

大数据与人文社会科学研究丛书

Big Data in Humanities and Social Sciences

Smarter University

智慧大学
基于大数据运用的智慧大学是
大学的必由之路

Big Data

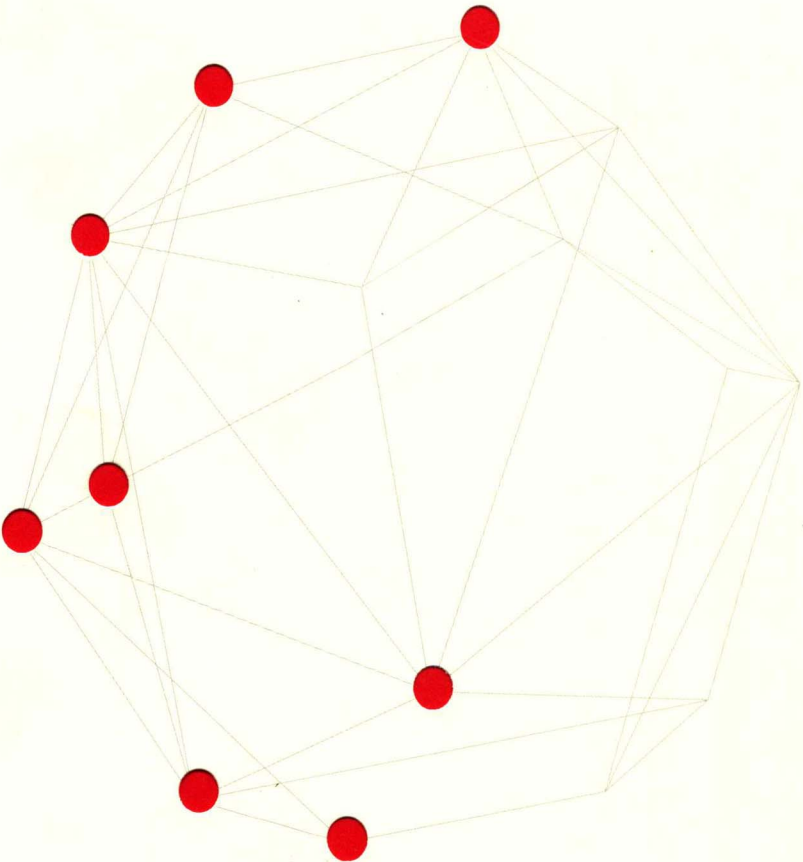
大数据
大数据的高维关联是重构大学
运作模式的根基

Higher Education
Management

高等教育管理
高等教育领域是大数据运动向
人文社会扩展的实验场

Learning Experience

学习体验
用大数据分析改善的学习体验
将重新定位高等教育的社会服
务模式



Building a Smarter University

Big Data, Innovation, and Analytics

建设更智慧的大学

大数据、创新与分析

[美] 杰森·E. 兰 (Jason E. Lane) 编
陆雨婷 等译

大数据与人文社会科学研究丛书

Big Data in Humanities and Social Sciences

Building a Smarter University

Big Data, Innovation, and Analytics

建设更智慧的大学

大数据、创新与分析

[美]杰森·E. 兰 (Jason E. Lane) 编

陆雨婷 等译



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

图书在版编目(CIP)数据

建设更智慧的大学:大数据、创新与分析/[美]
杰森·E.兰(Jason E. Lane)编.陆雨婷等译. —杭州:
浙江大学出版社, 2018.1

(大数据与人文社会科学研究丛书)

