

贝叶斯网在认知诊断属性层级结构确定中的应用*

喻晓锋^{1,2} 丁树良¹ 秦春影² 陆云娜^{1,3}

(¹ 江西师范大学计算机信息工程学院, 南昌 330027)

(² 安徽亳州师范高等专科学校计算机系, 亳州 233500)

(³ 江西省南昌市第一中学, 南昌 330003)

摘要 K. K. Tatsuoka 和她同事研究的规则空间模型(RSM)是一种在国内外有较大影响的认知诊断模型,但是 Tatsuoka 的 RSM 是由学科专家先从已编制的测验中抽取出属性,然后给出测验的关联 Q 阵,再由该 Q 阵导出属性间的层级关系。已有研究证明,这种做法所得到的属性间的层级关系难以保证是正确的,甚至难以保证属性间的层级关系是唯一确定的。这里利用贝叶斯网进行结构学习,从被试的属性掌握模式中挖掘出属性间的层级关系,学习所得到的层级关系可以用来验证由 RSM 中的方法得到的层级关系。模拟实验和实证研究的结果都显示了该方法所得到的属性层级关系是有参考价值的,可以为命题或测量专家带来有用的信息。

关键词 规则空间模型; 认知诊断; 贝叶斯网; 结构学习

分类号 B841

1 引言

认知诊断理论是新一代测量理论,在教育测量实践中具有广阔的应用前景;贝叶斯网络是基于概率的图模型,正逐渐成为人工智能、模式识别、机器学习和数据挖掘等众多领域中处理不确定性问题的重要方法之一,是十多年来在这些领域的一个研究热点(董立岩, 2007)。本文针对目前认知诊断中属性层级结构的确定有时会出现偏差或错误的现象,将贝叶斯网技术应用到认知诊断的属性层次结构的确定中,即通过贝叶斯网结构学习,从被试的属性掌握模式中得到属性间的层级关系,以验证使用 RSM 中的方法得到的层级关系或使用属性层次方法(AHM)(Leighton, Gierl, & Hunka, 2004)由专家给出的层级关系。实验结果显示了应用贝叶斯网对开发出高质量的认知诊断测验和对测验结果分析是十分有意义的。

2 综述

2.1 Q 矩阵

Q 矩阵是属性与项目的关联阵(incident matrix),它是一个测验规范(test specification),有人也认为它表示了一个认知模型(cognitive model)(Tatsuoka, 1995)。Q 矩阵对于大多数认知诊断模型都是必须的。规则空间模型(Rule Space Model, RSM)(Tatsuoka, 1995, 2009)、属性层级模型(Attribute Hierarchy Model, AHM)(Gierl, Leighton, & Hunka, 2000)明确要求有 Q 阵,而 DINA (deterministic input, noisy “and” gate model)(Junker & Sijtsma, 2001), NIDA (noisy input, deterministic “and” gate model)(Maris, 1999), FUSION (Hartz, Roussos, & Stout, 2002), LLTM (Linear Logistic Trait Model)(Fischer, 1973)等模型也必须要有 Q 阵。RSM 是较早提出的一种认知诊断模型,在国内外产生了较大的影响,一大批学者对它进行了深入的研究(如:戴

收稿日期: 2009-12-22

* 国家自然科学基金项目(编号 30860084 和 60263005), 全国教育考试“十一五”科研规划课题(编号 2009JKS2009), 教育部人文社科项目(编号 09YJCXLX012, 10YJCXLX049), 安徽省 2010 年度高校省级自然科学研究项目(编号 KJ2010B123)。

