 1。

3 EM 算法准确性模拟检验

3.1 检验目的

文章主要检验用于估计 R-DINA 模型参数的 EM 算法的正确性,包括项目参数返真性、比较由信息矩阵计算项目参数标准误与重复 100 次的经验标准误是否接近,并得出模拟条件下 R-DINA 模型的知识状态判准率。

3.2 检验设计

采用与(de la Torre, 2009) 类似的模拟条件,具体如下:

(1) 测验 Q 矩阵及项目参数模拟: 模拟的项目 Q 阵与(de la Torre, 2009) 研究相同,详见表 1。包括 30 个($m = 30$) 项目,每个项目的失误和猜测参数均固定为 0.2。

(2) 被试样本模拟: 由于测验 Q 阵中的 5 个属性之间相互独立,知识状态全集个数 32,每种知识状态模拟 60 人,共模拟 1920 个被试。

(3) 研究方法: 固定项目、项目参数和被试,模拟 100 个得分阵,共估计 100($R = 100$) 次,得到估计的项目参数、项目参数估计标准误及被试知识状态,若 $|Q(\beta^{(t+1)}) - Q(\beta^{(t)})| / |Q(\beta^{(t)})| < 0.0001$ 成立,EM 算法终止。

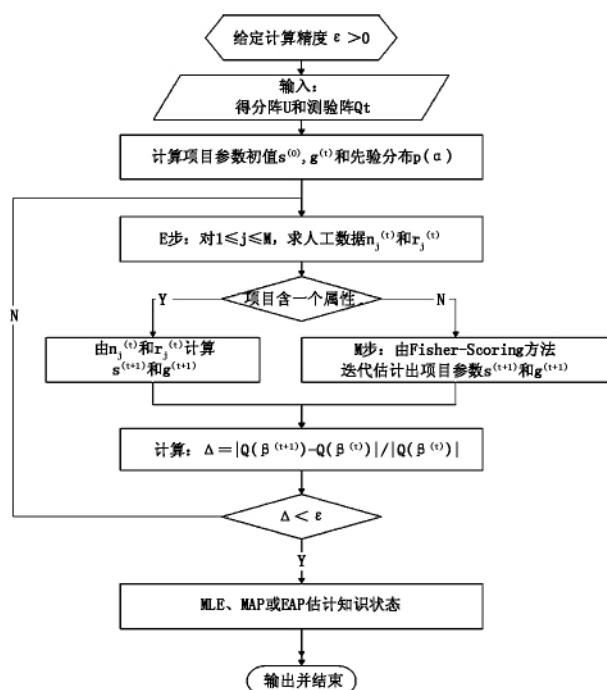


图1 R-DINA 模型 EM 参数估计流程图

表1 模拟实验 Q 阵

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

3.3 评价指标

(1) 模式判准率

模式判准率是从被试属性模式是否判断正确的角度进行考虑的。如果被试 i 的 α_i 与 $\hat{\alpha}_i$ 的各个分量均相同,意味着该判断正确。通过计算所有被试模式判断正确之和除以被试人数得到模式判准率。比如,测验共有 K 个属性且有 N 个被试参加测验,则可估计得到被试 i 的属性模式 $\hat{\alpha}_i$,而被试 i 的知识状态真值为 α_i ,如果 α_i 与 $\hat{\alpha}_i$ 的各个分量均相同,令 $h_i = 1$; 否则 $h_i = 0$,则属性模式判准率为 $\sum_{i=1}^N h_i / N$,记之为 PMR (Pattern Match Ratio)。

(2) 边际判准率

边际属性诊断判准率,也称为单个属性判准率,它是从被试属性模式中单个属性是否判断正确的角度进行考虑的。计算如下:对 K 个属性中第 k 个属性,考察对所有 N 个被试在第 k 个属性上的判准率即得到边际属性诊断判准率。如被试 α_i 掌握了(未掌握)第 k 个属性,模型正确地判断其掌握了(未掌握)该属性,则称为对第 k 个属性判准了一次,记为 $g_{ki} = 1$,否则为 0。令 $MMR(k)$ (Marginal Match Ratio(k)) 为第 k 个属性诊断判准率,也称为边际判准率; MMR 为 K 个属性的平均判准率,简称为属性平均判准率,则:

$$MMR(k) = \frac{1}{N} \sum_{i=1}^N g_{ki}$$

$$MMR = \frac{1}{K} \sum_{k=1}^K MMR(k)$$

PMR, MMR , $MMR(k)$ 三个指标其值越高,对被试的知识状态估计越准。

(3) 项目参数估计均值

$$\bar{M} = \frac{1}{Rm} \left(\sum_{j=1}^m \sum_{r=1}^R \hat{\beta}_{jr} \right)$$