书名原文:Building a Smarter University:Big
Data, Innovation, and Analytics

ISBN 978-7-308-17705-4

I. ①建… II. ①杰… ②陆… III. ①信息技术—
应用—高校管理 IV. ①G647-39

中国版本图书馆 CIP 数据核字 (2017) 第 315922 号

浙江省版权局著作权合同登记图字:11-2017-338 号

建设更智慧的大学:大数据、创新与分析

[美]杰森·E.兰(Jason E. Lane) 编
陆雨婷 等译

特约策划	陈 新
策划编辑	曾 熙
责任编辑	吴伟伟 weiweiwu@zju.edu.cn
责任校对	杨利军 张培洁
封面设计	程 晨
出版发行	浙江大学出版社 (杭州市天目山路 148 号 邮政编码 310007) (网址: http://www.zjupress.com)
排 版	杭州中大图文设计有限公司
印 刷	杭州日报报业集团盛元印务有限公司
开 本	710mm×1000mm 1/16
印 张	19.75
字 数	324 千
版 次	2018 年 1 月第 1 版 2018 年 1 月第 1 次印刷
书 号	ISBN 978-7-308-17705-4
定 价	58.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行中心联系方式:0571-88925591; <http://zjdxcbbs.tmall.com>

本书受中央高校基本科研业务费专项资金资助

supported by the Fundamental Research Funds for the Central Universities

献给凯莉和埃莫森
是你们告诉我生命的真谛。

序

南茜·L. 兹莫菲

XI

本书为纽约州立大学高等教育关键问题系列丛书的第三册,与前两册类似,本书同样是一次会议的产物——会议的主题为:建设智慧大学。

2013年10月29—30日,纽约州立大学邀请了数百位来自教育、商业、科技等各领域的精英来共同探讨当下最为火热的话题之一:大数据。与会人员 and 参与本书编写的作者们怀着一探21世纪数据爆炸可能性的目的,试图对大数据如何被利用于提升我们的教育形成更深入的理解。换言之,我们着手探索这样一个问题:如何才能驾驭并针对性地利用潜藏于不断膨胀的海量数据中的未知力量,来建设更具智慧、更有创造力、更灵活以及更高效的大学,以适应当下日益复杂的社会?

在赞助者支持下,会议召集了来自美国25个州,牙买加以及墨西哥等地400余名与会者;其中有50人来自纽约州立大学(SUNY校区),42人来自其他高等院校。会议一共包括65场讲座和会谈,这使我们得以推动一系列具有突破性的对话,比如“论大数据如何促使高等教育转型”,“利用大数据拓宽教育渠道”,“大数据所需警惕之处”以及“数据分析师:21世纪最性感的职业”,等等。同时,我们也很高兴在会议期间引荐首届SUNY大数据研究项目(SUNY Big Data Fellows)的八位成员,他们将是未来大数据科学在不同学科领域应用的领军人物。他们为我们展示了大数据在心理学、金融业、特殊教育、医护工作以及应用数学等领域所能发挥的极大潜力。

XII

然而在对大数据的这一系列探讨中,发生了一件出乎意料的事情。我们的特邀嘉宾哈佩尔·里德(Harper Reed)声称“大”数据这样的说法是不正确的,尽管“大数据”正是他在台上与我们探讨的主题。他强调,将“大”数据冠之以“大”其实并没有必要,它仅是推销这一普通词汇的噱头而已。接着,他更为深入地为我们解释其论点:所谓的“大”,并没有在描述一种新的数据类型或是

一种全新的现象,数据一直只是数据,无论规模大小。“大”这个形容词所能强调的只是我们才开始认识到的数据能被收集转译的潜能,以及对我们平日生活所积攒的大量数据加以应用,进而理解世界的可能。

从某种意义上来说,我并不同意哈佩尔的观点。如虚拟世界中飓风一样袭来的数据伴随着现代科技工具在日常生活中的应用而来,这些技术可以说是眨眼间兴起的,是全新的事物,而学着如何处理这些数据——获取、共享、存储、管理、转译、分析、并加以利用则是我们共同的挑战。2012 年度畅销书《大数据的人类面孔》(*The Human Face of Big Data*)提到,如今人类每两天所产生的数据量,相当于从文明诞生初始直至 2003 年所产生的数据量的总和,且数据产生速率还在不断加快。¹ 假如这样的数据还不能被称为“大”数据,我无法想象还有什么能称得上了。

