

叶子玉,秦春影,杨建芹,等.结合先验信息的多属性诊断测验分类研究[J].江西师范大学学报(自然科学版),2023,47(2):111-123.

YE Ziyu, QIN Chunying, YANG Jianqin, et al. The classification for high-dimensional cognitively diagnostic assessment based on prior information [J]. Journal of Jiangxi Normal University (Natural Science), 2023, 47(2): 111-123.

文章编号:1000-5862(2023)02-0111-13

结合先验信息的多属性诊断测验分类研究

叶子玉¹,秦春影^{1,2},杨建芹²,喻晓锋^{1,3*},付道轩¹

(1.江西师范大学心理学院,江西 南昌 330022;2.南昌师范学院数学与信息科学学院,江西 南昌 330032;

3.江西师范大学江西省心理与认知科学重点实验室,江西 南昌 330022)

摘要:在动态的学习过程中,随着学习的深入和知识属性的个数逐渐增加,学生的知识状态也会发生动态变化.在这样的应用场景下如何结合先验信息提高诊断测验的判准率具有较大的挑战.该文提出基于学生对已学习属性的掌握概率来预测包含新属性后的属性向量的先验信息.考虑2种实际的应用情境,并通过模拟和实证研究来评价该方法的表现.结果表明:在2种实际学习情境中,结合先验信息在多属性诊断测验中能起到提高判准率的作用,其中在基于个体先验信息时的表现更好.基于预测先验信息的方法也摆脱了在以往研究中不同学习阶段测验属性个数一致或数量的限制,使得分类精度有较大改进.实证数据分析进一步表明该方法具有较高的应用价值.

关键词:先验信息;诊断测验;多属性;纵向诊断测评

中图分类号:B 841 **文献标志码:**A **DOI:**10.16357/j.cnki.issn1000-5862.2023.02.01

0 引言

认知诊断测评(cognitive diagnosis assessment, CDA)^[1-3]是实现对学生内部心理加工成分和过程的形成性测量,是一种以提供学生认知加工诊断信息为目的而展开的测评.在通常情况下,学生在学习过程中对知识的掌握情况会随着时间动态变化,教师教授的知识点个数也会越来越多.在这个动态的过程中,对学生展开的诊断评估被称为纵向的认知诊断测评(longitudinal CDA)^[4-6],收集到的数据是纵向诊断数据^[7],通过在教学过程中进行频繁且高效的评估^[8],可以为学生提供及时、丰富的信息反馈,从而改善学习效果^[9-10].

近几年,研究者对纵向CDA表现出了极大的兴

趣,构建了许多可以处理纵向诊断数据的模型^[11-14],这些研究将潜在转换分析(latent transition analysis, LTA)与常规的认知诊断模型(如DINA、DINO、LCDM等)相结合.也有许多研究者构建了基于高阶潜在结构^[15]的纵向认知诊断模型^[16-17].张颖等^[5]在重复测量中对考生进行诊断评估的可行性进行探讨,利用CDA与在线个性化指导结合,将诊断结果与为学生量身定制的指导联系起来.詹沛达等^[17]对纵向诊断模型做了详细的介绍,并通过模拟研究展示纵向诊断评估的分析过程及结果的解读.

横断CDA更侧重于对单个时间点(阶段)的评估,而纵向CDA具有学习—评估不断循环的特征.在纵向CDA中单次测验的测验长度较短,涉及到的属性个数较少,而多次测验涉及到的总属性个数相对较多.那么如何在较短的测试长度中不断更新学

收稿日期:2022-11-19

基金项目:教育部教育考试院“十四五”规划支撑专项课题(NEEA2021050),江西省社会科学基金(22JY16,21JY06),南昌市教育大数据智能技术重点实验室课题(2020-NCZDSY-012),江西省教育厅科技课题(GJJ2202013, GJJ2202018, GJJ212602, GJJ191691)和江西省教育科学“十四五”规划课题(21YB257, 21YB027)资助项目.

通信作者:喻晓锋(1980—),男,安徽太湖人,副教授,博士,主要从事心理统计与测量方面的研究. E-mail:xyu6@jxnu.edu.cn

生的属性掌握模式(又称知识状态, knowledge state, KS)并保持较高的判准率是一个较大的挑战。

对于在多属性情况下的 CDA,许多研究者提出了不同的策略来解决多属性所带来的诊断数据分析等问题。当属性个数较多且允许 KS 数超过考生人数时,Gu Yuqi 等^[19]提出了一种基于惩罚性似然法从所有的潜在类中选择一些必要的 KS,理论上该方法允许潜在类数量为无穷大;Sun Yan^[20]提出一种手风琴程序(accordion procedure, AP)来解决多属性情况下的 CDM 参数估计问题;Wang Juntao 等^[21]结合先验信息,考虑了在应用于 CDM 多属性情况下的序列吉布斯抽样法,该方法在多属性情况下,可以有效地降低常规 MCMC 算法的计算成本。这些方法的提出也为研究在多属性条件下的 CDA 提供了更多的思路,但是它们也存在应用条件和约束较多的缺点,如需要较大的样本量、对专家的经验有依赖或属于运算密集型算法等,在一定程度上限制了它们的广泛使用。

基于已有的研究所存在的一些不足,本文考虑 2 种实际的应用场景,提出利用先验信息在多属性诊断测验中的分类算法,充分利用从每个学习阶段的测验中获得的学生信息,提高学生在下一阶段测验中的分类判准率。从而实现在较短的测验中,更新学生在不同学习阶段的 KS,并维持较高的分类判准率。

1 多属性诊断测验的考生分类

为了方便介绍,下面对本文使用到的符号进行列表说明,见表 1。

表 1 符号含义汇总表

符号	含义
i	学生编号
j	作为下标时表示项目 j , 作为上标时表示项目 $1, 2, \dots, J$
(t)	用带括号的上标表示学习阶段 t
k_t	在阶段 t 中的属性个数
$a_t (l=1, 2, \dots, 2^{k_t})$	在阶段 t 中拥有的可能的 KS
K	总属性个数
2^K	总属性下的 KS 个数
α_T	总属性下的 KS(长度为 K)
y_{ij}	学生 i 在项目 j 上的得分
$Y_i^j = (y_{i1}, y_{i2}, \dots, y_{ij})$	学生 i 作答 j 个项目的作答向量
$Y_i^{(t)}$	学生 i 在阶段 t 的作答向量
$d = \{1, 2, \dots, D\}$	属性子集,又称为域

1.1 相关概念

属性掌握概率,它描述了考生对属性的掌握程度。考生在测验属性上的掌握概率构成了属性掌握概率向量。如某学生的属性掌握概率向量为 $(0.66, 0.78, 0.23)$,表示该学生在这 3 个属性中的掌握概率分别为 0.66、0.78 和 0.23。

知识状态,它是一个由 0 和 1 组成的二值向量,其中向量中的 1 表示学生掌握了该位置对应属性,0 表示未掌握该属性。知识状态是通过属性掌握概率转换得到的,一般采用 0.5 作为截止点^[22],上文提到的学生的知识状态为 $(1, 1, 0)$ 。

1.2 基于多属性诊断测验的考生分类

从理论上讲,CDM 可以估计的属性数量是不受限制的。但在实践中,过多的属性在对数据进行分析时计算资源的要求非常高,因此导致无法完成数据分析的工作,而且也可能导致识别性问题或不合理的参数估计问题^[23-24]。在这种情况下对考生进行准确的分类也会越困难,因为允许的 KS 数量会随着属性的数量增加呈指数级增长。通常纵向的 CDA 存在总属性个数较多的情况,因此在多属性条件下对考生进行分类研究也是近期众多学者关注的问题,但是仍然存在一些亟待解决的问题,如调节参数的设定、依赖专家信息和属性个数的限制等。

