

基于大数据的教育技术研究新范式

祝智庭, 沈德梅

(华东师范大学 上海数字化教育装备工程技术研究中心, 上海 200062)

[摘要] 教育技术的发展包含了一系列不同的范式。随着大数据的崛起和数据密集科学的发展, 学习分析学(LA, Learning Analytics)和教育数据挖掘(EDM, Educational Data Mining)成为大数据在教育领域的具体应用, 基于数据的教学干预应用程式已出现并在实际教学中使用(如 Signals, Moodog 等)。文章探讨了科学范式, 大数据在教育技术领域的应用, 以及不同的教育技术范式, 提出由于能够更好地贯彻“以学习者为中心”的教育理念, 个性化自适应学习系统将成为以大数据为基础的新的教育技术范式。

[关键词] 科学范式; 大数据; 智慧教育; 教育技术范式; 个性化自适应学习

[中图分类号] G40-057

[文献标志码] A

[作者简介] 祝智庭(1949—), 男, 浙江衢州人。教授、博士生导师, 主要从事教育信息化理论、系统架构与技术标准、网络远程教育、教学与系统设计以及面向信息化的教师专业发展等研究。E-mail: ztzh@dec.ecnu.edu.cn。

一、科学范式 (Scientific Paradigms)

美国著名科学哲学家托马斯·库恩(Thomas Kuhn)在《科学革命的结构》(《The Structure of Scientific Revolutions》)一书中系统阐述了关于范式的概念和理论。所谓科学范式是指“在一定时间范围内, 能为研究者群体提供样板问题极其解决方案的普遍公认的科学成就”(Universally Recognized Scientific Achievements That, for a Time, Provide Model Problems and Solutions for a Community of Researchers)。^[1]

科学范式的概念是库恩范式理论的核心。库恩认为, 科学范式具备两个方面, 首先, 在科学范围内, 该术语指的是可以被复制或模拟的一组示例性的实验; 其次, 这组范例的基础是共享的先入之见(Preconceptions), 这些先入之见形成于证据收集之前, 并且影响证据的收集。先入之见体现于两方面, 一是其隐含的假定, 一是相关的形而上学的元素; 个体科学家对该范式的诠释可能会有所不同。^[2]因此范式界定了某一研究领域的研究方法, 即研究什么, 研究问题的提出, 如何针对研究问题进行研究活动, 以及如何对研究结果进行诠释等。同时, 范式具有哲学意义, 它暗示了某研究群体的研究遵循的基本理论和研究群体共享的信念和世界观等。

库恩同时认为, 范式不是一成不变的, 它在科学研究的进程中完善、发展, 最终可能退出。随着科学的发展, 新的科学范式会出现, 补充或者取代旧的范式, 这也就成为科学发展进程中的科学革命。在库恩看来, “科学革命”的实质就是“范式转换”; 在广泛接受的科学范式里, 发现现有理论或者范式无法解决的“例外”, 因此尝试用其他理论取而代之, 该理论得以发展最终成为新的范式。在自然科学领域, 范式的转换比较明显, 如伽利略的动力学相当于近代科学的初级阶段的范例, 爱因斯坦的相对论则为当代科学的研究发展提供了模式。

库恩本人认为范式这一概念不适合社会科学范畴。原因是当他在社会学者聚集的帕洛阿尔托学者中心写《科学革命的结构》一书时, 观察到社会科学学者们在诸多理论方面存在分歧。因此他在书的前言中特意指出, 他之所以提出范式的概念正是为了将社会科学从自然科学中区分开来, 他认为在社会科学中不可能存在任何范式。然而尽管社会科学不可能像自然科学那样在某一特定时期存在一个范式, 在相对较小范围的研究领域, 如社会学、人类学、教育学等或其下属领域内, 可能存在支持这些领域的研究范式、研究传统、研究计划等。这些较小领域的研究特征能够激发不同领域的研究, 界定什么是或不是研究证据, 以及

为控制与其他相似研究领域的学术争论。例如,斯金纳行为主义和个人建构理论同属于心理学和教育学研究范畴,这两个心理学子学科的一个最显著区别是对意义和意向的关注(Meanings and Intentions)。在个人建构理论中,这两个概念属于核心问题,但在行为主义中,它们不能作为科学证据,因为他们无法被直接观察到。^[3]另外,学者们认为,^[4]虽然社会科学内也存在明显的概念方面的改变,如从行为主义到认知方法,但是它们与自然科学范围内的科学革命不同,原有的理论一般不会被完全摒弃,而是仍旧在新的范式占据统治地位的情况下拥有一席之地。

Thagard 提出并阐述了理论(Theories)与方略(Approaches)的区别。理论是指“相关假定的集合,对大范围的实证研究结果和事实进行解释和归纳概括的基础”,方略是“实验研究方法和诠释风格的集合”。^[5]Thagard 认为,由于整个社会科学领域并不存在一个统一广泛的支持各个学科的理论基础,社会科学的变革更多的是由于研究方略的改变,而不是源自对理论统一性的评估。也就是说,社会科学范式,更多的偏重是指研究方略方面。本文之所以提出这点,是因为本文目的之一是评估新的研究方法,即大数据(Big Data)催生的数据密集科学(Data-Intensive Science)对教育技术研究范式,即教育技术研究方法的影响。

二、悄然兴起的大数据分析与应用

大数据一词出现于1997年,NASA 研究人员 Michael Cox 和 David Ellsworth 第一次用该词描述上个世纪90年代出现的数据方面的挑战,即超级计算机生成的巨大的信息数据量。当时,Cox 和 Ellsworth 对实验中产生于飞机周围的模拟气流数据无法进行处理或者将其可视化。“数据集相当大,对主内存、本地磁盘,甚至远程磁盘都造成挑战,”他们写道,“我们称此问题为大数据。”^[6]

