

量化学习

——数据驱动下的学习行为分析

刘三妍 杨宗凯 著



科学出版社

量化学习——数据驱动下的 学习行为分析

刘三妍 杨宗凯 著

科学出版社

北 京

内 容 简 介

信息技术助力下的教育已迎来量化时代。数据驱动下的学习行为分析研究跨越了教育领域的多个研究场景,以量化学习者在不同情境下的学习过程作为出发点,为解决数字化情境下个体和群体学习行为数据的采集、挖掘、分析和应用提供一系列方法和策略。本书系统地介绍了量化学习这一全新的教育科学研究范式,从研究的背景、应用领域以及研究方法三个方面展开了全面论述。在研究方法中,分别阐述了学习行为数据采集的场景、标准与规范,以及个体和群体学习行为分析的理论、方法与实践经验,同时在方法介绍中提供了具体的研究案例,并对案例所涉及的实证研究策略和结论展开了深入的探讨。

本书可为教育工作者以及从事教育研究、数据科学研究等相关研究的人员提供参考,亦可为在线教育服务领域的从业人员提供借鉴,相关专业的高年级本科生、研究生也能从本书中受益。

图书在版编目(CIP)数据

量化学习——数据驱动下的学习行为分析/刘三妍,杨宗凯著.

—北京:科学出版社,2016.6

ISBN 978-7-03-048234-1

I. ①量… II. ①刘… ②杨… III. ①学习过程—行为分析
IV. ①G442

中国版本图书馆 CIP 数据核字(2016)第 095459 号

责任编辑:陈 静 / 责任校对:郭瑞芝

责任印制:张 倩 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 6 月第 一 版 开本:720×1000 1/16

2016 年 6 月第一次印刷 印张:9 3/4

字数:185 000

定价:56.00 元

(如有印装质量问题,我社负责调换)

前言

量化学习

当前，以互联网为代表的信息技术已经成为社会进步和发展的重要推动力量，正在全面重塑传统行业，持续地颠覆人们的既有认知。面对汹涌而至的数字化浪潮，教育正面临着来自理论和实践两大层面的挑战：理论界和实践者对教育更加科学化和精确化的追求。

随着数字化时代的深入发展，人类开展科学研究和社会实践的思维方式正在不断演进，技术助力下的数据密集型科学研究范式和数据素养引起了人们的广泛关注，各领域在秉承既有认知模式（如基于实际经验的认知模式）的同时，高度重视基于量化信息的数据化认知模式。信息技术，特别是移动互联、物联网、云计算、人工智能等的快速发展和深度应用，为围绕学习者的学习场景和学习过程的全方位量化提供了可能，由此所产生的数据将有助于对学习者的进行更深刻的全景式洞察，并为回答新一轮教育革命中的两大核心命题（科学化和精确化）提供了新途径。

学习的本质是学习主体知识自我构建和人格不断完善的过程，而这一过程需要依托一系列的学习场景和学习活动来完成。量化学习的主旨是助力于学习主体在学习场景和学习活动中高效地实现知识的建构乃至人格的完善：通过帮助施教者（如教师）更准确地把握学习者的状态，继而开展有针对性的教学活动；通过帮助学习者更深刻地洞察自身的问题，继而采取相应的行动来完善自己。从这个意义上来讲，量化学习是伴随着学习者的一个“知行合一”的过程，具有天然的实践性。作为探索信息技术与教育深度融合的重要抓手，量化学习将助力于科学化和精确化的“学习者中心化”学习新生态的构建。

数据对教育的重要性

信息技术与经济社会的交汇融合引发了数据迅猛增长，大数据技术通过快速获取、处理、分析海量数据并从中发现新知识、创造新价值、提升新能力，已成为新一代信息技术和服务业态，并对全球生产、流通、分配、消费活动以及经济运行机制、社会生活方式和国家治理能力产生重要影响。全球范围内发达国家相

