

数据驱动的认知诊断理论在自适应测试的应用现状与启示

The Research of Application Status of Cognitive Diagnosis Theory in Adaptive Testing and Its

Enlightenment Driven by Data

刘妍，牛雨，祝嘉钰，戴静，石小恋，顾小清*

华东师范大学

* xqgu@ses.ecnu.edu.cn

【摘要】 随着科学技术的进步，计算机技术正在逐步与认知诊断理论结合形成计算机化自适应认知诊断测验（Computerized Adaptive Test for Cognitive Diagnosis，简称 CD-CAT）。本文运用内容分析法，以“计算机自适应认知诊断测验”为关键词开展搜索，研究结果表明：中国于 2006 年开始逐步开展 CD-CAT 研究，近几年关注点主要聚焦于理论概况（占总样本的比例为 15%）和现状综述（占总样本的比例为 74%）。用 UCINET 将“研究者派系”可视化，发现领域科研合作紧密，但是整体合作疏远，小团队更加集中。此外，阐述国内外 CD-CAT 实践案例，指出认知诊断评价在教育领域的发展方向 and 策略，期望对读者有所启发。

【关键字】 K-12；数据驱动；认知诊断（模型，Cognitive Diagnosis Model，CDM）；计算机化自适应认知诊断测验（CD-CAT）

Abstract: With the progress of science and technology, computer technology is gradually combining with cognitive diagnosis theory forming a computerized adaptive cognitive diagnostic test (computerized adaptive test for cognitive diagnosis, referred to as CD-CAT). Content analysis method was used in this paper and "computer adaptive cognitive diagnostic test" was the keyword to carry out the research. The research results show that: the research of CD-CAT in China began in the year of 2006 and gradually developed. In recent years, the main focuses are on theoretical overview (accounted for the proportion of the total sample of 15%) and review (accounted for the proportion of the total sample of 74%). The researcher factions were visualized by the UCINET revealing that research collaboration in the field is close but the overall cooperation from a small team is more concentrated. In addition, we elaborated the domestic and foreign CD-CAT practice and pointed out the development direction and strategy of Cognitive Diagnostic Assessment in the field of education, hoping to inspire readers.

Keywords: K-12, Big Data, CDM, CD-CAT

1. 研究背景

随着计算机技术、通讯设备和网络技术的迅速发展，教育信息化的发展面临着重大变革，由数据主导的“大时代”正在走入各个领域，同时渗透到教育发展与改革的战略中。明确教育的大数据、教育的个性化迫切需求，同时复杂的、数据驱动的个性化教学指导很多时候是非线性的辅导，需要能够提供满足学习者个别化、具有差异的认知水平和能力水平的诊断评估；以认知诊断测验为代表的新一代测验理论应运而生，能够解释学生在推理和解决问题过程中所用到的知识和技能，并提供关于学生在知识技能上的掌握情况的诊断信息，同时保证测验的“公平性”，将“因材施教”和“个性化”教学成为可能。

Frederiksen 等（1993）编著《Test theory for a new generation of tests》正式出版，标志着新一代测量理论的诞生，即认知诊断评估逐步走入历史舞台，强调在“能力水平”和“认知水平”同时能够被诊断，不再仅仅是单一的宏观维度的“知识水平”（涂冬波、蔡艳和丁树良，2012）。

Wu, Y.-T., Chang, M., Li, B., Chan, T.-W., Kong, S. C., Lin, H.-C.-K., Chu, H.-C., Jan, M., Lee, M.-H., Dong, Y., Tse, K. H., Wong, T. L., & Li, P. (Eds.). (2016). *Conference Proceedings of the 20th Global Chinese Conference on Computers in Education 2016*. Hong Kong: The Hong Kong Institute of Education.