Gu Yuqi 等^[19]基于结构化的潜在属性模型(structured latent attribute model, SALM),在似然函数上添加惩罚实现多属性诊断测验的分类问题,从结果来看,该方法的表现受调节参数的影响较大。

Sun Yan^[20]提出一种适用于在 CDA 中出现的属性个数比较多的情况,通过将属性进行部分“折叠”和“扩展”来解决多属性数据下的参数估计和分类问题。在这个方法里,属性子集的选择需要人为界定,在实际测评中使用会受到一定的限制。

当教学内容需要进行多次重复学习和测试(如教学内容相对于学习者来说比较难)时,即在不同学习阶段学习和测试的属性集合相同,此时可以使用前一阶段所得的后验信息作为后一阶段的先验信息^[25]。即当阶段 t 与阶段 $t+1$ 测量的属性完全一致时,可以使用阶段 t 的后验信息作为阶段 $t+1$ 的先验信息。在现实应用中,更常见的情形是阶段 t 与阶段 $t+1$ 测量的属性个数不一致,此时不能直接利用阶段 t 的后验信息作为阶段 $t+1$ 的先验信息,因为这 2 个阶段的 KS“长度”不一致。在这种情况下,Wang Chun^[25]基于所有阶段考察的总属性来计算后验概率,这样一来,每个阶段考虑的 KS 的“长度”就

相同的,但是它存在的问题是总属性个数不能太多.

2 基于 RRUM 利用先验信息的多属性诊断测验分类

RRUM^[26]是受到广泛欢迎的诊断模型之一,具有非常好的测量特性,比如它考虑了部分掌握项目考察的属性对于题目正确作答的影响,相对于 DINA 模型能更深入细致探讨被试问题解决的加工过程;同时有研究表明该模型有较好的测量性能,被很多研究者所推崇^[27-29];相对于 GDINA 模型,它的参数更少,比如考察 3 个属性的题目,在 RRUM 模型下这个题目只有 4 个参数,而在 GDINA 模型下该题目有 8 个参数,在多属性 CDA 中简洁的参数有利于提高参数估计的效率.因此,本文基于 RRUM 模型来考虑利用先验信息的多属性诊断测验分类.

2.1 诊断模型 RRUM

在 RRUM 模型中,属性掌握模式为 α_i 的被试正确作答第 j 题的概率为

$$P_{ij} = P(Y_{ij} = 1 | \alpha_i, \pi_j^*, r_j^*) = \pi_j^* \prod_{k=1}^K r_{jk}^{*q_{jk}(1-\alpha_{ik})}, \quad (1)$$

其中 $\pi_j^* = \prod_{k=1}^K P(Y_{ij} = 1 | \alpha_{ik} = 1)^{q_{jk}}$, π_j^* 为基线参数,也叫难度参数,即掌握项目 j 所有考察属性的被试 i 作答正确该题目的概率. π_j^* 越小,表示项目 j 越难; $r_{jk}^* = \frac{P(Y_{ijk} = 1 | \alpha_{ik} = 0)}{P(Y_{ijk} = 1 | \alpha_{ik} = 1)}$ 表示未掌握属性 k 与掌握属性 k 答对项目 j 的概率之比,其值越小,说明掌握属性 k 对答对该题的重要性越大,它也被称为项目 j 中属性 k 的区分度参数.

2.2 属性掌握模式的更新

按照常规的横断 CDA 的方式对纵向 CDA 进行处理是不可行的,因为将所有的属性及对应的题目合在一起进行分析,一方面题目太多且是在不同学习阶段中完成的,考生的 KS 在这个过程中也发生了变化;另一方面属性个数多需要更多的计算资源(如计算机内存等).为探究在纵向 CDA 中面临的不同阶段属性集变化、总属性个数多的问题,本文基于 RRUM,在 2 种可能的应用情境下分别提出考生 KS 的更新方法.

2.2.1 情境 1: 相同属性、多次学习和测量的后验概率更新 在动态的教学过程中会有许多的学习阶段.当阶段 t 测量的属性集合与阶段 $t+1$ 完全一致时,可直接将在阶段 t 中获得的后验概率密度作

为在阶段 $t+1$ 中的先验信息,首先计算考生 i 在阶段 t 中作答 j 个项目后的后验概率

$$P_i^j(\alpha_i | Y_i^j) = \frac{L(Y_i^j | \alpha_i)P(\alpha_i)}{\sum_{\alpha_i=1}^{2^{k_i}} L(Y_i^j | \alpha_i)P(\alpha_i)}, \quad (2)$$

其中 $P(\alpha_i)$ 为先验信息, Y_i^j 是学生 i 对 j 个项目的反应向量.在初始阶段时,如果没有考生的先验信息,将 α_i 按均匀分布处理,即 $P(\alpha_i) = 1/2^{k_i}$.当 t 与 $t+1$ 这 2 个学习阶段测试的属性集合相同时,便可直接将式(2)所得的 $P_i^j(\alpha_i | Y_i^j)$ 作为 $t+1$ 阶段的先验信息,即 $P(\alpha_{i,t+1})$,这种情况对应的学习内容相对较难且重要,需要多次对同一内容进行教学和测量.

2.2.2 情境 2: 不同属性、多次学习和测量、基于预测先验的后验概率更新 与上一种情况不同,这里考虑的是 α_t 与 α_{t+1} 的长度不同的情形.为了应对这种情况,考虑基于阶段 t 获得的后验密度来预测阶段 $t+1$ 的先验信息,即利用考生在阶段 t 的属性掌握概率来预测其在包含新属性的阶段 $t+1$ 中 KS 的先验信息.首先,可以通过计算考生在阶段 t 的属性掌握概率

$$P(\alpha_{ik_t} | y_i^j) = \sum_{l=1}^{2^{k_t}} P(\alpha_l | y_i^j) \alpha_{lk_t}, \quad (3)$$

其中 $P(\alpha_{ik_t} | y_i^j)$ 为阶段 t 考生 i 掌握属性 k 的概率,未掌握概率则为 $1 - P(\alpha_{ik_t} | y_i^j)$. $P(\alpha_l | y_i^j)$ 为在作答了 j 个项目后,考生 i 属于第 l 种属性掌握模式的后验概率, α_{lk_t} 表示在阶段 t 中第 l 种属性掌握模式下属性 k 的取值,可以为 0 或 1.通过式(3)可获得考生 i 在阶段 t 测量的 k 个属性上的掌握与未掌握概率.进一步可通过计算在阶段 t 与阶段 $t+1$ 中考察过的相同属性的掌握概率同时发生的概率均值,来预测考生在阶段 $t+1$ 中加入新属性后 KS 的先验信息

$$P_i^{t+1}(\alpha_i) = \prod_{K=1}^{K_t} P(\alpha_{ik_t} | y_i^j)^{\alpha_{ik_t}} (1 - P(\alpha_{ik_t} | y_i^j))^{1-\alpha_{ik_t}} / 2^{k_{t+1}-k_t}. \quad (4)$$

下面以 1 个具体的例子进行说明.假设学生 i 在 t 阶段中共学习了 2 个属性(α_{i1}, α_{i2}),在阶段 $t+1$ 中不仅复习了前 2 个旧属性,还学习了 1 个新的属性(α_{i3}).如表 2 所示,就可以通过 t 阶段已考察过的 2 个属性所获得的属性掌握概率来预测学生 i 在阶段 $t+1$ 中的先验信息.此时,预测阶段 $t+1$ 先验信息的计算方式如表 3 所示.