继制定实施大数据发展战略,纷纷在资金和政策上支持大数据研究,以大数据推动国民经济和社会发展。2012年3月,美国政府发布《大数据研究和发展计划》^①,旨在改进现有人们从海量和复杂的数据中获取知识的能力,从而加速美国在科学与工程领域发明的步伐,增强国家安全,转变现有的教学和学习方式。2014年,白宫发布题为《大数据:抓住机遇、保存价值》^②的白皮书。白皮书中指出,“新的数据类型的运用使得研究者对学习行为的研究能力有了质的飞跃”。2010年11月,欧盟通信委员会向欧洲议会提交了《开放数据:创新、增长和透明治理的引擎》^③的报告,并于2011年12月12日正式推行欧盟开放数据战略核心。该战略主要包括对数据处理、数据整合、数据门户网站、开放数据等方面的法律、政策、资金和项目支持,核心目的是在数据开放共享的基础上,提供创新工具和资料,形成有效共享和整合的公共数据池,更好地发挥大数据的价值^④。澳大利亚政府信息管理委员会于2013年8月发布了《公共服务大数据战略》^⑤,旨在推动公共行业利用大数据分析进行服务改革、制定更好的公共政策、保护公民隐私,使澳大利亚在该领域跻身全球领先水平。该战略强调数据的完整性与程序可复用性,数据开发过程中的技术与资源共享,公民隐私保护以及强化学界、业界合作和数据开放。2012年,印度政府科技与教育部颁布了《国家数据共享与获取政策》,其与知识委员会发布的《电子政务建议》和国家信息中心制定的《国家数据共享与获取政策实施指导意见》等文件一起,成为印度推进开放政府数据的指导性文件,确定了政府数据与信息的主动披露原则,促进政府拥有的数据和信息得到共享与使用^⑥。

在教育领域,“数据驱动学校、分析变革教育”的大数据时代已经来临,大数据掀起人类教与学的又一次变革。利用教育数据挖掘技术和学习分析技术,构建教育领域相关模型,探索教育变量之间的相关关系,为教育教学决策提供有效支持已经成为教育的发展趋势。美国教育部于2012年10月发布了《通过教育数据挖掘和学习分析促进教与学》^⑦报告,该报告指出,教育数据挖掘和学习分析技

① White House,《Obama administration unveils “big data” initiative: Announces \$200 million in new R&D investments》(http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf)。

② Executive Office of the President,《Big data: Seizing opportunities, preserving values》(https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf)。

③ 《Open data: An engine for innovation, growth and transparent governance》(<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF>)。

④ 中华人民共和国商务部,《欧盟大数据发展战略》(<http://www.mofcom.gov.cn/article/i/jyj/m/201412/20141200826137.shtml>)。

⑤ Australian Government,department of Finance,《The Australian Public Service Big Data Strategy》(http://www.finance.gov.au/sites/default/files/Big-Data-Strategy_0.pdf)。

⑥ 丁念,夏义堃,《发展中国家开放政府数据战略的比较与启示》。

⑦ Bienkowski M,Feng M, Means B,《Enhancing teaching and learning through educational data mining and learning analytics: An issue brief》。

术的出现,可以实现对于大量细粒度教育大数据的采集、处理和分析,必将改变传统教育数据应用的范式,也将为教育机构的教育决策、教师的教学决策和学习者的自我学习监控提供更加精细化的服务。2013年,美国教育部发布《扩展基于证据的方法在数字学习领域的应用》^①,指出基于大数据的教育评价可以充分利用技术手段采集、整合学习者的学习过程数据和学习结果数据,整合专家评价教师评价、学习者自评、同伴互评等多种评价数据,对学习者的进行多维、全面、深入而可靠的评价。美国教育部在2013年发布的《教育部战略规划(2014—2018)》中^②提出,将帮助各州构建数据系统和通用数据规范并帮助教育机构和教师提升使用数据来提高教学效果的能力。2013年,欧盟正式启动了为期七年的新的研究与创新框架计划——“地平线 2020”。该计划将在未来七年间投入 800 亿欧元,消除科学创新的障碍,确保欧洲产生世界顶级的科学成果。以学习者为中心,收集、存储教育数据并开展系统性的学习分析服务被列为该计划的研究议题之一^③。为了更好地促进基于大数据的教育应用,澳大利亚教育与培训部发布的五年规划^④提出:要构建基于证据和分析的文化,提高数据分析和分享能力,制定基于数据和证据的有效教育政策,推进针对教育评估的数据收集并建立面向决策制定者、教育机构、教师和家长的数据分享机制。印度政府发布的《第十二个五年计划:2012—2017》^⑤中指出将重点关注数据驱动的教育决策,帮助教师更有效地在课堂中使用信息技术,实施“因材施教”,提升教育质量。为实现这一目标,政府将构建学生档案,跟踪学生入学、出勤、退学等信息。同时,政府将和应用开发商合作,促进学生学情诊断、实时数据收集等方面的研究和开发工作。