通讯作者: 丁树良, E-mail: ding06026@163.com

海崎, 张青华, 2004; 丁树良, 汪文义, 杨淑群, 2009; 丁树良, 杨淑群, 汪文义, 2010; 丁树良, 祝玉芳, 林海菁, 蔡艳, 2009; Gierl et al., 2000; Leighton et al., 2004; 文剑冰, 2003; 辛涛, 焦丽亚, 2006; 余嘉元, 1995; 祝玉芳, 丁树良, 2009)。RSM 分为两大部分: Q 矩阵理论和判别分类。其中, Tatsuoka 的 Q 矩阵理论不仅在 RSM 中起着很重要的作用, 而且对其它认知诊断模型也起到重要作用(余娜, 辛涛, 2009)。Tatsuoka 的 Q 矩阵理论致力于确定不可观察的知识状态(Knowledge State, KS), 并通过可观察的项目反应模式(Ideal Response Pattern, IRP)把它们表示出来(Tatsuoka, 1995)。Q 矩阵建立了由属性掌握模式所表示的知识状态和由理想反应模式所表示的可测量的、可观察的知识状态之间的桥梁(Tatsuoka, 2009)。当然, Q 矩阵理论还应该研究如何将理想反应模式与知识状态的对应, 并力图给出一个好的测验蓝图, 使这种对应关系是一一对应(丁树良等, 2009a, 2010); Tatsuoka 在 1995 年提出了充分 Q 矩阵和充分题库的概念, 也想在 Q 矩阵理论中解决如何建立知识状态与测验蓝图之间的良好的对应关系的问题。RSM 中 Q 矩阵是先请学科专家从已编制的测验中抽取出属性, 然后给出测验的关联 Q 阵; 判别分类主要是将被试在测验项目上的作答反应(即观察反应模式)划归为某种与认知技能(属性)相联系的属性掌握模式。模型的一个基本假设是测验项目可以用特定的认知技能(属性)描述, 属性包括被试正确求解测验项目所必须具有的各方面的知识、技能和策略。对于认知诊断测验, 在确定了所测的属性后, 如果能正确的分析出属性间的层级关系, 就能够确定可达阵 R, 进而获得 Q 阵。因此, 要想得到 Q 阵, 只要能得到属性间的层级关系就可以。但实际上, 属性间的正确的层级关系并不容易获得, 原因主要有: ①使用的方法不太好, 比如 RSM 中使用事后分析的方法, 已有研究证明事后分析是不恰当的; ②专家认知的局限, 本文实测结果可以证明这一点; ③学科专家凭借自己的经验确定 Q 阵; 而被试的反应所对应的 Q 阵(潜在的)可能与专家的 Q 阵不一致; ④属性层级不明朗, 专家之间有分歧, 如不同的专家组对 2002 年美国 Texas 州小学三年级阅读评价测验就给出两个 Q 阵(Cheng, 2009, p.627); 2009 年我们与天津市教育招生考试院联合做小学分数加法认知诊断时, 也出现对 Q 阵争议十分激烈的情况, 因此有必要对专家给出的 Q 阵进行评估。

2.2 Q 矩阵理论的错误与改进方案

Gierl 等人(2000)及 Leighton 等人(2004)建议在测验之前就由专家给出欲测属性及这些属性间的层级关系(hierarchy relation), 这样实际上可以使认知理论对测验编制起到指导作用。RSM 的关联矩阵 Q 是通过分析测试项目得到的, Leighton 等人(2004)指出这样的 Q 矩阵的方法逻辑性不强。可以通过 Tatsuoka (1995, P. 337)的 Q 阵为例进行证明(丁树良等, 2009b): 通过测验分析导出的 Q 阵, 再由该 Q 阵导出的属性层级关系难以保证是正确的, 甚至难以保证它们的层级关系是唯一确定的。正因为 Q 矩阵理论存在着缺陷, 丁树良等人(2009a, 2009b, 2010)对 Q 矩阵理论某些不足之处提出了改进意见。丁树良等人(2009b)指出: Tatsuoka 认为 Q 矩阵的行(列)在布尔并, 交, 补运算下构成布尔代数是缺乏

理论依据的(比如 $Q = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ 的行或列就不可

能按 Tatsuoka 方法形成布尔格); 在布尔补运算之下获得的向量可能是不满足属性层级关系的; 为了建立起知识状态(KS)与理想反应模式(IRP)之间的对应关系, Tatsuoka 引入布尔描述函数(BDF)。引入 BDF 这种做法一方面没有理论保证, 另一方面也完全没有必要。为此, 丁树良等人(2009b)介绍了几种 KS 与 IRP 之间对应的方法。丁树良等人(2009a, 2010)还指出, KS 与 IRP 之间如果要建立双射, 则测验规范(即测验 Q 矩阵)必须且只须包含可达阵, 满足这样条件的 Q 矩阵与 Tatsuoka (1995, 2009)所定义的充分 Q 阵并不是一样的。

既然获得的属性层级关系可能存在问题或者不容易获得, 那么, 有没有一种方法不需要通过分析 Q 矩阵而得到属性间的层级关系呢? 属性间存在层级关系, 这个层级关系直接影响到被试对各个属性的掌握顺序, 进而影响到被试对项目的作答。也就是说, 被试对项目的作答反应中包含着他们对属性的掌握情况的信息。尽管被试的得分模式中包含随机因素(如失误、猜测等)的影响, 但得分模式大体上反映出被试的属性掌握模式(知识状态)。因此应尽量设计好一个认知诊断测验, 即好的测验蓝图, 使得具有不同知识状态的被试能够表现出不同的得分模式, 然后选取一种好的方法, 从被试的作答反应中挖掘出属性的层级关系。基于此, 本文利用贝叶斯网技术, 从被试的作答数据中挖掘出属性之间的层级关系。

2.3 贝叶斯网的相关概念

贝叶斯网络(Bayesian Networks, BN)又称信念网络(Belief Networks)。贝叶斯网络由两部分组成:有向无环图(Directed Acyclic Graph, DAG)和条件概率分布(Conditional Probability Distribution, CPD),其中有向无环图描述贝叶斯网络结点之间的定性关系,条件概率分布描述贝叶斯网络结点之间的定量关系。贝叶斯网络的理论依据是概率论,推理的基础是基于概率推理,并以图的形式来表达和描述数据实例中的关联或因果关系。贝叶斯网可以表示因果关系,但不局限于因果关系。适用于表达和分析不确定性的事物,能够对不完全、不精确或不确定的知识或信息做出有效的推理,是目前不确定知识表达和推理领域最有效的工具之一。