表 2 学生 i 在 t 阶段中的属性掌握概率

属性	掌握概率	未掌握概率
α_{11}	0.6	0.4
α_{12}	0.3	0.7

表 3 学生 i 在 $t+1$ 阶段中预测先验概率计算

KS	预测先验概率
(0 0 0)	$(0.4 \times 0.7)/2 = 0.14$
(1 0 0)	$(0.6 \times 0.7)/2 = 0.21$
(0 1 0)	$(0.4 \times 0.3)/2 = 0.06$
(0 0 1)	$(0.4 \times 0.7)/2 = 0.14$
(1 1 0)	$(0.6 \times 0.3)/2 = 0.09$
(1 0 1)	$(0.6 \times 0.7)/2 = 0.21$
(0 1 1)	$(0.4 \times 0.3)/2 = 0.06$
(1 1 1)	$(0.6 \times 0.3)/2 = 0.09$

从上面的计算过程可以看出,这里计算预测先验概率的实质是一种基于已知信息来预测考生在所有考察属性上的掌握情况的方法,这一点和常规的基于贝叶斯的估计思想相同,这里采用增强先验信息来提高估计的准确性。

在更新考生 KS 在学习过程中的变化时,通过考生对每个属性的掌握概率能直接地反映其在当前学习阶段的能力值大小.因此,利用考生在先前阶段中的属性掌握概率预测考生在下一阶段中的先验信息不仅可以充分利用获得的所有考生信息,从而缩短测验长度,减轻学生负担,还能为下一阶段估计考生 KS 的分类精度做贡献。

3 参数估计方法与评价指标

下面介绍 2 种常见的属性掌握模式估计方法:极大后验估计和期望后验估计. A. Huebner 等^[22]的研究表明在 DINA 模型下它们的表现非常接近,但不知道它们在 RRUM 模型下是否仍然有相同的性质,因此,这是笔者同时考虑这 2 种估计方法的原因。

3.1 极大后验估计

极大后验估计(maximum a posteriori, MAP)的计算方法为

$$\hat{\alpha}_i(M_{AP}) = \arg \max P(\alpha_i = a_i | Y_i^t). \quad (5)$$

3.2 期望后验估计

EAP(expected a posteriori, EAP)的计算步骤,只有最后一步与 MAP 不同.它的最后一步不是选

出具有最大后验概率的 KS,而是将所有可能的 KS 与其对应的概率相乘后求期望值:

$$\hat{\alpha}_i(E_{AP}) = \sum_{l=1}^{2^K} \alpha_l P(\alpha_i = a_l | Y_i^t); \quad (6)$$

然后对在这个向量中的每个元素进行“四舍五入”式取整,由此将各元素都转化为二分取值^[30].转化后的向量就是该考生 KS 的估计值。

3.3 评价指标

考生分类的评价指标采用属性边际判准率(average attribute match ratio, AAMR)和模式判准率(pattern match ratio, PMR),它们的计算方法分别为

$$A_{AMR} = \sum_{i=1}^N \sum_{k=1}^K I(\alpha_{ik}, \hat{\alpha}_{ik}) / (NK), \quad (7)$$

$$P_{MR} = \sum_{i=1}^N I(\alpha_i, \hat{\alpha}_i) / N, \quad (8)$$

其中 N 为学生总人数; A_{AMR} 表示测验对单个属性的平均估计准确概率; α_i 表示学生的真实属性掌握模式; $\hat{\alpha}_i$ 为真实掌握模式的估计值; I 为示性函数,当学生的属性估计值 $\hat{\alpha}_{ik}$ 或属性掌握模式的估计值 $\hat{\alpha}_i$ 与真实属性 α_{ik} 或真实掌握模式 α_i 完全相等时,示性函数 I 为 1,否则为 0; P_{MR} 为在所有学生中 KS 被正确估计的学生比例,即每个学生的 KS 正确估计的概率。

4 模拟研究

为验证使用不同类型的先验信息对于分类判准率的影响,考虑了 2 种先验信息,分别是个体先验(individual prior)和群体先验(population prior).个体先验信息指的是在每个学习阶段中各学生的 KS 都会得到更新.因此,对于每个学生来说都是独一无二的,他们的个体先验信息便是自己在前一阶段测试中所得到的后验概率;群体先验信息指的是在每个学习阶段中每一种 KS 下的学生占总体学生的比例,对于每名学生来说,他们在每一种 KS 下的先验信息都是一致的.根据前面所提到的 2 种学习情境考虑 2 个模拟研究,每个学习情境分 3 个学习阶段,以下是具体研究设计。

4.1 研究 1:个体先验和群体先验在情境 1 下的属性分类

4.1.1 研究目的 本研究考虑最简单的情境 1,即学生在 3 个阶段下学习和测量的属性是一致的.通过更新先验信息分析其对诊断测验分类精度的影响。

4.1.2 研究设计 在研究 1 中,一共有 3 个操纵变量,分别是项目质量水平、属性间的相关程度、被试 KS 的估计方法.共考虑 3 个学习阶段(即 $t = 3$),每个阶段都考察了相同的 5 个属性,各阶段测验的长度都为 30,被试人数 $N = 1\ 000$.项目质量设置分为高、低 2 个水平,其中,在低质量项目条件下, r 的取值范围为 $(0.25, 0.40)$, π 的取值范围为 $(0.6, 0.8)$;高质量项目的 r 从 $(0.05, 0.20)$ 中抽取, π 从 $(0.8, 1.0)$ 中抽取^[31].通过多元正态分布(multivariate normal distribution, MVN)模拟属性之间存在低、中等、高 3 种相关水平,即被试的高阶潜在能力 $\theta \sim MVN(0, \Sigma)$,对角线处均为 1,非对角线处 ρ 分别为 0.3、0.5、0.7^[32-33].被试 KS 采用 MAP 和 EAP 这 2 种估计方法.每种实验条件重复 30 次,取 30 次的平均值作为最终结果.

4.2 研究 2:个体预测先验和群体预测先验在情境 2 下的属性分类

4.2.1 研究目的 本研究考虑的是比情境 1 复杂的情境,情境 1 可以看作是它的特殊情况,即学生在

3 个学习阶段下学习的知识点个数是不一致的,进一步验证利用先验信息能否在复杂场景中达到提高判准率的目的.

4.2.2 研究设计 在研究 2 中,除研究 1 操纵的 3 个条件外,每阶段所测属性个数和测试长度与之前研究设计不一致,其中,总属性个数为 8.参考一些已有研究中的设置^[15],在阶段 1 考察了 3 个属性,测验长度为 20;在阶段 2 考察了 5 个属性,测验长度为 30;在阶段 3 考察了 8 个属性,测验长度为 30.共设置了 2 个基线水平,基线水平是指无先验信息(先验为均匀分布)、属性个数为已测阶段的测验属性总和时学生属性、模式判准率的估计结果,将阶段 2、阶段 3 使用预测先验得到的分类精度结果分别与 2 个基线进行比较可知预测先验的方法是否有效.

4.3 研究结果

表 4~表 6 和图 1~图 2 分别展示了研究 1 在不同实验设计下的结果;表 7~表 8 和图 3~图 4 分别展示了研究 2 在不同实验设计下的结果.