但我完全同意哈佩尔有关潜能的评论。就了解我们自身及其如何运作这一点来说,人性所面对的是几乎无法计算的庞大变量,但这一切又确实会改变我们整个人类,改变我们介入这个世界、介入未来的轨迹。这时,我们会庆幸于深入探索大数据的极限并尝试从中挖掘尽可能多的改善生活的讯息。

利用数据预测发展模式 and 趋势、改良商业产业已经不是什么新鲜事了。我们每个人每天都在印刻自己的电子足迹,比如刷信用卡、输入文字、发送推特(Tweet)消息、更新个人状态等。可以说,我们毫不费力地产生了这些数以亿计的数据等着这个世界对其加以利用。如今,各大公司正在尝试利用这些数据以提升用户体验。假如你在亚马逊购物,在结账前你会发现一些基于购物车商品以及其他顾客相似购物分析的促销信息推荐。奈飞公司(Netflix)会根据用户过往选择来提供电影或电视的个性化推荐。潘多拉(Pandora)则会通过复杂的算法挑选推送用户可能喜欢的歌曲。

在大数据的商业价值以外,我们同样可以看到各国政府、各个城市以及大大小小的医院,都在试图紧跟大数据的步伐并利用它来迎接自身所面临的各项挑战。民意调查是这一趋势的最早体现。而现在,越来越多的交通数据正被用来帮助疏导城市拥堵,为市民出行提供便利。医院和医疗系统利用电子病历使患者即使去不同的医院诊所问诊,其病情也可以受到持续监控。

一直以来,各大高校的专家学者都是推动大数据开发利用的主导力量。然而,却很少有人研究如何趁着这数据浪潮来改造高等教育机构。

高等教育肩负着改善人类生活,挖掘人类自身潜能,推动社会发展的重任。作为美国最大的综合性大学体系,纽约州立大学拥有超过 463000 名学生、88000 名教职工。它因此也成为纽约州学术、科研、服务、经济等方面的核心力量,放在全国也是如此。纽约州立大学 64 个不同院系就如同纽约州的中枢神经网络,随着知识创造、经济发展和文化繁荣的脉搏一同跳动。

对于 SUNY 来说,假如其他领域都能够如此精巧且创造性地利用数据,我们也一定能够且必须尝试将其用于高等教育领域。正是这一信念推动了我们校对大数据在高等教育应用领域的研究。利用数据,我们可以使我们的机械设备高效而又低成本地运行,扩展并加强从幼儿一直延伸到就业的教育渠道。数据使我们可以实现对高危学生的实时介入,帮助他们获得成功。

于是,和其他高校一样,SUNY 正向着这个目标努力着,同时,我们也在不断思索,如何突破现有观念来更好地完成这一使命。“去学习,去搜索,去服务。”让我们真正地利用数据来更深入地了解现在的年轻人并增加他们进入大学学习的机会,利用数据让学生们为入学做更充分的准备而减少后期补救的成本。我们还可以利用数据,让更多的学生按时毕业,背负更少的债务,收获更多对其有益的知识 and 人生经验——而当他们走出校园、步入工作岗位,这又将促进经济发展。

xiv

当然,我们对实现这一切积极成果的十足信心——即使还没有十足把握的话——也绝非无端而来。这也正是数据向我们展示的——如果我们能有效一致地处理数据的话。

回顾过去几年国家对高校经济补助透明度的重点关注,可以说,我们是完全仰仗了精巧的数据收集才实现了对目标人群的定位、信息传达和实际操作。如果我们的准大学生和他们的家人能获得更多格式统一、内容易读且全面的各大高校资料,他们就能在申请大学时做出更明智的选择,并为今后的花销提前做出预算,适度申请贷款。