大数据的产生与信息技术、互联网等密不可分并且以越来越多的方式产生,^[7]如多媒体内容、社会网络以及各类传感器,不论是传统的数据密集型行业如基因研究、药学,还是互联网新贵,都面临着储存分析大数据的问题。例如 Facebook 拥有超过9亿的用户,并且用户数量仍在增长;Google 每天有30亿的搜索查询,Twitter 每天处理4亿次的短信,相当于大约12TB 的数据量。

时至今日,大数据尚没有系统统一的定义和理论,学者们一般只是用该术语描述难以用传统软件

和方法分析的超大量的复杂的数据。^[8]Laney^[9] 首先提出用“3Vs”(Volume, Velocity, Variety)的概念,在此基础上,IBM^[10]用“4Vs”描述大数据,即大数据应该具备四个维度,大体量(Volume)、高速度(Velocity)、多样化(Variety)和真实性(Veracity)。大体量是指各种类型的不断增长的数据很容易积累到百万兆字节甚至千兆兆字节(Terabytes—Even Petabytes)的信息。高速度是指及时处理大数据的必要性,例如分析大量的当日呼叫详细记录可以实时预测客户流失的程度等。多样化是指数据形式的多样性,如可以分析多种数据的变化包括文本、图像、音频等来提高客户满意度等。真实性则意味着大数据提供信息的可信度,以及据此决策的可靠程度。还有些学者^[11](Quinn, 2012)认为应该加入另外两个V: Value (价值)和 Visualization (可视化)。关于类型,学者们认为数据,不论是否是大数据都分属三种类型:非结构化数据、半结构化数据、结构化数据。^{[12][13][14]}非结构化数据指没有格式的数据,如PDF、E-mail 和文档。结构化数据具备一定格式,便于存储、使用和从中提取信息,例如传统的事务型数据库。半结构化数据是指类似XML和HTML的有一定加工处理的数据。

大数据的应用和影响体现在各个领域。大数据不只意味着体量的大小,它同时意味着研究方法更倾向于利用新的多种类型的数据获取信息,以数据为基础进行研究,并作出决策。在天文研究方面,美国的 The Sloan Digital Sky Survey^[15](SDSS2008)成为天文学家的主要信息来源,同时,天文学家的主要工作也从包括拍摄星空图片等变为主要应用数据库查询和发现天象的变化。对企业来说大数据的应用则意味着更好的商业决策,有些公司如 Google、Amazon、Yahoo 等,分析利用此类数据,并将其结果作为扩张市场的依据或者提供个性化服务的方向,因此公司得以快速成长。大数据的出现和潜在的价值也引起了各国政府的注意。例如,奥巴马政府2012年宣布,每年将花费超过2亿美元在大数据研究应用方面,以致力于科学探索、环境、生物医学、教育和国家安全方面的研究。^[16]在教育领域,随着远程教育的发展和LMS(如Blackboard 和 Moodle 等)的应用,大数据的潜在应用也越来越广。这些系统每天都记录大量的学生交互信息、个人数据、系统数据等。^[17]这些也促进了教育界学习分析学(LA, Learning Analytics)和教育数据挖掘(EDM, Educational Data Mining)的发展应用,以及教育技术领域研究范式的变化。

三、“数据密集科学”作为科学研究第四范式所带来的机遇与挑战

大数据作为一个通用术语,实际描述着正在发生的影响到自然科学、工程学、医学、金融、商业、直至整个社会的科学革命。正是基于大数据的出现以及影响,Jim Gray^[18]在2007年提出了数据密集型科学(Data-Intensive Science)的概念。Gray认为,从进行科学研究的方法的角度来看,从古至今存在的科学研究方法范式包括:

1. 实证式(实验科学)(Empirical/Experimentation)分支,始于1000年前,主要的方法是对自然现象的描述论证,对自然现象进行系统归类,如对化学元素的分类;

2. 理论式(理论推演)(Theoretical)分支,当科学假设与预期结果一致时,则使得理论框架开始占有一席之地,出现于数百年前,主要采用建模方式,由特殊到一般进行推演;

3. 计算式(计算机仿真)(Computational)分支,始于几十年前,主要方法为利用计算方式模拟复杂现象,科学数据可以用模拟的方法获得,而不再依赖于单一的实验;

4. 数据密集型科学(Data-Intensive Science),在前三种方法的基础之上,采用IT技术获取、处理、存储、统计分析大数据,从中获取知识。

数据密集型科学被称之为科学研究的“第四范式”,与其他三种范式一起成为科学研究的方法,它的出现与大数据密切相关。因此,Gray提出的“范式”更接近于Thagard^[19]提出的“方略”。

Gray^[20]认为,数据密集科学包含三项针对数据的活动:获取、存储维护、分析。大数据给科学研究带来巨大改变的同时,也意味着多方面的挑战。学者们认为,整个数据获取到分析的过程都存在不同的困难和挑战。^[21]例如,在获取数据时,如何摒弃无用的数据,如何做到在数据收集的过程中过滤数据以免却储存之后再进行处理麻烦;同时,如何自动产生元数据(Meta Data)对数据进行描绘。其次是数据存储,即数据库存问题,事务性数据库不适合存储关系不明确的大数据。在分析方面,传统的统计算法的前提是数据的同质性(Homogeneity),大多数大数据不具备此特性。针对这些挑战(大数据的非结构化,具有多样性,同时数量巨大),传统的关系数据库无法满足要求;NoSQL(Not Only SQL)数据库则为存储和检索大数据提供了可能。