随着计算机技术的发展,贝叶斯网技术在很多领域都得到了成功应用,许多研究者正致力于将贝叶斯网应用到现代教育测量中,比如 Vomlel (2003)将贝叶斯网应用到技能诊断测量中, Almond (2007)等利用贝叶斯网对基于项目反应理论的认知诊断建模,并且将所得模型应用在细粒度的技能诊断测量上,郭伯臣等对贝叶斯网在教育测量中进行了理论和实践上的研究(Kuo, Hsieh, & Chang, 2006; Shih & Kuo, 2005; Shih, Kuo, & Liu, 2008)。Sewell 和 Shah (1967)利用贝叶斯网结构学习得到了影响高中学生报考大学的因素之间的关系。

在实际的应用中,利用贝叶斯网来处理问题的过程主要分为三步:首先是在领域专家的指导下确定合适的变量及变量的取值范围;其次是确定变量间的依赖关系,并以图的形式表示,即确定网络结构的有向无环图,简称结构学习;最后是确定变量间的分布函数,获得条件概率分布,简称参数学习。

贝叶斯网结构学习的目标是根据数据 D , 并结合先验信息 ζ , 学习出网络结点之间可能的因果关系图 S , 由于得到的因果关系图可能不止一个, 这就需要引入评分机制, 以挑选最优的那个关系图。一般采用 Heckerman (1995)的建议, 使用以贝叶斯统计理论作基础的贝叶斯后验评分函数:

$$P(S|D, \zeta) = P(S, D|\zeta) / P(D|\zeta) \quad (1)$$

选择使公式 1 中的 $P(S|D, \zeta)$ 值最大的结构作为学习所得的结构。其中 $P(D|\zeta)$ 是一个常数, 这样就只需要计算 $P(S, D|\zeta)$, 这个计算的详细推导可参见 Heckerman (1995)。

在现代教育测量中, 不容易获得知识属性之间

的层级关系, 由于贝叶斯网具备了从大量数据中学习出网络结点之间关系的能力, 这里将贝叶斯网作为工具来获得属性间的层级关系, 其中网络结点就是各知识属性, 数据就是被试的作答数据及项目, 通过某种结构学习算法来得到网络结点(属性)之间的层次关系, 由于被试对属性的掌握状态有两种:掌握和未掌握, 即网络中结点的状态都是二值的, 所以这里的网络都是离散的贝叶斯网。

本文主要是考虑在已知测验所考察的属性及被试的属性掌握模式情况下进行结构学习, 得到属性间的层级关系, 具体的做法将在下面介绍。

3 贝叶斯网技术在属性层级结构确定上的应用

将贝叶斯网技术应用到属性层级结构的确定上, 需要用到三个方面的数据, 分别是: 学科专家给出的测验中所考察的属性, 用来作为结构学习中的结点; 测验中各个项目所考察的属性及测验中各被试的作答数据, 它们用来确定被试的属性掌握模式(作为结构学习的训练数据)。利用贝叶斯网技术, 结合被试的反应完成结构学习, 可能提供和专家给出的属性层级结构不完全相同。这可以作为参考信息, 根据这个参考信息对给出的结构进行评估和修正。由于本研究中所采用的是 0, 1 评分的计分方式, 因此贝叶斯网中各个结点的状态有两个, 分别对应得 1 分和得 0 分的情况。

3.1 通过贝叶斯网结构学习得到属性间的层级关系

如果给出的属性之间的层级关系是不准确的, 必然影响到诊断测量的准确性和可靠性。由于属性之间的层级关系反映的是被试的认知过程和结构, 因此被试的属性掌握模式也能够反映属性间的层级关系。下面介绍贝叶斯网如何从被试的属性掌握模式中得到属性间的层级结构。

对于一个测验项目, 可以通过领域专家和有经验的老师来确定它所考察的属性(也可称为知识点)。对于测验的所有项目, 可以得到它们所考察的属性。因此, 根据被试对测验中的所有项目的得分向量与各个项目所考察的属性, 我们就可以得到所有被试的属性掌握模式。比如, 某个测验中考察了三个知识点(属性), 记为 A, B, C; 测验中共有 6 个项目, 记为(1 0 0) (1 1 0) (1 0 1) (0 1 0) (0 0 1) (1 1 1), 其中, 项目(1 0 0)表示该项目考察了属性 A, 没有考察属性 B 和 C; 有 5 个被试参加了测验, 作

答反应分别为(1 0 1 0 1 0) (0 0 0 0 0 0) (1 1 0 1 0 0) (1 1 1 0 1 0) (1 0 0 0 0 0), 其中, (1 0 1 0 1 0)表示该被试正确作答项目 1, 3, 5, 其余的项目作答错误, 而项目 1, 3, 5 分别包含属性 1, 3; 将该被试的作答反应向量中正确作答的项目的属性进行“或”运算, 得到这个被试的属性掌握模式为(1 0 1), 仿照这种做法, 得到各被试的属性掌握模式为(1 0 1) (0 0 0) (1 1 0) (1 1 1) (1 0 0), 这就是贝叶斯网结构学习的数据, 通过学习可得到属性结点 A, B, C 之间的关系图。