假如我们能在学生进入大学校园之前得到更多诸如其生源地、先前的学习经历等的资料,我们就能更好地帮助他们选择其最有可能成功的课程项目。在纽约州立大学,我们与哥伦比亚大学社区数据中心合作,开发了一套数据分析算法,这套算法通过分析学生的考试成绩,他们的听课模式,以及非认知素质评估指标等来帮助我们为这些准大学生提供更适合其未来发展的专业选择指导。

此外，大学如果能与美国国家劳动与经济发展部就数据形成合作，就能根据本地工作岗位和相关训练的需求来规划其学位项目类型。纽约州立大学在这方面已有多年的实践经验，比如根据相应需求，专门培养护理人员与工程师。而最近，我们在与合作伙伴进一步收集数据之后，对这个项目做了再评估与改进。这次，我们又加入了一些市场数据表明高需求的专业，并对相关院系加大了投入。这其中包括了信息工程、人力资源管理、临床医学以及卫生资源管理等。

XV 因此，本书的重任便是将大数据与教育结合起来，加以重点关注，从不同的视角仔细研究大学究竟如何才能利用数据来最大限度地提升其实力，服务水平以及教育质量，以满足当下乃至未来日益复杂的社会需求。纽约州立大学非常荣幸，也极其兴奋能成为大数据领域开拓者的一员，来面对探索大数据与信息革命之无限可能的挑战，并让大数据走入我们的校园，帮助我们建立智慧大学。

2014 年 2 月

注 释

1. 作者在这里引用了谷歌 CEO 埃里克·施密特(Eric Schmidt)的话(Smolán and Erwitte, 2012, p. 1)。

参考文献：

Smolan, R. & Erwitte, J. (2012). *The human face of Big Data*. Sausalito, CA: Against All Odds Productions.

无论情愿与否,我们都已迈入了一个全新的时代——如今的我们、我们的学生,无论做什么,都会留下数据,尤其是处于数字环境之中时。对于绝大多数的读者,这些数据尚未被开发提炼、无法利用。然而,越来越多的前沿学者逐渐意识到,数据正带来前所未有的机遇帮助我们完善学术研究,更好地理解教育渠道,全面提升大学院校的综合实力。而这些机遇正是我们称之为“大数据”的数据革命之卖点的一部分;许多公司也正利用这场数据革命来向教育机构推销其新产品和服务——因为许多教育机构领导明白他们必须要驾驭大数据,却并不完全清楚应该如何实施。

首先,面对大数据我们不应当畏惧。大数据的概念,就是一种既利用传统数据,又运用新兴大规模数据以提升我们娱乐、工作、生活体验水平的数据革命。事实上,科学家与行政管理人员已经在着手研究如何驾驭数据来建设更智慧的城市,更智慧的政府,甚至是更智慧的地球。然而奇怪的是,关于这次数据革命将如何影响高等教育转型的系统研究却少之又少。本书的目的就是向教育学术专家们介绍大数据的基本定义,大数据与传统数据的区别所在,以及这如何使两者共同作用于建立智慧大学。

在应对大数据浪潮的过程中,高等院校最重要的任务之一,就是要培养下一代的数据科学家,即《哈佛商业评论》(*Harvard Business Review*)中所说的“21世纪最性感的职业(人)”。部分院校已经开始领导大数据运用所必需的先进分析算法的发展,而这也改变着其教学和研究方法。随着数据科学人才需求的不断增长,高等院校必须始终走在最前线,为新生职业培养优秀人才。而贯穿本书的主线,便是关乎当下乃至未来管理人员成为数据科学家所必需的知识技能储备,以及对相关领域学术项目发展的深入理解。

本书将从以下三个基本方面论述如何建立智慧大学:(1)大数据概览,包括大数据时代的机遇与挑战;(2)改善学生交流渠道,提升学业完成率和成功

率；(3)改善政策制定和教育决策。值得注意的是，由于篇幅有限，本书并未涵盖所有大数据、数据与数据分析能为大学带来的益处——大学的功能领域实在不胜枚举，而大数据的影响范围又是如此无所不包。