由于各项目参数均固定为 0.2,项目参数估计均值越接近 0.2 表示项目参数估计返真性越好。

(4) 项目参数误差均方根 (root of the mean squared error, RMSE)

共估计 R 次,可计算项目参数 $\hat{\beta}$ 估计标准差为:

$$RMSE = \sqrt{\frac{1}{R} \left(\sum_{r=1}^R \hat{\beta}_r^2 - \frac{1}{R} \left(\sum_{r=1}^R \hat{\beta}_r \right)^2 \right)}$$

(5) 项目参数经验标准误 (the empirical standard error)

共估计 R 次,可计算项目参数 $\hat{\beta}$ 估计标准误为:

$$MSE = \frac{1}{R} \sum_{r=1}^R I(\hat{\beta}_r)^{-\frac{1}{2}}$$

其中 $I(\hat{\beta}_r)$ 为信息矩阵 $I(\hat{\beta}) = -E \frac{\partial^2 Q(\varphi | \beta | \varphi^{(i)} | \beta^{(i)})}{\partial \beta \partial \beta'}$ 中某个对角元素,对应于某项目的项目参数标准误。误差均方根与经验标准误小说明模型参数估计返真性好。

3.4 检验结果及结论

表 2 给出了模拟数据下 EM 算法的项目参数估计结果。结果显示:100 次重复模拟估计得到的各项目参数均值都非常接近真值 0.2,如从 30 个项目的均值来看,失误和猜测参数的估计均值分别为 0.1996 和 0.1999,与真值 0.2 相当接近;多数项目的误差均方根 RMSE 在 0.02 以下,失误参数的 RMSE 要大于猜测参数的 RMSE;绝大多数项目的经验标准误 MSE 在 0.02 以下,失误参数的 MSE 大于猜测参数的 MSE;误差均方根 RMSE 均值比较接近经验标准误 MSE 的均值,尤其在猜测参数上的 RMSE 和 MSE 较接近;RMSE 和 MSE 在失误与猜测参数上的大小关系与 DINA 模型 EM 算法估计结果类似 (de la Torre 2009)。

表 3 给出了模拟数据下知识状态 MLE 估计结果。各属性的边际判准率均在 0.895 左右,标准差在 0.007 左右;100 次重复的边际判准率为 0.8952,标准差为 0.003;模式判准率为 0.5915,标准差为 0.01。模式判准率不高是由于模拟实验中,在既定测验 Q 阵条件下,项目的失误和猜测参数值较大。这一结果与同等条件下 DINA 模型结果类似 (de la Torre & Douglas 2004)。

从项目参数估计结果和知识状态估计结果来看,显示 EM 算法估计程序结果稳定,项目参数返真性较好。

表 2 模拟数据下 EM 算法项目参数估计结果

重复试验	\bar{M}		MSE		RMSE	
	s	g	s	g	s	g
M	0.1996	0.1999	0.0174	0.0122	0.0221	0.0146
SD	0.0022	0.0016	0.0037	0.0005	0.0060	0.0013

表 3 模拟数据下知识状态 MLE 估计结果

重复试验	MMR(k)					MMR	PMR
	α_1	α_2	α_3	α_4	α_5		
M	0. 8948	0. 8948	0. 8959	0. 8955	0. 8953	0. 8952	0. 5915
SD	0. 0074	0. 0078	0. 0069	0. 0072	0. 0069	0. 0030	0. 0114

4 小结与讨论

文章首先简要介绍了 R - DINA 模型, 然后采用 EM 算法实现了 R - DINA 模型的参数估计。在 EM 算法的 M 步中采用 Fisher - Scoring 方法进行迭代估计出项目参数, 同时利用信息矩阵计算了项目参数的估计标准误。设计了模拟实验从项目参数返真性、项目参数标准误和经验标准误、知识状态模式和边际判准率五个指标, 评估了 EM 算法的表现。模拟研究结果显示: 综合五个指标来看, EM 算法表现不错。具体表现为: EM 算法的参数估计返真性较好; 估计得到的项目参数标准误和经验标准误基本相当, 但失误参数的估计标准误略高于经验标准误; 知识状态判准率较高。

文中 EM 算法未知参数的初值均是随机产生的, 为了加速收敛, 是否可以采用其它的初值的选取方法值得进一步探讨。知识状态分布建模和参数估计也有待进一步探讨。如果采用马尔可夫链蒙特卡罗(MCMC) 算法(Patz & Junker, 1999), EM 和 MC - MC 表现是否相当, 值得进一步探讨。另外, 项目参数大小、样本量等因素对参数估计准确性的影响如何? 文中假设属性间相互独立, 新模型在属性层级结构下(如 线性、发散、收敛结构等) 表现如何都是有待进一步探讨的问题。

参考文献

漆书青, 戴海崎, 丁树良. (2002). 现代教育与心理测量学原理. 北京: 高等教育出版社.

