

Q 矩阵理论探微

丁树良, 罗 芬, 汪文义, 熊建华

(江西师范大学 计算机信息工程学院 江西 南昌 330022)

摘要: Tatsuoka 认为 DINA 模型没有使用 Q 矩阵理论, 这个断言是否成立? 为了回答这个问题, 必须讨论 Q 理论的外延。Q 矩阵理论至少应该包括如何计算知识状态集合和理想反应模式集合、如何准确标识 Q 矩阵以及如何设计认知诊断测验蓝图等内容, 由此导出 DINA 使用了 Q 矩阵理论的结论, 并对元素为 0-1 的 Q 矩阵给出属性不可补偿条件下 0-1 评分优良认知诊断测验蓝图设计的定量描述, 并对今后研究方向进行相关的讨论。

关键词: Q 矩阵; Q 矩阵理论; 认知诊断; 测验蓝图设计

中图分类号: G449 **文献标识码:** A **文章编号:** 1000-579(2017)01-0071-09

Exploration of Q - matrix Theory

DING Shuliang, LUO Fen, WANG Wenyi, XIONG Jianhua

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

Abstract: Tatsuoka says that DINA model does not use the Q - matrix theory. Is it true or false? To answer this question, the extension of the Q - matrix theory must be discussed. The Q - matrix theory should at least include such methods as computing knowledge status collection and ideal response pattern collection, accurate identifying of the Q - matrix, design cognition diagnosis examination blueprint etc., so as to derive the conclusion for DINA to use the Q matrix theory. The paper gives a quantitative description of 0 - 1 grading fine cognition diagnosis examination blueprint design under the condition of non - compensatory attribute, and discusses future research directions.

Key words: Q - matrix; Q - matrix theory; cognition diagnosis; examination blueprint design

1 前言

Tatsuoka 是认知诊断方面的先行者, 在认知诊断方面做了开创性的工作, 创立了规则空间模型 (Rule Space Model, RSM), 发表一系列文章, 出版了相关专著, 给出了一系列的术语、定义, 特别使用了 Q 矩阵这个数学工具并且给出 Q 矩阵理论^{[1][2]}, 提出了一系列值得研究的问题, 并进行许多实证研究。Tatsuoka 认为认知诊断模型—确定性输入噪声与门模型 (deterministic input noisy AND - gate model, DINA)^[3] 只是应用偏系关系而没有使用 Q 矩阵理论。^{[2][16]} 然而, DINA 模型是不是真的没有使用 Q 矩

收稿日期: 2016 - 04 - 05

基金项目: 国家自然科学基金项目(编号: 31360237, 31500909, 31160203, 30860084); 国家社会科学基金项目(编号: 16BYY096); 全国教育科学规划教育部重点课题(编号: DHA150285); 教育部人文社会科学研究青年基金项目(编号: 13YJC880060); 江西省自然科学基金项目(编号: 20161BAB212044); 江西省教育科学 2013 年度一般课题(编号: 13YB032); 江西省教育厅科技计划项目(编号: GJJ13207, GJJ13208)

作者简介: 丁树良(1949 -), 男, 江西樟树人, 江西师范大学教授。研究方向为教育和心理测量。

阵理论?这就得研究 Q 矩阵理论到底包含些什么内容,本文试图对此进行讨论。以前曾经有研究揭示 Tatsuoka 的理论存在不足,^{[4][5]} 本文将进一步讨论和推广该理论。

1.1 Q 矩阵及 Q 矩阵理论

Tatsuoka 引入元素为 0 或 1 的 Q 矩阵表达属性(知识、技能、甚至策略,等等)和题目之间的关系。^{[1][2]} 因为有了 Q 矩阵,才能够清楚测验项目测查的属性,认知诊断测验才能够提供被试认知方面的缺陷和优势,而不是一个经典测量理论的总分或者项目反应理论中能力那样的概括性的指标值。

本文中 Q 矩阵行向量表示属性,列向量表示题目。然而 Q 矩阵的列也可表示被试的知识状态(属性掌握模式),如果将被试抽象为知识状态,则 Q 矩阵既是属性和项目的关联矩阵(incidence matrix)又是属性和被试的关联阵。

Tatsuoka 的规则空间模型包含两部分:Q 矩阵理论和统计分类方法。Q 矩阵理论是“to determining unobservable knowledge states and representing them by observable item response patterns(called Q - matrix theory)”。^{[1][p328]} 我们试译为“Q 矩阵理论确定不可观察的知识状态并用可观察项目反应模式描述它们”。她在其专著中说 Q 矩阵理论为模式识别中特征提取,^{[2][p8]} 提出 Q 矩阵是联系知识状态和可测量、可观察的理想反应模式的桥梁,而知识状态是可解释但不可观察的。^{[2][p83]}