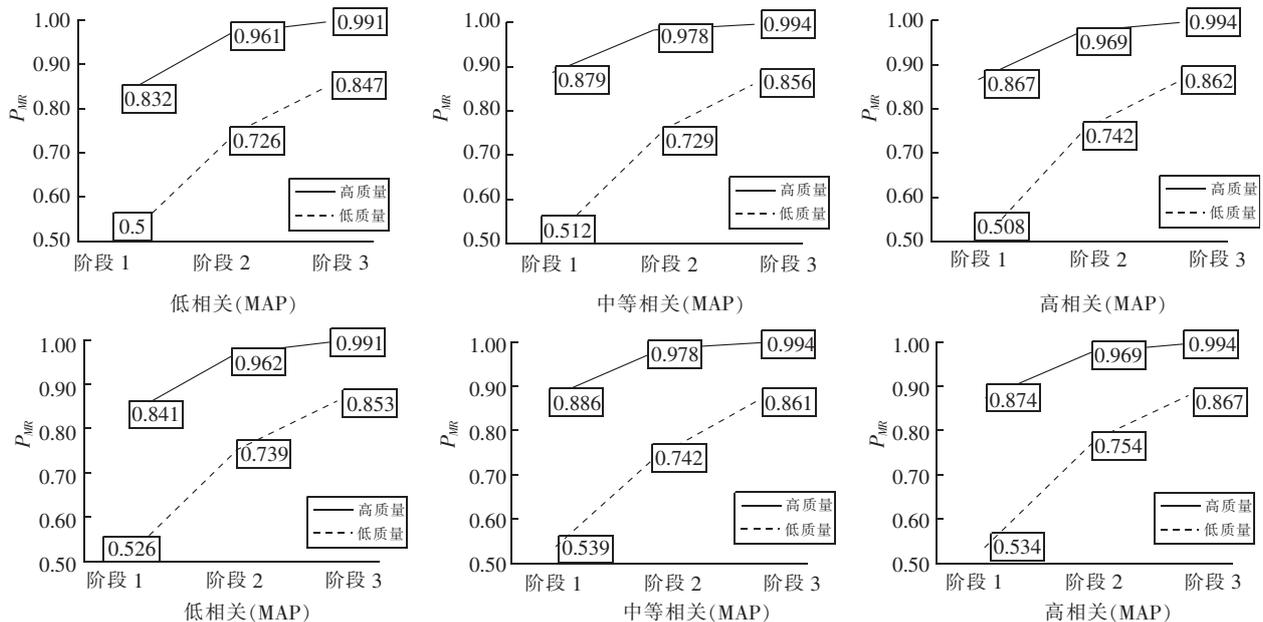


图 1 研究 1 使用个体在先验条件下 P_{MR} 值变化趋势图 ($K = 5$)

4.3.1 研究 1 结果 表 4~表 6 分别对应了属性之间存在低、中等、高相关情况下,在不同条件下属性与模式判准率的情况.根据表 4~表 6 与图 1~图 2 可以看出:从学习阶段来看,随着时间的推移,在 3 个学习阶段下的分类判准率均有了较大幅度的提高,如在 EAP 估计方法下,属性之间存在低相关关系,在项目质量低时,使用个体先验信息在阶段 3 相

较于阶段 1 的 A_{AMR} 值提高了 10% 左右, P_{MR} 值提高了 34% 左右,使用群体先验信息在阶段 3 相较于阶段 1 的 A_{AMR} 值提高了 1% 左右, P_{MR} 值提高了 3% 左右;在高质量项目下,使用个体先验信息的 A_{AMR} 值提高了 3% 左右, P_{MR} 值提高了 15% 左右,使用群体先验信息的 A_{AMR} 值提高了 1% 左右, P_{MR} 值提高了 3% 左右.

表 4 在模拟研究 1 中属性、模式判准率表(低相关)

属性 相关 程度	估计 方法	阶段 1 ($J=30$)				阶段 2 ($J=30$)				阶段 3 ($J=30$)				
		individual		Population		individual		Population		individual		Population		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
低	MAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.868	0.005	0.868	0.005	0.936	0.003	0.886	0.006	0.966	0.003	0.883	0.005
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.500	0.017	0.500	0.017	0.726	0.013	0.555	0.022	0.847	0.012	0.548	0.015
	EAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.878	0.005	0.878	0.005	0.940	0.004	0.890	0.007	0.968	0.003	0.889	0.004
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.526	0.018	0.526	0.018	0.739	0.014	0.564	0.024	0.853	0.013	0.560	0.013
高	MAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.963	0.004	0.963	0.004	0.992	0.002	0.972	0.003	0.998	0.001	0.972	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.832	0.018	0.832	0.018	0.961	0.007	0.870	0.012	0.991	0.003	0.874	0.011
	EAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.965	0.003	0.965	0.003	0.992	0.001	0.972	0.003	0.998	0.001	0.973	0.002
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.841	0.016	0.841	0.016	0.962	0.007	0.870	0.012	0.991	0.003	0.874	0.011

表 5 在模拟研究 1 中属性、模式判准率表(中等相关)

项目 质量	估计 方法	阶段 1 ($J=30$)				阶段 2 ($J=30$)				阶段 3 ($J=30$)				
		individual		Population		individual		Population		individual		Population		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
低	MAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.874	0.006	0.874	0.006	0.937	0.004	0.884	0.004	0.968	0.002	0.889	0.005
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.512	0.017	0.512	0.017	0.729	0.019	0.549	0.015	0.856	0.011	0.562	0.016
	EAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.883	0.006	0.883	0.006	0.941	0.004	0.888	0.005	0.970	0.002	0.894	0.005
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.539	0.016	0.539	0.016	0.742	0.016	0.558	0.015	0.861	0.011	0.573	0.016
高	MAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.974	0.003	0.974	0.003	0.995	0.001	0.976	0.003	0.999	0.001	0.977	0.002
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.879	0.012	0.879	0.012	0.978	0.005	0.892	0.013	0.994	0.003	0.893	0.010
	EAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.976	0.002	0.976	0.002	0.996	0.001	0.977	0.003	0.999	0.001	0.977	0.002
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.886	0.011	0.886	0.011	0.978	0.005	0.892	0.014	0.994	0.003	0.892	0.009

表 6 在模拟研究 1 中属性、模式判准率表(高相关)

项目 质量	估计 方法	阶段 1				阶段 2				阶段 3				
		Person		Population		Person		Population		Person		Population		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
低	MAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.872	0.005	0.872	0.005	0.940	0.003	0.890	0.006	0.969	0.002	0.888	0.006
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.508	0.014	0.508	0.014	0.742	0.012	0.571	0.019	0.862	0.009	0.562	0.020
	EAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.881	0.005	0.881	0.005	0.944	0.003	0.894	0.005	0.971	0.002	0.894	0.006
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.534	0.017	0.534	0.017	0.754	0.013	0.580	0.017	0.867	0.009	0.575	0.019
高	MAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.971	0.003	0.971	0.003	0.994	0.001	0.974	0.003	0.999	0.000	0.978	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.867	0.014	0.867	0.014	0.969	0.006	0.879	0.013	0.994	0.002	0.899	0.012
	EAP	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.973	0.003	0.973	0.003	0.994	0.001	0.974	0.003	0.999	0.000	0.978	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.874	0.012	0.874	0.012	0.969	0.006	0.879	0.012	0.994	0.002	0.899	0.013