计算机化自适应诊断测验 (Computerized Adaptive Test for Cognitive Diagnosis, CD-CAT) 是计算机化自适应测验理论和认知诊断理论结合的产物, 宗旨是“因材施教, 量体裁衣”式的自适应诊断, 根据学习当前状态, 选取与之相适应的题目进行测试, 进而诊断学习者能力水平, 得出诊断报告, 提供提升建议和策略。

综上所述, 为了从整体上把握 CD-CAT 在国内的发展脉络, 厘清发展现状和研究聚焦, 提出实施 CD-CAT 科研建议和实施策略; 分别从以下方面进行分析, 一方面以时间 (2005 年~2015 年) 为轴, 考察关于“认知诊断自适应测试”文献的研究情况, 采用内容分析方法对已有百余篇中英文文献进行内容分析, 并通过数据的可视化表征和社会网络分析工具, 探讨其研究重心和内容; 另一方面, 通过分析国内外已有 CD-CAT 平台和项目实践情况, 总结和借鉴国外实施经验, 为我国今后的 CD-CAT 研究提供发展策略。

2. 研究过程

2.1. 研究方法 with 工具

本研究主要采用内容分析法和数据可视化的方法, 对已有文本内容进行不同方面的, 由浅入深, 层层递进的线性分析。借助软件管理软件 NoteExpress, 数据处理软件 SPSS 20.0 和词云生成平台 Tagxedo。

2.2. 数据收集

采用内容分析法对 2005 到 2015 年国内发表的关于自适应认知诊断的文献进行比较、分析、综合, 从而对于研究内容和现状进行梳理和总结。以“认知诊断”、“认知诊断测验”为关键词进行搜索, 中文学术期刊检索到 140 篇, 学位论文 68 篇, 专业英文论文 66 篇。

2.3. 结果分析

2.3.1. 初步印象-有哪些研究关注点

本研究选取数据可视化-词云的表达方式展示样本研究的核心内容; 选择统计平台 Tagxedo (<http://www.tagxedo.com/app.html>), 中文词云中“认知诊断”、“规则空间模型”、“属性”等频次最高, 领衔关键词的说明了研究领域的聚焦, 其中研究问题主要关注于学业评估诊断等。为了中英文学生文献具有对照意义, 同样选取统计平台 Tagxedo 进行词频统计和可视化处理, 如下图 2 所示。英文词云中显示具体的应用方法 (蒙特卡罗模拟)、语言程序 (R 语言) 知识评估等。



图 1 中文学术论文词云



图 2 中文学术论文词云

2.3.2. 渐入佳境-局部剖析

2.3.2.1. CD-CAT 研究历程分析

借助 NoteExpress 统计学术论文和学位论文发表情况, 以折线图 3 表现发展历程:

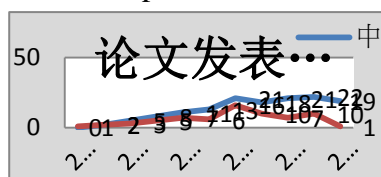


图 3 论文发表年度分布

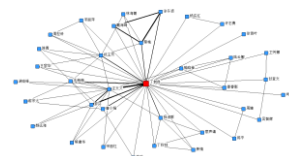


图 4 以“丁树良”为中心合作关系

从图 3 可以看出, 我国从 2008 年开始研究逐步增多, 但是发展起步于 2006 年, 例如 2006 年, 曾彦钧等人提出 CD-CAT 以认知结构为基础, 在题库建设、起始规则、选题策略、被试

Wu, Y.-T., Chang, M., Li, B., Chan, T.-W., Kong, S. C., Lin, H.-C.-K., Chu, H.-C., Jan, M., Lee, M.-H., Dong, Y., Tse, K. H., Wong, T. L., & Li, P. (Eds.). (2016). *Conference Proceedings of the 20th Global Chinese Conference on Computers in Education 2016*. Hong Kong: The Hong Kong Institute of Education.