本文只考虑 0-1 Q 矩阵,即 Q 矩阵为布尔矩阵的情形。如果感兴趣的诊断领域包含 K 个属性,那么 $Q = (q_{ij})$ 是一个 K 行 m 列 0-1 矩阵,其中 $q_{ij} = 1$ 表示第 j 个题目考查了第 i 个属性;否则 $q_{ij} = 0$ 。显然, Q 矩阵和 Q 矩阵理论是两个完全不同的概念。但是,没有 Q 矩阵就没有 Q 矩阵理论,没有 Q 矩阵的特殊构造,也就没有 Q 矩阵理论。

Tatsuoka 在叙述 Q 矩阵理论的时候,特别强调 Q 矩阵是认知模型;指出 Q 矩阵的行、列在布尔并、交运算下构成布尔格;介绍计算理想反应(理想反应模式)的方法和从 Q 矩阵中挖掘属性层级的方法;引入充分 Q 矩阵和充分题库等概念,指出充分 Q 矩阵和充分题库会提升测验构念效度,强调充分 Q 矩阵是知识结构的核心,由包含关系产生的偏序使得 Q 矩阵理论能够表示属性之间的先决关系;^{[1][2]} 特别在属性之间不存在先决关系(即独立结构)条件下,她指出单位矩阵可以使得知识状态和理想反应模式一一对应;^[1] 同时考虑到一般情况下,可能多个知识状态对应同一个理想反应模式,Tatsuoka 称这些知识状态为等价类。^[1] 根据这几个概念可知,从充分 Q 矩阵出发,希望指导题目编制和认知诊断测验设计,以提高测验构念效度,^{[1][2]} 这表明,她已经认识到 Q 矩阵理论必须包含认知诊断测验蓝图的设计,尽管她对此没有进行更加深入地研究,从而没有更加有效地解决这个重要的问题。

当然,Tatsuoka 认为 Q 矩阵的行、列在布尔并、交运算下构成布尔格的理论^{[1][2]} 是错误的,因为一般来说满足属性层级关系的 Q 矩阵的所有列在布尔并交下构成格,而不一定是布尔格,^{[4][5]} 特别认为 Q 矩阵的行在布尔并、交运算下构成布尔格(哪怕是格)的理论是错误的:首先 Q 矩阵的行表示属性,属性数目确定以后, Q 矩阵的行数随之确定,它们的并、交以后产生的新的行如何安排在 Q 矩阵的行集合之中?其次, Q 矩阵行的并、交的含义是什么?如何解释也不清楚。所以她上述论述难以成立。

1.2 知识状态和理想反应模式的定义

确定一个感兴趣的诊断范围以后(假设包含 K 个属性,并且这些属性已经编号),认知诊断要根据被试在测验上的反应,判断其在这个范围之内认知方面的强项和弱项,通常用一个 K 维 0-1 向量表示,等于 1 的分量表示掌握了相对应的属性,否则表示没有掌握这个属性。这个 K 维 0-1 向量称为知识状态(knowledge state),^{[2][p79]} 也称为属性模式或者属性掌握模式。稍微抽象一点讲,知识状态是 Q 矩阵的一个列(可以是零列)。

Tatsuoka 有时候用两种方式称呼知识状态,理想项目得分模式或者属性模式,^{[2][p80]} 这里理想项目分数模式就是理想反应模式,即在一组题目上既不猜测也不失误的反应向量。由于知识状态是认知诊断分类的重要的产品,所以有必要进一步阐明我们上述约定的知识状态的定义。将知识状态和理想反应模式混为一谈,这让人匪夷所思。对于同一个测验 Q 矩阵,可能多个知识状态对应一个理想反应模式,所以这两者并不等价;其次,知识状态由认知模型唯一确定,而理想反应模式由认知模型、知识状态和测验 Q 矩阵确定。测验 Q 矩阵当然和测验设计有关,但是受到认知模型的约束,比如一个线型结构,

不可能将单位矩阵作为测验 Q 矩阵。有的测验 Q 矩阵并不能很好地代表认知模型,这时候理想反应模式和知识状态不能够一一对应,可见知识状态和理想反应模式不是同一个概念。