表 7 模拟研究 2 中属性、模式判准率表 ($K=8$) EAP

项目 质量	学习 阶段	低质量						高质量						
		Baseline		Individual		Population		Baseline		Individual		Population		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
低 相 关	阶段 1	$A_{AMR}(\alpha_1 \sim \alpha_3)$		0.886	0.005	0.886	0.005			0.974	0.004	0.974	0.004	
		$P_{MR}(\alpha_1 \sim \alpha_5)$		0.703	0.015	0.703	0.015			0.925	0.010	0.925	0.010	
	阶段 2	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.880	0.005	0.916	0.004	0.884	0.005	0.973	0.003	0.985	0.002	0.974	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.533	0.018	0.649	0.018	0.545	0.019	0.872	0.012	0.927	0.011	0.880	0.011
	阶段 3	$A_{AMR}(\alpha_1 \sim \alpha_8)$	0.831	0.008	0.911	0.005	0.846	0.005	0.949	0.003	0.984	0.002	0.969	0.003
		$P_{MR}(\alpha_1 \sim \alpha_8)$	0.232	0.025	0.475	0.020	0.268	0.018	0.655	0.020	0.878	0.018	0.781	0.020
中 等 相 关	阶段 1	$A_{AMR}(\alpha_1 \sim \alpha_3)$		0.876	0.005	0.876	0.005			0.964	0.004	0.964	0.004	
		$P_{MR}(\alpha_1 \sim \alpha_3)$		0.681	0.014	0.681	0.014			0.900	0.010	0.900	0.010	
	阶段 2	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.875	0.005	0.926	0.004	0.892	0.004	0.972	0.003	0.988	0.002	0.981	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.515	0.017	0.683	0.014	0.572	0.016	0.870	0.014	0.944	0.010	0.909	0.013
	阶段 3	$A_{AMR}(\alpha_1 \sim \alpha_8)$	0.844	0.007	0.916	0.004	0.852	0.005	0.952	0.003	0.984	0.002	0.964	0.002
		$P_{MR}(\alpha_1 \sim \alpha_8)$	0.260	0.021	0.494	0.019	0.284	0.016	0.677	0.019	0.880	0.013	0.749	0.015
高 相 关	阶段 1	$A_{AMR}(\alpha_1 \sim \alpha_3)$		0.879	0.006	0.879	0.006			0.966	0.004	0.966	0.004	
		$P_{MR}(\alpha_1 \sim \alpha_3)$		0.683	0.015	0.683	0.015			0.906	0.011	0.906	0.011	
	阶段 2	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.883	0.005	0.914	0.004	0.880	0.004	0.973	0.002	0.985	0.003	0.976	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.540	0.017	0.638	0.014	0.528	0.014	0.872	0.012	0.925	0.012	0.886	0.014
	阶段 3	$A_{AMR}(\alpha_1 \sim \alpha_8)$	0.842	0.005	0.909	0.003	0.849	0.005	0.957	0.004	0.986	0.002	0.964	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.258	0.014	0.469	0.016	0.275	0.016	0.709	0.027	0.892	0.012	0.749	0.018

表 8 在模拟研究 2 中属性、模式判准率表 ($K=8$) MAP

项目 质量	学习 阶段	低质量						高质量						
		Baseline		Individual		Population		Baseline		Individual		Population		
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
低 相 关	阶段 1	$A_{AMR}(\alpha_1 \sim \alpha_3)$		0.967	0.003	0.967	0.003			0.972	0.004	0.972	0.004	
		$P_{MR}(\alpha_1 \sim \alpha_3)$		0.910	0.008	0.910	0.008			0.922	0.011	0.922	0.011	
	阶段 2	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.870	0.005	0.912	0.005	0.877	0.006	0.970	0.004	0.985	0.003	0.973	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.504	0.017	0.635	0.017	0.526	0.018	0.864	0.016	0.925	0.013	0.877	0.014
	阶段 3	$A_{AMR}(\alpha_1 \sim \alpha_8)$	0.821	0.008	0.909	0.006	0.842	0.006	0.955	0.005	0.983	0.003	0.968	0.003
		$P_{MR}(\alpha_1 \sim \alpha_8)$	0.205	0.024	0.465	0.026	0.258	0.020	0.638	0.022	0.873	0.022	0.777	0.022
中 等 相 关	阶段 1	$A_{AMR}(\alpha_1 \sim \alpha_3)$		0.876	0.005	0.876	0.005			0.962	0.004	0.962	0.004	
		$P_{MR}(\alpha_1 \sim \alpha_3)$		0.681	0.014	0.681	0.014			0.896	0.011	0.896	0.011	
	阶段 2	$A_{AMR}(\alpha_1 \sim \alpha_5)$	0.866	0.005	0.922	0.004	0.885	0.005	0.969	0.004	0.988	0.002	0.980	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$	0.490	0.018	0.673	0.016	0.553	0.019	0.859	0.016	0.943	0.012	0.907	0.014
	阶段 3	$A_{AMR}(\alpha_1 \sim \alpha_8)$	0.834	0.008	0.913	0.005	0.848	0.006	0.950	0.003	0.984	0.002	0.963	0.002
		$P_{MR}(\alpha_1 \sim \alpha_8)$	0.231	0.024	0.478	0.027	0.270	0.021	0.666	0.019	0.876	0.016	0.745	0.016

表 8(续)

项目	学习阶段	低质量						高质量							
		Baseline		Individual		Population		Baseline		Individual		Population			
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
高相关	阶段 1	$A_{AMR}(\alpha_1 \sim \alpha_3)$		0.864	0.006	0.864	0.006	0.965	0.004	0.965	0.004	0.965	0.004		
		$P_{MR}(\alpha_1 \sim \alpha_3)$		0.652	0.020	0.652	0.020	0.904	0.013	0.904	0.013	0.904	0.013		
	阶段 2	$A_{AMR}(\alpha_1 \sim \alpha_5)$		0.874	0.006	0.910	0.004	0.874	0.004	0.970	0.003	0.984	0.003	0.974	0.003
		$P_{MR}(\alpha_1 \sim \alpha_5)$		0.514	0.018	0.628	0.015	0.513	0.014	0.861	0.016	0.922	0.014	0.880	0.018
	阶段 3	$A_{AMR}(\alpha_1 \sim \alpha_8)$		0.835	0.006	0.907	0.003	0.845	0.005	0.955	0.005	0.985	0.002	0.963	0.003
		$P_{MR}(\alpha_1 \sim \alpha_8)$		0.238	0.020	0.458	0.020	0.266	0.019	0.698	0.031	0.890	0.014	0.743	0.020

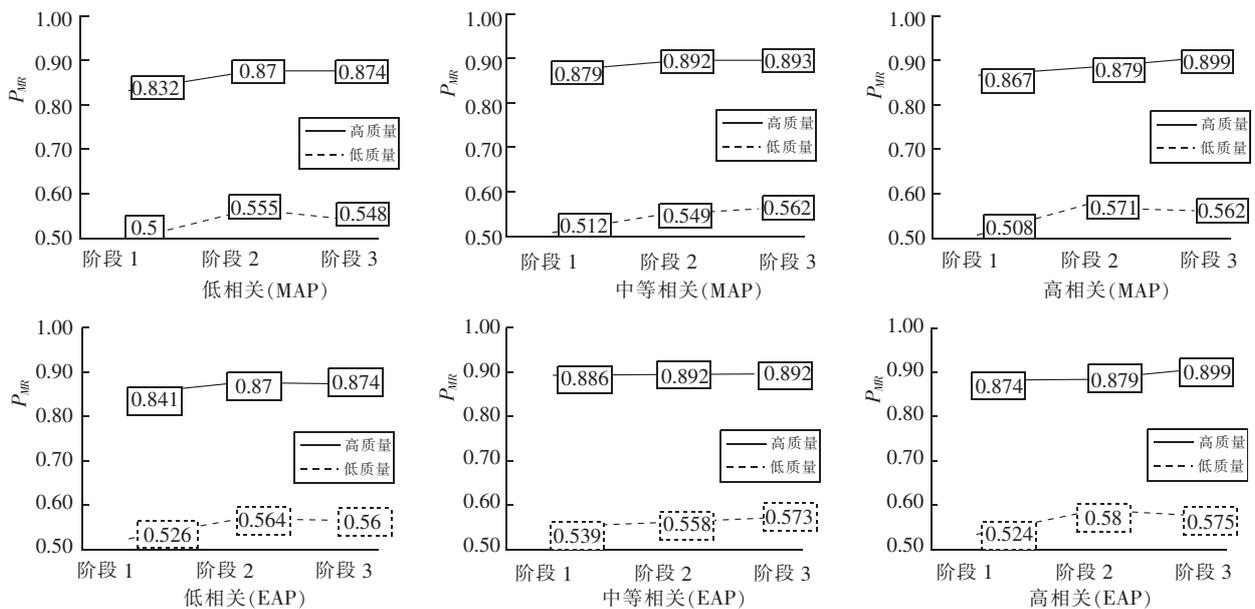


图 2 研究 1 使用群体在先验条件下 P_{MR} 值变化趋势图 ($K=5$)

