

文章编号:1000-5862(2020)05-0441-13

可以处理补偿作用的认知诊断模型:CDINA模型

罗照盛¹,杭丹丹¹,秦春影²,喻晓锋¹

(1. 江西师范大学心理学院,江西 南昌 330027;2. 南昌师范学院数学与计算机系,江西 南昌 330032)

摘要:对 DINA 和 DINO 模型进行改进,构建了一个比 DINA 和 DINO 模型更“一般”的模型,称为 CDINA 模型.除了猜测和失误参数之外,在 CDINA 模型中每个项目还包含 1 个补偿参数.构建的模型不仅保留了 DINA 模型简单和易于解释的特点,而且模拟和实证研究结果表明:(i) CDINA 模型可同时处理项目属性之间是连接、部分补偿或完全补偿的情况(可将 DINA 和 DINO 看成是 CDINA 的特例),且 CDINA 模型具有较高的参数估计精度和被试分类准确率;(ii)通过比较 CDINA、DINA、DINO 模型的分类准确率和相对拟合指标,发现 CDINA 不会比 DINA 和 DINO 的分类准确率更低,且在存在部分补偿的情况下 CDINA 模型更优于 DINA 和 DINO 模型.

关键词:补偿作用;参数估计;准确率;EM 算法

中图分类号:B 841 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2020.05.01

0 引言

随着社会的发展,人们期望从心理和教育测验中获得更多的信息,不仅仅是想知道被试在测验上的总体评价,还想了解被试在测验领域上的详细掌握状况以及对应的优势和弱点.也正是在这样的社会发展要求下,认知诊断评价(Cognitive Diagnosis Assessment, CDA)^[1-5]得到迅速发展.近年来,CDA 受到越来越多研究者们的关注, Q 矩阵和认知诊断模型(Cognitive Diagnosis Model, CDM)是 CDA 的重要组成部分.在 CDA 中,属性是指测验所要考查的目标潜质.属性粒度有粗有细,其粗细是相对而言的.CDM 的功能是在属性粒度较细的条件下,根据被试的作答反应去推断其属性掌握模式(也称为知识状态)^[2].若无特别说明,本文中的 CDM 特指潜在类模型,不包含多成分潜在特质模型^[6].

在 CDA 中,通常用 0 表示被试没有掌握某属性,1 表示被试掌握了该属性.基于此,被试在测验所考察属性上的掌握情况,即属性掌握模式,是一个

由 0 和 1 构成的向量.被试的属性掌握模式是潜在的,不可直接观察的.CDA 是通过被试的可观察的作答来推断其属性掌握模式,因此,必须建立起可观察的作答和不可观察的属性掌握模式之间的联系纽带,这个纽带就是 Q 矩阵^[7-8]. Q 矩阵描述的是测验项目和属性之间的关联关系,它为每个项目界定所考察的属性(即每个项目对应着一个属性向量).CDM 是通过被试在每个项目上的作答,借助 Q 矩阵,从而达到推断属性掌握模式的目的.不同的 CDM 在具体的推断机制上有所不同,使用时有不同的前提假设,这些推断机制和假设也就直接决定了 CDM 的应用范围.如 Deterministic Inputs Noisy “And” gate(DINA)^[9-10]模型,它是受众多研究者关注的 CDM 之一,不考虑属性间的相互关系,每个项目只有 2 个参数,即猜测参数(guessing parameter, g)和失误参数(slipping parameter, s).

DINA 最大的优点在于参数简单和易于解释,但是一个可能存在的缺点是:DINA 模型是建立在较强的假设条件上的,但在实际的测验中,这些假设很可能得不到满足,此时使用 DINA 模型就会得到

收稿日期:2019-09-06

基金项目:国家自然科学基金(31660279,31600909),教育部人文社会科学研究课题(17YJC190029),江西省教育厅科学技术研究课题(GJJ160309, GJJ191128, GJJ191691, GJJ181077, GJJ191129),江西省教育厅人文社科课题(XL1509),江西省社科规划课题(16JY11),江西省科技厅重点研发课题(20192BBEL50040),江西省教育科学“十三五”规划(20YB250),江西省高校人文社会科学青年课题(XL20202)和江西省高等学校教学改革研究课题(JXJG-19-2-13, JXJG-19-23-2)资助项目.

作者简介:罗照盛(1971-),男,江西万安人,教授,博士,博士生导师,主要从事心理统计与测量学的研究. E-mail:luozs@126.com

偏差比较大的结果. 在实际的测验领域中, 属性之间很可能不是独立的, 而是存在相关关系 (the correlation of attributes)^[11-12]、层级关系 (hierarchical relationship)^[13]、补偿关系 (compensatory interaction of skills)^[14]、或非补偿关系 (也称连接关系, conjunctive interaction of skills)^[14-15] 等. 若实际的属性之间存在某种关系, 而仍然使用 DINA 模型对结果进行分析, 则肯定会造成模型和数据的失拟, 进而造成项目参数估计不准确, 并且对被试的分类造成负面影响.

理想反应 (也称期望反应) 是指在不考虑猜测和失误时的反应. 在这个假设条件下, 对某个具体的题目, 被试只有在掌握这个题目所涉及的全部属性才能得分, 否则无法正确作答, 这就是所谓的属性连接 (conjunction)^[14, 16-18]; 另外, 虽然被试没有掌握这个题目所涉及的全部属性, 但是掌握部分属性也可以有较大可能在这个题目上得分, 因为考生所掌握的这部分题目所考察的属性可以对其他没有掌握的那部分属性起到“补偿”作用, 这种关系就是所谓的属性补偿 (compensation)^[14, 16-18], 更准确地说, 这里的补偿关系实际上是一种“部分补偿”关系, 即考生若掌握能够产生补偿作用的那部分属性, 则会提高考察被补偿属性那些题目的得分概率. 若这个得分概率在理想情况下可以提高到 1, 则这就是属性之间的“完全补偿”.

研究者们已经提出一些能够处理这些属性关系的模型, 比如 DINA 模型能够处理属性连接关系, DINO 模型能够处理属性间的完全补偿. 并且基于这些模型展开了一系列的研究, 比如基于 DINA 模型, Liu Hongyun 等^[19-20] 对英语科目诊断测验进行了尝试, J. L. Templin 等^[21-22] 基于 DINO 模型对病理性赌博的识别和错误类型的识别进行了研究.

而对部分属性关系的建模, 已有模型均无法较好地处理. 在实际的测验中, 项目属性之间存在部分补偿的情况也较多, 如在英语的阅读理解测验中, 一个对某个 (些) 单词无法识别的被试, 可以通过对上下文的语境因素 (contextual factors) 来推断出这个 (些) 单词的含义 (当然这个推断出的含义有可能正确), 即被试对上下文的理解和推断能力补偿了其在单词识别上的部分能力^[23]. 因此, 本文拟在 DINA 和 DINO 模型的基础上, 为每个项目增加 1 个参数, 即补偿参数, 将部分补偿作用对于作答反应带来的影响纳入模型中, 充分反映属性间的补偿关系. 修改后的模型不但保留了 DINA (或 DINO) 模型简单和易于解释的特性, 而且可以处理属性间存在补偿

(包括完全补偿和部分补偿) 和非补偿 (即连接作用) 的情况, 并且可以把 DINA 和 DINO 看成 CDINA 的特例.

1 DINA 与 DINO 模型

1.1 DINA 模型

DINA (Deterministic Inputs Noisy “And” gate) 模型假设题目的属性之间是“连接”型、非补偿的关系^[9], 它是较为“节俭”的模型, 每个题目只包含猜测和失误 2 个参数. 在 DINA 模型中, “确定性输入 (deterministic inputs)”是指 DINA 模型中被试的理想反应是由其知识状态和测验 Q 矩阵 (Q_i)^[24] 共同决定的, 通常用 η_{ij} 表示. η_{ij} 表示被试 i (知识状态为 α_i) 在项目 j (属性向量为 q_j) 上的理想反应, 其计算

公式为 $\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$, 其中 k 为测验所考察的属性个数.