Google的Google File System、Big Table、Map Reduce代表了这方面的技术创新。

在《第四范式:数据密集型科学发现》^[22]一书中,多位作者提出了各个科学领域的研究与大数据结合的必要性,和数据密集科学对不同领域科学研究方法的影响,包括地球与环境科学、生命与健康科学、数字信息基础设施和数字化学术信息交流等。他们也描述了大数据以及数据密集科学影响下不同领域的科研活动、过程、方法以及成果,拓宽了不同学科领域应用大数据的思路。例如,Robertson^[23]等在讨论发展中国家的医疗合作时,描述了他们的以计算机和手机结合为基础的NxKM(NxOpinion Knowledge Manager)系统。该系统包括一个有专家开发的知识库、一个医疗诊断引擎和一个手机界面,用来输入患者信息并根据该信息自动产生问题(多项选择),以从患者获得更多信息。因此,虽然患者信息可以由当地人连接输入系统,但该信息由远程专家分析,因此,多方面的合作以及信息数据的综合使用,使得诊断结果和诊治手段也将更加可靠。该书虽然涵盖了多方面内容,包括信息密集型科学研究范式对地球环境、医学、认知科学、学术信息交流等方面的深刻影响,但没有涉及大数据对教育,尤其是教育技术的影响。

四、教育技术研究范式演变轨迹

本文对教育技术不做定义方面的明确界定,它基本等同于英语的Educational Technology、Instructional Technology或者Instructional Development^[24](Dills & Romiszowski, 1997),其研究核心是应用技术支持教学和学习。教育技术研究的核心方略是设计研究,本文从研究范式的角度出发,主要目的是看教育技术的范式的特点、演变以及大数据对教育技术研究可能产生的影响。

首先,学者们认为,在教育技术领域,正如整个社会科学领域,从来都是多范式并存的。即从未有过只有一个范式存在而其他范式完全退出教育技术学术圈的情况,尽管也许在某个时期存在一个占据主要地位的范式。^[25]例如当斯金纳心理学占据统治地位的时候,与之对垒的非斯金纳行为主义地位次之,同时之前影响最大的弗洛伊德心理学的影响依然在某种程度上存在,而未来的范式(认知信息加工理论)也开始崭露头角。另外,Saettler^[26]认为20世纪的教育技术领域存在四种范式:物理科学或者媒体理论、传播学和系统理论、行为主义和新行为主义观点、认知理论。也有些学者描述了教育技术研究和评估领域内的范式

变化,如 Driscoll^[27]概括了八种教育技术研究范式,Clark 和 Sugrue^[28]描述了媒介研究(Media Research)中行为主义和认知理论范式对于研究设计和研究问题形成的影响。其次,Reigeluth^[29]认为,教育技术范式的改变是从整个社会的变化开始,而且其改变的速度随着人类知识库的迅速增长和科技的高速发展越来越快。大部分的教育技术范式研究中囊括了多种的技术应用,同时很多研究更偏重于理论方面的变迁,而 Koschmann 则详细论述了计算机为基础的教育技术范式演变。^[30]

Koschmann 认为:^[31]在计算机进入教育领域后,教育技术作为一个独立的研究领域才开始出现,因此针对教育技术范式的研究也应该集中在以计算机为基础的技术方面;同时他认为教育技术研究也经历了一系列范式转换。他应用库恩的理论,主要描述了四种涉及计算机的教育技术范式,即计算机辅助教学(CAI, Computer-Assisted Instruction)、智能教学系统(ITS, Intelligent Tutoring System)、Logo-as-Latin 以及计算机支持的协作学习(CSCL, Computer-Supported-Collaborative Learning)。

Koschmann 强调,CAI 主要是针对教学技术的设计和评估的研究范式,他在后来的研究中认为,^[32]CAI 并非与计算机同时出现,相反,它可能是桑代克(Thorndike)教育心理学研究范式的延伸和扩展。IBM 公司开发的 Coursewriter(一种课件著作软件)被认为是 CAI 开始的标志,即使没有任何编程经验,人们也可以用该系统开发自己的教学课件。由于 CAI 开发人员大部分具有教学背景,CAI 系统大多反映了教育界对教和学的认知,即学习是被动获取信息的过程,而教学则是知识传递的过程。CAI 系统一般贯彻如下学习策略和措施:确定学习目标,将学习目标分解为一系列学习任务,然后开发一系列学习活动,以达到预定学习目标。CAI 同样以行为主义和实证主义为理论基础,因此 CAI 研究人员认为学习是可测量的学习成绩或者能力水平的变化,学习是 CAI 研究中的因变量,而学习过程中引进的技术方面的创新成为干预措施和自变量。对照组的使用在研究中很常见,研究问题通常为:使用该项技术对教学有何影响?因此,教学效验(Instructional Efficacy)成为该范式下的核心研究问题。

第二个范式为 ITS,起源于人工智能,以 Carbonell^[33]的博士论文的出现为标志。ITS 理论认为,认知是一个计算过程,可以通过建立模拟人脑工作模式的智能型系统来研究。^[34]如果智能型行为可以通