从使用的研究估计方法来看,使用 EAP 比使用 MAP 会带来更高的判准率,如阶段 1 在低项目质量中使用个体先验信息的情况下,EAP 比 MAP 的 A_{AMR} 相差 1% 左右, P_{MR} 值相差了 2% 左右,随着项目质量的提高,2 种估计方法的判准率相差越小,但在总体情况下,EAP 的判准率均略高于 MAP.

从使用的先验信息来看,使用个体先验信息比

使用群体先验信息在 3 个学习阶段下的判准率提升得更多.如在表 9 中,当项目质量低、参数估计方法为 EAP 时,使用个体先验信息在阶段 3 比阶段 1 的 A_{AMR} 值提高了 19% 左右, P_{MR} 值提高了 33% 左右,使用群体先验信息在阶段 3 比阶段 1 的 A_{AMR} 值提高了 2% 左右, P_{MR} 值提高了 4% 左右.

表 9 实证数据的 Q 矩阵

项目	1	2	3	4	5	6	7	8	9	10	11	12	12	13	14	15	16	17	18	19	20	21	22	23	24	25
α_1	1	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1
α_2	1	0	1	0	1	0	0	1	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1
α_3	0	0	0	1	1	0	0	1	0	0	1	0	0	0	1	1	0	1	1	1	1	0	1	0	0	0
α_4	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
α_5	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
α_6	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
α_7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
α_8	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0

表 9(续)

项目	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
α_9	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
α_{10}	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
α_{11}	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
α_{12}	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
α_{13}	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0
α_{14}	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0
α_{15}	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0

总的来说,在情境 1 中,随着学习阶段的深入,使用个体先验与群体先验均会提高考生的属性、模式判准率,其中在使用个体先验的情况下对判准率的提高更有优势.使用 EAP 估计方法比使用 MAP 估计方法也会带来更好的属性、模式判准率,且这 2 种估计方法的估计标准差都较小.

4.3.2 研究 2 结果 根据表 7~表 8 与图 3~图 4 可以看出:从学习阶段来看,其研究结果与研究 1 结果一致.如在表 7 中,属性间存在低相关时,随着时

间的推移,由于考察属性个数越来越多,判准率受属性个数的影响存在下降趋势.然而,与基线水平相比较而言,使用先验信息仍能使判准率优于基线水平.如当属性间相关程度与项目质量均较低时,使用 EAP 估计方法在阶段 2 与基线水平一相比较而言 A_{AMR} 值提高 4% 左右, P_{MR} 值提高了 11% 左右,阶段 3 与基线水平 2 相较而言 A_{AMR} 值提高 8% 左右, P_{MR} 值提高了 24% 左右.