1.3 Q 矩阵理论的外延

我们先由定义看一看 Q 矩阵理论包含的内容。之所以讨论这个问题,是因为连 Tatsuoaka 都认为 DINA 模型没有使用 Q 矩阵理论。^{[2][p6]}

既然 Q 矩阵理论是确定不可观察的知识状态并且用可观察的理想反应模式表示它们,那么 Q 矩阵理论首要任务是确定知识状态的集合。给定认知模型(这是客观存在的)之后,和认知模型对应的知识状态的集合是什么值得研究;其次,既然 Q 矩阵理论要用可观察的理想反应模式表达知识状态,那么必须解决理想反应模式的计算问题;第三,如何设计测验 Q 矩阵(这是人为的),使得可以更好地用可观察反应模式描述知识状态,这也是十分重要的;而且因为牵涉测验的设计,就希望设计一个优良的 Q 矩阵,或者说优良的测验蓝图,而这便涉及到优良设计的标准讨论。这个标准,应该和 Q 矩阵理论定义中“描述”两个字紧密相关,这是优良设计如何构造的问题;第四,更进一步,如果测验 Q 矩阵标注不准确,那么或者知识状态可能标示错误,或者计算出来的可观察理想反应模式的集合不可能准确,从而很难准确描述知识状态,所以如何准确标注或者修正 Q 矩阵应该是 Q 矩阵理论包含的内容;第五, Q 矩阵理论的应用显然是 Q 矩阵理论的一个组成部分。我们认为 Q 矩阵理论至少包含上述五个内容。囿于本文主旨和篇幅,在此只能讨论前面三个问题。

研究和应用 DINA 模型时,至少要计算理想反应, DINA 模型的“确定性输入”部分,也就是 DINA 模型的潜在反应变量(latent response variable),它事实上是给定知识状态以后在给定题目上的理想反应;对于认知诊断测验而言,测验蓝图的设计是实验设计,至少对一个试验合理设计,比随心所欲安排的试验要科学;更何况有研究认为 DINA 模型考虑属性层级关系以后,分类准确率得以提高(请参见下文)。甚至对于“翻新”数据,应用 Q 矩阵理论可以推断 DINA 模型分类数目。^[6]

2 知识状态和理想反应模式的确定

2.1 记号约定

为了表述清晰,首先将不同的 Q 矩阵冠以不同的名称。K 维 0-1 向量的全体构成的矩阵称为全矩阵,记为 Q_a 。根据可达阵对 Q_a 的列进行筛选,^{[1][2]}获得的 Q 矩阵称为简化 Q 矩阵,记为 Q_r ;根据可达阵,使用扩张算法^{[4][5][7]}获得的矩阵,称为潜在 Q 矩阵,记为 Q_p ,因为其每一列都可以成为测验题目中的属性向量,即潜在的题目属性向量。事实上, Q_r 和 Q_p 相等。学生 Q 矩阵记为 Q_s ,是 Q_p 增加一个零列而成,实际上 Q_s 的列的全体即构成知识状态全集。而作为测验蓝图的 Q 矩阵,称为测验 Q 矩阵,记为 Q_t 。可达阵 R 是一种特殊的 Q 矩阵,可达阵 R 表示属性之间直接或者间接关系。邻接矩阵 A 仅仅表达属性之间直接关系,而邻接矩阵 A 不是 Q 矩阵。

2.2 认知诊断是模式识别

Tatsuoka 认为认知诊断是模式识别。统计模式识别包括聚类分析和判别分析。聚类分析又称为无监督分类,判别分析称为有监督分类。不少研究者使用判别分析^{[1][2][8][9]}进行认知诊断。和聚类分析完全不同,判别分析需要事先知道有多少类,每一个类的特征(分布特征)是什么,还需要好的训练样本。下面我们讨论使用判别分析方法进行的认知诊断。

Tatsuoka 将知识状态和理想反应模式进行变换,变换成为纯规则点,而将纯规则点作为类中心。有人提出使用理想反应模式作为类中心;^{[9][10][11][12][13]}但是,有人认为理想反应模式是确定性的,没有考虑随机误差,因而未充分利用总体分布信息,为了诊断分类效果和拓展诊断评估的实用性,建议使用给定知识状态下,观察反应模式的条件期望作为类中心。^[14]模拟研究结果发现,在观察反应模式与理想反应模式差异比较大的情况下,基于条件期望向量为类中心的欧氏距离判别方法优于基于理想反应模式为类中心的分类方法。^{[14][15]}如果能够建立知识状态和理想反应模式之间的一一对应(这对于测验蓝图的设计有比较高的要求),再将理想反应模式(或者给定知识状态,观察反应模式的条件期望向量)作为类中心,而认知诊断模型建立观察反应模式到类中心之间的对应,精心设计的测验 Q 矩阵建立类中心