我国政府同样注意到了大数据的价值和巨大潜力。2015年8月,国务院发布《促进大数据发展行动纲要》(以下简称《纲要》)^⑥,将促进大数据发展提高到国家战略层面。《纲要》提出,大数据是推动经济转型发展的新动力、重塑国家竞争优势的新机遇、提升政府治理能力的新途径,并进一步指出要通过高效采集、有效整合、深化应用政府数据和社会数据,建立“用数据说话、用数据决策、用数

① Cator K, Adams B, 《Expanding evidence approaches for learning in a digital world》。

② 《U.S. Department of Education Strategic Plan for Fiscal Years 2014–2018》(<http://www2.ed.gov/about/reports/strat/plan2014-18/strategic-plan.pdf>)。

③ 《HORIZON 2020-Work Programme 2016–2017 Information and Communication Technologies》(http://ec.europa.eu/research/participants/data/ref/h2020/wp/2016-2017/main/h2020-wp1617-leit_en.pdf)。

④ Australian Government Department of Education and Training, 《Department of Education and Training Corporate Plan 2015–2019》(https://docs.education.gov.au/system/files/doc/other/departments_of_education_and_training_corporate_plan_2015-19_0.pdf)。

⑤ Planning Commission, Government of India, 《Twelfth five year plan (2012/2017)》。

⑥ 《促进大数据发展行动纲要》(http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm)。

据管理、用数据创新”的管理机制，实现基于数据的科学决策。大数据技术将推动着教育领域发生更深层次的创新与变革，并为我国传统教育转型带来新的契机与挑战。在 2012 年发布的《教育信息化十年发展规划（2011—2020 年）》，教育部明确提出建设国家教育管理信息系统是支撑教育管理现代化的基础工程。国家教育管理信息系统建设包括基础数据库建设、数据共享、实现系统互联与数据互通，并建设纵向贯通、横向关联的教育管理信息化体系。在《纲要》中则进一步提到，要“完善教育管理公共服务平台，推动教育基础数据的伴随式收集和全国互通共享。建立各阶段适龄入学人口基础数据库、学生基础数据库和终身电子学籍档案，实现学生学籍档案在不同教育阶段的纵向贯通。推动形成覆盖全国、协同服务、全网互通的教育资源云服务体系。探索发挥大数据对变革教育方式、促进教育公平、提升教育质量的支撑作用。”

学习数据的隐私与安全

随着云计算、平板电脑、智能手环等信息技术和教育教学进一步融合，由教学过程所产生的海量数据和基于数据的精确分析将能进一步发挥大数据技术在教育领域的巨大潜力，帮助教学者实现人才培养个性化、教学评价多样化和教育决策科学化。与此同时，对教学数据的采集、存储和分析过程也对用户隐私保护提出了更高的要求。学习者个人信息、家庭背景、学习行为等数据如若不被妥善管理与使用，这些数据的隐私和信息安全将受到严重威胁。美国教育科技公司 inBloom 与州立教育部门合作，收集了大量学生信息数据并以此为基础向学生提供个性化的学习服务。虽然 inBloom 对这些数据有高级加密措施和严格的数据分享政策，但是该合作仍然遭到了家长的强烈反对，认为教育部门此举违反了美国“家庭教育权利和隐私法案（Family Education Rights and Privacy Act, FERPA）”，不应将这些数据提供给 inBloom 这样的第三方商业机构，inBloom 由此在 2014 年 4 月宣布停止服务^①。对学习数据隐私保护需要从法律法规层面加以管理，从数据标准层面加以规范，从技术层面通过一系列数据处理措施来实现对数据脱敏，在保护个人隐私的同时，促进数据资源的开放共享和合理利用。由于数据安全与隐私涉及技术、法律、伦理等一系列问题，需要专门予以研究与论述，故在本书中，我们重点介绍量化学习的方法和技术，不对数据隐私问题开展进一步的讨论。

① Young E, 《Educational privacy in the online classroom: FERPA, MOOCs, and the big data conundrum》。