过系统程序表现,那么具备经验和技能的教师的角色也可以设计出来。由于一对一教学被认为是金牌标准,^[35]因此可以推断出如果每个学生都有个人的导师,那么整个社会的教育水平都会相应得到提高,这也是智能教学系统研究范式的基本理念。信息加工理论是人工智能前提之一,它认为问题解决是定义问题空间的表征(Representations)的过程,包括初始状态、目标状态以及不同状态之间的一系列运作。在此基础上,表征成为解决问题和理解认知过程的中心问题,而学习则成为获取正确的问题空间的表征的过程,教学则是辅助学习者获取表征的活动。在此过程中技术的角色与其在 CAI 中并没有本质不同——然而人工智能系统更注重交互性,也更偏重于复杂技能的习得。与 CAI 不同,智能教学系统范式的核心研究问题是教学能力,即该系统是否完全能与娴熟的真正的导师相媲美。因此,研究问题更看重的是系统的效果,而不是学生的成绩。

计算机辅助教学与人工智能系统尽管有所不同,但从认知论的角度说,他们都属于现实主义和绝对主义(Realist and Absolutist),即认为学习是被动获取知识,而教师是绝对权威。^[36]

第三个范式为 Logo-as-Latin,意指像学习拉丁语一样看待 LOGO 语言的学习。其中 LOGO 是上个世纪 60 年代由 MIT 数学教育实验室 Papert 教授领衔开发的程序语言,主要供儿童在编程的过程中学习。该范式以建构主义为理论基础。建构主义起源于皮亚杰的发展心理学,认为学习是新的信息与已有的知识融合同化的过程。有学者认为,计算机编程可以成为建构主义学习方式的重要角色。^[37]例如学生可以建立模拟系统,在此过程中,学习者成为“教师”,而计算机则开创了一个新的教育技术在学习中的角色,即成为“被辅导者”。与 CAI 不同的是,CAI 研究关注教学效果,而 Logo-as-Latin 研究更专注于教学迁移。编程教学被看作干预手段,学习者在其他相关学习任务上的成绩被看作因变量。然而,在后期的文章中,Koschmann 认为,^[38]Logo-as-Latin 与 CAI 同样起源于传统教育心理学,与 CAI 密切相关,因此应该将它看作 CAI 的一个变种,而不是独立的教育技术研究范式。

Koschmann 提出的第四种范式是 CSCL。CSCL 与前三种范式有很多不同之处。首先,前三种研究范式都是建立在心理学基础之上,其本质是行为主义和认知主义。而 CSCL 的基础是人类学、社会学、语言学以及传播学等。具体说来,其理论基础包括(不限于)社会建构主义、社会文化理论以及情景认知理论等。社

会建构主义认为知识的建构本质上是社会性过程;社会文化理论以维果斯基的文化—历史理论为代表,强调语言在智力发展过程中的作用;情景认知理论认为,学习是进入实践共同体(Community of Practice)的过程,“要想学会如同真正的专业从业人员那样使用一个工具,一个学生就应该像一个学徒,必须融入该社区及其文化。因而,在相当大的程度上,学习是,我们相信,一个文化熏陶的过程”^[39]。因此 CSCL 范式中,学习的社会性和文化性成为核心问题。不同于前三种范式针对的问题(教学效果、教学能力、教学迁移),CSCL 被称为“演绎的实践性教学”(Instruction as Enacted Practice)。CSCL 研究范式有几个特点:(1)研究问题比较集中在学习过程而不是结果;(2)研究多倾向于描述性,而不是实验性;(3)很多研究者乐于以参与者(CSCL 成员)的角度研究合作学习的过程。因此 CSCL 研究着眼于参加者的谈话,合作过程中使用的工具,合作小组的成果等。CSCL 范式下的研究问题包括:学习如何在学习者的语言中表现出来?社会性因素如何影响学习过程?技术如何在合作学习中应用?Koschmann 将这四种范式做了简单对比,见表 1。

表 1 教育技术研究范式

| 教育技术研究范式 | 学习理论 | 教学模型 | 研究问题 |
|---------------|-----------|------------|-------|
| CAI | 行为主义 | 程序化教学/教学设计 | 教学效果 |
| ITS | 信息加工理论 | 一对一培训、交互性 | 教学能力 |
| Logo-as-Latin | 认知主义、建构主义 | 发现性学习 | 教学迁移 |
| CSCL | 社会文化理论 | 合作学习 | 实践性教学 |

五、数据密集科学影响下的教育技术研究范式:个性化自适应学习

数据密集型研究方法捕捉了整个信息时代带来的大数据的基本整体影响。在不同的领域,研究方法的侧重和目的不同,因此各有特点。如在工业界,商业智能系统(Business Intelligence System)体现了大数据对决策的影响。在教育领域,美国教育部在一份简报中指出,^[40]大数据在教育领域的具体应用主要为学习分析学(LA, Learning Analytics)和教育数据挖掘(EDM, Educational Data Mining)。EDM 和 LA 之间没有明确的分界线,但它们的起源、理论和目标不尽相同,并且逐渐成为泾渭分明的两个研究领域。

EDM 的目的是研究和利用统计学、机器学习和数据挖掘方法来分析教和学的过程中产生的数据。学者们认为,EDM 的研究目的包括以下方面:^[41](1)应用多方信息如学生的知识程度、动机、元认知、学习态度等建立学生模型,并以此预测学生的学习行为;(2)发现或改进学习内容展现和最佳教学序列的领域模型;(3)研读由学习软件提供的不同的教学支持的效果;(4)建立包括学生、领域模型和教学软件的计算模型,推动关于学习和学习者的科学研究。

美国教育部的简报中总结了 EDM 针对和所要回答的问题:^[42](1)什么样的教学顺序(不同学习主题)对不同特点的学生最有效?(2)什么样的行为与更好的学习成绩相关(如较高的课程学习成绩)?(3)什么样的学生的行为指标预示了学生的满意程度、参与度和学习进步,等等?(4)什么特点的在线学习环境能导致更好的学习成绩?(5)什么因素能够预测学生取得成功?