大数据概览。大数据这个概念经常被使用，却并非那样被了解。这种理解的缺失就使得想在大数据领域有所作为的学者们很难做出有意义的研究工作。在第一章，杰森·兰(Jason Lane)和亚力克士·芬赛尔(Alex Finsel)给出了什么是大数据，大数据与传统数据的区别，以及这两者在建立智慧大学的过程中同等重要的原因。同时，他们也谈到了在利用数据提升学生体验过程中潜在而易犯的错误。在第二章中，杰弗里·孙(Jeffrey Sun)从隐私保护，数据管理以及知识产权三方面探究了与大数据相关的法律问题。接下来的第三章，讨论了大数据与那些规模更大却尚未被开发的数据的区别。这一章由伊丽莎白·布林思杰德(Elizabeth Bringsjord)和赛尔摩·布林思杰德(Selmer Bringsjord)执笔，着重强调了会使用新型分析模型所产出的数据与真正理解开发这个模型所使用的数学方法之间的重大区别。

学生交流渠道、学业完成率和成功率。几十年来，教育研究机构一直在收集学生数据来统计分析返校率以及毕业率等。随着科技的日益进步，学校可收集的数据丰度和深度也不断增加。学生校园卡的使用使得学校能追踪学生使用图书馆资源、饮食习惯、医疗健康、课程活动出勤率以及进出教室或宿舍等情况。课程管理软件能监控学生的成绩，能够看到哪个学生借走了哪些课程资料，使用了多久，还能够提供实时的学习评估。一些大学院校还借鉴许多城市的经验，生成了相同形式的校内信息基础设施以提高效率。其中就包括了停车管理、行人情况、车流量、能耗，以及回收工作等信息。大量的数据使得更为精妙的设计有着无限的可能。

进而，杰·高夫(Jay Goff)和克里斯·夏福(Chris Shaffer)在第四章中，从大数据能够如何改良高等院校招生招聘的视角展开了全书的第二部分。弗雷德·方塞卡(Fred Fonseca)和迈克尔·马克沁科夫斯基(Michael Marcinkowski)在第五章则是运用了哲学思辨的方式来思考什么是“大数据学生”。第六章，讨论了学院保留计划(college retention)，本·威尔达夫斯基(Ben Wildavsky)分析了不少正试图利用大数据来提升学生的结业率的新兴企业及其产品。最后在第七章，达雅·欧文斯(Taya Owens)和丹·诺克斯

(Dan Knox)通过对纽约州立大学(SUNY)这一案例的分析,强调了数据分析为政策制定提供信息支持以改善学生转校、交流以及帮助其大学生生活获得最终成功的能力。

政策发展与教育决策。近年来,对于如何更好地保持教育渠道畅通,完善学生交流方式,提高其结业率、成功率的研究在学校内外日益受到关注。对于生源情况、学生的大学生活体验以及其毕业后的发展,我们已经有了不少的探索。并且,通过校内外开发的新型分析算法,我们正尝试进一步改善学生体验和教育渠道。

本书第三部分,丽萨·海尔敏·福斯(Lisa Helmein Foss)从一份对国内高校系主任的调查报告提炼其策略共性开始,鼓励学者在决策中使用大数据分析(第八章)。接着,约翰·切斯洛克(John Cheslock)、罗德尼·休斯(Rodney Hughes)以及马克·乌姆里奇(Mark Umbricht)在第九章指出,整合现有数据集和制定数据统一标准能够帮助高等院校更好地利用现有的数据来支持学院的决策。而第十章中,杰森·兰和拉吉卡·班达里(Rajika Bhandari)则将焦点转向了国际高等教育领域,讨论了在各大高校都寻求国际化发展之际,国际化数据是如何发挥着越来越大的作用。最后,在第十一章中,布莱恩·普莱斯科特(Brian Prescott)通过举例为我们介绍了融合教育和劳动部门的国家级数据库可以更清晰地理解、预测学生毕业后的职业取向。