本书导读

本书主要分为四个部分。第一部分讲述了量化学习及数据驱动下学习行为分析的概念和发展由来，以及它在教育领域的应用状况与前景，从实际应用需求层面来把握本书的内容。第二部分从物理和虚拟两种不同的空间来阐述学习过程的量化方法，并介绍了在量化过程中所涉及的行为数据规范与标准。第三、四部分分别从个体、群体的视角来探讨学习行为分析的研究场景、研究方法与实验手段。通过这四个部分的阅读，读者可以全面了解数据驱动下的学习行为分析所需的数据、方法以及相关领域的研究概况，为读者日后的学术科研、教学实践以及学习服务系统研发提供理论依据和材料支持。

致谢

我要感谢以下人员：孙建文、郑年亨、刘智、程小鸥、彭晔，没有他们的帮助，这本书是不可能完成的。我还要感谢我们的研究团队，以及研究生在资料收集和整理方面的工作。衷心感谢所有为本书做出贡献的人，感谢你们的付出和努力！也由衷感谢正在阅读此书的读者们，你们的积极参与才能将此书的内容更好地运用于研究与实践，促进数据科学在教育领域的应用，推动教育领域的创新与变革。

作 者

2016年4月

目 录

前言

第一部分 概 述

第 1 章 学习行为分析	3
1.1 教育与大数据同行	4
1.2 学习分析方兴未艾	6
1.3 从量化自我到量化学习	7
1.4 数据驱动的学习行为分析	9
参考文献	11
第 2 章 学习行为分析应用	12
2.1 个性化学习诊断	12
2.2 学习路径规划	13
2.3 个性化学习推荐	14
2.4 学习状态可视化	15
2.5 学习干预	16
参考文献	18

第二部分 学习过程的量化

第 3 章 量化方法	21
3.1 物理空间	21
3.1.1 课堂学习	21
3.1.2 户外学习	24
3.2 虚拟空间	25
3.2.1 学习管理系统	26
3.2.2 大规模在线开放课程	27
3.2.3 智能导师系统	28

3.2.4 面向资源交互的学习过程量化	29
3.2.5 游戏化学习情境	30
参考文献	32
第4章 数据标准与规范	36
4.1 学习分析标准和规范	36
4.2 学习活动规范	38
4.2.1 学习经历规范	38
4.2.2 面向学习情境的交互规范	39
4.3 面向不同学习系统的学习经历规范	41
4.3.1 面向 LMS 的跨平台数据规范	41
4.3.2 面向大规模在线开放课程的数据规范	42
4.3.3 面向智能导师系统的数据规范	43
4.4 面向资源交互的数据规范	44
参考文献	45

第三部分 个体学习行为分析

第5章 学习效果预测	51
5.1 简介	51
5.2 基于统计分析的学习效果预测	53
5.3 基于统计机器学习的学习效果预测	55
5.4 基于时间序列分类的学习效果预测	58
参考文献	61
第6章 知识能力分析	63
6.1 简介	63
6.2 基于标准化测验的知识等级定位	64
6.3 基于交互式测验的高等能力评量	65
6.4 基于概念构图法的知识结构建立	67
6.5 基于自动文章评分的文本回应分析	68
参考文献	69
第7章 认知风格分析	73
7.1 简介	73

