

基于扩展知识空间的诊断模型^①

钟 舒, 杨贯中

(湖南大学 软件学院, 长沙 410002)

摘 要: 传统的诊断模型并没有考虑到受测者在测试过程中存在着非理性的失误, 但是在实际测试中, 这却是影响学生测试结果不可忽略的因素。已有的诊断模型将知识空间和诊断相结合, 获得受测者的知识状态。但随着测量理论的不不断发展, 人们越来越关注受测者的技能掌握水平。将诊断运用到扩展知识空间中, 提出一种扩展知识空间中的学习诊断模型, 该模型使用项目反应理论的认知诊断方法, 当受测者出现猜测或失误时, 该模型仍可对受测者的技能水平进行有效的诊断。

关键词: 扩展知识空间理论; 认知诊断; 项目反应理论; 概率

Diagnosis Model Based on Extension of Knowledge Space Theory

ZHONG Shu, YANG Guan-Zhong

(School of Software, Hunan University, Changsha 410002, China)

Abstract: The traditional diagnostic model has not considered that there exists the possibility of irrational mistakes when testing. While it is an indispensable factor in the actual test to effect the result of one's test. The existing diagnostic model has combined the knowledge space theory and diagnosis to get the state of knowledge subjects. But with the continuous development of measurement theory, there is growing concern about the skill level of testee. This article applied diagnosis to Extension of Knowledge Space and proposed a diagnosis model with extension of the knowledge space that uses the formula in item response theory. The diagnosis model could still effectively diagnose the skill level of testee when those mistakes happened.

Key words: extending-knowledge space theory; cognitive diagnosis; item response theory; probability

传统的测验通常只给出学生的分数, 而对于学生的认知结构(即学生已掌握哪些知识点, 未掌握哪些知识点)显得无能为力。而测试者希望通过测验能够提供学生的诊断信息, 并认为考察学生的认知结构, 区分相同分数的个体差异对于测验来说, 是非常重要的。随着认知心理学与心理计量学的发展, 新一代测量理论将认知与测量结合起来, 不仅对学生的整体水平做出评估, 同时对学生的认知结构模式化, 利用合适的计量模型进行诊断, 确定受测者的能力水平以及受测者的认知缺陷。认知诊断理论被视为新一代测量理论的核心。

自从 1985 Doignon 和 Falmagne 发表知识空间理

论^[1]的文章以来, 知识空间理论在自适应测试中起着越来越重要的作用。然而随着测量理论的不不断发展, 人们越来越关注受测者的技能掌握水平, 而基于知识空间的测试最终是以确定的试题评价学生的知识结构为目的, 没有普遍意义。如果将测试中的问题映射为技能的集合, 则可以转化通过试题考察学习者的技能水平, 更具有实际意义。文献[2]在知识空间理论上做了技能的扩展, 称为“扩展知识空间理论”。

同时, 上述基于知识空间的诊断模型和具有诊断性功能的自适应测试重点在于选题策略, 认为学生在答题过程中是完全理性的, 不会出现失误的过程。但在实际过程中, 学生由于心理因素或者其它影响, 出

① 收稿时间:2011-03-09;收到修改稿时间:2011-04-05

现失误也是完全可能的。所以，本文将在知识空间理论的基础上，结合扩展知识空间理论和认知诊断方法，提出新的的认知诊断模型，对受测者的技能水平进行诊断。

1 扩展知识空间理论：

文献[3]对知识结构和技能结构的关系作出了明确的定义，定义了技能函数、问题函数，完成知识结构到技能结构的映射，并给出了自适应技能测试的过程。

对于本文将涉及到的扩展知识空间理论中的内容，本节将介绍相关理论和概念。

概念 1 技能结构 (S,T): 令技能域 S 是一个非空有限技能集合，S 的子集组成的集合 T 被称为技能状态，当且仅当：

① K 集合中至少包含空集 \emptyset 和 Q，即 $\emptyset \in T, S \in T$ 。

② 对于任何试题的子集 T1, T2 属于 T, 它们的并 $K1 \cup K2$ 也属于 K。

概念 2 技能映射：技能函数 $\gamma: Q \rightarrow 2^S$ 表示知识域 Q 到技能域 S 的映射关系(显然，对于包含 S 个元素的技能集合，其子集数为 2^S)。 $\gamma(q)$ 表示解决试题 q 所需的最小技能集合的集合。