根据被试对测验项目的得分向量与项目所考察的属性, 可以得到所有被试的属性掌握模式。将得到的属性掌握模式作为贝叶斯网结构学习的训练数据集进行结构学习, 可以得到属性之间的层级关系。这里采用的结构学习算法是 K2 算法(Cooper & Herskovits, 1992), 其中用到的评分方法是贝叶斯评分方法(Heckerman, 1995)。K2 算法进行结构学习的具体过程为: 对给定的拓扑顺序(在这里, 指的是各知识属性的掌握顺序, 比如要想掌握属性 B,

必须先掌握属性 A, 则在拓扑顺序中 A 排在 B 的前面)中的每个结点, 首先计算每个结点在其父结点集合为空时网络结构的评分(用来评估该结构的优劣), 然后将该拓扑顺序中其前面的结点的不同组合依次作为其父结点集合, 计算新网络结构的评分, 找出评分最高的网络结构, 并在评分最高的网络结构中得到该结点的父结点集合, 这个过程直到所有结点的父结点集合都找到。

3.2.1 模拟实验 为了验证对被试的属性掌握模式进行贝叶斯网结构学习得到属性层级结构的准确性, 我们进行 Monte Carlo 模拟研究。为比较不同层级结构下的结构学习的准确性, 采用 Cui, Leighton 和 Zheng (2006)中的 4 种属性层级结构(如图 1 所示)来进行模拟。由图 1 所示的 4 种属性层级结构用 Tatsuoka (1995)或者丁树良等人(2009b)介绍的方法依次求出邻接阵、可达阵、关联阵、简化关联阵和期望属性掌握模式, 可计算得到非零的期望模式个数: 发散型是 25 个, 收敛型是 8 个, 线型是 7 个, 无结构型是 64 个。

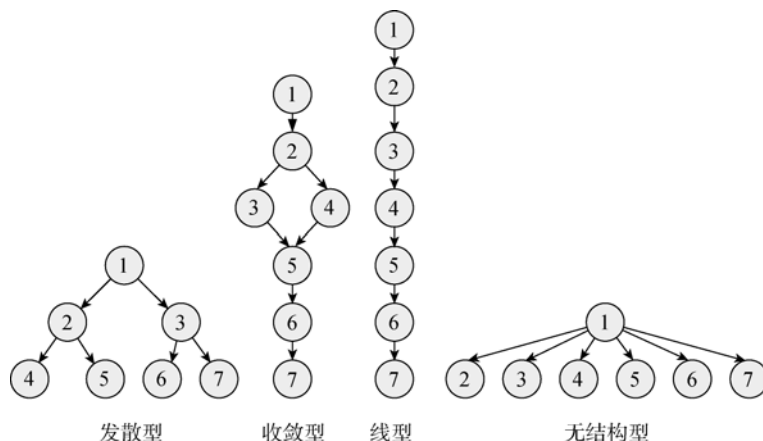


图 1 四种属性层级结构

由于实际的测验中, 我们可以由专家或老师得到每个项目所考察的属性及它们之间的拓扑顺序。此处按照 Tatsuoka 的方法先给出测验中的项目所考察的属性及它们之间的拓扑顺序, 这可以作为属性之间层级关系的一个“初值”或“参考值”。为了考察贝叶斯网络训练出的属性层级结构是否符合实际情况, 我们开展 Monte Carlo 模拟, 其做法如下。

第一步: 模拟含有“失误”或“猜测”的被试反应。把期望反应模式按总分从小到大排序, 然后使具有这些得分的被试人数满足标准正态分布, 产生 5000 个人进行分配, 其中总分相同的期望反应模式平均分配人数, 这样保证期望反应模式的总得分

服从正态分布。为了产生发生失误或猜测的观察反应模式, 我们按如下的方法模拟: 如果要模拟概率值为 β 的失误(slip)或猜测(guess)(如 $\beta=0.05$), 对任一期望反应模式 X_α 中任一分量 X_{aj} , 产生一个服从 $U(0,1)$ 的随机数 r , 如果 $r > 1-\beta$, 则该期望反应模式 X_α 在第 j 题目上的得分 X_{aj} 变成 $1-X_{aj}$ 。如果 $r < 1-\beta$, 则该期望反应模式在该题上的得分不变。对所有期望反应模式中的所有分量都按这个方法操作, 这样就模拟产生一个有失误或猜测概率为 β 的观察反应模式阵, 然后把这些观察反应模式分类到期望反应模式中, 用发生失误前的属性模式作为真值, 然后计算属性模式分类的正确率来比较方法

的好坏。用这个方法产生包含 5%、10%、15% 的失误或猜测的概率的数据。

第二步：结构学习。根据模拟得到的观察反应模式和测验项目，将被试正确作答的项目包含的属性向量进行“或”运算，得到所有被试的属性掌握模式，将它作为结构学习的数据集。根据四种层级结构下得到的被试的属性掌握模式(包含不同的猜测率和失误率)，进行结构学习。