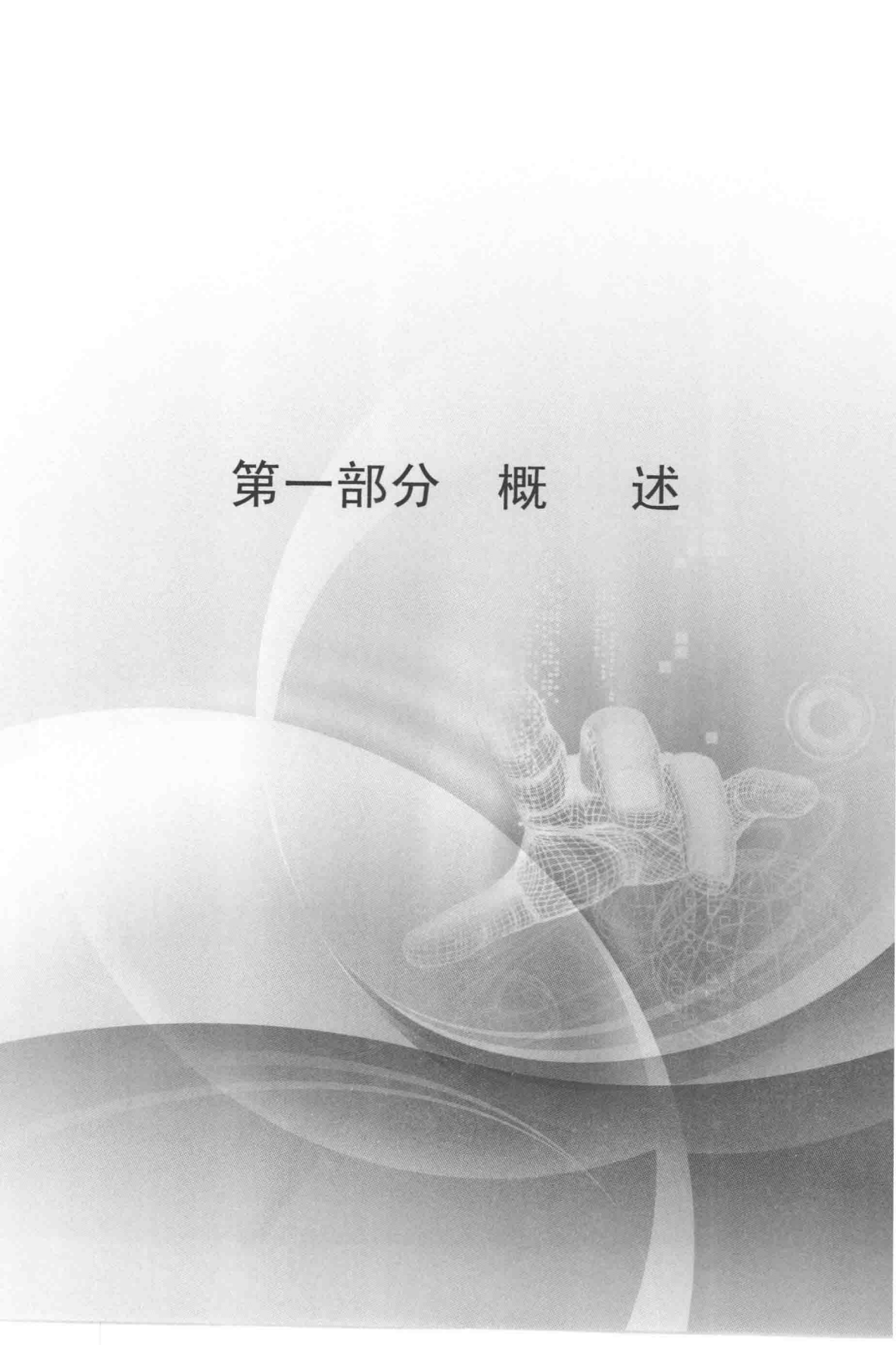
7.2 基于问卷和心理计量测验的认知风格分类	74
7.3 基于神经网络方法的认知风格识别	75
7.4 基于模糊分类树的认知风格预测	77
参考文献	78
第 8 章 学习兴趣挖掘	81
8.1 简介	81
8.2 基于学习者自报告的学习兴趣挖掘	82
8.3 基于在线点击行为的学习兴趣挖掘	84
8.4 基于在线文本内容的学习兴趣挖掘	85
参考文献	86
第 9 章 情感状态识别	89
9.1 简介	89
9.2 基于量表的情感测评方法	90
9.3 基于生理信号的情感检测方法	91
9.4 基于面部表情的情感识别方法	92
9.5 基于语音信号的情感检测方法	94
9.6 基于文本数据的情感分析方法	96
参考文献	98

第四部分 群体学习行为分析

第 10 章 群体学习动力学分析	103
10.1 简介	103
10.2 数据化的群体动力学研究——社会网络分析	104
10.3 基于学习者自报告的社会网络分析法	106
10.4 基于在线行为统计的社会网络分析方法	110
参考文献	113
第 11 章 群体学习行为模式分析	116
11.1 简介	116
11.2 基于聚类方法的学习行为模式识别	116
11.3 基于序列分析的学习行为分析	117
11.4 基于序列模式挖掘的学习行为分析	119