一样,布尔格是特殊的有界分配格,要求对并、交和补三个运算封闭,而 Tatsuoka 忽视了对“补”运算封闭条件,所以她的这个断言是错误的。^{[4][5]}

2.3.3 渐增式扩张算法(The incremental algorithm)

渐增式扩张算法^[16]如下:

- 1) 给出空矩阵 Q ;
- 2) 给出可达阵 R 第一列,植入 Q ;
- 3) 依次取出 R 第 j 列 r_j ,分别和 Q 矩阵中的每一列作布尔并,如果布尔并的结果和 Q 矩阵中已经存在的列都不相同(下面称之为“新列”),则将新列添加在 Q 矩阵最右边,仍然记为 Q ;
- 4) $j \leftarrow j + 1$,如果 $j \leq K$,则转 3); 否则停止。

扩张算法就像是统计中向前回归,而 Tatsuoka 的缩减算法好像是向后回归。渐增式扩张算法相当于逐步回归。

2.3.4 理想反应模式的计算

在属性之间不可以补偿条件下,给出知识状态集合和测验 Q 矩阵,理想反应模式的计算特别简单,^{[4][5]}只需要用知识状态减去项目 j 对应的属性向量(即 Q 矩阵的第 j 列),如果所得到的差向量的每一个分量均非负,那么在第 j 个题目的理想反应正确,即得 1 分,否则就得 0 分,这只需要用非补偿的定义和朴素集合论就可以,而用不着像使用诸如布尔格^{[1][2]}(请注意,她的相关理论是错误的)这样更加高深数学理论,^{[4][5]}也不用定义布尔描述函数。当然也可以使用 DINA 模型中计算潜在反应的方法,其实, DINA 模型的潜在反应实际上就是理想反应。由于 Tatsuoka 将理想反应的计算复杂化,^{[1][2]}使得她看不清 DINA 模型中潜在反应就是理想反应,从而下了 DINA 模型不使用 Q 矩阵理论的错误断言。

3 Q 矩阵可以表示认知模型吗?

3.1 认知模型的代数和几何表示

Tatsuoka 认为 Q 矩阵表示认知模型;而 Leighton 等人不仅认为 Q 矩阵表示认知模型,而且认为属性层级关系(hierarchical relation)表示认知模型。^[8]属性层级关系是属性的先决关系。属性集合(S)上面定义先决关系(r)以后, r 满足自反、反对称和传递,因此是偏系关系, $\langle S, r \rangle$ 就是偏序集(poset)。可以认为,Leighton 等人所说的属性层级关系为认知模型,严格意义上是说 $\langle S, r \rangle$ 是认知模型。^[8]

前面已经假定 Q 矩阵行表示属性,Tatsuoka 提出 Q 矩阵行的逐对比较可以挖掘属性层级关系。Tatsuoka 的 Q 矩阵的行的逐对比较相应的操作步骤和方法是:先进行 Q 矩阵行的逐对比较,获取属性之间的先决关系,由导出的属性层级关系(可以得出哈斯图),给出邻接矩阵,再给出可达阵。比如一个 3 阶 Q 矩阵,第一行是(101),第二行是(110),第三行是(011),逐对比较之后,发现它们之间不存在先决关系,于是,邻接矩阵是零矩阵,可达阵是单位矩阵。

由于 Tatsuoka 的简化 Q 矩阵 Q_r (也就是潜在 Q 矩阵 Q_p)包含可达阵,所以通过 Q_r 的逐行比较,一定可以挖掘出隐藏的属性层级结构,获得可达阵。丁树良、罗芬还给出由偏序关系的关系矩阵导出邻接矩阵的清洗算法。^[21]可达阵实际上是偏序关系的关系矩阵。这表明,可达阵和邻接阵是一一对应的。因此 Leighton 等人所认为的“由 Tatsuoka 的简化 Q 矩阵 Q_r (也就是本文所说的潜在 Q 矩阵 Q_p)不可能识别唯一的属性层级”的说法是不成立的。^[8]

当然潜在 Q 矩阵包含可达阵,它可以代表正确的属性层级结构,其实 Q 矩阵是认知模型的代数表示,属性层级是认知模型的几何表示。理论上,这两个模型应该统一。但是,不是精心设计的测验 Q 矩阵(Q_t)不一定可以完整地表示属性层级关系。下面讨论 Q_t 可以在多大程度上代表属性层级结构,这是一个重要的问题。