知识状态估计、终止规则和曝光率等部分与传统的 CAT 很大区别(曾彦钧、张雅媛和郭伯臣, 2006)。在 2011 年分别是学术论文和学位论文发展的高峰,之后学位论文研究数量所有下滑,期刊论文仍旧保持、变化浮动不大。

2.3.2.2. CD-CAT 研究内容历程分析

团队详细阅读 140 篇学术文献,发现很多文献聚焦于诊断模型、Q 矩阵以及概况综述和自适应认知诊断流程描述的研究,缺少完整的实践应用的成功案例。统计数据表明:

(1) 相比较于国外 CD-CAT 研究进程较晚,并且研究内容集中在理论概述和小规模的实验证明(陈秋梅和张敏强,2010),缺乏大规模应用 CD-CAT 相关的中文文献支持。

(2) 基础理论研究占文献总数的 74.2%,有 64 篇文献(45.7%)对认知诊断模型进行介绍,例如截止 2007 年,常见的认知诊断模型有补偿模型 LLTM(线性逻辑斯蒂克测验模型)、MIRT-C 等,非补偿模型 RSM(规则空间模型)、DINA(确定性输入噪音与门模型)、HO-DINA(高阶 DINA 模型)、RUM(融合模型)、AHM(属性层级模型)。

(3) 试验实践比例为 30%,其中有 10%的文献叙述了相应的教育干预措施,例如数学、物理、化学和英语。

此外,余娜和辛涛(2009)提出诊断模型的评估问题由信度和效度决定。认知诊断评估中信度采用校正之后的模型产生两组平行的模拟考生,分布估计每组掌握情况,计算相同考生被划分到相同属性掌握状态的百分比;效度分为内部效度和外部效度。

2.3.3. 升华-合作关系追踪

科学领域的研究往往存在学术带头人,也会以学术带头人为核心形成研究领域的网络关系。本研究借鉴魏顺平提出的新的 n-派系定义,利用研究者合作关系网络研究,建立“作者 A-作者 B”关联关系。在社会网络分析中,派系(subgroup)是社群中的一小群人关系特别紧密,以至于结合成一个次团体。

首先,借助 EXCEL 统计研究者与文献之间的关联表。通过论文作者机构分析发现,大部分文献作者来源于高校、研究所,并且主要来源于江西师范大学心理学院和计算机信息工程学院、北京师范大学心理学院;本研究中 CD-CAT 文献量大于 5 篇的作者列表 2 如下。究其原因,自适应认知诊断测验研究专业性较强、理论基础扎实,高校师生是新理念和方法的发起者和倡导者,一方面说明真正得到了价值的体现;另一方面,认知诊断测试的最终目的在于以 K-12 教育为代表的评估教育,能够实现学习者的自适应学习和诊断评价。文献分布说明研究内容处于高校“实验室”阶段,缺乏社会的广泛关注、实践和推广应用。

接着,通过 UCINET 软件呈现 51 个节点,以文献篇数最多的“丁树良”派系做为研究范例,将上表的矩阵导入 UCINET 中,绘图功能(DRAW)得到科研合作关系网络图,如图 4。

采用 UCINET 社会网络分析软件将这些派系进行可视化处理。研究结果表明:

(1) 目前教育技术研究领域的合作类型主要为师生合作,较少跨校、跨区域合作;也存在少数的跨区域合作,因此形成了星状辐射型。

(2) CD-CAT 研究应实现走出团队、校门,甚至跨越国界;开展跨校、跨区域、跨学科的科研合作,优化研究团队构成,从而进一步提高研究质量,加快知识扩散。

3. CD-CAT 研究项目现状

由北京师范大学心理学院刘红云教授、美国伊利诺伊香槟分校张华教授研究以国家基础教育课程教材发展中心-学生学业质量评价项目为依托,属于阐述了在北京、大连等地进行计算机认知诊断的大规模测试(Liu, et al., 2013)。由于软件和硬件以及专业技能的限制,将 CD-CAT 应用于学校项目是成为一个巨大的挑战,该项目克服传统 Client/Server 框架,将 B/S(Browser/Server)架构设计应用于国内的英语二级考试中。2010 年,美国也针对 K-12 提出了全美教育标准(Common Core States Standards,简称 CCSS),用于评估各个年级学生的数

Wu, Y.-T., Chang, M., Li, B., Chan, T.-W., Kong, S. C., Lin, H.-C.-K., Chu, H.-C., Jan, M., Lee, M.-H., Dong, Y., Tse, K. H., Wong, T. L., & Li, P. (Eds.). (2016). *Conference Proceedings of the 20th Global Chinese Conference on Computers in Education 2016*. Hong Kong: The Hong Kong Institute of Education.

