

宋丽红,袁思玉,汪文义. 面向学习测评的纵向认知诊断模型的泛化性能比较研究 [J]. 江西师范大学学报(自然科学版), 2023, 47(4): 384-392.

SONG Lihong, YUAN Siyu, WANG Wenyi. The comparative study on the generalization performance of longitudinal cognitive diagnostic models for learning assessment [J]. Journal of Jiangxi Normal University (Nature Science), 2023, 47(4): 384-392.

文章编号: 1000-5862(2023) 04-0384-09

面向学习测评的纵向认知诊断模型的 泛化性能比较研究

宋丽红¹, 袁思玉², 汪文义^{2*}

(1. 江西师范大学教育学院, 江西 南昌 330022; 2. 江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 为了研究纵向认知诊断模型适应新数据的能力, 该文主要考查 3 种纵向认知诊断模型在不同类型纵向数据上的泛化性能. 这 3 种纵向认知诊断模型分别为模式级别上的潜在转换分析模型 Patt-DINA、属性级别上的潜在转换分析模型 Att-DINA 和基于高阶潜在结构的 sLong-DINA 模型. 借助被试知识状态的属性判准率、模式判准率、绝对拟合指标和相对拟合指标等 4 个指标, 评价这 3 种模型的表现. 研究结果表明: Att-DINA 模型和 sLong-DINA 模型在大多数条件下更具优势, 即泛化性能相对较好, Patt-DINA 模型因待估计参数较多而优势较小, 但 Patt-DINA 模型在样本量较大时仍具有优势并且它能估计的知识状态类间转移概率有更大的变化空间.

关键词: 纵向认知诊断模型; 属性转换; 模式转换; 高阶模型; 泛化性能

中图分类号: B 841 **文献标志码:** A **DOI:** 10. 16357/j. cnki. issn1000-5862. 2023. 04. 08

0 引言

“为学习而测评”理念^[1-3]强调测评是为了学生更好的发展. 认知诊断测评(cognitive diagnostic assessment, CDA) 因其能够为学生和教师提供诊断反馈报告而日益受到人们的重视. 学生随时间变化的测评数据可以视为纵向数据, 利用好这些纵向数据有助于分析学生学习反馈和补救教学成效. 由于传统的横向认知诊断模型无法有效分析这些纵向数据, 因此国内外研究者们开发了不同的纵向认知诊断模型.

詹沛达等^[4]按照建模逻辑将纵向认知诊断模型分为 2 个大类: 1) 基于潜在转换分析^[5](latent transition analysis, LTA) 的纵向认知诊断模型. 在该

类型中, Li Feiming 等^[6]将确定性输入噪声与门模型(deterministic input, noisy “and” gate model, DINA) 融入 LTA 模型中, 通过建立属性级别上的转换概率分析学生随时间变化的知识掌握情况, 本文称其为 Att-DINA 模型. Y. Kaya 等^[7]将 DINA 模型融入 LTA 模型中, 通过建立属性向量级别上的转换概率分析学生随时间变化的知识掌握情况, 本文称其为 Patt-DINA 模型. Chen Yinghan 等^[8]建立了学习轨迹模型对学生的进行学习状态进行追踪. 此外, 还有一些更为细致的研究, 例如, 在高阶隐 Markov 模型中引入协变量, 以分析导致学生学习变化的因素^[9], 又如, 将学生作答反应时间(response time, RT) 纳入考虑^[10]. 2) 基于高阶潜在结构的纵向认知诊断模型. 在该类型中, S. Y. Lee 等^[11]提出一种纵向成长曲线认知诊断模型用于分析学生成长的纵向数据.

收稿日期: 2022-12-11

基金项目: 国家自然科学基金(62267004, 62067005, 61967009) 和江西省高等学校教学改革研究课题(JXJG-22-2-44) 资助项目.

通信作者: 汪文义(1983—), 男, 湖南衡山人, 教授, 博士, 主要从事教育测量与信息处理的研究. E-mail: wenyiwang@jxnu.edu.cn

Huang Hungyu^[12]提出了基于潜在增长模型的多水平认知诊断模型. Zhan Peida 等^[13]提出了一种基于多元正态分布的纵向高阶认知诊断模型.

总体来看,纵向认知诊断模型已有一定的发展.研究者们均从不同的角度建立模型对纵向数据进行分析,并且在模型中考虑的因素也越来越细致.然而,目前还未有这些纵向认知诊断模型的对比研究,这不利于实践者选择合适的模型^[4].因此,本文选取其中3个纵向认知诊断模型进行对比研究,探究不同纵向数据下适合的纵向认知诊断模型.