在实际应用过程中，技能函数与问题函数的构建已经由教师或者领域专家完成。

问题函数 $\delta: 2^S \rightarrow 2^Q$ ，满足关系 $\delta(\emptyset) = \emptyset, \delta(S) = Q, \delta(t)$ 表示技能集合 t 能解决的试题的集合。

在实际应用过程中，技能函数与问题函数的构建已经由教师或者领域专家完成。

概念 3 试题的技能结构：试题 q 包含的所有技能的组合构成的集合称为试题 q 的技能结构。受测者要答对试题 q，则必须掌握 $\gamma(q)$ 中的至少一个技能集合。 $|\gamma(q)|=1$ 时表示解决这个问题 q 的方法只有一个。

例 1: 当 $\gamma(q) = \{\{a, b\}, \{b, c, d\}\}$ ，表示受测者必须掌握技能 a 和 b，或者掌握技能 b、c 和 d，才可以答对试题 q。

概念 4 试题间的前提关系：

与关系 $\wedge: q_0 \wedge (q_1, q_2, \dots, q_n)$ ，如果会做试题 q_0 ，则一定会 q_1, q_2, \dots, q_n 中的所有试题。

或关系 $\vee: q_0 \vee (q_1, q_2, \dots, q_n)$ ，如果会做试题 q_0 ，则至少会 q_1, q_2, \dots, q_n 中的一道试题。

例 2: 某次测验需要考察 5 个技能(a,b,c,d,e)，针

对这 5 个技能设计的 5 个试题的技能结构为：

$\gamma(q1) = \{\{a\}\}, \gamma(q2) = \{\{b\}\}, \gamma(q3) = \{\{a,c\}, \{b,c\}\}, \gamma(q4) = \{\{d,e\}\}, \gamma(q5) = \{\{a,c,d,e\}, \{b,c,d,e\}\}$ 。
由此建立的试题间前提关系的与或图如图 1 所示。

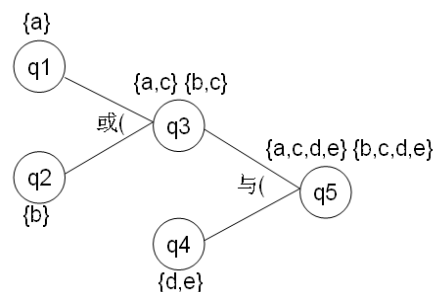


图 1 试题间前提关系的与或图

概念 5 答题状态：对于每个试题，受测者有答对或者答错两种结果，对一组试题的答题结果称为一个答题状态，答题状态用 L 表示。这里用 0 表示受测者答错试题，1 表示受测者答对试题。易知对于 n 个试题有 $2^n - 1$ 个可能的答题状态，而这些答题状态可用一串二进制数标识。本文中用这串二进制对应的十进制数值 i 来标识此答题状态，用符号 L_i 表示（见例 3）。

概念 6 相容答题状态：由于试题之间存在前提关系的约束，所以理想情况下，某些答题状态不能出现。不与试题之间的前提关系发生冲突的答题状态称为相容答题状态；反之则称为不相容答题状态。

例 3: 对例 2 中的 5 道试题 {q1q2q3q4q5} 进行测试，则有 $2^5 = 32$ 种可能的答题状态。如答题状态 $L_1(00001)$ 代表只答对了 q5； $L_{10}(01010)$ 代表只答对了 q2 和 q4。其中， L_1 是一个不相容答题状态，因为根据试题间的前提关系，答对 q5 一定会答对 q3 和 q4；而 L_{10} 是一个相容状态。根据试题间的前提关系，可得例 2 中所有的相容答题状态： $\{L_0, L_2, L_8, L_{10}, L_{12}, L_{14}, L_{15}, L_{16}, L_{18}, L_{20}, L_{22}, L_{23}, L_{24}, L_{26}, L_{28}, L_{30}, L_{31}\}$ ，答题状态 $\{L_1, L_3, L_4, L_5, L_6, L_7, L_9, L_{11}, L_{13}, L_{17}, L_{19}, L_{21}, L_{25}\}$ 为不相容答题状态。

2 认知诊断方法

据上文中在例 3 中通过 5 道试题的前提关系得到所有相容答题状态之后，当受测者作答试题状态为相容试题状态时，我们可以直接推断出他的真实技能状态。而判断当受测者出现某个不相容状态时候他的实