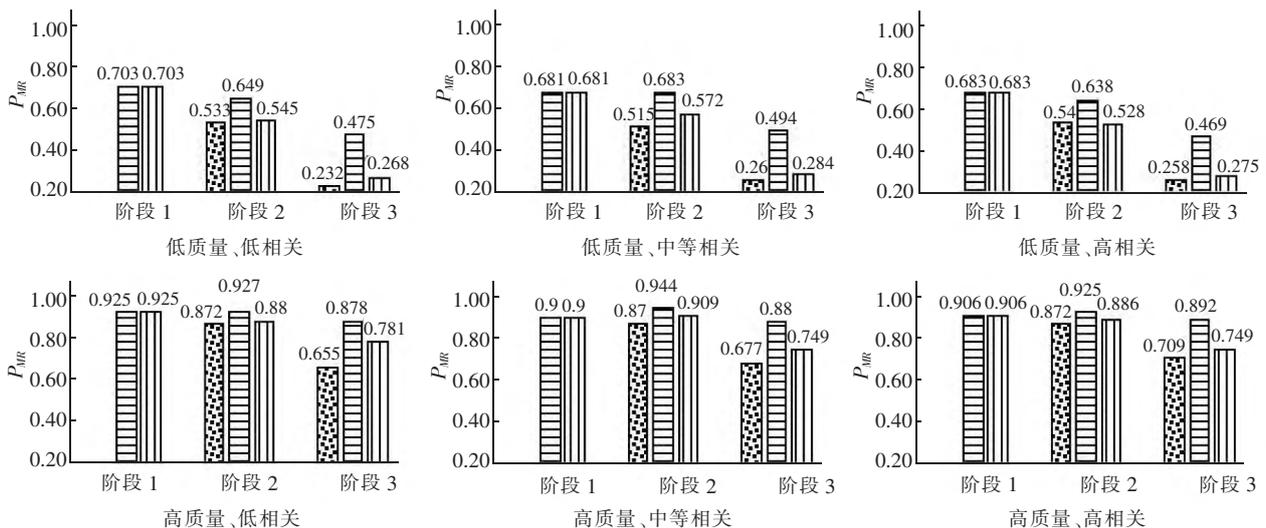


图 3 研究 2 使用 EAP 估计方法下 P_{MR} 值变化趋势图

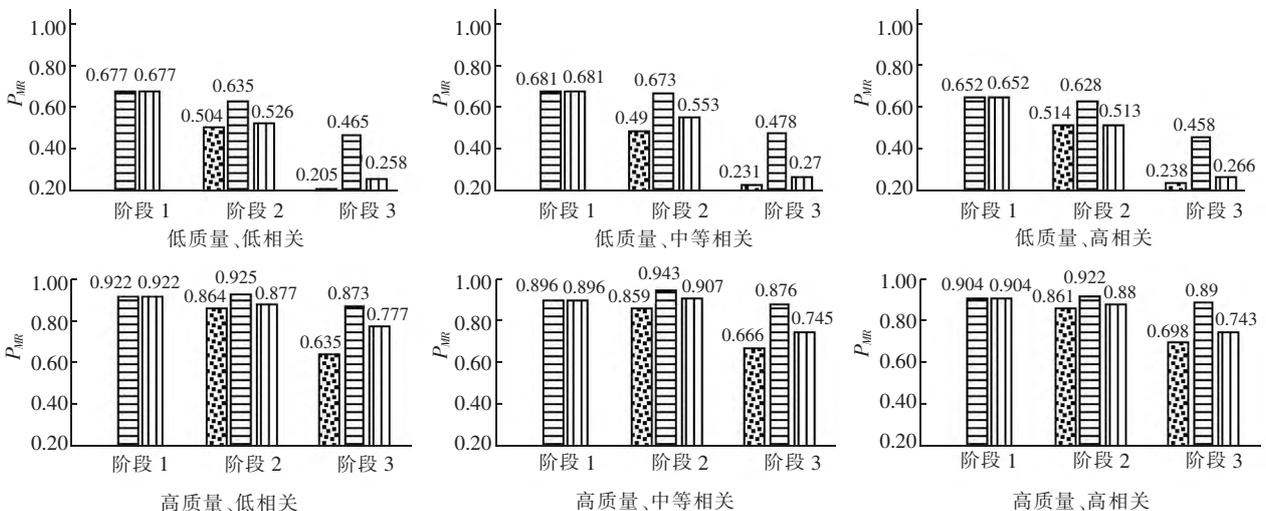


图 4 研究 2 在使用 MAP 估计方法下 P_{MR} 值变化趋势图

从使用的研究方法来看,EAP方法比MAP方法会带来更高的判准率,项目质量和属性间相关均低时,EAP方法比MAP方法的 A_{AMR} 提高了1%左右, P_{MR} 值提高了3%左右.从属性间相关程度来看,可以发现在项目质量高的条件下使用预测先验的方法,随着属性间相关程度的提高会表现得更好,如在项目质量高时使用EAP估计方法的情况下,在阶段3属性相关程度为高时使用个体先验条件下 A_{AMR} 比属性低相关条件下的 A_{AMR} 值高了0.2%左右, P_{MR} 高了1.5%左右.

总的来说,基于预测先验、不同内容、多次学习和测量的情境2中,使用预测的个体、群体先验信息仍能保持与情境1相同的结果.在复杂的应用情景中也能取得较好的结果.

5 实证数据分析

5.1 数据说明

实证数据来源于TIMSS 2007年2 619名四年级学生的数学测验^[20].该测验从数、几何图形和测量以及数据的表达3个知识领域对学生进行考察,3个知识领域分别测量了8、4、3个属性,共包含25道题目,15个属性.该数据具有测验长度短、总属性个数较多的特点,因此可以将学生在前5道测量了8个属性(知识领域:数)的题目作答情况作为时间点1的信息,来预测学生在后20道测量了15个属性(知识领域:数、几何图形和测量、数据的表达)的题目上的掌握情况,即时间点2的先验信息.使用RRUM模型、个体和群体2种先验信息、MAP和

EAP这2种估计方法分析该数据,并计算2个时间点下,每种方法下学生不同属性上的掌握占比.

5.2 实证结果

表10呈现了不同方法下属性的掌握占比随时间的发展趋势.首先,整体而言,学生不同属性下有不同的发展趋势.其次,不同的先验信息和估计方法对部分属性变化趋势的追踪略有差异.具体而言,对属性2、属性5、属性6、属性8,使用不同先验在2种估计方法下的追踪结果较为一致;而对于属性1、属性3、属性4、属性7,使用不同先验在2种估计方法下的追踪结果存在一定差异,这与属性测量次数的多少存在一定相关,如属性2和属性5在时间点1测量次数较多,因此,这些属性在时间点1与时间点2间的差异较小.另外,在时间点1使用2种估计方法所得结果差异较大时,在时间点2的这种估计方法所得结果之间差异将会缩小,如属性2和属性7.对于时间点2出现的新属性,大部分属性在不同方法下所得分析结果一致,属性13、属性14在不同方法下所得结果存在略微差异.遗憾的是,由于实证数据中无法获得学生的属性掌握真值,无法像模拟研究一样计算不同方法下学生的分类准确性,因此,无法比较不同方法在追踪数据时能对分类准确性带来多大程度上的提高.总之,实证研究表明结合先验信息的方法可用于分析属性个数多,且相邻时间点所测属性个数不一致的情况,利用前一时间点所获得的信息可以丰富后一时间点的先验信息,为分类准确性提供更可靠的学生信息.

表10 属性的掌握占比随时间发展趋势

项目	阶段1($J=5, K=8$)				阶段2($J=20, K=15$)			
	individual		Population		individual		Population	
	MAP	EAP	MAP	EAP	MAP	EAP	MAP	EAP
α_1	0.754	0.754	0.754	0.754	0.308	0.308	0.313	0.313
α_2	0.640	0.593	0.640	0.593	0.612	0.584	0.618	0.593
α_3	0.552	0.552	0.552	0.552	0.670	0.669	0.669	0.669
α_4	0.298	0.298	0.298	0.298	0.506	0.506	0.354	0.354
α_5	0.390	0.390	0.390	0.390	0.390	0.390	0.374	0.374
α_6	0.552	0.552	0.552	0.552	0.549	0.545	0.374	0.374

表 10(续)

项目	阶段 1($J=5, K=8$)				阶段 2($J=20, K=15$)			
	individual		Population		individual		Population	
	MAP	EAP	MAP	EAP	MAP	EAP	MAP	EAP
α_7	0.000	0.472	0.000	0.472	0.188	0.188	0.188	0.188
α_8	0.593	0.593	0.593	0.593	0.595	0.595	0.595	0.595
α_9					0.102	0.102	0.102	0.102
α_{10}					0.726	0.726	0.726	0.726
α_{11}					0.465	0.465	0.465	0.465
α_{12}					0.568	0.568	0.568	0.568
α_{13}					0.685	0.690	0.685	0.693
α_{14}					0.480	0.462	0.480	0.474
α_{15}					0.786	0.786	0.786	0.786

6 结论和讨论

本文通过对以往利用先验信息进行诊断测验分类的研究回顾发现已有方法在实际应用中受限较多. 基于此, 本文从已有的方法出发, 在 2 种测验情境下分别考察 2 种不同利用先验信息的方法. 已有研究^[22]表明: MAP 和 EAP 在 DINA 模型下的表现非常接近, 在 RRUM 模型下它们的表现如何也是未知的. 因此, 为验证使用先验信息在多属性诊断测验分类研究中的有效性, 再将项目质量与属性相关程度作为变量后开展了 2 项模拟研究, 分别对应了简单到复杂的 2 种实际应用情境, 最后通过实证研究展示了如何结合先验信息应用于多属性诊断测验.

研究表明: 基于预测先验信息的方法在多属性诊断测验中能起到提高分类准确性的作用, 尤其是在使用个体先验信息的条件下, 其表现优于基线水平. 实际上, 将在阶段 t 中所获得的历史信息和在阶段 $t+1$ 中的作答数据结合起来, 丰富了阶段 $t+1$ 的先验信息, 从而能达到提高分类判准率的效果. 而在基于个体先验、群体先验、基线水平 3 种条件下, 使用个体先验可以给阶段 $t+1$ 提供最丰富的先验信息. 因此, 更新个体先验信息有利于分类判准率的提高. 在 2 种估计方法下, 无论项目质量和属性相关程度的高低, EAP 估计方法均比 MAP 估计方法表现得更好. 其中 A_{AMR} 值差距较小, P_{MR} 值差距较

大, 因此在 RRUM 模型下, 使用 EAP 估计方法会带来更高的属性分类判准率.

本文验证了结合先验信息提高诊断测验的分类准确率, 而以往利用先验信息的诊断测验分类研究在实际的应用场景中会受到较大限制, 一般限制每个学习阶段测量属性是一致的, 本文提出一种预测先验信息的方法, 既能提高诊断分类的精度, 同时也能扩大结合先验信息在诊断测验分类研究中的应用范围.