1 纵向认知诊断模型

1.1 Patt-DINA 模型

LTA 模型为潜在类别分析(latent class analysis, LCA)模型^[14]的纵向应用提供了一种建模方式,即 Patt-DINA 模型为 LTA 模型与 DINA 模型的结合. Patt-DINA 模型为

$$P(Y = y) = \sum_{c_1=1}^C \sum_{c_2=1}^C \cdots \sum_{c_T=1}^C \delta_{c_1} \tau_{c_2, c_1} \tau_{c_3, c_2} \cdots \tau_{c_T, c_{T-1}} \prod_{t=1}^T \prod_{j=1}^J \prod_{r_j=1}^{R_j} ((1-s_j)^{\eta_{j,t}(1-\eta_{j,t})} g_j^{I(y_{j,t}=r_j)})$$

其中 C 为潜在类别数量, T 为时间点数量, J 为项目数量, R_j 为反应类别的数量, δ_{c_1} 为被试初始的知识状态分布情况, $\tau_{c_T, c_{T-1}}$ 为被试知识状态随时间转变的转移概率. 该模型建立了在被试类与类(模式与模式)转移概率,且测量模型选择的是 DINA 模型,因此将该模型称为 Patt-DINA 模型.

1.2 Att-DINA 模型

与 Patt-DINA 模型不同的是, Att-DINA 模型的转移概率定义在属性级别上,且该模型假设属性之间的转换概率相互独立. Att-DINA 模型为

$$P(Y = y) = \sum_{\alpha_{k1}=1}^C \sum_{\alpha_{k2}=1}^C \cdots \sum_{\alpha_{kT}=1}^C \prod_{k=1}^K P(A_{k1} = \alpha_{k1}) \prod_{t=2}^T \prod_{k=1}^K P(A_{kt} = \alpha_{kt} | A_{k(t-1)} = \alpha_{k(t-1)}) \cdot \prod_{t=1}^T \prod_{j=1}^J \prod_{r_j=1}^{R_j} ((1-s_j)^{\eta_{j,t}(1-\eta_{j,t})} g_j^{I(y_{j,t}=r_j)})$$

其中 C 表示潜在属性的类别数量, K 为属性数量, T

为时间点数量, $\prod_{k=1}^K P(A_{k1} = \alpha_{k1})$ 为在初始状态下每种认知属性的掌握概率, $\prod_{t=2}^T \prod_{k=1}^K P(A_{kt} = \alpha_{kt} | A_{k(t-1)} = \alpha_{k(t-1)})$ 为认知属性的转移概率,且允许不同认知属性的转移概率有所不同.

设有 2 个时间点,可将某个认知技能的转移概率矩阵表示为

$$\begin{pmatrix} p_{n|n}^{21} & p_{m|n}^{21} \\ p_{n|m}^{21} & p_{m|m}^{21} \end{pmatrix},$$

其中 m 表示被试掌握了该技能, n 表示被试没有掌握该技能. 转移概率矩阵由 4 个部分组成: $p_{n|n}^{21}$ 表示从时间点 1 到时间点 2 被试保持非掌握状态, $p_{m|n}^{21}$ 表示被试保持掌握状态, $p_{n|m}^{21}$ 表示被试从掌握到非掌握状态(遗忘了该技能), $p_{m|m}^{21}$ 表示被试从非掌握到掌握状态(新学会了该技能).

1.3 sLong-DINA 模型

Long-DINA 模型是一种基于多元正态分布策略和潜在增长模型策略的纵向认知诊断模型. sLong-DINA 模型为 Long-DINA 模型的简化版本, Long-DINA 模型原本有 3+1 阶, sLong-DINA 模型将不同时间点上锚题之间的局部题目依赖性设置为 0,从而去掉这一特殊维度. 下面将详细介绍 sLong-DINA 模型各个阶.

第 1 阶模型:

$$\text{Logit}(P(y_{nit} = 1 | \alpha_{n, c_t}, \lambda_{i0t}, \lambda_{ilt})) =$$

$$\lambda_{i0t} + \lambda_{ilt} \prod_{k=1}^K \alpha_{nk, c_t}^{q_{ikt}}.$$

第 1 阶模型用于刻画各时间点上属性与题目作答之间的关系,它为测量模型. 其中 λ_{i0t} 和 λ_{ilt} 分别为时间点 t 上题目 i 的截距和载荷; α_{n, c_t} 表示被试 n 在时间点 t 下的知识状态, α_{nk, c_t} 表示被试 n 在时间点 t 下对于属性 k 的掌握情况; q_{ikt} 表示时间点 t 下项目 i 的 k 属性是否被测量.

第 2 阶模型:

$$\text{Logit}(P(\alpha_{nk, c_t} = 1 | \theta_{nt}, \delta_{0kt})) = \delta_{kt} \theta_{nt} - \beta_{kt},$$

$$\theta_n = (\theta_{n1}, \theta_{n2}, \cdots, \theta_{nT})^T.$$

第 2 阶模型用于刻画各时间点上一般潜在能力与属性之间的关系,它为高阶潜在结构模型. 其中, θ_{nt} 为时间点 t 上被试 n 的一般潜在能力, δ_{kt} 和 β_{kt} 分别为时间点 t 上属性 k 的截距和载荷. 同时,由于该模型假设潜在结构存在时间不变性,因此不同时间点上属性 k 的截距参数需保持一致,即 $\delta_{kt} = \delta_k$.

