

“十三五”国家重点图书出版规划项目

大数据技术与应用

丛书主编
朱扬勇 吴俊伟

Bio Data

Technology and Application Series

肖君
主编

教育
大数据

上海科学技术出版社



大数据技术与应用

教育大数据

肖 君
主编

上海科学技术出版社



图书在版编目(CIP)数据

教育大数据 / 肖君主编. —上海: 上海科学技术出版社, 2020. 1

(大数据技术与应用)

ISBN 978 - 7 - 5478 - 4709 - 1

I. ①教… II. ①肖… III. ①教育工作—信息化
IV. ①G43

中国版本图书馆 CIP 数据核字(2019)第 278661 号

教育大数据

肖君 主编

上海世纪出版(集团)有限公司 出版、发行
上海科学技术出版社

(上海钦州南路 71 号 邮政编码 200235 www. sstp. cn)

印刷

开本 787×1092 1/16 印张 22. 75

字数 450 千字


2020 年 1 月第 1 版 2020 年 1 月第 1 次印刷

ISBN 978 - 7 - 5478 - 4709 - 1/G · 935

定价: 128. 00 元

本书如有缺页、错装或坏损等严重质量问题, 请向工厂联系调换

内容提要



《教育大数据》从教育大数据的重要意义、国内外研究现状以及政策支持出发,对其相关概念和内涵进行了阐述说明,介绍了教育大数据的技术、技术标准、治理与开放、风险管理与法律政策,重点阐述了大数据在教育领域的典型应用,包括课堂教育、校园管理、在线教育、开放远程教育、教育培训行业以及学生综合素质评价等方面的应用,最后提出教育大数据的发展趋势和机遇挑战。

本书的读者对象包括教育信息化研究人员,企业中从事教育大数据产品研发、运营和管理的相关人员,学校中从事教育大数据决策部署的相关人员等;可作为教育领域从事教育大数据研究工作者的参考用书。

大数据技术与应用

学术顾问



中国工程院院士 邬江兴

中国科学院院士 梅 宏

中国科学院院士 金 力

教授, 博士生导师 温孚江

教授, 博士生导师 王晓阳

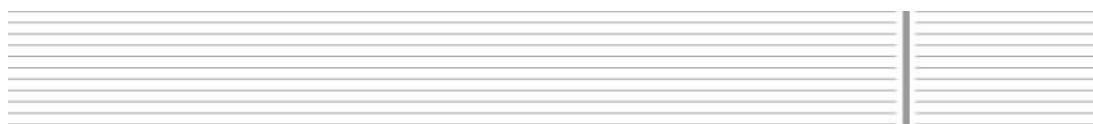
教授, 博士生导师 管海兵

教授, 博士生导师 顾君忠

教授, 博士生导师 乐嘉锦

研究员 史一兵

大数据技术与应用
编撰委员会



丛书指导

干 频 石 谦 肖 菁

主 任

朱扬勇 吴俊伟

委 员

(以姓氏笔画为序)

于广军 朱扬勇 刘振宇 孙景乐 杨 丽 杨佳泓 李光亚
李光耀 吴俊伟 何 承 邹国良 宋俊典 张 云 张 洁
张绍华 张鹏翥 陈 云 武 星 宗宇伟 赵国栋 黄冬梅
黄林鹏 韩彦岭 童维勤 楼振飞 蔡立志 熊 贇 糜万军