Siemens 将 LA 定义为“关于学习者以及他们的学习环境的数据测量、收集、分析和汇总呈现,目的是理解和优化学习以及学习情境”。^[43]LA 的一个重要应用是监测和预测学生学习成绩,及时发现潜在问题,并据此作出干预,以预防学生在某一科目或者院系课程学习中产生风险。^[44]相比于 EDM,LA 借鉴了更广泛的学科,除了计算机科学、统计学、心理学、学习科学,还引进并应用信息学和社会学的理念和技术。^[45]LA 回答的问题如下:(1)什么时候学生可以进行下一个学习主题?(2)什么时候学生可能在某一门课程中落后?(3)什么时候某个学生可能有完不成一门课程的风险?(4)如果没有干预补救措施,学生可能得到什么样的成绩?(5)对特定学生来说,下一个最好的课程是什么?学生是否需要特殊帮助?

美国教育部^[46]的简报中总结了应用 EDM 和 LA 的范畴:(1)用户知识模拟、用户行为分析、用户体验分析;(2)用户分类/分组(Profiling);(3)知识域模拟如学习课题分类排序等,知识元素与相应的教学原则分析;(4)趋势分析;(5)自适应和个性化学习。

应用 LA 和 EDM 数据分析结果,教师可以更好地了解学生,理解和观测学生的学习过程,发现最适合的教学方法和顺序,及时发现问题并进行干预,以提供个性化的学习服务为主旨。现在已经研发出的应用系统案例有普渡大学的“课程信号系统”(Course Signals System,以下简称 Signals)^[47]、在美国加州大学圣巴巴拉分校以及阿拉巴马大学使用的 Moodog,^[48]以及美国西部州际高等教育委员会教育技术合作部

(WICHE, WCET, Western Interstate Commission for Higher Education, Cooperative for Educational Technologies)的教育大数据分析项目——预测分析报告(PAR, Predictive Analytics Reporting)系统。^[49]

Signals 系统通过数据挖掘和统计预测模型,根据多个变量(表现指标包括:现有平均分和努力程度,如学生 LMS 的交互频率;个性特点指标包括学术准备,如高中平均分和各项标准考试成绩;学生特点,如是否为美国居民、年龄和选修学分)来预测学生是否能够完成/通过该课程。Signals 在课程进行的过程中,以交通信号指示灯的方式,让学生了解自己的学习状况:课业良好(绿色),课业中度危急(黄色),或者课业严重危急(红色)。同时教师可以提供给学生有效的反馈信息,引导学生使用合适的资源等来提高成绩。^[50] Moodog 的主要功能是跟踪记录学生在课程管理系统(CMS, Course Management System)上的学习活动,其基本目标有两个:(1)为教师提供学生与在线学习材料交互情况;(2)帮助学生将自己的学习行为和进程与其他学生相比较。^[51]

PAR^[52]的主要目的是应用 EDM 技术,分析跨越多所高等院校的学生数据,以期发现并确认影响学生退学/以及是否能够毕业的因子,并据此实施有效的教学干预。参与 PAR 的高校包括两年制和四年制高校,有公共学校和私立学校,亦有传统高校和非传统高校如网络大学。已经有 16 个 WCET 成员机构提交了 1,700,000 条匿名和去身份标识的学生记录以及 8,100,000 条课程级别数据记录。所有高校使用统一的数据模型,该数据模型包括以下核心数据元素。(1)总体元素:基本框架,描述所有 PAR 数据的基本概况。(2)学生一般元素:描述学生人口数据和学术背景信息。(3)学生课程元素:描述学生参加的课程和学生的课程成果。(4)学生的学术元素:学生级别的数据。(5)课程目录的元素:教育机构开办的 PAR 学生就读的课程细节信息。(6)学校元素:学术单位的具体信息。

应用描述、推理和预测分析技术,PAR 项目初步发现 32 个影响学生学习以及退学的普通变量(多为学生特点变量),包括性别、种族、学位种类、多种专业、课程数量、班级人数等。其他发现如学生的性别、年龄以及种族与该生是否会退出某门课没有关系。该研究仍在继续。

这些系统目前的主要功能是分析学生的网上学习活动,判断实施干预措施的时间以及方法等;其潜在功能则可能包括教师可以根据学生使用学习资源

的情况发现哪些最受学生欢迎或者哪些活动影响学习成绩,提供适合不同学生需要的学习材料;调整学习顺序或者学习活动等;最终,完善的学习系统可以根据学生的特点(学习风格、已有知识、动机情况等)引导学生使用适合自己特点的学习材料和学习路径。因此 EDM 和 LA 在教育技术领域内的应用最终指向个性化学习和自适应学习环境的研究和开发。