“噪音输出 (noisy output)”是指被试在题目上的作答存在猜测和失误^[25], 也就是说, 被试完全掌握了某题目所考察的属性未必能正确作答, 被试未完全掌握题目中所有的属性, 未必会错误作答. “与门 (‘and’ gate)”是指理想反应 η 需要被试只有完全掌握题目中的所有属性才能正确作答, 若缺少任何一个必需的属性则均无法正确作答, 这实际上也是“连接”的含义. 在 DINA 模型下, 被试 i 在题目 j 上的正确作答概率为

$$P(x_{ij} = 1 | \alpha_i, q_j, \eta_{ij}) = (1 - s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}, \quad (1)$$

其中 x_{ij} 表示考生 i 在题目 j 上的得分, 取 0 或 1, 其中 0 表示错误作答, 1 表示正确作答. (1) 式中每个题目包含 2 个参数, 分别是失误参数 s_j 和猜测参数 g_j , 其中失误参数表示考生掌握题目所考察的属性但是在题目上没有得分的概率, 猜测参数表示考生没有掌握题目所考察的所有属性但仍然得分的概率. 即 s_j 和 g_j 的含义为

$$s_j = P(x_{ij} = 0 | \alpha_i, q_j, \eta_{ij} = 1),$$

$$g_j = P(x_{ij} = 1 | \alpha_i, q_j, \eta_{ij} = 0).$$

1.2 DINO 模型

DINO (Deterministic Inputs Noisy “Or” gate) 描述的是另一种情形, 即属性之间是“非连接”的补偿关系^[21]. 与 DINA 模型一样, DINO 模型中的题目 j 也只包含 2 个参数: 猜测参数 g_j 和失误参数 s_j , 但是它们的含义较 DINA 模型有所不同. 在 DINO 模型中的 g_j 表示考生没有掌握题目 j 的任何属性但仍然得

分的概率; s_j 表示被试至少掌握了题目 j 的1个属性但不得分的概率.对于理想得分 ω ,DINO模型和DINA模型也不相同,可以按 $\omega_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik})^{q_{jk}}$ 方法计算.这表明,在理想情况下,DINO模型假设考生 i 只要至少掌握题目 j 考察的1个属性就可以得1分,否则得0分.题目 j 的各个属性之间是可以相互“完全补偿”的.DINO模型的题目反应函数为

$$P(x_{ij} = 1 | \alpha_i, q_j, \omega_{ij}) = (1 - s_j)^{\omega_{ij}} g_j^{(1-\omega_{ij})}.$$

需要注意的是,即使属性之间存在“完全补偿”作用,也并不意味着属性之间可以相互代替或重复定义,如J. L. Templin等^[21]根据美国精神病学学会发布的精神疾病诊断与统计手册(DSM-IV-TR),界定出了赌博研究量表(Gambling Research Instrument, GRI)的 Q 矩阵,其中包含能够判断被试是否患有病理性赌博的10个属性,被试只要掌握其中的任何一个属性,就倾向于在题目上回答“是”(题目作答采用“是”和“否”2种情况),并根据被试在GRI上的作答反应,运用DINO模型检测被试是否患有病理性赌博(Pathological gambling)障碍.罗兴南^[22]将DINA和DINO模型应用到“水溶液”的知识诊断测验中,其中DINA模型用来诊断被试对知识的掌握情况,而DINO模型用于诊断被试所犯的错误类型(测验有选择题和问答题2类,每种错误类型对应了某些特定的作答结果,只要被试在作答中出现了这些结果之一,就可以确定错误类型).

DINA模型和DINO模型描述了2个“极端”的情况:题目的所有属性之间非补偿和题目的所有属性相互完全补偿,但实际的属性之间可能会存在另外的情形,比如题目的属性之间存在部分补偿的作用.

2 补偿作用

根据前面的介绍,属性补偿是指考生在没有掌握题目中的一部分属性,但是如果该考生掌握另外一部分属性,并且掌握的这部分属性对未掌握的那部分属性会产生一定程度的补偿作用,即这个考生仍然有较大机会在这个题目上得分^[14,17,23-34].一般来说,有2个因素会对补偿作用产生影响:被试和题目.在不影响理解的情况下,把能够产生补偿的属性称为“补偿属性”,而会受到“补偿属性”补偿的属性称为“被补偿属性”.基于此,补偿作用能够产生作用的情形是:(i)从考生的角度来看,必须是掌握了“补偿属性”的考生;(ii)从题目的角度来看,必

须是题目考查了“被补偿属性”,这2个条件必须同时满足才行.

用题目水平的参数 c (即compensating parameter)对题目属性间的补偿作用进行建模.下面用一个简单的例子来说明,设一个测验考察3个属性,并且第1个属性会对第2个属性产生一定的补偿作用,若一个考生只掌握了第1个属性,即 $\alpha_{i_1} = (1\ 0\ 0)$.这个考生在 $q_j = (1\ 1\ 0)$ 的题上仍然会有较高的得分概率;而另一个考生 $\alpha_{i_2} = (0\ 0\ 1)$ 在这个题目上的得分就不会受到影响.进一步,设这个题目的补偿参数大小0.35,这意味着被试 i_1 在题目 j 上得分的概率为0.35,而被试 i_2 在这个题目上要想得分只能通过猜测.这个描述性的例子表明属性补偿只会影响“特定的考生”,即那些掌握“补偿属性”但是不掌握“被补偿属性”的考生在“特定的题目”(即考察了“被补偿属性”的题目)上的得分.

3 CDINA 模型及其识别

下面首先介绍几个会用到的定义.

定义1 称考生掌握了题目所考察的所有属性为考生完全掌握某题目.

定义2 称考生掌握题目所考察的部分属性(不含全部属性)为考生部分掌握某题目.

定义3 称考生未掌握题目所考察的任何属性为考生完全未掌握某题目.

3.1 CDINA 模型

为了方便说明,记可以处理属性间部分补偿作用的模型为CDINA(Compensated DINA).在这个模型中,每个题目有3个参数: s 、 g 和 c (compensation parameter).其中失误参数 s 的含义与DINA模型中的相同,猜测参数 g 是指考生完全未掌握某题目考察的属性但是在该题目上仍得分的概率,补偿参数 c 是指考生部分掌握某题目并且在该题目上得分的概率.相对于DINA模型,CDINA模型将没有完全掌握某题目的考生划分成2部分:纯粹由于猜测所导致的得分概率和由于补偿作用所导致的得分概率.

因此,CDINA模型与2个常见的模型(DINA和DINO)不同之处在于:根据考生在每个题目上的得分,前者将考生分成3类(完全掌握该题目属性的被试、部分掌握该题目属性的被试和完全未掌握该题目的被试),而后2者(DINA和DINO)将考生分成2类(完全掌握该题目属性的被试和未完全掌握该题目属性的被试,或至少掌握该题目1个属性的被试

和完全未掌握该题目属性的被试)。

对于 CDINA 模型,仍然采用 η_{ij} 来表示考生 i 对于题目 j 的属性掌握情况,不同于 DINA 和 DINO 模型, $\eta_{ij} = 1$ 表示考生 i 完全掌握题目 j 考察的属性; $\eta_{ij} = 0$ 表示考生 i 完全未掌握题目 j 考察的属性; $\eta_{ij} = 2$ 表示考生 i 掌握题目 j 考察的部分属性. η_{ij} 可按如下方式计算

$$\eta_{ij} = \begin{cases} 0, \alpha_i^T q_j = 0, \\ 1, \alpha_i^T q_j = q_j^T q_j, \\ 2, \text{其他}, \end{cases} \quad (2)$$

其中 α_i 是考生 i 的知识状态, q_j 对应题目 j 的属性向量, T 表示矩阵转置. 可以看出, 与 2 个受限模型 DINA 和 DINO 不同的是, 基于考生的知识状态, CDINA 模型把考生在题目中的得分按 3 种情况来处理, 分别是考生没有掌握题目的任何属性 ($\eta_{ij} = 0$)、考生拥有题目的所有属性 ($\eta_{ij} = 1$) 和考生拥有题目的部分属性 ($\eta_{ij} = 2$). 下面通过一个描述性的例子来进行说明, 设某个题目所考察的属性向量为 (1 0 1 1), 则 $q_j^T q_j = 3$. 由 (2) 式得, 有 3 类被试 (1 1 1 1)、(0 1 0 0) 和 (1 0 1 0), 他们对应的 $\alpha_i^T q_j$ 值为 3、0 和 2, 在这个题目上他们的理想作答分别是 1、0 和 2.