际技能状态成了我们主要考虑的问题。此时，我们认定受测者出现了某些原因的失误，为了推测出受测者真实的知识状态，本文以项目反应理论为基础，将与不相容的答题状态拥有“最大相似概率”的某相容答题状态看做受测者的诊断答题状态，然后将此相容状态对应的技能状态作为诊断结果反馈给测试者。

2.1 项目反应理论模型的确定和项目反应概率的估计

项目反应理论 (Item Response Theory, IRT) 是一种测量理论，它反应被测试者能力与他们对项目的反应 (正确作答概率) 之间存在的关系并以数学函数的形式表示出来。

首先定义项目反应的数学模型，三参数的 IRT 模型如下： a 是题目 j 的区分度参数， b 是题目 j 的难度参数， c 是题目 j 的伪机遇水平参数， a, b, c 参数可以预先估计求得。该模型表示能力为 θ 的学生答对题目 j 的概率。当 $c_j=0$ ，就是二参数 IRT 模型，本文中采用二参数的 IRT 模型，参数 a 和 b 可以通过极大似然估计法获得。

$$p_j(\theta) = c_j + \frac{1 - c_j}{1 + e^{-1.7a_j(\theta - b_j)}}$$

2.2 最大相似概率

当受测者出现不相容答题状态时，可以推断该受测者在答题过程中出现了猜测或者失误。此时，将被测者的某不相容答题状态与相容答题状态逐个比较，求出不相容答题状态与每个相容答题状态相似的概率，最后将这个不相容答题状态归于有最大相似概率的相容答题状态，并将这个相容状态对应的知识状态作为诊断结果反馈给测试人员。同时能反应受测者的真实答题情况以及其最可能掌握的对知识状态。

2.3 相似概率的具体求法

用某相容答题状态向量减去被测不相容答题状态，得到一个向量 d ，当 d 的某个分量为 -1 时，称为出现了 $0 \rightarrow 1$ 型失误，即某相容答题状态在此分量 (即题目) 上为 0 ，被测不相容答题状态在此分量上为 1 ，也就是说，受测者预计在此分量上错误作答，而被测时却回答正确。根据公式 1，能力值为 θ 的受测者在第 j 个分量对应的题目上答对的概率为 $P_j(\theta)$ 。当 d 的某个分量为 1 时，称为出现了 $1 \rightarrow 0$ 型失误，即某相容答题状态在此分量 (即题目) 上为 1 ，被测不相容答题状态在此分量上为 0 ，也就是说，受测者预计在此分量上正确作答，而被测时却回答错误。根据公式 1，

能力值为 θ 的受测者在第 j 个分量对应的题目上答对的概率为 $1 - P_j(\theta)$ 。由此可以得出，一个被测者在 n 个题目上连续出现 k 个 $0 \rightarrow 1$ 型失误和 m 个 $1 \rightarrow 0$ 型失误的概率为：

$$P_{result}(\theta) = \prod_{k=1}^K p_j(\theta) \prod_{m=1}^M [1 - p_j(\theta)]$$

其意义为，如果将这个不相容答题状态看做某个相容答题状态的“相似状态”，根据受测者的能力所求得的一系列失误的概率。当相应的 $P_{result}(\theta)$ 为最大时，我们就判定该受测者拥有取得最大值的相容答题状态。

3 认知诊断过程

3.1 求出所有相容答题状态

根据所考察的技能结构确定试题间的前提关系，计算出所有相容答题状态 L 和不相容答题状态 L' 。本文中例 2 中试题构成的相容答题状态如表 3 第 2 列所示。

3.2 项目参数和能力参数的估计

本文采用通用 IRT 分析软件 BILOG 估计区分度参数 a 和难度参数 b ，参数精度为 0.001。估计受测者能力参数如表 1。

表 1 区分度参数和难度参数表

区分度参数 a	难度参数 b
-0.042	-0.054
-0.042	-0.054
0.364	0.457
0.310	0.392
0.236	0.325

3.3 受测不相容答题状态的最大相似概率答题状态求取

在测试中，如果受测者被测得到不相容试题状态 $L7$ (00111)，将不相容试题状态 $L7$ (00111) 与所有相容试题比较，求出它与每一个相容试题状态的相似概率，取得最大相似概率的相容答题状态可以看做受测者真实的答题状态，其对应的技能状态则为受测者真实的技能水平。