且 $\beta_{kt} = \beta_k$.

第 3 阶模型:

$$\theta_n = (\theta_{n1} \ \theta_{n2} \ \cdots \ \theta_{nT})^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

第 3 阶模型用于刻画不同时间点上一般潜在能力的变化情况,它为纵向发展阶. 其中,均值向量 $\boldsymbol{\mu} = (\mu_1 \ \mu_2 \ \cdots \ \mu_T)^T$ 对称方差协方差矩阵 $\boldsymbol{\Sigma}$ 为

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1T} & \sigma_{2T} & \cdots & \sigma_T^2 \end{pmatrix}.$$

sLong-DINA 模型一般假设 $\mu_1 = 0$ 且 $\sigma_1^2 = 1$,即被试在时间点 1 上的一般潜在能力满足标准正态分布 σ_{1T} 为被试的一般潜在能力在时间点 1 和时间点 T 上的协方差. 使用该模型不仅能够得到有关学生能力的信息,也能得到被试关于属性的掌握信息.

表 1 模拟研究中的 Q 矩阵

属性	题目											
	1	2	3	4	5	6	7	8	9	10	11	12
pb	1	0	0	0	1	1	1	1	1	0	1	1
cp	0	1	0	0	1	1	0	0	0	1	1	0
un	0	0	1	0	0	0	1	1	0	0	0	1
id	0	0	0	1	0	0	0	0	1	1	1	1

注: pb 表示事件的概率, cp 表示补事件的概率, un 表示 2 个互不相容事件并的概率, id 表示 2 个独立事件的概率.

表 2 模拟研究中题目的参数

参数	题目											
	1	2	3	4	5	6	7	8	9	10	11	12
s	0.093	0.037	0.042	0.036	0.147	0.046	0.062	0.046	0.226	0.193	0.303	0.189
g	0.224	0.321	0.107	0.124	0.232	0.247	0.351	0.547	0.346	0.063	0.060	0.031

当生成基于 Patt-DINA 模型的纵向数据时,被试在时间点 1 处的知识状态分布如表 3 所示. 被试知识状态由时间点 1 向时间点 2 转变的转移概率矩

2 模拟研究

2.1 研究目的

探究在不同实验条件下 Patt-DINA 模型、Att-DINA 模型和 sLong-DINA 模型的泛化性能.

2.2 研究设计

本文基于 3 种纵向认知诊断模型分别生成 3 种类型的纵向数据,时间点设置为 2 个,其中每种类型的数据均考虑被试总人数为 100、500 和 1 000 的情况. 对于生成的 3 类纵向数据,分别使用这 3 种模型对模拟数据进行拟合分析,以探究不同纵向模型在不同类型纵向数据上的表现. 本文使用的 Q 矩阵和题目参数如表 1 和表 2 所示,该数据来自文献 [15] 使用的示例数据集.

阵,采用文献 [8] 中所得的结果(见表 4). 接着,通过 DINA 模型可以分别模拟得到被试在 2 个时间点处的作答矩阵,记为数据 1.

表 3 被试在时间点 1 时的知识状态分布

序号	α_i	π_i	序号	α_i	π_i
1	(0 0 0 0)	0.31	9	(1 0 0 0)	0.03
2	(0 0 0 1)	0.02	10	(1 0 0 1)	0.01
3	(0 0 1 0)	0.01	11	(1 0 1 0)	0.02
4	(0 0 1 1)	0.01	12	(1 0 1 1)	0.02
5	(0 1 0 0)	0.05	13	(1 1 0 0)	0.05
6	(0 1 0 1)	0.02	14	(1 1 0 1)	0.02
7	(0 1 1 0)	0.02	15	(1 1 1 0)	0.03
8	(0 1 1 1)	0.01	16	(1 1 1 1)	0.37

表 4 模式转移概率矩阵

知识 状态	时间 2															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
时间 1	1	0.81	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.01	0.02	0.01	0.01	0.01
	2	0.00	0.13	0.00	0.11	0.00	0.18	0.00	0.10	0.00	0.11	0.00	0.12	0.00	0.12	0.00
	3	0.00	0.00	0.12	0.11	0.00	0.00	0.13	0.13	0.00	0.00	0.16	0.11	0.00	0.00	0.12
	4	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.21	0.00	0.00	0.23
	5	0.00	0.00	0.00	0.00	0.37	0.15	0.11	0.08	0.00	0.00	0.00	0.09	0.07	0.07	0.06
	6	0.00	0.00	0.00	0.00	0.00	0.72	0.00	0.12	0.00	0.00	0.00	0.00	0.09	0.00	0.07
	7	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.17	0.00	0.00	0.00	0.00	0.00	0.18	0.15
	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.22
	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.09	0.17	0.10	0.19	0.10	0.12
	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.25	0.00	0.29	0.00	0.24
	11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.73	0.08	0.00	0.00	0.12	0.07
	12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.57
	13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72	0.08	0.09	0.11
	14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.53
	15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.45
	16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