学能力和英语能力，并且已经在 50 个州中的 46 个州和哥伦比亚区施行。Smarter Balanced 评估系统利用强制性的终结性评价和可选的中期评估来提升计算机自适应测试(CAT)的效率。综上所述，国外 CD-CAT 发展已经将智能学习系统、评估系统和教育测评融入其中。虽然从应用角度来说，“自适应”的成分居多，“认知”的成分较少，但是将诊断功能结合到计算机自适应测验中，已经成为研究热点之一，吸引了众多学者将 CD-CAT 的实施应用继续推进。虽然国内很多研究者已经意识到了 CD-CAT 的优势，也有研究者提出了中国教育进展评估基于 CD-CAT 的改进与展望。但是国内采取的依旧是传统的纸笔测试和基于计算机的考试形式，简单提供一个考试分数，不能给出学生知识掌握的诊断信息。

4. 对我国基础教育中评价诊断的启示

计算机自适应认知诊断测试在教育评估中具有重要的意义，目前混合式的学习环境中需要个性化的、实时的、及时的诊断和评估报告。随着网络技术的发展，MOOC 等一大批在线课程蜂拥而至，但纵观众多网络课程和实践研究，其中缺少对学习者的“量体裁衣”式的诊断评估。同时，在我国基础教育的测验考试中，纸笔考试仍旧保留其重要的地位，然而缺少有效甄别、准确评估的和学习知识建议等内容。根据我国国情和实际需求，提出相应发展启示：

1. 科研创新方面：打造精锐团队，延伸研究深度。从文献数据分析说明，目前我国研究“派系”较为集中，能够拓展实践应用的驱动项目较少。因此加强跨团队、跨区域、甚至跨国界的学术交流和合作，共享资源与经验，共同丰富和深化 CD-CAT。

2. 实践应用方面：

(1) 利用现有资源进行校际合作，将纸质阅卷向电脑阅卷转化，利用计算机平台等，开发计算机自适应系统。根据国情，考虑具体学科先行策略，借鉴国外经验从不同学科、学段探究学业水平发展趋势，对影响学生能力发展的各种因素进行调研，从认知模型中能够与影响因素结合，为学习者提供可选择的、具有层级关系的学习路径；联动家长、教师和学校管理者共同组建“诊断评估生态圈”，为每一个学生的个性化和均衡发展付诸努力！

(2) 区域化合作，共同推进 CD-CAT 试验。开发大量题库，保密性，安全性，机器的差异能够被云计算所克服，从而能够满足异地同时，同地异时考试的进行。通过云平台持续地将优质题目添加到题库，进而保证了陈题被淘汰，新题被“涌出”。云计算可以通过网络使储存在“云”上的教育与资源通过终端设备传递给每个有需要的人，这样共享优质的教育资源，促进教育公平；而且把教育资源储存在“云”上也可以节省购买这些硬件资源所需要的成本，减少服务器及所需基础设施的更新维护、人工管理和能源消耗费用，降低教育成本。

参考文献

余娜和辛涛(2009)。认知诊断理论的新进展。*考试研究*, 3, 22-34。

陈秋梅和张敏强(2010)。认知诊断模型发展及其应用方法述评。*心理科学进展*, 3, 522-529。

涂冬波、蔡艳和丁树良(2012)。*认知诊断理论，方法与应用*。北京：北京师范大学出版社。

曾彦钧、张雅媛和郭伯臣(2006)。基于认知结构之适应性诊断测验系统的防猜测选题策略。*测验统计年刊*, 14, 37-51。

Liu, H. Y., You, X. F., & Wang, W. Y., et al.(2013). The development of computerized adaptive testing with cognitive diagnosis for an English achievement test in China. *Journal of Classification*, 30(2), 152.