3.2 理论构念效度

可以用简便的方法计算理论构念效度:对测验 Q 矩阵形式上使用扩张算法,获得 Q 矩阵,记为 Q_1 ,假设 Q_1 的列数为 M_1 ,而基于可达阵扩张出来的列数为 M_2 ,那么测验 Q 矩阵的理论构念效度等于 $(M_1 + 1) / (M_2 + 1)$ 。^[6]应用理论概念效度的概念可以考察和评价认知诊断测验蓝图或者使用“翻新”手段进

行认知诊断评估的有效性。

例 3(引自 [6]) 引用 DeCarlo(2011)对 Tatsuoka(1990)的分数减法数据重新分析的例子。这个诊断测验包含 8 个属性 20 个项目,这个测验可以用一个 Q 矩阵(Q_i)来表示:对 Q_i “形式上”使用扩张算法,可以导出除零向量外的 57 种知识状态,而由 R 通过扩张算法可以导出除零向量外的 65 种知识状态。纵使 Q_i 挖掘出来的层级关系为真,这时 Q_i 的理论构想效度也只有 $58/66=0.879$ 。命题专家和测量专家需要对测验的“完整性”(即测验 Q 阵是否包含了 Q 挖掘出的可达矩阵 R)进行考察,如果 Q 阵缺乏 R 阵中的列,则应该修改测验,使之包含可达矩阵 R 中所有的列对应的项目。我们可以将这一过程小结为四步:

第一步:由测验 Q 阵(Q_i)挖掘出层级关系,导出可达矩阵 R ;

第二步:对 Q_i 形式上使用扩张算法,扩张出所有的知识状态 Q_s ;

第三步:由 R 扩张出所有知识状态 $Q_s(R)$;

第四步:比较 Q_s 和 $Q_s(R)$;如果 $Q_s(R)$ 的列数多于 Q_s 的列数,则应修改 Q_i ,使之包含 R 中所有列。

由此也可以看出,不论使用什么认知诊断模型,只要 Q 矩阵表示的层级关系确定,属性之间是不可补偿的,知识状态的等价类就确定了。无论采用 DINA 模型还是其他非补偿模型,实际上是没有关系的。

3.3 属性层级关系

3.3.1 基本属性层级关系的重新划分

Leighton 等人认为 Tatsuoka 对现存的试卷进行认知诊断的“翻新”做法在逻辑上不严密。他们主张先分析欲诊断的范围内的属性及其层级关系,然后开发题目,实施测验,再进行认知诊断分析。他们将基本的属性层级结构归纳成为 4 类:线型、收敛型、发散型和无结构型,更加复杂的层级结构可以由这 4 种基本层级结构复合出来。^[8]但是,不少研究者使用独立型层级结构,即属性之间不存在先决关系。

其实 Leighton 等人所说的基本层级结构并不“基本”,因为收敛型实际上是线型和“菱形”(rhomb)的复合;根据图论(Graph Theory),Leighton 等人给出的基本属性结构可以归并为根树型(rooted tree)和“菱形”。而根树型包括线型、发散型和无结构型。^[22]当然这里菱形的定义比较复杂:如果一个属性层级结构至少包含 4 个属性,其中有一个最大元(它是所有属性的先决属性)和一个最小元(所有属性都是它的先决属性),删除最小元以后,该菱形退化为 $K-1$ 个属性的根树型,那么这个属性层级结构称为菱形。

于是,通常的基本属性层级结构可以归纳为 3 大类:根树型、菱形和独立型。对于根树型和菱形层级结构, $\langle S, r \rangle$ 仅仅形成格而不是布尔格。当且仅当独立型层级结构在布尔并、交、补运算之上构成布尔格。

3.3.2 层级相合性指标

被试对认知诊断测验的反应,也就是观察反应模式和属性层级是不是吻合,可以用层级相合性指标(HCI)^{[6][23]}或者新的层级相合性指标(NHCI)^[6]衡量。层级相合性指标是假定属性层级关系正确并且测验 Q 矩阵准确的基础上开发的。由于属性及其层级代表认知模型,准确的测验 Q 矩阵能够代表认知模型,能够表达属性层级,所以 HCI 衡量的是被试反应(观察反应模式)是否和认知模型相吻合。如果吻合, HCI 的值就比较大,否则就比较小;同时若 HCI 的值太小,有理由怀疑 Q 矩阵的正确性。