当生成基于 Att-DINA 模型的纵向数据时,被试在时间点 1 处的知识状态分布和上述相同.对于生成时间点 2 的被试知识状态,每个属性的转移概率矩阵为

$$\tau_{k_1} = \begin{pmatrix} 0.778 & 0.222 \\ 0.000 & 1.000 \end{pmatrix}, \tau_{k_2} = \begin{pmatrix} 0.674 & 0.326 \\ 0.000 & 1.000 \end{pmatrix},$$
$$\tau_{k_3} = \begin{pmatrix} 0.725 & 0.275 \\ 0.000 & 1.000 \end{pmatrix}, \tau_{k_4} = \begin{pmatrix} 0.679 & 0.321 \\ 0.000 & 1.000 \end{pmatrix}.$$

该概率矩阵由上述基于 Patt-DINA 模型的纵向数据统计得出.同样地,通过 DINA 模型可以分别模拟得到被试在 2 个时间点处的作答矩阵,记为数据 2.

当生成基于 sLong-DINA 模型的纵向数据时,首先,生成被试的能力 2 维向量 $\theta \sim N(\mu, \Sigma)$, $\mu = \begin{pmatrix} 0.0 \\ 0.5 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1.000 & 0.00 & 1.006 & 23 \\ 1.006 & 23 & 1.250 & 00 \end{pmatrix}$.然后,模拟被试的知识状态,其中 $\beta = (-1 \ 0 \ 0 \ 1)^T$, $\delta = (1.00 \ 1.25 \ 1.25 \ 1.50)^T$.最后,通过 DINA 模型计算被试的正确作答概率并得到被试的作答矩阵(记为数据 3),其中 $\lambda_{i0i} = -2.197$, $\lambda_{i1i} = 4.394$.

本研究使用 Zhan Peida 等^[16]提出的在 R 中使用 JAGS(just another Gibbs sampler)的方法并自编程序实现模型的参数估计.

2.3 评价指标

2.3.1 判准率 A_{AMR} 为属性判准率, A_{PMR} 为模式判准率.这 2 个指标值越高代表对被试的知识状态

估计越准确,其计算公式分别为

$$A_{AMR} = \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K I(\hat{\alpha}_{ik}, \alpha_{ik}),$$
$$A_{PMR} = \frac{1}{N} \sum_{i=1}^N I(\hat{\alpha}_i, \alpha_i).$$

2.3.2 基于后验预测模型检查的绝对拟合指标 基于后验预测模型检查的绝对拟合指标 p_{PPP} ^[17],计算方法为

$$p_{PPP} = P(D(u_i^{rep}, \omega_i) \geq D(u, \omega_i) | y) = \sum_{i=1}^T I(D(u_i^{rep}, \omega_i) \geq D(u, \omega_i)) / T.$$

该指标的值越接近 0.5 表示模型与数据拟合越好,越接近 0 与 1 说明拟合越差.

2.3.3 模型相对拟合指标 这一类拟合指标常用于对多个模型进行比较,从而选择一个拟合最优的模型.常用指标主要有偏差、AIC、BIC、DIC 和贝叶斯因子等.

偏差(f_{-2LL})是似然函数值的自然对数的负 2 倍,用来反映模型的拟合程度,其值越小表示拟合程度越好.其表达式为 $f_{-2LL} = -2\log p(X|\theta)$, θ 为模型参数, X 为被试观察得分, $p(X|\theta)$ 为观测数据的概率模型(如边际似然函数).

AIC^[18]计算公式为 $f_{AIC} = f_{-2LL} + 2d$,其中 d 为模型的估计参数个数. AIC 在偏差指标的基础上,考虑了模型参数带来的影响,参数多的模型将受到惩罚. AIC 值越小表示拟合越好.

BIC^[19]计算公式为 $f_{BIC} = f_{-2LL} + d\ln N$,其中 N 为

样本量. 该指标同样对复杂模型进行了惩罚,且惩罚力度比 AIC 值更大. BIC 值越小表示拟合越好.

DIC^[20] 计算公式为

$$f_{DIC} = \overline{D(\theta)} + p_D = \overline{D(\theta)} + 2p_D = 2\overline{D(\theta)} - D(\theta) = -4E_{\theta}(\log p(X|\theta) | X) + 2\log p(X|\theta) ,$$

其中 $\overline{D(\theta)}$ 是偏差的后验均值, $D(\theta)$ 是后验均值偏差 $p_D = \overline{D(\theta)} - D(\theta)$ 为待估计模型的参数个数. 若 2 个模型的 DIC 相差 10 或以上,则 DIC 小的模型拟合更好. 若 DIC 相差 5 以下,则认为 2 个模型的拟合无显著差异.

2.4 实验结果

表 5~表 10 分别为当被试数量为 100、500 和 1 000 时的判准率与拟合统计量结果. 共有 3 种类型的纵向数据,分别为基于 Patt-DINA 模型、Att-DINA 模型和 sLong-DINA 模型生成的模拟数据. 拟合统计量除 Patt-DINA 模型外,基本上生成数据模型与拟合数据模型相同,模型拟合指标最佳,在此不再详细讨论. 下面主要比较各模型的判准率.