虽然本研究均取得了较好的结果, 但仍存在一些不足之处. 首先, 学生在每个学习阶段间的 KS 是会发生变化的, 如对知识点有了进一步的学习或遗忘, 结合学习模型可以研究更接近实际应用的场景, 如 Chen Yinghan 等^[12]提出了用于追踪学习轨迹的 1 阶隐马尔可夫模型. 其次, 本文使用的预测先验的方法是基于相邻学习阶段中所测量的属性具有部分重合, 未来可以尝试当将该方法拓展应用在更多不同的应用情景. 另外, 在多属性情况下, 除了利用先验信息可以提高判准率, 同样可以尝试考虑结合 Sun Yan 等^[34]提出的一种适用于多属性下的潜在 4 步回归法. 最后, 在模型选择上, 本文仅考虑了 RRUM 模型, Zhan Peida 等^[16]和 Huang Hungyu^[15]利用了高阶潜在结构模型来建立纵向的 CDM, 在将来可以考虑使用一些纵向的 CDM 来进行研究. 很多已有研究都表明结合反应时的诊断测验可以更准确地测量被试^[35], 结合反应时的诊断测验也是未来需要深入探讨的方向.

7 参考文献

- [1] LEIGHTON J P, GIERL, M J, HUNKA, S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuoaka's rule-space approach [J]. *Journal of Educational Measurement*, 2004, 41(3):205-237.
- [2] 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):191-210.
- [3] VON DAVIER M, LEE Y S. *Handbook of diagnostic classification Models* [M]. New York: Springer, 2019.
- [4] 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):369388.
- [5] 张颖, 边玉芳. 探索认知诊断研究的新思路: 追踪研究中的诊断分析 [J]. *考试研究*, 2017, 13(5):72-77.
- [6] 詹沛达, 潘艳方, 李菲著. 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.
- [7] LEE S Y. *Growth curve cognitive diagnosis models for longitudinal assessment* [D]. Berkeley: University of California, 2017.
- [8] KINGSBURY G, FREEMAN E, NESTERAK M. The potential of adaptive assessment [J]. *Educational Leadership*, 2014, 71(6):18-19.
- [9] HANNA G S. Effects of total and partial feedback in multiple-choice testing upon learning [J]. *The Journal of Educational Research*, 1976, 69(5):202-205.
- [10] KLUGER A N, DENISI A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory [J]. *Psychological Bulletin*, 1996, 119(2):254-284.
- [11] LI Feiming, COHEN A, BOTTGE B, et al. A latent transition analysis model for assessing change in cognitive skills [J]. *Educational and Psychological Measurement*, 2016, 76(2):181-204.
- [12] 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.
- [13] WANG Shiyu, YAN Yang, 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.
- [14] ZHANG Susu, WANG Shiyu. Modeling learner heterogeneity: a mixture learning model with responses and response times [J]. *Frontiers in Psychology*, 2018, 9:2339.
- [15] HUANG Hungyu. Multilevel cognitive diagnosis models for assessing changes in latent attributes [J]. *Journal of Educational Measurement*, 2017, 54(4):440-480.
- [16] ZHAN Peida, HONG Jiao, LIAO Dandan, et al. A longitudinal higher-order diagnostic classification model [J]. *Journal of Educational and Behavioral Statistics*, 2019, 44(3):251-281.
- [17] DE LA Torre J, DOUGLAS J A. Higher-order latent trait models for cognitive diagnosis [J]. *Psychometrika*, 2004, 69(3):333-353.
- [18] WU Hueymin. Online individualised tutor for improving mathematics learning: a cognitive diagnostic model approach [J]. *Educational Psychology*, 2018, 39(10):1218-1232.
- [19] GU Yuqi, XU Gong Jun. Learning attribute patterns in high-dimensional structured latent attribute models [J]. *Journal of Machine Learning Research*, 2019, 20(1):1-58.
- [20] SUN Yan. *Strategies for addressing high-dimensional cognitively diagnostic assessment problems* [D]. New Jersey: Rutgers University, 2019.
- [21] WANG Juntao, SHI Ningzhong, ZHANG Xue, et al. Sequential gibbs sampling algorithm for cognitive diagnosis models with many attributes [J]. *Multivariate Behavioral Research*, 2021, 57(5):1-37.
- [22] HUEBNER A, Wang Chun. A note on comparing examinee classification methods for cognitive diagnosis models [J]. *Educational and Psychological Measurement*, 2011, 71(2):407-419.
- [23] HUEBNER A. An overview of recent developments in cognitive diagnostic computer adaptive assessments [J]. *Practical Assessment*, 2010, 15(3):1-7.
- [24] VON DAVIER M. A general diagnostic model applied to language testing data [J]. *British Journal Mathematical and Statistical Psychology*, 2008, 61(2):287-307.
- [25] WANG Chun. On interim cognitive diagnostic computer-

- ized adaptive testing in learning context [J]. *Applied Psychological Measurement*, 2021, 45(4):235-252.
- [26] HARTZ S M. A Bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality [D]. Urbana Champaign: University of Illinois, 2002.
- [27] 戴步云,张敏强,焦璨,等. 基于 CD-CAT 的多策略 RRUM 模型其选题方法开发 [J]. *心理学报*, 2015, 47(12):1511-1519.
- [28] CULPEPPER S A, Hudson A. An improved strategy for Bayesian estimation of the reduced reparameterized unified model [J]. *Applied psychological measurement*, 2018, 42(2):99-115.
- [29] KIM Y H. Diagnosing EAP writing ability using the reduced reparameterized unified model [J]. *Language Testing*, 2011, 28(4):509-541.
- [30] FENG Yuling, HABING B T, HUEBNER A. Parameter estimation of the reduced RUM using the EM Algorithm [J]. *Applied Psychological Measurement*, 2014, 38(2):137-150.
- [31] SUN Xiaojian, LIU Yanlou, XIN Tao, et al. The impact of item calibration error on variable-length cognitive diagnostic computerized adaptive testing [J]. *Frontiers in Psychology*, 2020, 11:575141.
- [32] HENSON R, DOUGLAS J. Test construction for cognitive diagnosis [J]. *Applied Psychological Measurement*, 2005, 29(4):262-277.
- [33] SINHARAY S. How often do subscores have added value? Results from operational and simulated data [J]. *Journal of Educational Measurement*, 2010, 47(2):150-174.
- [34] SUN Yan, DE LA TORRE J. Innovative psychometric modeling and methods: Improving attribute classification accuracy in high dimensional data: A four-step latent regression approach [M]. Charlotte: Information Age, 2020.
- [35] MARIANT S, FOX J P, AVETISYAN M, et al. Testing for aberrant behavior in response time modeling [J]. *Journal of Educational and Behavioral Statistics*, 2014, 39(6):426-451.

The Classification for High-Dimensional Cognitively Diagnostic Assessment Based on Prior Information

YE Ziyu¹, QIN Chunying^{1,2}, YANG Jianqin², YU Xiaofeng^{1,3*}, FU Daoxuan¹

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

2. School of Mathematics and Information Science, Nanchang Normal University, Nanchang Jiangxi 330022, China;

3. Key Laboratory of Psychology and Cognition Science of Jiangxi Province, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: In a dynamic learning process, the student's knowledge state changes dynamically as learning progresses and the number of knowledge attributes increases. In such a scenario, it is challenging to combine a priori information to improve the accuracy of diagnostic assessments. The probability of student mastery of learned attributes is used to predict the prior information of the attribute vector after the inclusion of new attributes. Two realistic application contexts are considered and the performance of the method is evaluated through simulations and empirical studies. The results show that combining a priori information in two real-world learning contexts can be useful in improving accuracy in high-dimensional attribute diagnostic tests, with better performance when based on individual a priori information. The method based on predictive a priori information also breaks away from the limitations of previous studies in terms of the number of attributes that are consistent across different learning stages of the test, resulting in a greater improvement in classification accuracy. The empirical data analysis further demonstrates the applied value of the methods used in the methodology.

Key words: prior information; diagnostic assessment; high-dimensional attributes; longitudinal diagnostic assessment

(责任编辑:冉小晓)