4 Q 矩阵设计理论

Q 矩阵设计的问题,是认知诊断测验蓝图的设计的问题,也就是认知诊断测验设计的问题。Gierl 在文章的注释中指出,2006 年 Erlbaum 出版的,由 Steven M. Downing and Thomas M. Haladyna 编辑的测验开发手册(Handbook of test development)编辑对测验开发的概念,实践和创新给出出色的评述,但是没有认知诊断方面的内容。^[24]可见这方面的研究的缺失。

4.1 充分 Q 矩阵的作用有限

Tatsuoka 提出充分 Q 矩阵的概念、充分题库的概念,希望更好地表示知识结构和提高认知诊断测验

的效度。但是充分 Q 矩阵概念对于知识结构的表示和对于提高认知诊断测验效度的作用不充分,甚至非充分 Q 矩阵的效度高于充分 Q 矩阵。为明此理,第一,先考察充分 Q 矩阵定义^{[1][2]}对测验 Q 矩阵行在包含关系下的逐对比较产生可达阵,就称之为表达论域对应的认知模型是充分的^{[1][p78]};第二,考察“产生”可达阵(yields the reachability matrix R)的“产生”的含义。我们认为 Q 矩阵在包含关系下的逐对比较,只能得到属性的层级关系,而不可能直接获得可达阵,所以定义中“产生可达阵”的说法实质上是“行的逐对比较获得层级关系,在此基础上获得可达阵”;另外,根据 Tatsuoka 对独立型结构的处理来看,她认为测验 Q 矩阵包含单位矩阵,单位阵对应的题目太简单,而且这句话出现在充分 Q 矩阵定义之后,这意味着她说的充分 Q 矩阵只不过是行的逐对比较“蕴涵”出可达阵就可以,而不是说要求测验 Q 矩阵要以可达阵为其子矩阵。第三, Tatsuoka 指出甚至是充分 Q 矩阵也不一定能够很好地表示所考察的领域潜在的认知任务(如果以可达阵为子矩阵就不存在这个问题!);第四,容易举出反例,说明充分 Q 矩阵导出的知识状态可能少于非充分 Q 矩阵导出的知识状态,从而对于知识结构的表示不成功,也容易举出反例,说明充分 Q 矩阵的理论概念效度小于非充分 Q 矩阵。

4.2 测验蓝图设计

认知模型或属性及其层级关系确定之后,要充分发挥认知诊断评估的作用,一个优良的认知诊断测验希望能够比较准确地探测每一个被试的知识结构,所以仅仅有打磨好的题目还远远不够,必须要求测验结构合理,即要合理编制测验。下面简要介绍属性之间不可相互补偿条件下 0-1 评分条件下认知诊断测验蓝图的设计原则。理想反应模式不带有随机误差,不同知识状态应该对应不同的理想反应模式,这是对认知诊断测验蓝图基本要求;理论上这时候只要测验 Q 矩阵包含可达阵,就可将所有属性掌握模式对应的理想反应模式一一区分开。这时,好的认知诊断测验的设计应该满足的必要条件是,对任意两个不同的知识状态,它们在这个测验之下的理想反应模式不相等。

定理 1 若属性之间不可相互补偿,在 0-1 评分条件下,测验蓝图中包含可达矩阵(R)当且仅当知识状态与理想反应模式一一对应。^{[17][18]}

称以可达阵作为子矩阵的测验 Q 矩阵为必要 Q 矩阵。

如果一个 Q 矩阵可以使得知识状态集合和理想反应模式集合一一对应,并且包含的列数最少,那么这个 Q 矩阵称为完备(perfect) Q 矩阵。^[22]显然在课堂评估当中,不可能使用太多题目,因此完备 Q 矩阵的概念显得十分重要。在属性之间不可补偿、0-1 评分条件下,可达阵是完备 Q 矩阵。

对于 Leighton 等人^[8]的 4 种属性层级结构,设计认知诊断测验,使用 DINA 模型进行分析,植入可达阵的测验模式判准率超过不植入可达阵的模式判准率,而且植入可达阵越多,模式判准率越高。^[18]这也说明认知诊断测验的设计(即 Q 矩阵理论)对于 DINA 模型的分析结果有非常大的影响。

5 结论与讨论

5.1 结论

从 Q 矩阵理论的定义,可以挖掘出它除了计算知识状态集合和理想反应模式集合之外,还应该包含认知诊断测验蓝图设计、Q 矩阵元素的标注和修正、Q 矩阵质量的评价等等内容。由此可知,DINA 模型实际上使用了 Q 矩阵理论。