用 P_{ij} 来表示被试 i 在题目 j 上的正确作答概率, 当 $\eta_{ij} = 1$ 时, $P_{ij} = 1 - s_j$; 当 $\eta_{ij} = 0$ 时, $P_{ij} = g_j$; 当 $\eta_{ij} = 2$ 时, $P_{ij} = c_j$. 在 CDINA 模型中, 被试 i 在题目 j 上的正确作答概率可以表示为 $P(x_{ij} = 1 | \alpha_i, q_j, s_j, g_j, c_j, \eta_{ij}) = (1 - s_j)^{I(\eta_{ij}, 1)} (g_j)^{I(\eta_{ij}, 0)} (c_j)^{I(\eta_{ij}, 2)}$, 其中 I 为示性函数, 当 $x = y$ 时, $I(x, y) = 1$; 否则 $I(x, y) = 0$.

3.2 CDINA 模型的假设

在测验中, 属性间补偿关系的确定是非常具有挑战的工作. 从题目的角度看, 其考察的属性之间可能会存在某种补偿作用. 但是这个作用在不同的题目上又不相同, 它会通过考生在题目上的作答数据体现出来. 因此, 根据估计出来的 c 参数可以来判断是否存在补偿作用及其大小, 而不需要事先来确定.

对于 CDINA 模型中题目 j 的补偿参数的理解, 可以有 2 种解释, 第 1 种是考生所掌握的那部分属性 (题目也考察了) 对其没有掌握的那部分属性 (题目也考察了) 起到补偿作用; 第 2 种是考生使用“其他能力”在该题目上得分的概率, 就像 E. Maris^[35] 提到的那样.

CDINA 模型基于如下 2 个假设.

假设 1 将由于题目 j 考察的属性所产生的补偿作用 c_j 的大小定义为 $c_j \in (0, 1)$, 换句话说, 一个

较小的 c_j 表示补偿作用不存在或补偿作用较小; 一个接近于 1 的 c_j 表示补偿较大或完全补偿; 除此之外均表示存在部分补偿. 由于随机因素所带来的影响, 在实际数据中即使题目所考察的属性之间不存在补偿, 估计得到的 c_j 可能也不为 0.

假设 2 若考生掌握了某题目所考察的所有属性, 则其在这个题目上的得分概率由参数 s 来决定; 若考生掌握了某题目所考察的部分属性, 则其得分概率由参数 c 来决定; 若考生没有掌握题目考察的任何属性, 则其得分概率由参数 g 来决定.

3.3 CDINA 与 2 个受限模型之间的关系

由前文可知, 当题目的 c_j 与 $1 - s_j$ 相等时, CDINA 就退化成 DINO 模型. 当题目的 c_j 与 g_j 相等时, CDINA 就退化成 DINA 模型. 可以看出, DINA 和 DINO 都是 CDINA 模型的特例, 下面举例来说明这一点.

设有一个考察 3 个属性的测验, 并且测验的 Q 矩阵为

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad (3)$$

其中 (3) 式中的行对应题目, 列对应属性. 基于 CDINA 模型, 相应的 η^{CDINA} (η 的上标表示所采用的模型) 和 P^{CDINA} (CDINA 模型下的得分概率) 矩阵为

$$\eta^{\text{CDINA}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 000 \\ 100 \\ 010 \\ 001 \\ 110 \\ 101 \\ 011 \\ 111 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \\ 0 & 2 & 2 \\ 1 & 2 & 2 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \end{matrix},$$

$$P^{\text{CDINA}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 000 \\ 100 \\ 010 \\ 001 \\ 110 \\ 101 \\ 011 \\ 111 \end{matrix} & \begin{pmatrix} g_1 & g_2 & g_3 \\ 1 - s_1 & g_2 & c_3 \\ g_1 & c_2 & g_3 \\ g_1 & c_2 & c_3 \\ 1 - s_1 & c_2 & c_3 \\ 1 - s_1 & c_2 & 1 - s_3 \\ g_1 & 1 - s_2 & c_3 \\ 1 - s_1 & 1 - s_2 & 1 - s_3 \end{pmatrix} \end{matrix} \quad (4)$$

当基于 DINO 模型 (即考虑完全补偿) 时, 相应的 η^{DINO} 和 P^{DINO} 可以表示为

$$\boldsymbol{\eta}^{\text{DINO}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 000 \\ 100 \\ 010 \\ 001 \\ 110 \\ 101 \\ 011 \\ 111 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{matrix},$$

$$\boldsymbol{P}^{\text{DINO}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 000 \\ 100 \\ 010 \\ 001 \\ 110 \\ 101 \\ 011 \\ 111 \end{matrix} & \begin{pmatrix} g_1 & g_2 & g_3 \\ 1-s_1 & g_2 & 1-s_3 \\ g_1 & 1-s_2 & g_3 \\ g_1 & 1-s_2 & 1-s_3 \\ 1-s_1 & 1-s_2 & 1-s_3 \\ 1-s_1 & 1-s_2 & 1-s_3 \\ g_1 & 1-s_2 & 1-s_3 \\ 1-s_1 & 1-s_2 & 1-s_3 \end{pmatrix} \end{matrix}.$$

进一步,若限定(4)式中所有的 $c_j = 1 - s_j$,则 CDINA 模型等价于 DINO 模型.

当考虑 DINA 模型时,相应的 $\boldsymbol{\eta}^{\text{DINA}}$ 和 $\boldsymbol{P}^{\text{DINA}}$ 应为

$$\boldsymbol{\eta}^{\text{DINA}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 000 \\ 100 \\ 010 \\ 001 \\ 110 \\ 101 \\ 011 \\ 111 \end{matrix} & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \end{matrix},$$

$$\boldsymbol{P}^{\text{DINA}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 000 \\ 100 \\ 010 \\ 001 \\ 110 \\ 101 \\ 011 \\ 111 \end{matrix} & \begin{pmatrix} g_1 & g_2 & g_3 \\ 1-s_1 & g_2 & g_3 \\ g_1 & g_2 & g_3 \\ g_1 & g_2 & g_3 \\ 1-s_1 & g_2 & g_3 \\ 1-s_1 & g_2 & 1-s_3 \\ g_1 & 1-s_2 & g_3 \\ 1-s_1 & 1-s_2 & 1-s_3 \end{pmatrix} \end{matrix}.$$

可以看出,若限定(4)式中的 $c_j = g_j$,则 CDINA 模型就转化为 DINA 模型.

综上所述,CDINA 模型可以看作是比 DINA 和 DINO 模型更一般的模型.由 $\boldsymbol{\eta}^{\text{CDINA}}$ 可知,

CDINA 模型对于题目属性间存在补偿作用(不同程度的补偿作用)、连接或2者的组合的情况均可以处理.

3.4 模型的参数估计

参照 de la Torre^[9] 在 2009 年介绍的 EM 算法,可以得到 CDINA 模型的参数估计过程,为节省篇幅,这里重点对 2 者不一样的地方予以说明.