XX

作为纽约州立大学高等教育关键问题系列丛书的第三本书,本书旨在探索大数据应用于学术产业的机遇与挑战。希望本书能引人深思,让读者理解本书所分析的当下关于利用大数据建立智慧大学最紧迫、复杂的问题,培养学术创新、创造力和下一代的数据科学家。本书的作者们正是基于其学术研究和实践理解,在书中为我们勾勒出这些国际问题的框架。

纽约州立大学高等教育关键问题系列丛书的最初目的,是希望通过每年一次的国家级会议召集具有深思卓见的领导者,共同探寻未来公共高等教育重要问题的解决方案,并将结果编辑出版。最开始的两年,我们就近年来高等教育某些新的公共功能进行了探讨。2011年,我们研究了高等院校作为社会经济驱动力的角色。而2012年,我们从全新视角探讨了高等教育系统在帮助完成国家公共事务,为建设分校区提供支持的角色。无论是教育系统本身,还是高等教育对经济建设起到的作用都已在先前的会议和学术著作中阐述清楚了。这些已有成果为我们之后在这些方向上的研究提供了坚实的基础。

2013年的会议我们决定探讨大数据。大数据这个话题,近来已经引起了媒体的极大关注,然而,在高等教育领域却罕有大数据实践。我们已经开始逐渐认识到,信息时代产生的海量数据能让我们对人类生活、工作、娱乐都产生新的理解。

会议刚开始,我们便发现了一些有趣的课题。比如奥斯汀佩伊州立大学(Austin Peay State University)就有这样一个系统,能根据学生过去的经历,为他们推荐大学课程(这与亚马逊电子商务平台基于顾客消费记录向其推荐商品的系统有些相似);又比如亚利桑那州立大学(Arizona State University)和Knewton自适应教育平台合作,尝试通过学生的电子足迹来分析他们是否有辍学的危险,并加以实时干预以助其获得成功。一些关注高等教育讯息的媒体,比如高等教育纪事出版社(The Chronicle of Higher Education and Inside Higher Ed^①),曾出版发行了一系列专题报道,其中就有这样一篇关于信息时代下人文学科信息检索方式如何转型的文章。然而,即使把这些都算

^① The Chronicle of Higher Education and Inside Higher Ed 是一家总部位于华盛顿,就各大高校相关讯息和学生事务进行网站和报纸新闻报道的出版社。——译者注

上,我们也不能说已经系统地探索了如何利用大数据改进高等院校运作的问题。

所以,我们试图去调研其他行业是如何利用大数据的。在此过程中,我们发现信息革命的实质其实是思考这样一个问题:如何利用现有的各种类型和规模的数据,来使我们更好地理解人类的行为,并将其转化为更好的服务、更优的决策,以及更高的办事效率。正是这一切促成了本书的问世——“纽约州立大学高等教育关键问题系列丛书 2013”《建设更智慧大学:大数据、创新与分析》。在本次大会上,我们召集了各个领域的专家学者,共同讨论这前所未有的数据获取量究竟能给高等教育带来怎样的益处。我们讨论的结果都被记录于此书中,希望这不仅仅是一个开始,更是对高等教育如何利用大数据使学生获得更好发展的长期探索。我希望,本书能够提升我们对大数据助益高等教育的认识,并为学术管理人员提供理解大数据机遇与挑战的工具。