本文认为将可达阵作为测验 Q 矩阵的子矩阵是属性之间不可补偿、0-1 评分的认知诊断测验设计的重要原则。

设计认知诊断测验蓝图时要考虑很多因素,比如认知模型(补偿或者非补偿,呈现什么样的层级关系)、评分方式(0-1 还是多级)、Q 矩阵采用的元素(0-1 还是多值)、是否要充分利用选择项中的诊断信息^[21]、是否必须使用短测验(如课堂诊断测验)、测验的长度、是否有非统计约束、甚至还要考虑使用什么样的认知诊断模型分析:因为有的认知诊断模型对于认知模型的容错能力比较强(如 RUM),有的认知诊断模型对于认知模型相当严格,比如广义距离判别方法。^[9]列出这些因素,仅供参考,而绝对不是包括了方方面面。

5.2 讨论

5.2.1 Q 矩阵理论的其他相关研究

Leighton 等人对于规则空间模型的认知诊断分类的逻辑顺序做出重要修改,提出应该先进行认知诊断测验设计;Chui & Douglas, Samejima, Tatsuoka 和 Madison & Bradshaw 对于独立型属性结构提出测验蓝图植入单位阵的设计的看法;甘朝红等人认为必要 Q 矩阵对于测验 Q 矩阵元素标注失误有一定的补救作用。^[28]杨淑群认为属性层级关系只不过是蕴涵关系的特例。^[29]她所说的蕴涵关系可以是集合 A 是集合 B(或者属性 B)的先决。

5.2.2 属性之间补偿与非补偿的问题的探讨

必须掌握项目中所有属性才能够获得对项目正确反应,属性之间这样的关系是非补偿(即连接)的;如果在某个题目上成功的概率,某些属性的低水平能够被其他属性的高水平补偿,那么属性之间的关系是补偿型的。^[30]Stout 认为连接比补偿倾向于更加流行是因为求解一个题目的认知加工往往看上去是连续执行一系列步骤的每一步;而有时候将补偿模型看成是在被试之间使用多个策略的可能性进行模型化。^[30]

不论属性之间补偿与否,由可达阵经扩张后,均可获得所有的非零知识状态。但是,给定知识状态集合和测验 Q 矩阵,理想反应模式就必须按照补偿还是非补偿分别计算。然而什么条件下出现补偿功能,有必要仔细考虑。比如,如果 A 是 B 的先决属性, A 不可以补偿 B,更不可能 B 补偿 A。但是对于根树型层级结构,两片树叶之间不可以比较,它们之间反而可以相互补偿。是否可以更加广泛一点,任何两个不可以比较(即不存在先决关系)的属性,才可以相互补偿?

5.2.3 属性之间有补偿作用,如何设计认知模型测验?

我们知道,属性 A 是属性 B 的先决,是说掌握 A 是掌握 B 的必要条件。如果说,属性 A 可以弥补属性 B,则掌握 A 是掌握 B 的充分条件;从逻辑上可知两种说法的差异。

如果说,在属性之间不可以相互补偿条件下,知识状态集合和理想反应模式集合之间一一对应,是优良测验蓝图的特征,这个特征不可能用于评价属性之间可以补偿条件下测验蓝图的优劣,我们必须寻找其他的评价标准。补偿作用条件下,认知诊断测验蓝图的设计是一个困难的富有挑战性的课题,值得研究。

5.2.4 其他问题

尽管 Q 矩阵理论已经用于认知诊断模型的开发,用于计算机化自适应诊断测验选题策略的制定,用于 Q 矩阵的计算机辅助标定,但是还应该深入讨论其应用,在推广应用中发现问题,解决问题。

本文限于篇幅,是在 Q 矩阵标注正确的前提下进行讨论,而不可能讨论 Q 矩阵修正这个重要论题,希望研究者注意。

参考文献:

- [1] Tatsuoka, K. K. Architecture of knowledge structure and cognitive diagnosis: a statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 327–361). Hillsdale, NJ: Erlbaum, 1995.
- [2] Tatsuoka, K. K. *Cognitive Assessment: An introduction to the Rule Space Method*. New York: Routledge, Taylor & Francis Group, 2009.
- [3] Haertel, E. H. An application of latent class models to assessment data. *Applied Psychological Measurement* 8, 333–346. 1984.
- [4] Ding, S. L., Luo, F., Cai, Y., Lin, H. J., & Wang, X. B. Complement to Tatsuoka's Q matrix theory. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 417–423). Tokyo: Universal Academy, 2008.
- [5] 丁树良, 祝玉芳, 林海菁, 蔡艳. Tatsuoka Q 矩阵理论的修正[J]. *心理学报* 2009, 41(2).
- [6] 丁树良, 毛萌萌, 汪文义, 罗芬, Cui Ying. 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).
- [7] 杨淑群, 蔡声镇, 丁树良, 林海菁, 丁秋林. 求解简化 Q 矩阵的扩张算法[J]. *兰州大学学报(自然科学版)* 2008, 44(03).