美国教育部简报认为应用 LA 和 EDM 技术达成预测学生学习和干预学习过程的自适应学习环境应该包括六个部分:^[53](1)自适应学习内容,通过与学生的交互活动,可以辨别学生的水平和能力,因此可以管理、维护和呈现适合特定学生的学习内容;(2)学生学习数据库,用来获取存储学生与学习内容的互动,包括时间和行为等;(3)预测模型,应用学生学习数据和人口统计数据(储存在另外的数据库中,如年龄种族等),追踪学生学习过程,预测未来行为以及成绩,如课业成绩,是否有可能辍学等;(4)可视化报告,将预测模型产生的结果用仪表盘形势表现出来;(5)自适应引擎,用来操控学习内容,确保学习内容适合学生的能力和特点;(6)干预引擎,教师、管理员、系统开发人员等可否决系统提供的建议进行人为干预。除了六个自适应系统内部组成部分外,还包括一个外部的学生信息系统。这个信息系统是由学校学区或者地区教育部门持有维护的学生背景信息资料,如年龄、性别、所学过的课程、成绩、学习风格等。预测模型可以从中获取数据作为预测学生行为成绩等的部分依据。图 1 提供了自适应学习系统的概况。

图 1 中的箭头和数字表示自适应学习环境中的数据流的方向和顺序。整个自适应学习系统包括三个信息反馈回路数据流(Feedback Loop)。数据流的第一步是学生与学习内容的交互,交互内容被储存于学生学习数据系统里(第二步),第三步则是预测模型抽取学生学习数据和背景数据,应用 LA 和 EDM 技术进行分析,然后将结果传递至自适应引擎(第四步),自适应引擎据此针对特定学生作出学习方面的调整,而这些调整则通过学习内容策略等的改变表现出来。同时,预测结果也可以通过数据仪表盘(数据可视化面板)传达给教师和管理人员(第五步)。当学生、教师、管理人员等得到相关信息时整个反馈回路得以完成。学生通过自适应引擎获取的反馈信息包括学习活动情况、学习目标/技能完成程度、测评结果等,学生可据此更好地进行自我调节学习(Self-Regulated Learning),如在自己尚未完全掌握的学习内容上投入更多时间、改变学习策略等。教师方面获得的信息则

包括学生整体学习情况以及每个学生的学习情况,教师可以根据这些信息做出课程内容以及进度方面的调整,例如决定是否对个别学生进行干预,提供更多学习资料等。管理人员获取的信息则是包括多门课程、多个学生和教师的情况。根据整体信息,管理者可以知道哪门课程的学生成绩不尽如人意、哪些特点的学生成绩更出色等。管理层可以据此作出决策,如增加某门课程作为另一课程的先决条件(Prerequisite)等。

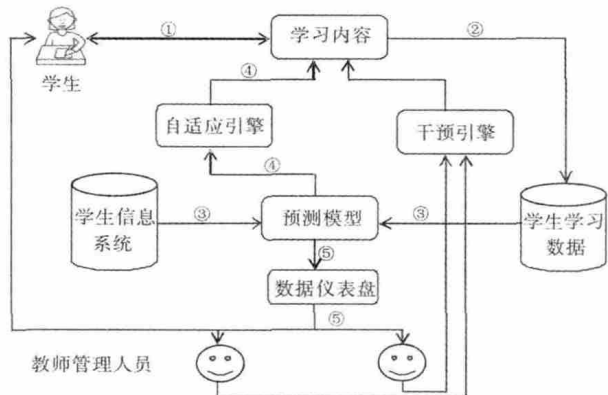


图1 自适应学习系统的构成以及数据流程图

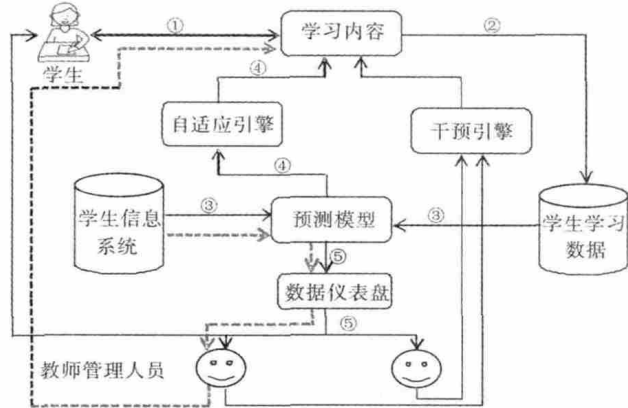


图2 个性化自适应学习(PAL)系统的构成

美国教育部简报中的自适应学习系统通过 LA 和 EDM, 引导学生了解自己的学习状况, 为教师的教学干预提供依据, 也使得管理层更好地进行决策。然而, 以数据密集科学为基础, LA 和 EDM 技术能够更好地分析学生的需要和特点, 从而使得学习更倾向于个性化。因此我们认为基于数据密集科学的自适应学习系统应该体现个性化的学习特点, 下一个教育技术研究范式是个性化自适应学习(Personalized Adaptive learning, 简称 PAL), 即在自适应基础之上, 学习内容更体现学生特点和需求。根据学生的特点(已有知识, 学习风格等)和其他信息(年龄, 性别, 兴趣等)数据可以将学生分组(Profiling), 学习系统可以根据学生特点和需要推荐学习内容, 教师针对不同特点的学生提

供丰富的学习材料, 学生同时可以自己选择学习材料、测评方式等。图2提供了 PAL 系统的基本结构。

图2中虚线部分代表了学习内容生成的过程: 学生背景数据(以往成绩、所学课程、学习风格等)导入预测模型, 分析生成可视化数据, 教师据此设计适合不同特点学生组的不同学习内容。图2的实线箭头部分代表了自适应学习过程和数据流。自适应学习过程与图1一样, 由三个反馈回路组成。

不同于以往的个性化学习和自适应学习的是, PAL 环境将以大数据为基础, 纳入 EDM 和 LA 数据分析和结果, 因此能够提供更适合特定学生的学习内容, 获取更多和更精确的学习者信息和学习活动信息, 更好地分析学习过程模式和学习活动有效性, 更准确地进行学习评估等。