在 CDINA 模型中,当考生掌握了某题目考察的所有属性时的得分概率为 $P(x_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{q}_j, \boldsymbol{\eta}_{ij} = 1) = 1 - s_j$,当考生未掌握该题目的任何属性时,其得分概率 $P(x_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{q}_j, \boldsymbol{\eta}_{ij} = 0) = g_j$,当考生只是部分掌握该题目时,其得分概率 $P(x_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{q}_j, \boldsymbol{\eta}_{ij} = 2) = c_j$.

下面是模型中各参数计算过程.首先是计算 I_l :

$$I_l = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i),$$

这里 I_l 是总体中知识状态为 $\boldsymbol{\alpha}_i$ 的考生人数的期望值,其中 I_l 包括 3 部分,它们分别是 $I_{jl}^{(0)}$ 、 $I_{jl}^{(1)}$ 和 $I_{jl}^{(2)}$, $I_{jl}^{(0)}$ 为完全没有掌握题目 j 的考生人数的期望值, $I_{jl}^{(1)}$ 为完全掌握题目 j 的考生人数的期望值, $I_{jl}^{(2)}$ 为部分掌握题目 j 的考生人数的期望值.它们的计算公式分别为

$$I_{jl}^{(0)} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i),$$

$$I_{jl}^{(1)} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i),$$

$$I_{jl}^{(2)} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i).$$

接下来给出 R_{jl} 的计算公式:

$$R_{jl} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i) X_{ij},$$

其中 R_{jl} 是知识状态为 $\boldsymbol{\alpha}_i$ 的考生在题目 j 上得分的人数的期望值,它也包括 3 个部分: $R_{jl}^{(0)}$ 、 $R_{jl}^{(1)}$ 和 $R_{jl}^{(2)}$,其中 $R_{jl}^{(0)}$ 是指完全未掌握题目 j 但仍然得分的考生人数的期望值, $R_{jl}^{(1)}$ 是指完全掌握题目 j 并且得分考生人数的期望值, $R_{jl}^{(2)}$ 是指部分掌握题目 j 但仍然得分的考生人数的期望值,它们的计算公式分别为

$$R_{jl}^{(0)} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i) X_{ij},$$

$$R_{jl}^{(1)} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i) X_{ij},$$

$$R_{jl}^{(2)} = \sum_{i=1}^N P(\boldsymbol{\alpha}_i | X_i) X_{ij}.$$

因此, s_j 的计算公式为 $\hat{s}_j = (I_{jl}^{(1)} - R_{jl}^{(1)})/I_{jl}^{(1)}$, 可以看出, 这个公式与 DINA 中相同. g_j 可以按 $\hat{g}_j = R_{jl}^{(0)}/I_{jl}^{(0)}$ 计算, 它与 DINA 中不相同. c_j 的计算公式可以表示为 $\hat{c}_j = R_{jl}^{(2)}/I_{jl}^{(2)}$.

4 研究 1: CDINA 模型识别及考生分类

由前文可知, 为探讨提出的 CDINA 模型的参数估计算法的性能(题目参数和考生参数的估计精度), 采用蒙特卡罗模拟方法.

4.1 Q 矩阵、题目参数、考生的知识状态

在本研究中所使用的 Q 矩阵如图 1 所示, 其中 Q_1 、 Q_2 和 Q_3 分别对应 3、4 和 5 个属性, 测验长度有 2 个水平, 即 15 和 20(长度为 15 时的测验 Q 矩阵取 Q_1 、 Q_2 和 Q_3 的前 15 个题目). 模型的参数从均匀分布中抽取, 其中猜测和失误参数的取值区间均为 $[0.05, 0.25]$, 而补偿参数的取值区间为 $(0, 1)$. 设考生的知识掌握状态服从均匀分布, 样本量取 4 个水平, 分别为 500、1 000、2 000 和 4 000, 因此共有 $3 \times 2 \times 4 = 24$ 种实验条件, 为降低随机抽样误差, 每种实验条件重复 100 次, 最后报告 100 次的平均值.

4.2 作答模拟

基于题目参数和被试参数, 根据 CDINA 模型来生成包含补偿参数的作答得分, 详细过程如下: (i) 首先由考生的属性掌握模式和题目考察的属性, 得到 η_{ij} 的值; (ii) 根据 CDINA 模型, 得到考生在该题目上的得分概率 P_{ij} ; (iii) 随机生成一个服从均匀分布, 并在 $0 \sim 1$ 之间的随机数 r_1 , 当 $P_{ij} \geq r_1$ 时, $X_{ij} =$

1, 否则 $X_{ij} = 0$.

4.3 评价指标

这里的评价指标分为题目参数估计指标和考生分类指标, 其中题目参数的精度指标采用平均绝对离差 A_{BSE} , 按下式计算:

$$A_{BSE} = (\sum_{i=1}^N |x_i - \hat{x}_i|)/N,$$

其中 x_i 为真值, \hat{x}_i 为估计值, A_{BSE} 越小表明估计越准确.

考生分类的指标采用模式判准率 (Pattern Match Ratio, P_{MR}) 和属性边际判准率 (Average Attribute Match Ratio, A_{AMR})^[26]. P_{MR} 和 A_{AMR} 按下式计算:

$$P_{MR} = (\sum_{i=1}^N I(\alpha_i, \hat{\alpha}_i))/N,$$

$$A_{AMR} = (\sum_{i=1}^N \sum_{k=1}^K I(\alpha_{ik}, \hat{\alpha}_{ik}))/N,$$

其中 N 为考生总人数; P_{MR} 为在所有考生中属性模式被正确估计的考生比例, 即指对每个考生的模式正确估计的概率; I 表示示性函数; A_{AMR} 表示测验对单个属性的平均估计准确概率; α_i 表示考生的真实属性掌握模式; $\hat{\alpha}_i$ 为属性掌握模式的估计值.

4.4 实验结果

项目参数和被试参数对应的结果如表 1 和表 2 所示, 图 2 描述了测验 Q 矩阵分别为 Q_1 、 Q_2 、 Q_3 在不同样本量和不同测验长度情况下 CDINA 模型的表现, 其中图 2(a) 对应测验长度为 20, 图 2(b) 对应测验长度为 15.

表 1 项目参数估计结果 (CDINA)

Q 矩阵	被试人数	A_{BSE}		
		s	g	c
Q_1	500	0.022 9/0.026 2	0.018 1/0.024 5	0.020 4/0.015 5
	1 000	0.016 6/0.018 4	0.017 8/0.019 6	0.012 0/0.012 4
	2 000	0.013 5/0.014 0	0.013 4/0.013 9	0.007 9/0.006 8
	4 000	0.008 5/0.010 9	0.009 7/0.009 8	0.006 4/0.007 3
Q_2	500	0.028 4/0.028 8	0.032 0/0.035 1	0.018 0/0.019 1
	1 000	0.026 0/0.018 5	0.026 3/0.027 9	0.013 1/0.016 6
	2 000	0.015 7/0.017 4	0.011 2/0.016 4	0.007 5/0.015 3
	4 000	0.013 6/0.009 4	0.007 8/0.011 9	0.005 6/0.010 7
Q_3	500	0.028 0/0.027 1	0.032 5/0.039 9	0.026 2/0.020 5
	1 000	0.019 8/0.022 0	0.013 0/0.025 1	0.018 2/0.018 7
	2 000	0.012 5/0.011 6	0.011 0/0.020 2	0.011 8/0.017 4
	4 000	0.008 0/0.007 8	0.009 7/0.013 5	0.007 8/0.008 8

注: 在 A_{BSE} 中“/”前面的数字对应的是在测验长度为 20 时的结果, 后面的数字对应的是在测验长度为 15 时的结果.