- [8] Leighton J. P., Gierl M. J. & Hunka S. M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuo-ka's rule space approach [J]. *Journal of Educational Measurement* 2004 41(3).
- [9] 孙佳楠, 张淑梅, 辛涛和包珏. 基于 Q 矩阵和广义距离的认知诊断方法 [J]. *心理学报* 2011 43(9).
- [10] Chiu J. Y. & Douglas J. A. A nonparametric approach to cognitive diagnosis by proximity to ideal response patterns. *Journal of Classification* 2013 30 225 - 250.
- [11] 丁树良, 罗 芬, 汪文义. 认知诊断分类中心的确定 [J]. *心理学探新* 2013 33(5).
- [12] Sun J. N., Xin T., Zhang S. M., de la Torre. A polytomous extension of the generalized distance discriminating method [J]. *Applied Psychological Measurement*. 2013 37(7).
- [13] 罗照盛, 李喻骏, 喻晓锋, 高椿雷, 彭亚风. 一种基于 Q 矩阵理论朴素的认识知诊断方法 [J]. *心理学报* 2015 47(2).
- [14] 汪文义, 丁树良, 宋丽红. 认知诊断中基于条件期望的距离判别方法 [J]. *心理学报* 2015 47(12).
- [15] 汪文义, 宋丽红, 丁树良. 基于探索性因素分析的 Q 矩阵标定方法 [J]. *江西师范大学学报(自然科学版)* 2015 39(2).
- [16] Yang Shuqun, Ding Shuliang, Ding Qiulin. Incremental augment algorithm based on reduced Q - matrix [J]. *Transactions of Nanjing University of Aeronautics & Astronautics*. 2010 27(2).
- [17] 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).
- [18] 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).
- [19] 丁树良, 罗 芬, 汪文义, 熊建华. 0 - 1 和多值可达阵的性质及应用 [J]. *江西师范大学学报(自然科学版)* 2015 39(1).
- [20] 汪文义, 丁树良, 游晓锋. 计算机化自适应诊断测验中原始题的属性标定 [J]. *心理学报* 2011 8.
- [21] 丁树良, 罗 芬. 求偏序关系 Hasse 图的算法 [J]. *江西师范大学学报(自然科学版)* 2005 29(2).
- [22] 丁树良, 汪文义, 罗芬. 多级评分认知诊断测验蓝图的设计—根树型结构 [J]. *江西师范大学学报(自然科学版)*, 2014 38(2).
- [23] Cui Y. & Leighton J. P. The hierarchy consistency index: evaluating person fit for cognitive diagnostic assessment. [J] *Journal of Educational Measurement* 2009 46(24).
- [24] Gierl M. J. Making diagnostic inferences about cognitive attributes using the Rule - Space model and Attribute Hierarchy Method. [J] *Journal of Educational Measurement* 2007 44(4).
- [25] 李瑜. 多选题认知诊断测验编制及多策略的多选题认知诊断模型的开发 [D]. 江西师范大学 2014.
- [26] Samejima F. A cognitive diagnosis method using latent trait models: competency space approach and its relationship with Di-Bello and Stout's unified cognitive - psychometric diagnosis model. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively Diagnostic Assessment* (pp. 391 - 410). Hillsdale, NJ: Erlbaum. 1995.
- [27] Madison M. J. & Bradshaw L. P. (2015). The effects of Q - matrix design on classification accuracy in the log - linear cognitive diagnostic diagnosis model [J]. *Educational and Psychological Measurement* 2015 75(3).
- [28] 甘朝红, 汪文义, 丁树良. 项目属性标错时可达阵补救作用的研究 [J]. *江西师范大学学报(自然科学版)* 2014 38(6).
- [29] 杨淑群. 基于属性蕴含的 Q 矩阵理论 [J]. *江西师范大学学报(自然科学版)* 2015 39(6).
- [30] Stout, W. Skills diagnosis using IRT - based continuous latent trait models [J]. *Journal of Educational Measurement*. 2007, 44(4).

(责任编辑: 余小江)