这里以收敛型结构为例来描述贝叶斯网结构学习的过程：涉及到七个结点(对应七个属性)，这里分别是用①, ..., ⑦来表示，每个结点分别有两个状态，这里用 1 和 0 表示，可以理解为考生对该结点掌握/未掌握两种情况；测验中所用到的项目(共 8 个)是事先准备好的，各个项目所考察的属性也是已知的(如图 2 所示)再加上每个被试的作答向量，根据前面所介绍的“或”运算得到所有被试的属性掌握模式，把属性掌握模式作为结构学习的训练集。将结点和训练集作为输入，经过结构学习，将学习所得的结构作为输出。

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
A_1	1	1	1	1	1	1	1	1
A_2	0	1	1	1	1	1	1	1
A_3	0	0	1	0	1	1	1	1
A_4	0	0	0	1	1	1	1	1
A_5	0	0	0	0	0	1	1	1
A_6	0	0	0	0	0	0	1	1
A_7	0	0	0	0	0	0	0	1

图 2 测验中的项目(收敛型结构)

注： $I_1, \dots, I_8$ 表示项目， $A_1, \dots, A_7$ 表示属性

结构学习的结果如表 1(表 1 表头 0%、5%、10%、15% 分别表示用不存在猜测和失误的数据进行学习，用存在 5% 的猜测和失误的数据进行学习，等等)。从结果中可以看出，在不存在猜测和失误时，结构学习能得到正确的结构(即与原始结构一致)；当存在 5% 的猜测和失误时，结构学习仍能得到正确的结果；当存在 10% 和 15% 的猜测和失误时，收敛型结构都缺少了从属性 2 到属性 3 之间的一条边；而当存在 15% 的猜测和失误时，发散型结构少了一条连接属性 1 和属性 3 之间的边。

3.2.2 实证研究

(1) 属性及层级关系的确定

本研究诊断高二学生对不同进位计数制及其相互转换的技能(陆云娜, 2008)掌握情况。首先，根据多个相关的学科专家一起对这部分的知识技能进行了界定，确定了七种属性，并对这七种属性间的关系进行分析，理清它们之间的层级关系(如图 3)。根据属性的层级关系模式编制了 16 道项目来测试这七种属性，每题所含属性均清晰明了。

图 3 中 A_1 表示数的概念， A_2 表示基数， A_3 表示位权， A_4 表示十进制转换成其它进制， A_5 表示其它进制转换成十进制， A_6 表示二进制转换成八进制或十六进制， A_7 表示八进制或十六进制转换成二进制。

(2) 试题的编制

根据所要考查的属性编写测验项目，分别含有 1 至 7 种属性的测验项目应该有 $2^7 - 1 = 127$ 种，但由于属性间存在着图 3 规定的层级关系，故只存在 19 种符合属性间层级关系的项目，这 19 个项目所对应的 Q 阵可以根据可达阵，通过扩张算法(丁树良等, 2009b)获得。Q 阵中的 1 表示该项目考察了该属性，0 表示没有考察该属性(图 4)。

根据已确定的 7 个属性及其层级关系，我们编制了 16 道试题来测试被试在本研究所指向领域的知识结构。

选取高二的学生，用一节信息课进行施测。发下试卷 240 份，回收 236 份，因为当时使用规则空间模型分析，满分试卷不便使用，扣除满分试卷后可用于分析的有效试卷 189 份。

(3) 分析步骤

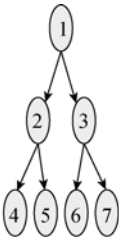
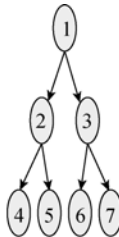
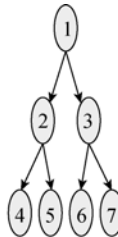
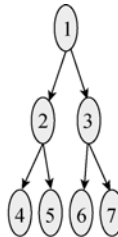
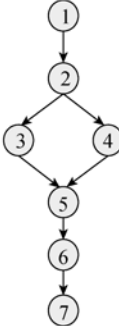
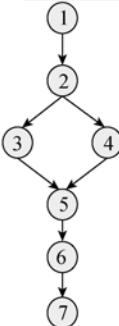
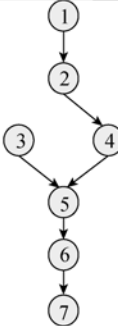
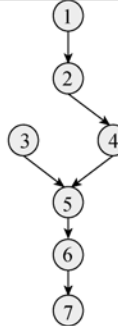
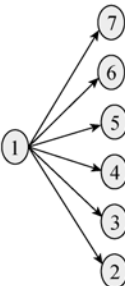
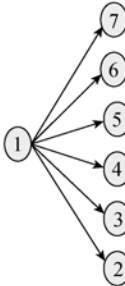
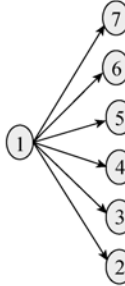
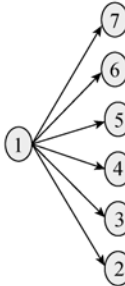
① 得到所有被试的属性掌握模式