当被试数量为 100 时,模拟结果如表 5 和表 6 所示. 3 种纵向认知诊断模型的表现均优于横向认知诊断模型的表现. 对于 Patt-DINA 模型生成的纵向数据,时间点 1 的 Att-DINA 模型表现最佳,时间点 2 的 sLong-DINA 模型表现最佳. 总体来看 sLong-DINA 模型在判准率上表现更优,但从表 6 可以看出 sLong-DINA 模型的拟合效果较差. 对于 Att-DINA 模型生成的纵向数据,2 个时间点均为 Att-DINA 模型

判准率结果最佳且拟合效果也最优. 对于 sLong-DINA 模型生成的纵向数据,时间点 1 的 Patt-DINA 模型表现最佳,时间点 2 的 sLong-DINA 模型表现最佳,但由于 Patt-DINA 模型待估计参数较多导致相对拟合指标较高,因此在此数据下 sLong-DINA 模型表现最优,Att-DINA 模型次之.

当被试数量为 500 时,模拟结果如表 7 和表 8 所示. 总体来看 3 种纵向认知诊断模型的表现均优于横向认知诊断模型,但在某些时间点或某些数据上判准率表现不佳,如时间点 1,3 个模型在 sLong-DINA 的模拟数据上的判准率结果不佳. 对于 Patt-DINA 模型生成的纵向数据,总体来看, Patt-DINA 模型和 sLong-DINA 模型均表现较好. 对于 Att-DINA 模型生成的纵向数据, Att-DINA 模型在判准率和拟合指标上都表现最佳,其他模型的判准率均在时间点 2 上表现较差. 对于 sLong-DINA 模型生成的纵向数据, Att-DINA 模型和 sLong-DINA 模型的判准率均表现较好,但 Att-DINA 模型在拟合指标上更具优势.

当被试数量为 1 000 时,模拟结果如表 9 和表 10 所示. 总体来看 3 种纵向认知诊断模型的表现均优于横向认知诊断模型. 对于 Patt-DINA 模型生成的纵向数据, Patt-DINA 模型和 sLong-DINA 模型表现较好. 对于 Att-DINA 模型生成的纵向数据, Patt-DINA 模型和 Att-DINA 模型表现较好,但 Att-DINA 模型在拟合指标上更具优势. 对于 sLong-DINA 模型生成的纵向数据, sLong-DINA 模型表现最优.

表 5 当被试数量为 100 时判准率结果

数据	模型	时间 1					时间 2				
		AMR(1)	AMR(2)	AMR(3)	AMR(4)	PMR	AMR(1)	AMR(2)	AMR(3)	AMR(4)	PMR
1	Patt-DINA	0. 930	0. 920	1. 000	0. 960	0. 840	0. 930	0. 950	0. 940	0. 950	0. 810
	Att-DINA	0. 940	0. 930	1. 000	0. 960	0. 870	0. 940	0. 950	0. 930	0. 950	0. 800
	sLong-DINA	0. 930	0. 920	0. 990	0. 960	0. 840	0. 940	0. 950	0. 960	0. 970	0. 840
	DINA	0. 900	0. 860	0. 920	0. 960	0. 720	0. 930	0. 920	0. 960	0. 960	0. 800
2	Patt-DINA	0. 930	0. 920	1. 000	0. 960	0. 820	0. 930	0. 930	0. 900	0. 950	0. 750
	Att-DINA	0. 930	0. 960	0. 980	0. 980	0. 860	0. 950	0. 960	0. 910	0. 920	0. 770
	sLong-DINA	0. 880	0. 880	0. 980	0. 960	0. 750	0. 900	0. 940	0. 900	0. 890	0. 710
	DINA	0. 900	0. 860	0. 940	0. 960	0. 740	0. 880	0. 870	0. 970	0. 910	0. 690
3	Patt-DINA	0. 960	0. 950	0. 960	0. 940	0. 840	0. 970	0. 970	0. 950	0. 920	0. 840
	Att-DINA	0. 950	0. 950	0. 920	0. 900	0. 760	0. 970	0. 980	0. 940	0. 940	0. 860
	sLong-DINA	0. 950	0. 970	0. 950	0. 950	0. 830	0. 970	0. 980	0. 950	0. 960	0. 870
	DINA	0. 950	0. 890	0. 910	0. 950	0. 750	0. 990	0. 960	0. 920	0. 960	0. 840