杨智为, 卓淑瑜, 郭伯臣, 陈亭宇. (2011). DINA 与 G - DINA 模式参数不变性探讨. 测验统计年刊, 第十九辑.

de la Torre J. & Douglas J. (2004). Higher - order latent trait models for cognitive diagnosis. *Psychometrika* 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, 343 - 362.

de la Torre J. (2009). DINA model and parameter estimation: A

didactic. *Journal of Educational and Behavioral Statistics* 34, 115 - 130.

de la Torre J. (2011). The generalized DINA model framework. *Psychometrika* 76, 179 - 199.

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 la Torre J. & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: a simulation study. *Journal of Educational Measurement* 46, 450 - 469.

de la Torre J. & Lee, Y. - S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement* 47, 115 - 127.

Huebner, A. & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement* 71(2), 407 - 419.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement* 25, 258 - 272.

Liu, H. - Y., You, X. - F., Wang, W. - Y., Ding, S. - L., & Chang, H. - H. (in press). The development of computerized adaptive testing with cognitive diagnosis for english achievement test in China. *Journal of Classification*.

Patz, R. J., & Junker, B. W. (1999). A Straightforward Approach to Markov Chain Monte Carlo Methods for Item Response Models. *Journal of Educational and Behavioral Statistics* 24, 146 - 178.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 51, 337 - 350.

von Davier, M. (2005). *A general diagnostic model applied to language testing data*(ETS RR - 05 - 16). Princeton, NJ Educational Testing Service.

(下转第 422 页)

- Shigemasu ,A. Okada ,T. Imaizumi ,& T. Hoshino (Eds.) , *New Trends in Psychometrics* (pp. 417 – 424) . Universal Academy Press: Tokyo.
- Gierl M. J. ,Leighton J. P. ,& Hunka S. M. (2000) . Exploring the logic of Tatsuoka ' s rule – space model for test development and analysis. *Educational Measurement: Issues and Practice* 19 34 – 44.
- Leighton J. P. ,& Gierl M. J. (2007) . Why cognitive diagnostic assessment? In J. P. Leighton & M. J. Gierl (Eds.) , *Cognitive diagnostic assessment for education: Theory and applications* (pp. 3 – 18) . Cambridge ,UK: Cambridge University Press.
- Leighton J. P. ,Gierl M. J. ,& Hunka S. M. (2004) . The attribute hierarchy method for cognitive assessment: a variation on Tatsuoka ' s rule – space approach. *Journal of Educational Measurement* 41(3) 205 – 237.
- 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) . Erlbaum: Hillsdale.
- Tatsuoka K. K. (2009) . *Cognitive assessment: an introduction to the rule space method* (pp. 9 ,79 ,99 ,107) . New York: Taylor & Francis Group.

Extension to Tatsuoka ' s Q Matrix Theory

Ding Shuliang Luo Fen Wang Wenyi

(School of Computer and Information Engineering Jiangxi Normal University Nanchang 330022)

Abstract: Tatsuoka ' s Q matrix theory and its modification and extension were reviewed briefly. In order to describe the situation that there is a pair of attributes and at least there are two roads connected with them ,the concept of basic matrix ,which included the reachability matrix ,is proposed. The test Q matrix is a mapping from the set of knowledge states to the set of the expected response patterns. For a kind of polytomous scoring ,we proposed a principle of the design of the test blueprint. The principle is that the rank of the test Q matrix equals to the number of the attributes ,the mapping determined by the test Q matrix ,is one – to – one and onto under the scoring mode. It should be noted that this condition is a sufficient only.

Key words: cognitive diagnosis; Q matrix theory; design of the test blueprint; basic matrix; polytomous scoring

(上接第 413 页)

Assessing the Performance of EM Algorithm for Parameter Estimation of the R – DINA Model

Song Lihong¹ ,Dai Haiqi² ,Wang Wenyi³ ,Ding Shuliang³

(1. Elementary Educational College Jiangxi Normal University Nanchang 330027;

2. College of Psychology Jiangxi Normal University Nanchang 330022;

3. Computer Information Engineer College Jiangxi Normal University Nanchang 330022)

Abstract: The deterministic inputs noisy “and” gate (DINA) model is an extensively focused cognitive diagnostic model. However ,it can only partitions examinees into two groups on each item. The revised deterministic inputs noisy “and” gate (R – DINA) model proposed in this paper can obtain more refined partition of examinees. After a brief introduction of the R – DINA model ,this paper achieves model parameter estimation with EM algorithm and then conducts simulation experiment to verify the estimate accuracy of EM algorithm under R – DINA model. Results of simulation analysis indicate that the EM algorithm can be used to obtain stable and accurate parameter estimation of the R – DINA model.

Key words: cognitive diagnostic model; the revised deterministic inputs noisy “and” gate model; model estimation; EM algorithm