在此,我要真诚感谢无数为此书辛勤工作、深入思考、提供建议与支持的人。由于篇幅有限,我在此特向未点名致谢的各位先行致歉。我要感谢亚力克士·芬赛尔(B. Alex Finsel)为本书提供调研支持,他帮助我们在数以百计的媒体和学术文献中寻找并识别关键词及演讲者,同时,他也是本书第一章合著者之一。我还要感谢莎拉·富勒·克里伯格(Sarah Fuller Klyberg)在编排审稿时所给予的极大帮助。她对写作风格以及对高等教育的深刻认识对完善本书有极大的帮助。接着我要感谢凯特琳·甘布里奥(Kaitlin Gambrill)、约翰娜·肯德里奇-霍尔姆斯(Johanna Kendrick-Holmes),以及朱丽特·普莱斯(Juliette Price)。我们一同承办了这次会议,策划本书。我们花费了大量时间打电话联系相关人员,确定会议日程,处理后勤工作。我还要感谢劳伦·麦凯伯(Lauren McCabe),是他杰出的组织能力,将如此繁杂的事务安排得井井有条,使得这次会议能够圆满完成。当然,我必须要感谢纽约州立大学出版社的每一个人,很高兴能与你们一同工作。其中,我尤其要感谢贝丝·波洛克斯(Beth Bouloukos)、唐娜·狄克松(Donna Dixon)、弗兰·肯内斯顿(Fran Keneston)、詹姆斯·佩尔茨(James Peltz),以及莱恩·莫里斯(Ryan Morris),是他们的贡献让本书的成书时间大大缩短并使该书传播得更广。我要感谢洛克菲勒政府研究所(Rockefeller Institute of Government)的员工们,特别是汤姆·盖斯(Tom Gais)、鲍勃·布洛克(Bob Bullock)、迈克尔·库伯

(Michael Cooper)、希瑟·特雷拉(Heather Trela)、米歇尔·沙伯努(Michelle Charbonneau)、希瑟·斯通(Heather Stone),以及帕蒂·凯德雷特(Patty Cadrette),是他们帮助整合整个系列丛书并在各大场合进行推广。我由衷地感谢本书所有的作者们,是你们的努力,照亮了这片尚未被开发的领域。最后,我尤其要感谢纽约州立大学的校长南茜·兹莫菲(Nancy Zimpher),正是因为她认识到对公共高等教育进行研究和写作的重要性,才有了本系列丛书。

我谨以本书献给我的妻子凯莉以及我们刚出生的女儿埃莫森,她来到这个世界的时候,恰逢本书即将完成。她的到来是对完成工作的全新挑战,然而就算是整个世界(或者所有的数据)也无法换走我对她的爱!

杰森·E. 兰

写于纽约奥尔巴尼

目 录

第一部分 概 论	1
第一章 建设更智慧的大学:数据、大数据及其分析 杰森·E. 兰, B. 亚力克士·芬赛尔	3
第二章 高等教育中和大数据相关的法律问题:伦理问题与警示案例 杰弗里·C. 孙	24
第三章 关于教育,大数据与未开发大数据之辩 伊丽莎白·L. 布林思杰德, 赛尔摩·布林思杰德	50
第二部分 获取信息,达成目标,走向成功	77
第四章 大数据对于大学录取工作和选拔策略的影响 杰·W. 高夫, 克里斯·M. 夏福	79
第五章 什么是“大数据学生”? 弗雷德·方塞卡, 迈克尔·马克沁科夫斯基	105
第六章 助推改变国家:数据使用新方式 ——帮助学生进入大学并完成学业 本·威尔达夫斯基	128
第七章 高等教育体系中意料之外的数据导向革新 ——从学生成功到课程等价 达雅·L. 欧文斯, 丹尼尔·J. 诺克斯	142

第三部分 政策发展与机构决策制定	165
第八章 让数据分析融入高等教育机构	
——提升组织性和学生成功性	
..... 丽萨·海尔敏·福斯	167
第九章 大学中决策操作导向型(大)数据的机遇、挑战和策略	
..... 约翰·切斯洛克,罗德尼·P.休斯,马克·乌姆里奇	188
第十章 衡量高等教育的国际化:数据,大数据及其分析	
..... 杰森·E.兰,拉吉卡·班达里	211
第十一章 大数据与人力资本的发展及其流动	
..... 布莱恩·T.普莱斯科特	232
作者介绍	256
索引	264
译后记	298