根据被试对项目的作答及各项目所考察的属性，利用前面介绍的“或”运算得到被试的属性掌握模式

② 得到属性之间的层次关系

将①中得到的数据作为贝叶斯网结构学习的训练数据集，通过结构学习所得到的属性层次结构如图 5 所示：通过对比结构学习所得到的结果和编制测验前界定的层级关系的对比可以看出，两种层次结构差距很小，结构学习所得到的层级结构多了一条从属性 2 到属性 3 的弧，通过对项目、属性 2 和属性 3 之间的关系进行进一步分析，结构学习所得到的层次关系可能更符合实际的情况。比如一个二进制数 $(10111)_2$ ，它对应的十进制数为 23，而对应

表 1 模拟实验中结构学习的结果

结构	失误/猜测率			
	0%	5%	10%	15%
发散型				
				
				
				

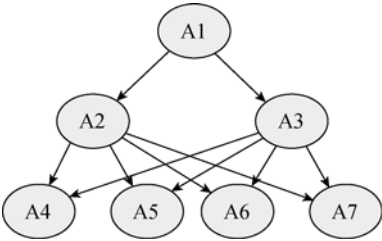


图 3 七个属性之间的层级关系

的八进制数为 $(27)_8$ 和 16 进制数为 $(17)_{16}$ 。对应于

$(27)_8$ 中十位数 2 的位权为 8, $(17)_{16}$ 中十位数 1 的位权为 16, 可见位权和基数有关, 但原来命题人员并没有考虑到这一点, 故 16 个项目中没有出现相应的测试项目。可见应用贝叶斯网结构学习是很重要的, 它可以从被试反应数据中挖掘出有用的信息, 反馈给命题专家参考。有了更加符合实际的层级结构, 就应该对 Q 阵进行修改, 重新作认知诊断。国外十分注意对 Q 阵的修改工作(de la Torre, 2008; Rupp & Templin, 2008)

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	1	0	1	1
0	0	0	0	0	1	0	0	1	0	0	1	1	0	1	1	1	0	1
0	0	0	0	0	0	1	0	0	1	0	1	0	1	1	0	1	1	1
0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	1	1	1

图 4 可能的测验项目

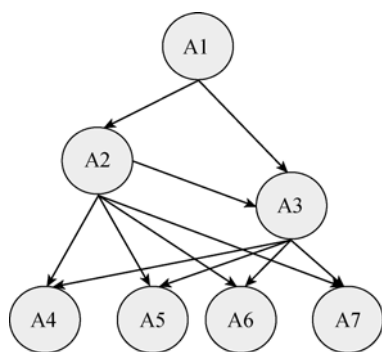


图 5 通过结构学习所得到的属性层级关

图 3 与图 5 的区别在于图 5 中多了从属性 2 到属性 3 的弧, 由于贝叶斯网结构学习会将有关联的属性之间加上弧, 用图 5 作为哈斯图(Hasse diagram; 左孝凌, 李为鑑, 刘永才, 1982)是不对的, 因此, 要将属性 1 和属性 3 之间的弧删除, 将图 5 改成哈斯图(参见图 6), 用贝叶斯网结构学习所得到的结构在教育测量中还要进行稍微的“修匀”。

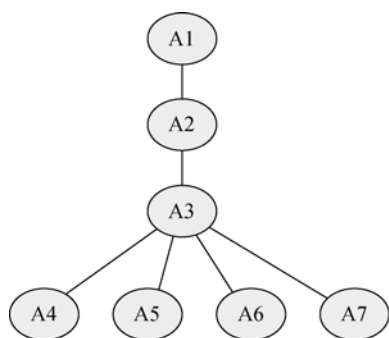


图 6 图 5 所对应的哈斯图系

从对模拟实验及实证研究的结果中可以看出, 利用贝叶斯网对被试的属性掌握数据经结构学习后得到的结果是有参考价值的, 可以为命题或测量专家带来有用的信息, 即如果将贝叶斯网络训练出

的结构反馈给命题和测量专家, 对考试结果的分析、以后的测验及教学计划制定提供有用的信息。属性间的层次关系反映了属性之间的内在联系, 反映了被试对属性的掌握顺序, 如果层级关系设置错误, 则会导致诊断结果与实际情况严重不一致的情况。正确的属性间的层次关系, 对于编制优质的测验项目是非常有必要的; 同样, 对于教师进行补救教学也是很关键的, 因为, 在诊断测验中, 根据考生的作答数据及正确的层次关系, 可以得出考生的属性掌握模式, 进而教师可以进行有针对性的补救教学。