表 6 当被试数量为 100 时的拟合统计量

数据	模型	-2LL	NP	AIC	BIC	DIC	p_{PPP}
1	Patt-DINA	1 833. 920	288	2 409. 920	3 160. 209	2 260. 471	0. 662
	Att-DINA	1 841. 889	56	1 953. 889	2 099. 779	2 205. 492	0. 624
	sLong-DINA	1 848. 524	59	1 966. 524	2 120. 229	2 347. 197	0. 742
	DINA(time1)	951. 663	39	1 029. 663	1 131. 265	1 329. 085	0. 510
	DINA(time2)	938. 320	39	1 016. 320	1 117. 921	1 223. 943	0. 622
2	Patt-DINA	1 856. 040	288	2 432. 040	3 182. 329	2 430. 395	0. 578
	Att-DINA	1 864. 264	56	1 976. 264	2 122. 153	2 268. 678	0. 494
	sLong-DINA	1 937. 322	59	2 055. 322	2 209. 027	2 563. 365	0. 542
	DINA(time1)	952. 923	39	1 030. 923	1 132. 525	1 273. 470	0. 484
	DINA(time2)	906. 209	39	984. 209	1 085. 810	1 238. 140	0. 612
3	Patt-DINA	1 526. 511	288	2 102. 511	2 852. 800	2 122. 429	0. 636
	Att-DINA	1 526. 676	56	1 638. 676	1 784. 566	1 995. 916	0. 668
	sLong-DINA	1 504. 879	59	1 622. 879	1 776. 584	1 978. 262	0. 660
	DINA(time1)	814. 050	39	892. 050	993. 651	1 255. 460	0. 554
	DINA(time2)	720. 521	39	798. 521	900. 123	1 001. 725	0. 512

表 7 当被试数量为 500 时判准率结果

数据	模型	时间 1					时间 2				
		AMR(1)	AMR(2)	AMR(3)	AMR(4)	PMR	AMR(1)	AMR(2)	AMR(3)	AMR(4)	PMR
1	Patt-DINA	0. 954	0. 908	0. 946	0. 970	0. 826	0. 954	0. 936	0. 968	0. 972	0. 862
	Att-DINA	0. 944	0. 938	0. 960	0. 988	0. 848	0. 920	0. 918	0. 964	0. 968	0. 792
	sLong-DINA	0. 956	0. 914	0. 956	0. 978	0. 844	0. 946	0. 934	0. 970	0. 962	0. 846
	DINA	0. 894	0. 896	0. 926	0. 952	0. 724	0. 952	0. 904	0. 944	0. 948	0. 800
2	Patt-DINA	0. 938	0. 922	0. 960	0. 966	0. 822	0. 910	0. 930	0. 958	0. 942	0. 798
	Att-DINA	0. 952	0. 928	0. 976	0. 982	0. 846	0. 926	0. 940	0. 954	0. 954	0. 814
	sLong-DINA	0. 940	0. 906	0. 950	0. 964	0. 800	0. 896	0. 926	0. 866	0. 882	0. 670
	DINA	0. 908	0. 898	0. 926	0. 952	0. 736	0. 924	0. 896	0. 940	0. 934	0. 748
3	Patt-DINA	0. 956	0. 904	0. 908	0. 900	0. 728	0. 972	0. 946	0. 918	0. 934	0. 804
	Att-DINA	0. 958	0. 912	0. 908	0. 924	0. 750	0. 974	0. 944	0. 906	0. 964	0. 816
	sLong-DINA	0. 954	0. 938	0. 916	0. 932	0. 778	0. 980	0. 944	0. 916	0. 964	0. 832
	DINA	0. 966	0. 940	0. 926	0. 950	0. 812	0. 974	0. 928	0. 902	0. 948	0. 796

表 8 当被试数量为 500 时的拟合统计量

数据	模型	-2LL	NP	AIC	BIC	DIC	P_{PPP}
1	Patt-DINA	9 097. 575	288	9 673. 575	10 887. 383	12 156. 745	0. 550
	Att-DINA	9 088. 024	56	9 200. 024	9 436. 042	10 672. 972	0. 522
	sLong-DINA	9 159. 600	59	9 277. 600	9 526. 262	11 838. 464	0. 562
	DINA(time1)	4 606. 257	39	4 684. 257	4 848. 627	8 147. 589	0. 488
	DINA(time2)	4 541. 105	39	4 619. 105	4 783. 474	6 930. 591	0. 444
2	Patt-DINA	9 206. 039	288	9 782. 039	10 995. 846	12 675. 764	0. 486
	Att-DINA	9 206. 425	56	9 318. 425	9 554. 443	11 185. 556	0. 510
	sLong-DINA	9 545. 773	59	9 663. 773	9 912. 435	12 637. 074	0. 518
	DINA(time1)	4 602. 914	39	4 680. 914	4 845. 284	8 512. 935	0. 502
	DINA(time2)	4 529. 899	39	4 607. 899	4 772. 269	7 057. 262	0. 492
3	Patt-DINA	8 399. 008	288	8 975. 008	10 188. 815	12 850. 858	0. 492
	Att-DINA	8 439. 294	56	8 551. 294	8 787. 312	12 899. 573	0. 496
	sLong-DINA	8 272. 276	59	8 390. 276	8 638. 938	12 929. 834	0. 524
	DINA(time1)	3 977. 268	39	4 055. 268	4 219. 637	6 784. 834	0. 430
	DINA(time2)	3 979. 258	39	4 057. 258	4 221. 628	5 336. 097	0. 466