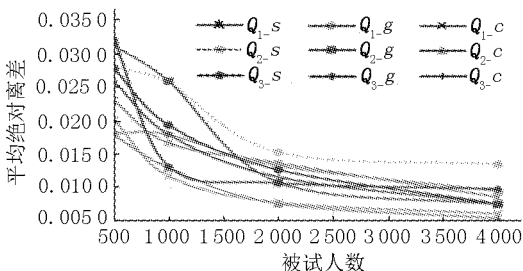
$$Q_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \quad Q_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

图1 Q 矩阵

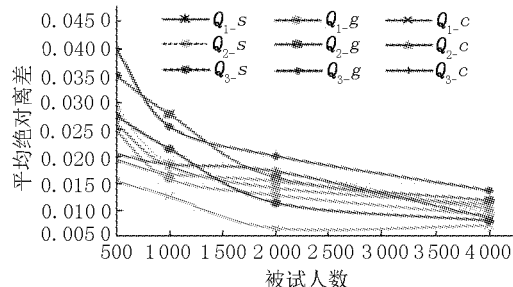
由表1和图2表明,CDINA模型在测验长度为15和20时均有较好的项目参数估计.以 Q_1 为例,测验长度为15,被试样本量为500,3个参数 s 、 g 、 c 的 A_{BSE} 指标分别为0.0262、0.0245和0.0155;当被试人数增加到1000时, s 、 g 、 c 的 A_{BSE} 指标均有所下降,分别为0.0184、0.0196和0.0124;当人数继续增加时, s 、 g 、 c 的 A_{BSE} 指标会继续下降,这表明随着被试人数的增加,题目参数的估计精度会越来越好.这

一结论在测验属性测验矩阵为 Q_2 和 Q_3 时也同样适用.

由表2和图3~图4可知,这3个被比较的模型,在考生分类指标上,占优的模型是CDINA.并且在不同的条件下,在图3和图4中显示出CDINA模型的考生分类指标更稳定.这3种测验Q矩阵下平均的考生PMR指标分别为95.1%、86.7%和79.4%.这表明了考生分类参数的准确率会随着测验属性的个数增多而下降.通过比较测验长度是15和20这2种情况下的结果发现题目参数和考生分类会随着测验长度的增加而增加.



(a) 测验长度为20



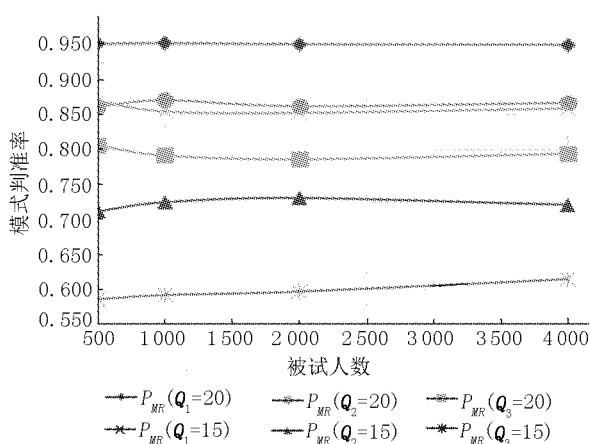
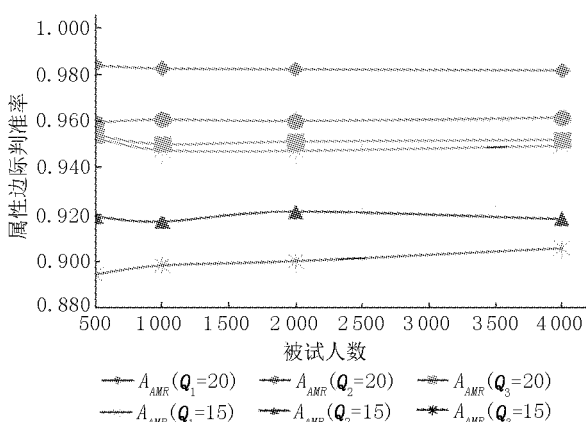
(b) 测验长度为15

图2 CDINA模型的项目参数精度 A_{BSE}

表2 考生分类指标: P_{MR} 和 A_{AMR}

Q 矩阵	被试人数	分类判断率	
		P_{MR}	A_{AMR}
Q_1	500	0.951/0.871	0.983/0.952
	1 000	0.952/0.854	0.984/0.946
	2 000	0.953/0.855	0.983/0.946
	4 000	0.952/0.858	0.983/0.951
Q_2	500	0.861/0.711	0.961/0.918
	1 000	0.873/0.727	0.962/0.918
	2 000	0.864/0.732	0.961/0.922
	4 000	0.868/0.722	0.963/0.919
Q_3	500	0.804/0.584	0.953/0.892
	1 000	0.791/0.591	0.953/0.897
	2 000	0.786/0.596	0.952/0.901
	4 000	0.794/0.616	0.953/0.907

注:在分类判断指标中,斜线“/”前面的数字是对应测验长度为20的结果,后面的数字是对应测验长度为15的结果.

图3 CDINA 模型的 P_{MR} 指标图4 CDINA 模型的 A_{AMR} 指标

5 研究2:DINA、DINO 和 CDINA 模型交叉比较

为了比较所涉及的3个模型在不同条件下的表现,按如下方式展开模拟实验:首先分别按其中的一个模型来模拟数据,然后分别使用3个模型来分析数据,对考生进行分类,比较3种模型在考生分类上的表现.因此,基于3个模型来模拟数据(DINA、DINO 和 CDINA)、被试样本量有4种(500、1 000、2 000 和 4 000)、测验中考察的属性个数有3种(3、4 和 5)和基于3个模型(DINA、DINO 和 CDINA)来分析数据,这样模拟的实验条件一共有 $3 \times 4 \times 3 \times 3 = 108$ 种情况.

5.1 模拟数据的过程

(i) Q 矩阵:3个 Q 矩阵具体内容如图1所示;

(ii) 题目参数的模拟:对于 CDINA 模型中的 s 参数和 g 参数按均匀分布 $U[0.05, 0.25]$ 模拟,而对于 c 参数按 $U(0,1)$ 模拟. DINA 和 DINO 模型中的 s 参数和 g 参数按 CDINA 中相同的方法模拟;

(iii) 被试参数的模拟:被试总体按均匀分布模

拟属性掌握模式;

(iv) 作答数据的模拟:基于前面步骤模拟的数据,模拟作答数据.对每种实验条件重复100次,取平均值作为结果,评价指标仍然采用 P_{MR} 和 A_{AMR} .

5.2 模型的比较

为了考察3个模型在分析不同类型数据上的表现,按 Huo Yan 等^[27]类似的方法,即模型与数据的交叉分析.具体过程是:为了比较2个可供选择的模型 A 和 B ,先根据模型 A 来生成数据,然后分别用它们来分析数据.由于数据是基于模型 A 生成的,因此,数据与模型 A 应该是拟合的.在这种情况下,得出模型 B 比 A 更好的结论是基于如下事实出现:根据模型 A 生成的数据,模型 B 对数据的拟合情况较好;而根据模型 B 生成的数据,模型 A 的拟合情况不好.

根据这3个模型来生成数据,进行交叉分析,并比较拟合的指标.这里考虑相对拟合指标 $-2LL$ 、AIC 和 BIC,相对拟合评价指标通常在对“竞争”模型进行比较时使用^[28].这3个相对拟合指标均为似然函数(Maximum Likelihood, ML)的函数,其中似然函数可以按如下公式计算:

$$M_L = \prod_{i=1}^N \sum_{l=1}^L L(X_i | \hat{\beta}, \alpha_l) P(\alpha_l),$$

其中 N 为考生数量, L 为可能的属性掌握模式数, X_i 为第 i 个考生的得分模式, α_l 为第 l 种属性掌握模式, $\hat{\beta}$ 为和题目相关的参数估计值, $L(X_i | \hat{\beta}, \alpha_l)$ 为第 i 个考生对应的得分似然函数, $P(\alpha_l)$ 为第 l 种属性掌握模式的先验概率.因此,3个相对拟合指标的计算方法可以表示为

$$L_{-2LL} = -2\ln(M_L), A_{IC} = L_{-2LL} + 2P, B_{IC} = L_{-2LL} + p\ln(N),$$

其中 p 为模型总的参数个数,在 DINA 模型和 DINO 模型中, p 为 $(2J + 2^K - 1)$,而对于构建的 CDINA 模型, p 为 $(3J + 2^K - 1)$.这3个相对拟合指标取值越小,模型与数据的拟合性越好.