4 总结与展望

利用贝叶斯网结构学习得到属性之间的层级关系, 是非常有实际意义的, 对 RSM 来说, 可以验证从 Q 阵中分析得到的属性层级关系; 对于 AHM 来说, 也可对事先确定的属性层级关系进行验证。利用贝叶斯网进行结构学习不需要以项目反应理论中的项目参数和被试参数作为基础, 只需要有项目和作答数据就可以完成。将这个办法在更多的实测数据中验证, 是迫切需要做的。应用本方法需要事先确定测验中所考察的属性及它们之间的拓扑顺序, 这是本方法的一种局限, 因此研究直接从作答数据中得到测验所考察的属性及它们之间的拓扑顺序和层次关系的方法就显得十分重要。本方法对属性层级关系是否有较大的参考价值, 受到被试作答反应质量的影响, 即作答反应所含误差(slip)越小, 贝叶斯网训练出的结构越准确; 数据所含误差越大, 所得到的结果的准确性会降低, 因此, 对作答反应数据中所含误差大小进行判断是需要进行深入研究的问题。另外, 将贝叶斯网技术应用到现代教育测量的其它方面: 如认知诊断测验的编制、认知诊断的分类等也值得深入研究。

规则空间模型分析和 IRT 分析一般需要大样本, 虽然本文并非应用规则空间模型进行分析, 也不是应用 IRT 进行参数分析, 但是贝叶斯网络在认知诊断中的应用也对样本大小是较敏感的。本文中的实证研究部分样本量相对较小。一般来说, 由相对较小的样本量导出的结果和对该结果的分析只能作为参考。

参 考 文 献

- Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J. D. (2007). Modeling diagnostic assessment with Bayesian networks. *Journal of Educational Measurement*, 44(4), 341–359.
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74(4), 619–632.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, 309–347.
- Cui, Y., Leighton, J. P., & Zheng, Y. G. (2006). Simulation studies for evaluating the performance of the two classification methods in the AHM. Paper presented at NCME, San Francisco, CA.
- Dai, H. Q., & Zhang, Q. H. (2004). An applied research of the Rule Space Model in the identification of statistic study pattern (in Chinese). *Psychological Science*, 27(4), 949–951.
- [戴海崎, 张青华. (2004). 规则空间模型在描述统计学习模式识别中的应用研究. *心理科学*, 27(4), 949–951.]
- de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: development and applications. *Journal of Educational Measurement*, 45(4), 343–362.
- Ding, S. L., Wang, W. Y., & Yang, S. Q. (2009a). A principle of construction of Cognitive Diagnostic Test. *Chinese Science paper Online*.
- [丁树良, 汪文义, 杨淑群. (2009a). 认知诊断测验编制的原则. *中国科技论文在线*.]
- Ding, S. L., Yang, S. Q., & Wang, W. Y. (2010). Importance of reachability matrix in constructing Cognitively Diagnostic Testing. *Journal of Jiangxi Normal University*, 34(5), 490–494.
- [丁树良, 杨淑群, 汪文义. (2010). 可达矩阵在认知诊断测验编制中的重要作用. *江西师范大学学报*, 34(5), 490–494.]
- Ding, S. L., Zhu, Y. F., Lin, H. J., & Cai, Y. (2009b). Modification of Tatsuoaka's Q matrix theory. *Acta Psychologica Sinica*, 41(2), 175–181.
- [丁树良, 祝玉芳, 林海菁, 蔡艳. (2009b). Tatsuoaka Q 矩阵理论的修正. *心理学报*, 41(2), 175–181.]
- Dong, D. Y. (2007). Research of application foundation on Bayesian networks. *Doctoral dissertation*. Jilin University.
- [董立岩. (2007). 贝叶斯网络应用基础研究. *博士学位论文*, 吉林大学.]
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37(6), 359–374.
- Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2000). Exploring the logic of Tatsuoaka's Rule-Space Model for test development and analysis. *Educational Measurement: Issues and Practice*, 19, 34–44.
- Hartz, S., Roussos, L., & Stout, W. (2002). A Bayesian framework for the unified model for assessing cognitive abilities: blending theory with practicality. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Heckerman, D., Geiger, D., & Chickering, M. (1995). Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20, 197–243.
- Junker, B. M., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258–272.
- Kuo, B. C., Hsieh, T. Y., & Chang, Y. Y. (2006). Combining multiple Bayesian networks for modeling students' learning bugs and skills (pp.13–15). The 7th International Conference on Intelligent Technologies, Taipei, Taiwan.
- Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoaka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205–237.
- Lu, Y. N. (2008). Application of Rule Space Model in diagnosing the different carrying notation and the interconversion skill. *Master's Dissertation*. Jiangxi Normal University.
- [陆云娜. (2008). 规则空间模型在进位计数制诊断性测验中的应用. *硕士学位论文*, 江西师范大学.]
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64(2), 187–212.
- Rupp, A. A., & Templin, J. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68(1), 78–96.
- Sewell, W. H., & Shah, V. P. (1967). Socioeconomic status, intelligence, and the attainment of higher education. *Sociology of Education*, 40(1), 1–23.
- Shih, S. C., & Kuo, B. C. (2005). Using Bayesian networks for modeling students' learning bugs and sub-skills. *Lecture Notes in Artificial Intelligence*, 3681, 69–75.
- Shih, S. C., Kuo, B. C., & Liu, Y. L. (2008). Building and applying the Bayesian networks based adaptive test system-using rounding & estimating with decimals unit in the fifth grade as a example. The 16th International Conference on Computers in Education ICCE 2008 (pp.27–37), Taipei, Taiwan.
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: a statistical pattern recognition and classification approach, In P. D. Nichols, S. F. Chipman and R. L. Brennan. *Cognitively Diagnostic Assessment* (pp.327–361). Lawrence Erlbaum Associates, Hillsdale.
- Tatsuoka, K. K. (2009). Cognitive assessment: an introduction to the rule space method (pp.75–83). Routledge Academic, USA.
- Vomlel, J. (2003). Bayesian networks in educational testing. *International Journal of Uncertainty, Fuzziness and Knowledge-Based System*, 1, 83–100.
- Wen, J. B. (2003). Application of the rule space model in computerized adaptive testing for diagnostic assessment (in Chinese). *Doctoral's dissertation*. Chinese University of Hong Kong.
- [文剑冰. (2003). 规则空间模型在诊断性计算机自适应测验中的应用. *博士学位论文*, 香港中文大学]
- Xin, T., & Jiao, L. Y. (2006). The new progress in the theory of measurement: the Rule Space Model (in Chinese). *Journal*