11.5 基于眼动追踪方法的眼动行为分析	120
参考文献	121
第 12 章 协作学习及参与行为分析	125
12.1 简介	125
12.2 基于物理课堂的小组协作学习	126
12.3 基于计算机辅助教学的协作学习	127
12.4 基于移动技术支持的课堂协作学习	128
12.5 基于游戏场景下的协同学习	129
参考文献	130
第 13 章 互动话语行为分析	133
13.1 简介	133
13.2 基于语音记录的话语分析方法	134
13.3 基于姿势识别的话语分析方法	135
13.4 基于文本交互的话语分析方法	136
13.5 基于 ITS 人机智能对话分析	138
参考文献	139
后记	141

第一部分 概 述



第 1 章 学习行为分析

近年来，大数据（Big Data）已成为一个大众耳熟能详的词汇。何为大数据？麦肯锡全球研究院认为大数据是一种规模大到在获取、存储、管理、分析等方面大大超出传统数据库软件工具处理能力范围的数据集合，具有数据规模海量（Volume）、数据流转快速（Velocity）、数据类型多样（Variety）和价值密度较低（Value）等 4V 特征。知名咨询公司 Gartner 定义大数据是需要新处理模式才能具有更强的决策力，洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。

2012 年年初，《纽约时报》撰文宣告“大数据时代”已然降临，并指出大数据正在对各个领域造成影响。无论是金融、通信、商业，还是军事、交通、医疗，基于大数据的决策已成为或正在成为现代社会各行业运行的基础。以商业为例，大数据已经帮助无数商家或厂家实现了柔性化生产、社会化供应链、巨型网络零售以及个性化营销。例如，天猫上一家销售家居的公司，为提高个性化服务能力，不断收集消费者数据，包括各个年龄阶段消费者的偏好、习惯以及全国各大城市的楼盘和户型等信息，形成多侧面、多维度的数据，然后通过进一步整合与分析形成完整的消费者画像。在生产和流通环节，该公司也通过大数据应用，整合企业上下游的供应链信息，优化生产工艺与流程，调整生产策略，以适应市场变化，提高企业竞争力。

再看医疗领域，Apixio 是一家总部位于美国加利福尼亚州的人工智能公司，目标是在实践研究与数据分析的基础上为患者提供个性化医疗服务。Apixio 先从各种来源广泛收集患者的医疗资料，如电子健康档案、医疗记录，以及医生的咨询记录、书面笔记、病理分析结论等非结构化数据，然后经过处理和分析创建出患者的数据模型，汇总大量这样的模型后，就能更为准确地发现疾病的患病率和治疗模式。Apixio 首席执行官 Schulte 说到：“我们的系统可以从医学实践中学习更多的东西，通过计算处理来改进我们的临床护理方法。这其实更像一个‘医疗保健学习系统’，对于什么方法有用，什么方法没用，我们会从真实世界不断更新的数据中获得，如此会更加准确。”

如 Schulte 所言，我们正处在一个全新的世界，以数据洞察力为驱动，可以使我们变得更加精准熟练，这是数据价值的最大体现。国际数据公司（International Data Corporation, IDC）预计到 2020 年，全球将拥有 35ZB 的数据量，若把其全

部刻录到容量为 9GB 的光盘上,叠加后的高度相当于在地球与月球之间往返三次的距离。随着数据量的激增,在各行各业深化大数据的应用是未来社会的发展趋势,从信息技术 (Information Technology, IT) 时代迈向数据处理技术 (Data Technology, DT) 时代则是未来可预见的发展趋势。

1.1 教育与大数据同行

如上所述,大数据正在对社会生产和生活产生广泛的影响,这在教育以外的行业已十分明显。然而,迄今为止,我们的教育系统几乎依然是在沿袭远古教育的范式,管理者仍然习惯根据经验或惯性来制定决策,教师则主要靠个人经验来了解学习者和实施教学。假如孔子走进今天的教室,他也许仍会觉得相当熟悉,他仍可按照他在 2000 多年前就提倡的“因材施教”的理念来指导学习者。这是一种完全由个人经验驱动的施教模式,也是今天我们绝大多数教师正在践行的方式。对此,乔布斯生前也曾感叹为什么信息技术几乎改变了所有领域,却唯独对教育的影响小得令人吃惊?