表 9 当被试数量为 1 000 时判准率结果

数据	模型	时间 1					时间 2				
		AMR(1)	AMR(2)	AMR(3)	AMR(4)	PMR	AMR(1)	AMR(2)	AMR(3)	AMR(4)	PMR
1	Patt-DINA	0. 959	0. 923	0. 958	0. 954	0. 828	0. 945	0. 938	0. 950	0. 969	0. 850
	Att-DINA	0. 958	0. 947	0. 969	0. 974	0. 867	0. 931	0. 916	0. 940	0. 966	0. 797
	sLong-DINA	0. 951	0. 929	0. 951	0. 954	0. 825	0. 950	0. 944	0. 948	0. 967	0. 854
	DINA	0. 928	0. 894	0. 946	0. 929	0. 768	0. 959	0. 920	0. 944	0. 959	0. 835
2	Patt-DINA	0. 954	0. 929	0. 958	0. 966	0. 830	0. 929	0. 923	0. 957	0. 949	0. 797
	Att-DINA	0. 967	0. 931	0. 974	0. 973	0. 853	0. 928	0. 922	0. 957	0. 954	0. 793
	sLong-DINA	0. 955	0. 875	0. 945	0. 942	0. 765	0. 916	0. 904	0. 898	0. 908	0. 709
	DINA	0. 919	0. 892	0. 945	0. 936	0. 763	0. 920	0. 884	0. 915	0. 962	0. 730
3	Patt-DINA	0. 953	0. 923	0. 916	0. 926	0. 762	0. 975	0. 939	0. 937	0. 927	0. 810
	Att-DINA	0. 954	0. 915	0. 903	0. 945	0. 764	0. 973	0. 940	0. 937	0. 935	0. 820
	sLong-DINA	0. 964	0. 925	0. 927	0. 941	0. 797	0. 968	0. 937	0. 939	0. 942	0. 811
	DINA	0. 946	0. 897	0. 909	0. 924	0. 740	0. 966	0. 944	0. 930	0. 946	0. 816

表 10 当被试数量为 1 000 时的拟合统计量

数据	模型	-2LL	NP	AIC	BIC	DIC	p_{PPP}
1	Patt-DINA	18 430. 955	288	19 006. 955	20 420. 389	23 622. 181	0. 512
	Att-DINA	18 502. 065	56	18 614. 065	18 888. 899	21610. 015	0. 502
	sLong-DINA	18 743. 880	59	18 861. 880	19 151. 438	23 960. 053	0. 504
	DINA(time1)	9 159. 288	39	9 237. 288	9 428. 690	17 216. 960	0. 510
	DINA(time2)	9 086. 300	39	9 164. 300	9 355. 702	14 706. 780	0. 478
2	Patt-DINA	18 673. 617	288	19 249. 617	20 663. 051	26 734. 982	0. 514
	Att-DINA	18 791. 014	56	18 903. 014	19 177. 849	22 746. 425	0. 502
	sLong-DINA	19 448. 441	59	19 566. 441	19 855. 999	26 442. 526	0. 496
	DINA(time1)	9 155. 055	39	9 233. 055	9 424. 457	19 544. 910	0. 504
	DINA(time2)	9 226. 383	39	9 304. 383	9 495. 785	14 470. 110	0. 502
3	Patt-DINA	16 116. 202	288	16 692. 202	18 105. 635	26 085. 371	0. 538
	Att-DINA	16 205. 263	56	16 317. 263	16 592. 097	27 720. 499	0. 526
	sLong-DINA	15 800. 340	59	15 918. 340	16 207. 897	23 972. 535	0. 528
	DINA(time1)	7 754. 028	39	7 832. 028	8 023. 431	15 501. 560	0. 516
	DINA(time2)	7 934. 669	39	8 012. 669	8 204. 071	12 009. 350	0. 452

3 结论与讨论

总体来看 ,Att-DINA 模型和 sLong-DINA 模型在大多数条件下表现较好 ,Patt-DINA 模型在某些条件下也具有优势 ,但是该模型待估计参数较多 ,导致拟合指标较差. 从样本量的变化来看 ,Patt-DINA 模型的优势会随着样本量的增多而增加. 实践者在选用模型时 ,在样本量较大时(如 1 000) ,可以考虑选择 Patt-DINA 模型 ,否则推荐优先选择估计参数较少的 Att-DINA 模型. 若需考虑被试能力信息 ,则优先考虑 sLong-DINA 模型.

本文使用的约束型 Att-DINA 模型并没有允许属性遗忘 ,而 sLong-DINA 模型因为采用的高阶概率模型是允许属性遗忘. 从模拟结果来看 ,至少 sLong-DINA 模型不能较好地拟合 Att-DINA 模型生成的数据. 本文为模型的泛化性能对比提供了一些思路和方法 ,但本研究仅考虑了 3 种纵向认知诊断模型 ,对于其他纵向认知诊断模型的性能如何 ,值得继续研究. 另外 ,本文仅从评价指标的角度对 3 种模型进行了对比研究 ,对于导致这些结果差异的原因也可以继续深入探究.