考虑到篇幅,这里只考察当考生人数为500, Q 矩阵为 Q_1 时的情况.分别基于前面提到的3个模型来模拟得分数据,对每批数据都分别使用3个模型来进行分析,然后通过拟合指标的比较来判断模型的表现.

5.3 实验结果

表3~表4列出了各个模型的分类结果,图5~图6描绘了各个模型在不同 Q 矩阵下的分类表现.

表 3 DINA 为真实模型,3 种模型的交叉比较结果

Q 矩阵	被试人数	DINA	DINO	CDINA	DINA	DINO	CDINA
		P_{MR}	P_{MR}	P_{MR}	A_{AMR}	P_{MR}	A_{AMR}
Q_1	500	0.931 *	0.231	0.931	0.974 *	0.704	0.975
	1 000	0.912 *	0.260	0.906	0.969 *	0.713	0.969
	2 000	0.913 *	0.265	0.913	0.968 *	0.718	0.968
	4 000	0.917 *	0.259	0.921	0.968 *	0.706	0.968
Q_2	500	0.797 *	0.345	0.801	0.944 *	0.792	0.941
	1 000	0.799 *	0.301	0.796	0.935 *	0.785	0.933
	2 000	0.798 *	0.383	0.794	0.940 *	0.816	0.941
	4 000	0.801 *	0.322	0.798	0.941 *	0.793	0.938
Q_3	500	0.811 *	0.291	0.813	0.955 *	0.815	0.954
	1 000	0.772 *	0.242	0.775	0.942 *	0.806	0.941
	2 000	0.773 *	0.315	0.768	0.943 *	0.828	0.942
	4 000	0.781 *	0.286	0.786	0.948 *	0.817	0.946

注:带 * 的数据列对应的模型为模拟数据采用的模型.

表 4 DINO 为真实模型,3 种模型的交叉比较结果

Q 矩阵	被试人数	DINA	DINO	CDINA	DINA	DINO	CDINA
		P_{MR}	P_{MR}	P_{MR}	A_{AMR}	P_{MR}	A_{AMR}
Q_1	500	0.257	0.923 *	0.921	0.720	0.972 *	0.968
	1 000	0.378	0.935 *	0.935	0.785	0.974 *	0.975
	2 000	0.426	0.945 *	0.946	0.803	0.981 *	0.981
	4 000	0.389	0.931 *	0.933	0.788	0.975 *	0.976
Q_2	500	0.533	0.843 *	0.847	0.858	0.952 *	0.954
	1 000	0.343	0.835 *	0.838	0.806	0.952 *	0.954
	2 000	0.481	0.842 *	0.843	0.845	0.953 *	0.955
	4 000	0.427	0.831 *	0.835	0.832	0.951 *	0.952
Q_3	500	0.381	0.763 *	0.746	0.847	0.947 *	0.943
	1 000	0.408	0.799 *	0.795	0.851	0.955 *	0.954
	2 000	0.388	0.806 *	0.807	0.845	0.952 *	0.954
	4 000	0.435	0.797 *	0.797	0.854	0.952 *	0.954

注:带 * 的数据列对应的模型为模拟数据采用的模型.

综合表 3 和图 5 的结果,当数据的真实模型为 DINA 时,DINO 的表现最差,在这 3 种测验数据下, P_{MR} 的平均值只有 30% 左右.而另 2 个模型的表现大致相当,图 5 显示出了这一点,因为 CDINA 和 DINA 对应的曲线非常接近重合.这表明了 CDINA 对于属性“连接”的情况可以较好地处理.

综合表 4、图 6 可以看出,当数据的真实模型为 DINO 时,在分类准确率上,表现最差的为 DINA 模型.在这 3 种测验数据下, P_{MR} 平均值低于 50%.与表 3 和表 4 的结果一致,CDINA 和 DINO 表现相当,图 6 也清楚地显示了这一点.这表明了 CDINA 模型对于属性“完全补偿”的情况可以较好地处理.

综合表 5、图 7 可以看出,当真实模型为 CDINA 时,表现最好的是 CDINA,其次是 DINA,表现最差的是 DINO,图 7 中的曲线较好了解释了这一点.当测验 Q 矩阵为 Q_1 时,考虑平均的 P_{MR} 指标,CDINA 模型比

DINA 和 DINO 分别高出 5.4% 和 11.1%;当测验 Q 矩阵为 Q_2 时,CDINA 模型比 DINA 和 DINO 分别高出 11.0% 和 18.7%;当测验 Q 矩阵为 Q_3 时,CDINA 模型比 DINA 和 DINO 分别高出 15.6% 和 18.2%.这些结果表明 CDINA 模型的适应性更好,对题目属性之间的补偿作用可以较好地处理,并且优势随着属性个数的增加而增加.

从表 6 所提供的拟合指标来看,无论数据是基于 DINA 还是基于 DINO 模型来模拟的,采用 CDINA 模型来分析,得到的拟合统计量 $-2LL$ 比数据对应的实际模型(即模拟数据采用的模型)还更好,如采用 DINA 模型数据, $-2LL$ 统计量为 10 556.85,当采用 DINO 模型分析时, $-2LL$ 统计量的值为 11 714.04 > 10 556.85,当采用 CDINA 模型分析时, $-2LL$ 统计量的值为 10 550.34 < 10 556.85.当考虑了模型的参数个数时,CDINA 模型的 AIC 和 BIC 指标与数据的实

际模型的对应指标非常接近,这表明 CDINA 模型可以较好地拟合 DINA 和 DINO 模型的数据.而当采用 CDINA 模型来模拟数据时,DINA 和 DINO 模型的拟

合指标均要远大于 CDINA 的指标,这表明 DINO 和 DINA 对包含部分补偿作用的数据拟合性较差.