PAL 与 Koschmann 的四个教育技术范式相比较, 其独特之处在于它倾向于利用多方面数据: 一方面根据已经存在的数据, 提供适合学生特点和需要的学习内容; 另一方面分析已有数据和学习过程中产生的数据, 根据结果发现问题, 并采取个性化干预措施。因此其基础为数据密集科学, 同时体现了以学生为中心, 根据学生的个性特点, 发展潜能来进行教学的人本主义的教学观念。

六、机遇与挑战

个性化自适应学习能够体现“以学习者为中心”的学习理念, 并且与智慧教育^[54]的主张不谋而合, 成为教育技术的一个新的研究范式。智慧教育主张借助信息技术的力量, 创建具有一定智慧特性(如感知、推理、辅助决策)的学习时空环境, 旨在促进学习者的智慧全面、协调和可持续发展, 通过对学习和生活环境的适应、塑造和选择, 以最终实现对人类的共善(对个人、他人、社会的助益)。智慧教育充分体现了“以学习者为中心”的思想, 强调学习是一个充满张力和平衡的过程, 揭示了“教育要为学习者的智慧发展服务”的深刻内涵。

智慧学习环境的一个基本特征是: 基于学习者的个体差异(如能力、风格、偏好、需求)提供个性化的学习服务; 并记录分析学习历史数据, 便于数据挖掘和深入分析, 数据结果用于评估学习过程、预测未来表现和发现潜在问题, 并以数据分析结果进行干预。因此以大数据为基础的个性化自适应学习将成为智慧学习环境的重要组成部分。同时, EDM 和 LA 能够为高校提供有效信息, 对学生进行干预, 最终能够提高学生成就, 降低辍学率, 提升毕业率。

另外,大数据除了支持学习过程分析外,还可以在知识表征(概念提取、本体建立、可视化)与利用(自动翻译、答疑)方面大有可为,因此,除 PAL 方式外,大数据能够促进个性化学习服务与社会智慧发展,促进人本主义教育理念的实现,并且成为社会知识生态发展模式的重要组成部分。

总而言之,大数据为教育技术的发展带来很多可能性,例如创建个性化自适应学习环境、知识发现工具、管理决策平台等,同时它的应用面临诸多挑战。首先是来自数据方面的挑战:如何储存海量的非结构化数据,例如学生的讨论等文本数据? 如何分析这些复

杂数据? 如何真正理解数据结果并传达给非数据专业人员? 只有当大数据管理技术、数据分析以及数据可视化工具方面取得突破性进展,才有可能真正实现以学习者为中心,满足不同学习者的需要的个性化自适应学习环境。其次,数据密集型科学的应用在教育方面的体现主要是 LA 和 EDM,通过数据使得学习过程透明化,并以数据为基础分析学生的行为和学习成绩。这些数据可以传达“发生了什么”,而不能回答“为什么”,尤其是那些数据中没有体现出来的原因。因此如何将无法从数据中观察到的因素,如学生的学习动机、情感等纳入干预设计,仍有待进一步研究。

[参考文献]

- [1] Thomas S.Kuhn. The Structure of Scientific Revolutions(3rd Edition)[M]. Chicago: University of Chicago Press, 1996:10.
- [2] T.S.Kuhn.The Structure of Scientific Revolutions(2nd Edition) [M]. University of Chicago Press, 1970: 88 and 41 respectively.
- [3] Wikipedia, the free encyclopedia[EB/OL].<http://en.wikipedia.org/wiki/Paradigm>.
- [4] [5] [19] P.Thagard. Conceptual Revolutions[M]. Princeton, NJ: Princeton University Press, 1992.
- [6] U.Friedman(2013). Big Data: A Short History[DB/OL].[2013-05-10].http://www.foreignpolicy.com/articles/2012/10/08/big_data?page=0,1.
- [7] [12] Purcell, B.. The emergence of “big data” Technology and Analytics. Journal of Technology Research [DB/OL].[2013-05-10].<http://www.aabri.com/manuscripts/121219.pdf>.
- [8] C.Snijders, U.Matzat, U.-D.Reips. ‘Big Data’: Big Gaps of Knowledge in the Field of Internet Science[J]. International Journal of Internet Science, 2012, (7): 1-5.
- [9] Doug Laney. 3-D Data Management: Controlling Data Volume, Velocity and Variety [DB/OL].[2013-04-10].<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>.
- [10] IBM[DB/OL].[2013-05-10]<http://www-01.ibm.com/software/data/bigdata/>.
- [11] E.Quinn. The 6 Vs: The BI/Analytics Game Changes so Microsoft Changes Excel [DB/OL].[2013-05-10].<http://www.esg-global.com/blogs/the-6-vs-the-bianalytics-game-changes-so-microsoft-changes-excel/>.
- [13] C.Coronel, S.Morris, P.Rob. Database Systems: Design, Implementation, and Management (10th Ed.)[M]. Boston: Cengage Learning, 2013.
- [14] P.Baltzan. Business Driven Information Systems, (3rd ed.) [M]. New York: McGraw-Hill, 2012.
- [15] SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems [DB/OL].[2013-01-10].<http://www.sdss3.org/collaboration/description.pdf>.
- [16] White House OSTP. Obama Administration Unveils “Big Data” Initiative: Announces \$200 Million In New R&D Investments[DB/OL].[2012-03-29]http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf.
- [17] R.Mazza, C.Milani. GISMO: A Graphical Interactive Student Monitoring Tool for Course Management Systems [DB/OL].[2012-01-10].http://linux3.dti.supsi.ch/~mazza/Web_area/Pubblicazioni/TEL04/TEL04.pdf.
- [18] [20] J.Gray. Jim Gray on eScience: A Transformed Scientific Method[R]. The Fourth Paradigm: Data-Intensive Scientific Discovery, 2009.
- [21] Agrawa, et al.. Challenges and Opportunities with Big Data [DB/OL].[2013-05-11].<http://cra.org/ccc/docs/init/bigdatawhitepaper.pdf>.
- [22] T.Hey, S.Tansley, K.Tolle (Eds.). The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington. UNT Digital Library[DB/OL].[2013-04-23].<http://digital.library.unt.edu/ark:/67531/metadc31516/>.
- [23] J.Robertson, D.DeHart, K.Tolle, D.Heckerman. Healthcare Delivery in Developing Countries: Challenges and Potential Solutions [A]. T. Hey, S.Tansley, K.Tolle.(Eds.). The Fourth Paradigm: Data-Intensive Scientific Discovery[C]. Redmond, Washington, 2009: 65-73.
- [24] [25] C.R.Dills, A.J.Romiszowski. The Instructional Development Paradigm: An Introduction [A]. C.R.Dills, A. J. Romiszowski (Eds.). Instructional Development Paradigms[C]. Englewood, NJ: Educational Technology Publications, Inc, 1997.