本书编委会

主 编

上海开放大学 肖 君

副主编

上海开放大学 王腊梅

上海图书馆上海科学技术情报研究所 曹 磊

上海市电化教育馆 张 治

华东师范大学 吴永和

编 委

北京邮电大学 李 青

上海师范大学 江丰光

北京奥鹏远程教育中心有限公司 龚亚勋

华东师范大学—上海开放大学博士后科研工作站 杜 华

华东师范大学—上海开放大学博士后科研工作站 李绚兮

上海开放大学 朱晓晓


上海大学 朱文浩

腾讯云计算(北京)有限责任公司 潘鸿雁

上海远大学习广场教育科技有限公司(上海教育服务园区) 钱 辉

智园教育研究院 郑旭成

丛书序



我国各级政府非常重视大数据的科研和产业发展,2014年国务院政府工作报告中明确指出要“以创新支撑和引领经济结构优化升级”,并提出“设立新兴产业创业创新平台,在新一代移动通信、集成电路、大数据、先进制造、新能源、新材料等方面赶超先进,引领未来产业发展”。2015年8月31日,国务院印发了《促进大数据发展行动纲要》,明确提出将全面推进我国大数据发展和应用,加快建设数据强国。前不久,党的十八届五中全会公报提出要实施“国家大数据战略”,这是大数据第一次写入党的全会决议,标志着大数据战略正式上升为国家战略。

上海的大数据研究与发展在国内起步较早。上海市科学技术委员会于2012年开始布局,并组织力量开展大数据三年行动计划的调研和编制工作,于2013年7月12日率先发布了《上海推进大数据研究与发展三年行动计划(2013—2015年)》,又称“汇计划”,寓意“汇数据、汇技术、汇人才”和“数据‘汇’聚、百川入‘海’”的文化内涵。

“汇计划”围绕“发展数据产业,服务智慧城市”的指导思想,对上海大数据研究与发展做了顶层设计,包括大数据理论研究、关键技术突破、重要产品开发、公共服务平台建设、行业应用、产业模式和模式创新等大数据研究发展的各个方面。近两年来,“汇计划”针对城市交通、医疗健康、食品安全、公共安全等大型城市中的重大民生问题,逐步建立了大数据公共服务平台,惠及民生。一批新型大数据算法,特别是实时数据库、内存计算平台在国内独树一帜,有企业因此获得了数百万美元的投资。

为确保行动计划的实施,着力营造大数据创新生态,“上海大数据产业技术创新战略联盟”(以下简称“联盟”)于2013年7月成立。截至2015年8月底,联盟共有108家成员单位,既有从事各类数据应用与服务的企业,也有行业协会和专业学会、高校和科研院所、大数据技术和产品装备研发企业,更有大数据领域投资机构、产业园区、非IT

领域的的数据资源拥有单位,显现出强大的吸引力,勾勒出上海数据产业的良好生态。同时,依托复旦大学筹建成立了“上海市数据科学重点实验室”,开展数据科学和大数据理论基础研究、建设数据科学学科和开展人才培养、解决大数据发展中的基础科学问题和技术问题、开展大数据发展战略咨询等工作。

在“汇计划”引领下,由联盟、上海市数据科学重点实验室、上海产业技术研究院和上海科学技术出版社于2014年初共同策划了《大数据技术与应用》丛书。本丛书第一批已于2015年初上市,包括了《汇计划在行动》《大数据评测》《数据密集型计算和模型》《城市发展的数据逻辑》《智慧城市大数据》《金融大数据》《城市交通大数据》《医疗大数据》共八册,在业界取得了广泛的好评。今年进一步联合北京中关村大数据产业联盟共同策划本丛书第二批,包括《大数据挖掘》《制造业大数据》《航运大数据》《海洋大数据》《能源大数据》《大数据治理与服务》等。从大数据的共性技术概念、主要前沿技术研究和当前的成功应用领域等方面向读者做了阐述,作者希望把上海在大数据领域技术研究的成果和应用成功案例分享给大家,希望读者能从中获得有益启示并共同探讨。第三批的书目也已在策划、编写中,作者将与大家分享更多的技术与应用。

大数据对科学研究、经济建设、社会发展和文化生活等各个领域正在产生革命性的影响。上海希望通过“汇计划”的实施,同时也是本丛书希望带给大家一个理念:大数据所带来的变革,让公众能享受到更个性化的医疗服务、更便利的出行、更放心的食品,以及在互联网、金融等领域创造新型商业模式,让老百姓享受到科技带来的美好生活,促进经济结构调整和产业转型。