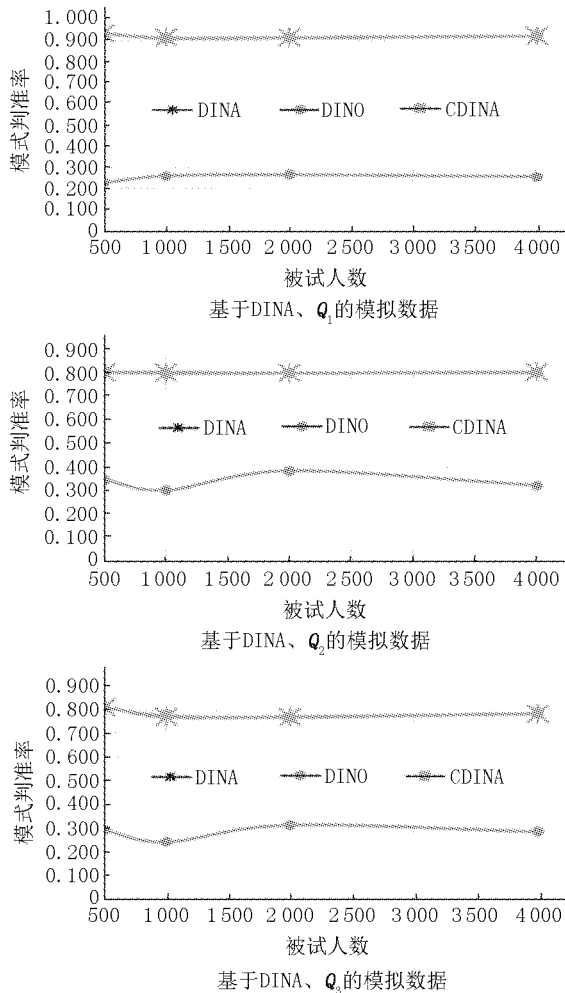


图5 真实模型为 DINA,3 种模型的分类表现

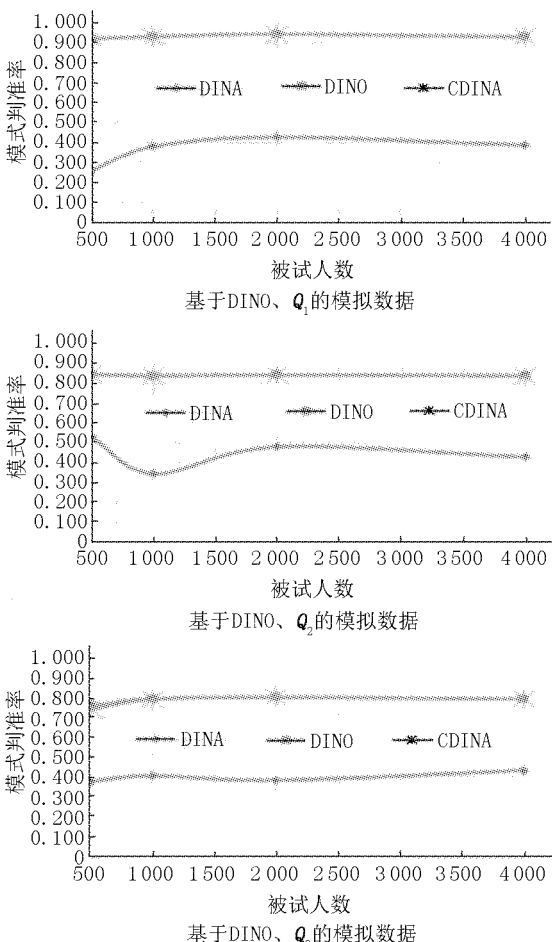


图6 真实模型为 DINO,3 种模型的分类表现

表5 CDINA 为真实模型,3 种模型的交叉比较结果

Q 矩阵	被试人数	DINA	DINO	CDINA	DINA	DINO	CDINA
		PMR	PMR	PMR	AAMR	PMR	AAMR
Q_1	500	0.901	0.836	0.942 *	0.962	0.901	0.978 *
	1 000	0.877	0.839	0.936 *	0.959	0.876	0.977 *
	2 000	0.876	0.822	0.935 *	0.957	0.876	0.973 *
	4 000	0.876	0.811	0.935 *	0.956	0.875	0.975 *
Q_2	500	0.634	0.563	0.758 *	0.888	0.633	0.924 *
	1 000	0.665	0.564	0.766 *	0.895	0.665	0.925 *
	2 000	0.682	0.614	0.793 *	0.908	0.682	0.938 *
	4 000	0.689	0.624	0.797 *	0.908	0.687	0.936 *
Q_3	500	0.597	0.515	*0.715 *	0.903	0.597	0.934 *
	1 000	0.545	0.527	0.709 *	0.892	0.549	0.933 *
	2 000	0.554	0.526	0.705 *	0.893	0.551	0.931 *
	4 000	0.526	0.552	0.718 *	0.886	0.525	0.935 *

注:带 * 的数据列对应的模型为模拟数据采用的模型.

表6 Q_1 下,500 人,模型和数据交叉分析对应的拟合统计量

拟合统计量		-2LL			AIC			BIC		
分析数据的模型		DINA	DINO	CDINA	DINA	DINO	CDINA	DINA	DINO	CDINA
模拟数	DINA	10 551.47 *	11 712.67	10 550.52	10 650.79 *	11 807.94	10 683.87	10 847.99 *	12 005.90	10 966.45
据所用	DINO	11 833.36	10 694.28 *	10 688.02	11 928.77	10 788.95 *	10 820.92 *	12 128.33 *	10 986.86 *	11 102.96
的模型	CDINA	11 904.55	11 929.31	11 391.83 *	11 997.53	12 021.76	11 526.48 *	12 196.56	12 221.15	11 808.31 *

注:带 * 的数据表示分析数据和模拟数据所使用的模型一致,在这种情况下计算出来的拟合统计量可以作为基准.

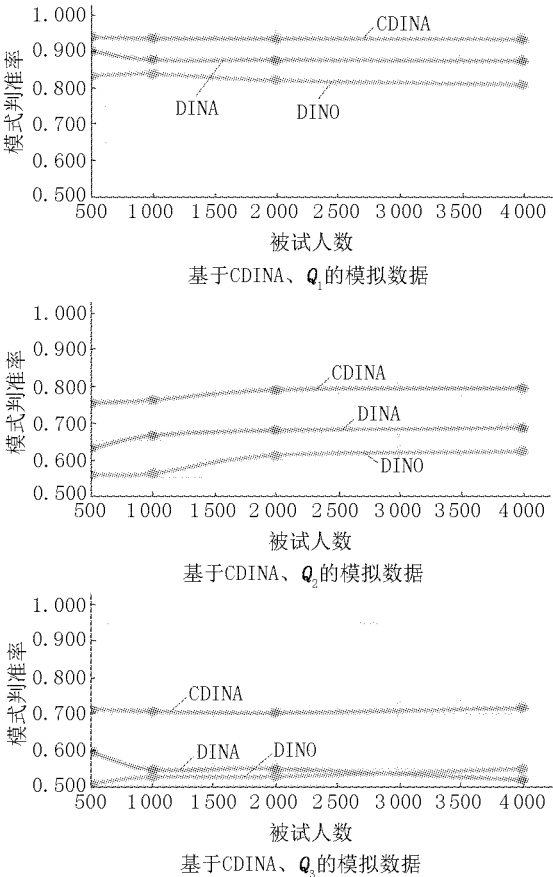


图7 真实模型为 CDINA,3 种模型的表现

6 结论和讨论

6.1 结论

本文基于 DINA 和 DINO 模型,通过引入题目水平的补偿参数,构建了可以处理属性间部分补偿作用的 CDINA 模型,对于模型的假设、参数的定义以及参数的估计过程进行了详细介绍.研究表明,CDINA 模型可以区分出题目属性间的部分补偿作用,每个题目能够将被试群体分成 3 类,即掌握这个题目的所有属性而得分的被试、掌握这个题目的部分属性而得分的被试、未掌握这个题目的任何属性而得分的被试.这也正是它与 DINA 和 DINO 模型不同的地方.

模拟实验结果显示了 CDINA 模型可以较好地处理属性之间的补偿作用,模型在参数的估计、考生

的分类和数据拟合上均有较好的表现.在对包含“连接”关系的题目属性的数据进行处理时,CDINA 不差于 DINA 模型;在对包含“完全补偿”关系的题目属性的数据进行处理时,CDINA 不差于 DINO 模型;CDINA 模型的优势在于它还可以处理包含“部分补偿”关系的数据,因此,CDINA 模型有更好的适应性.

6.2 讨论

本文构建了一个可以处理部分补偿关系的 CDINA 模型,它假定题目的属性之间可能存在部分补偿作用. CDINA 模型将补偿参数定义在题目水平上.这样考虑的主要目的是将补偿作用所带来的效果当作待估计的参数,而不需要事先对属性作出连接(如 DINA)或非连接(如 DINO)的假设.从模型的角度来看,CDINA 模型与 DINA 和 DINO 一样,也是一个简洁的模型,对每个题目,它只有 3 个参数.在实际应用中,根据教师的知识或经验界定属性之间的关系是主观的,并且是一件很困难的工作.更重要的是,CDINA 模型无须事先确定题目的属性之间的具体补偿关系(如“完全补偿”关系、“连接关系”还是“部分补偿”关系),可以根据补偿参数的大小进一步来确定;当然,CDINA 模型的构建,只是用来处理属性关系的一种探索.真正确定属性之间的关系非常困难,已有的研究中还没有较好的方法,这种情况下,使用 CDINA 来分析实际数据是一个可用的选项.