求取概率过程如下：首先将 (00111) 与相容状态 $L0$ (00000) 相比，用 $L0$ 向量减去 $L0$ ，得到一个向量 $d(0,0,-1,-1,-1)$ ，即可看做受测者在题目 $q3, q4, q5$ 上出现了 3 个 $0 \rightarrow 1$ 型失误。使用二参数的 IRT 模型的 IRT 模型，得到相似概率为 0.0260。将 $L7$ (00111) 与所

有相容答题状态进行比较后,分别求出它们相似概率,可知与 L2 (00111) 拥有最大相似概率的相容答题状态为 L23 (10111), 相容概率为 0.5288。即当受测者测试得到答题状态为 (00111) 时,我们认定他由于某些原因,在答题状态 (10111) 上出现了失误,答错了第 1 题,他的实际技能状态为 (a,c,d,e)。

表 2 相容答题状态集

L_i	$q_1q_2q_3q_4q_5$	θ	答题状态
L_0	00000	-1.2675	Φ
L_2	00011	-1.2800	{d,e}
L_8	01000	-0.4625	{b}
L_{10}	01010	-0.2215	{b,d,e}
L_{12}	01100	-0.2299	{b,c}
L_{14}	01110	0.6881	{b,c,d,e}
L_{15}	01111	1.5620	{b,c,d,e}
L_{16}	10000	-0.4625	{a}
L_{18}	10010	-0.2215	{a,d,e}
L_{20}	10100	-0.2299	{a,c}
L_{22}	10110	0.6881	{a,c,d,e}
L_{23}	10111	1.5620	{a,c,d,e}
L_{24}	11000	-0.4303	{a,b}
L_{26}	11010	0.4422	{a,b,d,e}
L_{28}	11100	0.4904	{a,b,c}
L_{30}	11110	1.4355	{a,b,c,d,e}
L_{31}	11111	2.4400	{a,b,c,d,e}

4 结论

本文结合扩展知识空间理论和认知诊断理论建立

新的认知诊断模型,在认知诊断过程中,考虑到受测者的答题过程并不是完全理性的,通过计算“最大相似概率”来得到受测者的真实技能状态。根据项目反应理论中的方法可知,试题的技能结构可影响 IRT 模型中的公式结果,也就是影响相似概率,如何通过建立试题的技能结构提高诊断精度是本文未解决的问题。

参考文献

- 1 Doignon, JP, Falmagne JCL. Knowledge Spaces. Berlin: Springer Verlag,1999.50-55.
- 2 孙波,傅骞.扩展知识空间理论研究.中国电化教育,2004,(25)4:74-77.
- 3 刘艳花.基于扩展知识空间理论的技能自适应测试过程.计算机系统应用,2010,19(7):69-74.
- 4 林伯成.利用知识地图诊断数学问题之研究[硕士学位论文].台湾:中原大学资讯工程所,2002.
- 5 Steinberg L, Thissen D. Item response theory in personality research. Personality research, methods, and theory: A Festschrift honoring Donald W.Fiske. Hillsdals NJ: Lawrence Erlbaum Associates,1995,161-181.
- 6 胡麒,何华灿.基于试题空间的学习诊断方法.微计算机信息,2007,23(10):238-240.
- 7 Albert D, Hockemeyer C, Wesiak G. Current Trends in e-Learning based on Knowledge Space Theory and Cognitive Psychology. Psychologische Beiträge, 2002,44(4):478-494.

(上接第 26 页)

- 5 赵林海,李晓风,谭海波.基于 CACTI 的分布式 ORACLE 监控系统的设计与实现.计算机系统应用,2010,19(9):134-137,133.
- 6 Honser C.OpenVPN and the SSL VPN revolution. Sans Institute, 2004.
- 7 Yonan J. Understanding the User-Space VPN: History, Conceptual Foundations, and Practical Usage. Linux Fest Northwest, 2004.

- 8 郭学超,翟正军.OpenVPN 体系安全性研究.科学技术与工程,2007,7(8):1742-1745.
- 9 岑锐坚.使用 Cacti 监测系统与网络性能.开放系统世界,2006,(7):69-72.
- 10 Case J, McCloghrie K, Rose M, et al. Textual conventions for Version2 of the simple network management protocol (SNMPv2). RFC2579. IETF, 1999.