- [26] P.L.Saettler. The Evolution of American Educational Technology [M]. Englewood, CO: Libraries Unlimited, 1990.
- [27] M.P. Driscoll. Paradigms for Research in Instructional Systems [A]. In C. R. Dills, and A. J. Romiszowski (Eds.). Instructional Development Paradigms[C]. Englewood, NJ: Educational Technology Publications, Inc, 1995.
- [28] R.E.Clark, B.M.Sugrue. Research on Instructional Media, 1978–1988[A]. G.J. Anglin (Ed.). Instructional Technology: Past, Present and Future (2ed ed.) [C]. CO: Libraries Unlimited.
- [29] C.M. Reigeluth. Educational Systems Development and Its Relationship to ISD[A]. G.J. Anglin (Ed.). Instructional Technology: Past, Present and Future (2ed ed.)[C]. CO: Libraries Unlimited.
- [30] [31] [38] T.Koschmann. Paradigm Shifts and Instructional Technology [A]. CSCL: Theory and Practice of An Emerging Paradigm [C]. Mahwah, NJ: Lawrence Erlbaum, 1996; 1~23.
- [32] T.Koschmann. Revisiting the Paradigms of Instructional Technology [A]. Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education[C]. 2001: 15~22.
- [33] J. Carbonell. Mixed-Initiative Man-Computer Instructional Dialogues[D]. Massachusetts Institute of Technology, 1970.
- [34] Z. Pylyshyn. Computing in Cognitive Science[A]. M. Posner (Ed.). Foundations of Cognitive Science[C]. Cambridge, MA: MIT Press, 1989: 51~91.
- [35] B.S. Bloom. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring [J]. Educational Researcher, 1984, 13(6): 4~16.
- [36] M. Schommer. Effects of Beliefs about the Nature of Knowledge on Comprehension [J]. Journal of Educational Psychology, 1990, 82: 498~504.
- [37] S. Papert. Mindstorms[M]. New York: Basic Books, 1980.
- [39] J.S. Brown, A. Collins, P. Duguid. Situated Cognition and the Culture of Learning[J]. Educational Researcher, 1989, 18(1): 32~42.
- [40] [42] [45] [46] [53] U.S. Department of Education, Office of Educational Technology. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C. [DB/OL]. [2013-05-20]. <http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf>.
- [41] R. S. J. D. Baker, K. Yacef. The State of Educational Data Mining in 2009: A Review and Future Visions[J]. Journal of Educational Data Mining, 2009, 1(1): 3~17.
- [43] G. Siemens. Learning Analytics A Foundation for Informed Change in Higher Education [DB/OL]. [2013-01-17]. <http://www.slideshare.net/gsiemens/learning-analytics-educause>.
- [44] L. Johnson, R. Smith, H. Willis, A. Levine, K. Haywood. The 2011 Horizon Report. Austin, TX: The New Media Consortium. [DB/OL]. [2013-05-20]. <http://net.educause.edu/ir/library/pdf/HR2011.pdf>.
- [47] K. E. Arnold. "Signals: Applying Academic Analytics [DB/OL]. [2013-04-18]. <http://www.educause.edu/EDUCAUSE+Quarterly/EDUCAUSEQuarterlyMagazineVolum/SignalsApplyingAcademicAnalyti/199385>.
- [48] EDUCAUSE. 2010. Next Generation Learning Challenges: Learner Analytics Premises [DB/OL]. [2013-04-18]. <http://www.educause.edu/Resources/NextGenerationLearningChalleng/215028>.
- [49] [52] WCET Predictive Analytics Reporting (PAR) Framework [DB/OL]. [2013-05-11]. <http://wcet.wiche.edu/advance/par-framework>.
- [50] R. Ferguson. The State of Learning Analytics in 2012: A Review and Future Challenges [DB/OL]. [2012-12-11]. <http://kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf>.
- [51] H. Zhang, K. Almeroth, A. Knight, M. Bulger, R. Mayer. Moodog: Tracking Students' Online Learning Activities[A]. World Conference on Educational Multimedia, Hypermedia and Telecommunications [C]. 2007: 4415~4422.
- [54] 祝智庭, 贺斌. 智慧教育: 教育信息化的新境界[J]. 电化教育研究, 2012, (12): 1~13.