另一方面,基于 CDINA 的补偿参数为题目属性间的补偿作用有一定的参考意义,即透过补偿参数的大小来判断哪些属性之间可能存在补偿作用及作用的大小.进一步结合属性的实际意义来确认这个补偿作用的存在,这还需要结合领域专家的意见.对于处理多策略^[34]的作答数据以及属性之间存在层级关系^[13]的数据,CDINA 模型的表现,需要进一步研究.并且还需要和测验蓝图的设计^[35-37]结合起来综合考虑新模型的实际应用.

7 参考文献

[1] Yu Xiaofeng, Cheng Ying. Data-driven Q -matrix validation

- using a residual-based statistic in cognitive diagnostic assessment [EB/OL]. [2019-08-17]. British Journal of Mathematical and Statistical Psychology, <http://dx.doi.org/10.1111/bmsp.12191>.
- [2] Leighton J P, Gierl M J. Cognitive diagnostic assessment for education: theory and applications [M]. Cambridge: Cambridge University Press, 2007: 19-60.
- [3] Rupp A A, Templin J, Henson R. Diagnostic measurement: theory, methods and application [M]. Guilford: Guilford Press, 2010: 10-28.
- [4] 罗照盛. 项目反应理论基础 [M]. 北京: 北京师范大学出版社, 2012: 134-139.
- [5] 涂冬波, 蔡艳, 丁树良. 认知诊断理论、方法与应用 [M]. 北京: 北京师范大学出版社, 2012: 15-47.
- [6] Embretson S E, Yang Xiangdong. A multicomponent latent trait model for diagnosis [J]. Psychometrika, 2013, 78 (1): 14-36.
- [7] Tatsuoka K K. Rule space: an approach for dealing with misconceptions based on item response theory [J]. Journal of Educational Measurement, 1983, 20 (4): 345-354.
- [8] Tatsuoka K K. Cognitive assessment: an introduction to the rule space method [M]. Abingdon: Taylor and Francis Group, 2009: 47-79.
- [9] de la Torre J. DINA model and parameter estimation: a didactic [J]. Journal of Educational and Behavioral Statistics, 2009, 34 (1): 115-130.
- [10] Junker B W, Sijtsma K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory [J]. Applied Psychological Measurement, 2001, 25 (3): 258-272.
- [11] Wang Chun, Chang Huahua, Douglas J. Combining CAT with cognitive diagnosis: a weighted item selection approach [J]. Behav Res, 2012, 44 (1): 95-109.
- [12] Wang Chun. Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length [J]. Educational and Psychological Measurement, 2013, 73 (6): 1017-1035.
- [13] Leighton J P, Gierl M J, 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): 205-237.
- [14] Roussos L A, Templin J L, Henson R A. Skills diagnosis using IRT-based latent class models [J]. Journal of Educational Measurement, 2007, 44 (4): 293-311.
- [15] 涂冬波, 蔡艳, 戴海崎. 几种常用非补偿型认知诊断模型的比较与选用: 基于属性层级关系的考量 [J]. 心理学报, 2013, 45 (2): 243-252.
- [16] DiBello L V, Roussos L A, Stout W. Review of cognitively diagnostic assessment and a summary of psychometric models [M]. North Holland: Elsevier, 2007.
- [17] Stout W. Skills diagnosis using IRT-based continuous latent trait models [J]. Journal of Educational Measurement, 2007, 44 (4): 313-324.
- [18] Xiang Rui. Nonlinear penalized estimation of true Q -Matrix in cognitive diagnostic models [D]. Columbia: Columbia University, 2013.
- [19] Liu Hongyun, You Xiaofeng, Wang Wenyi, et al. Large-scale applications of cognitive diagnostic computerized adaptive testing in China [C]. the Annual Meeting of National Council on Measurement in Education, Denver, CO, 2010.
- [20] Liu Hongyun, You Xiaofeng, Wang Wenyi, et al. The development of computerized adaptive testing with cognitive diagnosis for an english achievement test in China [J]. Journal of Classification, 2013, 30 (2): 152-172.
- [21] Templin J L, Henson R A. Measurement of psychological disorders using cognitive diagnosis models [J]. Psychological Methods, 2006, 11 (3): 287-305.
- [22] 罗兴南. DINA 与 DINO 模式于国小自然科电脑化诊断之应用: 以五年级“水溶液”单元为例 [D]. 台中: 亚州大学, 2012.
- [23] Stanovich K E. Toward an interactive-compensatory model of individual differences in the development of reading fluency [J]. Reading Research Quarterly, 1980, 16 (1): 32-71.
- [24] 丁树良, 毛萌萌, 汪文义, 等. 教育认知诊断测验与认知模型一致性的评估 [J]. 心理学报, 2012, 44 (11): 1535-1546.
- [25] Liu Jingchen, Xu Gongjun, Ying Zhiliang. Data-driven learning of Q -matrix [J]. Applied Psychological Measurement, 2012, 36 (7): 548-564.
- [26] Cui Ying, Leighton J P, Zheng Yinggan. Simulation studies for evaluating the performance of the two classification methods in the AHM [EB/OL]. [2019-09-16]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.385.8437>.
- [27] Huo Yan, de la Torre J. Estimating a cognitive diagnostic model for multiple strategies via the EM algorithm [J]. Applied Psychological Measurement, 2014, 38 (6): 464-485.
- [28] Chen Jinsong, de la Torre J, Zhang Zao. Relative and absolute fit evaluation in cognitive diagnosis modeling [J]. Journal of Educational Measurement, 2013, 50 (2): 123-140.
- [29] Tatsuoka K K. Toward an integration of item-response theory and cognitive error diagnosis [M] // Frederiksen N, Glaser R, Lesgold A, et al. Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, New Jersey: Erl-

- baum,1990:453-488.
- [30] de la Torre J. An empirically based method of Q -matrix validation for the DINA model: development and applications [J]. Journal of Educational Measurement,2008,45(4):343-362.
- [31] de la Torre J. The generalized DINA model framework [J]. Psychometrika,2011,76(2):179-199.
- [32] de la Torre J,Douglas J A. Higher-order latent trait models for cognitive diagnosis [J]. Psychometrika,2004,69(3):333-353.
- [33] DeCarlo L T. On the analysis of fraction subtraction data: the DINA model, classification, latent class sizes, and the Q -matrix [J]. Applied Psychological Measurement,2011,35(1):8-26.
- [34] de la Torre J,Douglas J A. Model evaluation and multiple strategies in cognitive diagnosis: an analysis of fraction subtraction data [J]. Psychometrika,2008,73(4):595-624.
- [35] Maris E. Estimating multiple classification latent class models [J]. Psychometrika,1999,64(2):187-212.
- [36] 丁树良,汪文义,杨淑群. 认知诊断测验蓝图的设计 [J]. 心理科学,2011,34(2):258-265.
- [37] 丁树良,杨淑群,汪文义. 可达矩阵在认知诊断测验编制中的重要作用 [J]. 江西师范大学学报:自然科学版,2010,34(5):490-494.

The Cognitive Diagnostic Model Which Can Deal with Compensation and Noncompensation Effects: CDINA

LUO Zhaosheng¹, HANG Dandan¹, QIN Chunying², YU Xiaofeng¹

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

2. Department of Mathematics and Computer Science, Nanchang Normal University, Nanchang Jiangxi 330032, China)

Abstract: The DINA and DINO models are improved to build a more “general” model that is called CDINA model. In addition to the guessing and slipping parameters, each item in the CDINA model contains a compensation parameter. The model built preserves the parsimonious and easy-to-explain characteristics of the DINA model, and the simulation and empirical results show that the CDINA model can handle the item with its attributes are conjunctive, partially compensated, or fully compensated. That is, CDINA can be regarded as a general case of DINA and DINO, and it has relatively high parameter estimation accuracy and classification accuracy. The classification accuracy of the CDINA model is evaluated by comparing the three models of interest based on the relative fit indicators, CDINA is found to be no less accurate than DINA and DINO classifications. In the case of partial compensation, CDINA outperforms the DINA and DINO models.

Key words: compensate; parameter estimation; match ratio; EM algorithm

(责任编辑:冉小晓)