- of East China Normal University (Educational Sciences), 24(3), 50–61.
- [辛涛, 焦丽亚. (2006). 测量理论的新进展: 规则空间模型. 华东师范大学学报(教育科学版), 24(3), 50–61.]
- Yu, N., & Xin, T. (2009). New progress in cognitive diagnostic theory. *Examinations Research*, 5(3), 22–34.
- [余娜, 辛涛. (2009). 认知诊断理论的新进展. 考试研究, 5(3), 22–34.]
- Yu, J. Y. (1995). Application of Rule Space Model for identifying cognitive errors in problem solving (in Chinese). *Acta Psychologica Sinica*, 27(2), 196–203
- [余嘉元. (1995). 运用规则空间模型识别解题中的认知错误. 心理学报, 27(2), 196–203.]
- Zhu, Y. F., & Ding, S. L. (2009). A polytomous extension of attribute hierarchy method based on graded response model. *Acta Psychologica Sinica*, 41(3), 267–275
- [祝玉芳, 丁树良. (2009). 基于等级反应模型的属性层级方法. 心理学报, 41(3), 267–275.]
- Zuo, X. L., Li, W. J., & Liu, Y. C. (1982). Discrete mathematics (pp. 280–356). Shanghai Scientific and Technological Publishing Press.
- [左孝凌, 李为鑑, 刘永才. (1982). 离散数学(pp.280–356). 上海: 上海科学技术文献出版社.]

Application of Bayesian Networks to Identify Hierarchical Relation Among Attributes in Cognitive Diagnosis

YU Xiao-Feng^{1,2}; DING Shu-Liang¹; QIN Chun-Ying²; LU Yun-Na^{1,3}

(¹ Computer Information Engineer College, Jiangxi Normal University, Nanchang 330027, China)

(² Computer Department, Bozhou teachers College, Bozhou 233500, China) (³ Nanchang No.1 high school, Nanchang 330003, China)

Abstract

It's very significant to correctly identify the hierarchical relation among attributes correctly when constructing a diagnosis test. As we know, there are various methods to identify the hierarchical relation among attributes in different cognitive diagnosis models (CDMs). Rule Space Model (RSM) is a kind of great influence CDM which was developed by Tatsuoaka and her associates. In RSM, the task of attribute identification is performed after the test items have already been developed. And then an incidence Q matrix can be determined which reflects hierarchical relation. However, in Leighton, Gierl and Hunka's (2004) Attribute Hierarchy Method (AHM) the organization of attributes(attributes, number of attributes and attribute hierarchical relation) should be determined before developing the test items. In RSM and AHM the importance of correctly identifying the attributes and their hierarchical relation cannot be overstated, and the attributes and their hierarchical relations serve as the most important input variables to the models because they provide the basis for interpreting the results. The hierarchical relation among attributes describes the domain knowledge structures. Understanding the domain knowledge structures in highly specific detail provides a rational basis for proposing and evaluating potential improvements in the measurement of general proficiency. Otherwise, improvement remains largely a trial-and-error process. It has been proved that is not reliable to get the hierarchical relation by analyzing the test items, even the resulting hierarchical relations are not unique.

Now, structure learning of Bayesian Networks would be introduced to study the hierarchical relation from the examinees' attribute mastery patterns. On the one hand, this paper conducted a simulation study in which Cui, Leighton and Zheng (2006)'s attribute hierarchical relation was employed, and the sample size of 5000 was used. In the simulated study, different slipping rates was considered for the purpose of testing the robustness of Bayesian structure learning. On the other hand, this paper conducted an empirical study to research the attributes which were commonly used in different numeration representation system converting, and the sample size was 189. Also the hierarchical relation generated by Bayesian structure learning and the hierarchical relation extracted by domain expert was compared.

The result of simulation study and empirical study all showed the Bayesian Networks was a very useful tool in determining the hierarchical relation among attributes in the realm of education measurement and psychometrics. It is proved that the result for reference is valuable through structure learning of Bayesian Networks in simulated experiments and practical application. Meanwhile, the Bayesian Networks also can be used to test the hierarchical relation in RSM.

Key words rule space model; cognitive diagnosis; Bayesian networks; structure learning