4 参考文献

[1] WILIAM D. What is assessment for learning? [J]. Studies in Educational Evaluation ,2011 ,37(1) :3-14.

[2] DE LA TORRE J. The generalized DINA model framework [J]. Psychometrika ,2011 ,76(2) :179-199.

[3] TEMPLIN J L , HENSON R A. Measurement of psychological disorders using cognitive diagnosis models [J]. Psychological Methods ,2006 ,11(3) :287-305.

[4] 詹沛达 ,潘艳方 ,李菲茗. 面向“为学习而测评”的纵向认知诊断模型 [J]. 心理科学 ,2021 ,44(1) :214-222.

[5] COLLINS L M , WUGALTER S E. Latent class models for stage-sequential dynamic latent variables [J]. Multivariate Behavioral Research ,1992 ,27(1) :131-157.

[6] LI Feiming , COHEN A , BOTTGE B , et al. A latent transition analysis model for assessing change in cognitive skills [J]. Educational & Psychological Measurement ,2016 ,76(2) :181-204.

[7] KAYA Y , LEITE W L. Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance [J]. Educational and Psychological Measurement ,2017 ,77(3) :369-388.

[8] CHEN Yinghan , CULPEPPER S A , WANG Shiyu , et al. A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills [J]. Applied Psychological Measurement ,2017 ,

- 42(1): 5-23.
- [9] WANG Shiyu, YANG Yan, CULPEPPER S A, et al. Tracking skill acquisition with cognitive diagnosis models: a higher-order, hidden Markov model with covariates [J]. *Journal of Educational and Behavioral Statistics*, 2018 43(1): 57-87.
- [10] WANG Shiyu, ZHANG Susu, DOUGLAS J, et al. Using response times to assess learning progress: a joint model for responses and response times [J]. *Measurement: Interdisciplinary Research and Perspectives*, 2018, 16(1): 45-58.
- [11] LEE S Y. Growth curve cognitive diagnosis models for longitudinal assessment [D]. California: University of California 2017.
- [12] HUANG Hungyu. Multilevel cognitive diagnosis models for assessing changes in latent attributes [J]. *Journal of Educational Measurement* 2017 54(4): 440-480.
- [13] ZHAN Peida, JIAO Hong, LIAO Dandan, et al. A longitudinal higher-order diagnostic classification model [J]. *Journal of Educational and Behavioral Statistics*, 2019 44(3): 251-281.
- [14] COLLINS L M, LANZA S T. Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences [M]. Hoboken, NJ: John Wiley 2010.
- [15] PHILIPP M, STROBL C, DE LA TORRE J, et al. On the estimation of standard errors in cognitive diagnosis models [J]. *Journal of Educational and Behavioral Statistics* 2018 43(1): 88-115.
- [16] ZHAN Peida, JIAO Hong, MAN Kaiwen, et al. Using JAGS for Bayesian cognitive diagnosis modeling: a tutorial [J]. *Journal of Educational and Behavioral Statistics*, 2019 44(4): 473-503.
- [17] GELMAN A, CARLIN J B, STERN H S, et al. Bayesian data analysis [M]. 3rd ed. London: Chapman and Hall/CRC 2003.
- [18] AKAIKE H. A new look at the statistical model identification [J]. *IEEE Transactions on Automatic Control*, 1974, 19(6): 716-723.
- [19] SCHWARZ G. Estimating the dimension of a model [J]. *Annals of Statistics*, 1978 6(2): 461-464.
- [20] SPIEGELHALTER D J, BEST N G, CARLIN B P. Bayesian measures of model complexity and fit [J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002 64(4): 583-639.

The Comparative Study on the Generalization Performance of Longitudinal Cognitive Diagnostic Models for Learning Assessment

SONG Lihong¹, YUAN Siyu², WANG Wenyi^{2*}

(1. School of Education, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

2. School of Computer and Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: To investigate the ability of longitudinal cognitive diagnostic models to adapt to fresh data, the generalization performance of three longitudinal cognitive diagnostic models on different types of longitudinal data is investigated. The first longitudinal cognitive diagnostic model is Patt-DINA, a model of latent transition analysis at the attribute pattern level. The second is the Att-DINA, a model of latent transition analysis at the attribute level. And the sLong-DINA model is based on higher-order latent structures. The performance of these three models is evaluated with the correct classification rates of attribute and pattern of students' knowledge states, the absolute model fit index and the relative model fit index. The results of the simulation study show that the Att-DINA model and the sLong-DINA model are more advantageous in most conditions, which means that their generalization performance is relatively better. The Patt-DINA model is less advantageous due to the larger number of parameters to be estimated, but the model still has advantages when the sample size is large and it can estimate transition probabilities of knowledge states with more space for variation.

Key words: longitudinal cognitive diagnostic model; attribute transition; pattern transition; higher-order model; generalization performance

(责任编辑: 冉小晓)