过去,在班级授课制主导的年代,受技术发展水平的限制,很难在以教室为主阵地的教学活动中收集到足够丰富、精细和客观的数据来刻画学习者或辅助教学决策。如今,随着云计算、大数据、移动通信和物联网等新兴技术的普及,以及大规模开放在线课程 (Massive Open Online Courses, MOOC)、翻转课堂等新的教学模式的冲击,学习者终于可以完全摆脱时空限制,教室或学校不再是教学活动发生的唯一场所或首选项,在线课程、电子书和计算机支持测验等工具使学习过程的跟踪和学习数据的采集不再困难,大数据的收集、处理和分析技术也已成熟,这为数据驱动教育创新与变革、实现教育规模化与个性化的有机统一以及最终实现人们对未来教育理想形态的追求创造了条件。结合迈尔-舍恩伯格等 (2014) 在《与大数据同行:学习和教育的未来》一书中所介绍的如下 3 个正在当前教育领域发生且极具代表性的故事,有理由相信我们正在见证一场新的教育革命,一场教育与大数据同行、通往未来教育之路的历史性变革。

故事 1 Coursera 上有一门热门的机器学习课,主讲人是 Coursera 的联合创始人吴恩达教授,选课人数超过百万。吴教授通过学习平台尽可能全面地收集学习者的行为数据,如通过追踪学习者浏览课程视频的行为,他可轻松发现学习者多次回看同一个视频,或者后退至上一个视频浏览课程内容的行为。此外,在视频课程中穿插突击测验,能够发现学习者是否正在领会教学材料,以及是否卡在

了某个问题上。通过追踪学习者完成在线作业或测验的情况，吴教授能够确定学习者需要额外帮助的具体内容。除对学习者个体的追踪，他也可以统计整个班级的数据，分析群体学习情况，并以此为依据对课程进行调整。吴恩达教授在 Coursera 上对于大规模学习行为数据的分析和运用的尝试吸引了广泛的关注，预示着大数据对教学模式的变革。

故事 2 Duolingo 是一个帮助人们免费学习外语的系统，其创始人 Luis von Ahn 认为针对最佳语言学习方法的实证研究很少，且相关研究所采用的用户规模也较小，将其研究发现作为普遍规律加以推广是不可靠的。为什么不以数以千万的学习者作为研究对象而得出结论呢？Duolingo 使之成为可能。例如，Luis von Ahn 发现母语为西班牙语的人刚开始学习“he”、“she”和“it”等英语代词时，“it”一词容易引起学习者的迷惑和焦虑，原因是其难以翻译成对应的西班牙语。于是，Luis von Ahn 做了几次试验，先只教“he”和“she”，直到数周后坚持学习而不放弃的人数显著增加时再开始教“it”，这样就能显著提高坚持学习的人数。Duolingo 的故事为我们呈现了大数据重塑教育的一种方式，反映了大数据改善学习的要素和潜力。

故事 3 可汗学院（Khan Academy）的创始人可汗先生在上传第一个使其闻名于世的视频之前，便设计了如何收集学习者的行为数据并从中获取信息的机制。可汗认为有了这些数据，他才得以处理那些在过去甚至都难以表达的问题：如学习者在答对问题上花费的时间多，还是在答错问题上花费的时间多？学习者答错是因为对内容理解不够，还是由于身心疲惫？他们的学习更多依靠的是勤奋还是灵感？其次，通过数据分析，可汗学院能为学习者规划最适合他们所学主题的学习路径，教师和家长也可通过数字面板获取学习者的学习进度报告，学习者也可借此发挥自身在学习中的积极作用。可汗学院的故事突显了大数据时代教与学的变化，学校、班级、课本和课程等都可被视为数据收集和分析的平台，分析的结果则可用于教学改良。

当前，由于大数据带来的积极影响，类似的故事正在教育领域不断涌现，如同迈尔-舍恩伯格等（2014）所言：“我们能够理解在最大规模情况下学习者是如何学习的，理解在任何给定的学年中数以百万计的各种数据。我们能够理解在最小规模情况下学习者是如何学习的，理解每一个个体在几分钟的课程中是如何学习的，而不只是每一个个体是如何学习的。不同于旧有的调查世界和样本，我们能够连接上述两类规模——大数据是数以亿万计的小数据的汇集，以进行个体学习行为和群体学习行为的分析和研究。”