上海市科学技术委员会副主任

2015年11月

前言

随着新一代信息技术的发展,人类正在进入大数据时代,数据已经成为世界基础性的战略资源,大数据越来越受到各国家和各领域的重视。2012年,联合国在发布的《大数据促进发展:挑战与机遇》白皮书中指出:“大数据时代已经到来,大数据的出现将会对社会各个领域产生深刻影响。”2015年,国务院印发了《促进大数据发展行动纲要》,其中明确指出将全面推进我国大数据的发展和应用,加快数据强国建设。毫无疑问,大数据正在逐步深入人心,快速融入各行各业,正在对人类的工作、生活与思维方式产生巨大的变革。大数据的“威力”也强烈地冲击着教育系统,正在给教育带来全新景观,成为推动教育系统创新与变革的颠覆力量,展现出未来教育的无限魅力。

本书正是在这样的背景下编写,力求把握教育大数据应用的前沿,以教育大数据的应用为特色,阐述大数据在教育各个应用领域的发展和现状。本书共12章:第1章介绍教育大数据的基本概念、内涵、研究现状以及政策支持等;第2、3章介绍教育大数据的技术以及相关技术标准;第4、5章介绍教育大数据的治理与开放、风险管理与法律政策。第6~11章阐述了教育大数据在教育领域的典型应用,包括教育大数据在课堂教育、校园管理、在线教育、开放远程教育、教育培训行业以及学生综合素质评价等方面的应用;第12章提出教育大数据的发展趋势和机遇挑战。

本书由肖君主编,由肖君、王腊梅研究确定主要内容结构,并组织上海开放大学、上海图书馆上海科学技术情报研究所、上海市电化教育馆、华东师范大学、北京邮电大学、上海师范大学、北京奥鹏远程教育中心有限公司、华东师范大学—上海开放大学博士后科研工作站、上海大学、腾讯云计算(北京)有限责任公司、上海远大学习广场教育科技有限公司、上海教育软件发展有限公司、智园教育研究院等单位专家合作组织编写。其中,第1章1.1节、1.2节、1.3节、1.6节,第7章7.1节、7.2节,第9章9.1节、9.2节由

杜华、孙艳超编写；第1章1.4节、1.5节，第6章6.3节由王腊梅编写；第2章、第9章9.4节由龚亚勋编写；第3章由吴永和、朱丽娟编写；第4章由李青、韩俊红、王海兰编写；第5章由曹磊编写；第6章6.1节、6.2节由孙林宏证、郑旭成编写，6.4节由张治、徐冰冰编写；第7章7.3节由朱文浩编写，7.4节由潘鸿雁、丁游、卢冬梅编写；第8章8.1节、8.2节、8.3节由李绚兮编写，8.4节由冯伟、顾方好编写；第9章9.3节由朱晓晓编写；第10章由钱辉、郁凌虚、袁家宏编写；第11章11.1节、11.2节、11.3节由张治、徐冰冰编写，第11.4节由吴蓉瑾编写，11.5节由张志敏编写；第12章由黄子杰、江丰光编写。全书由肖君、王腊梅统稿，潘志敏、李梦圆等硕士研究生协助。此外，本书的研究工作得到了上海市科学技术委员会科研计划项目课题“上海开放远程教育工程技术研究中心”（课题编号：13DZ2252200）和全国教育科学规划课题“大数据下在线学习用户画像的构建其应用研究”（课题编号：BCA160053）资助。

本书受限于作者的学识和水平，尽管已做了最大努力多次修改成稿，但仍难免存在疏漏与不足之处，欢迎广大读者批评指正。在成书过程中参考了大量文献，已在书中尽可能标注，如有疏忽未标注，也敬请谅解。

作者

2019年5月

目 录



第1章 教育大数据概述	1
• 1.1 大数据的定义与特征	2
1.1.1 大数据的定义	3
1.1.2 大数据的特征	5
• 1.2 教育大数据的定义与特征	5
1.2.1 教育大数据的定义	5
1.2.2 教育大数据的特征	8
• 1.3 教育大数据的类型	9
1.3.1 根据数据结构化的程度划分	10
1.3.2 根据数据产生的环节划分	11
1.3.3 根据主体划分	12
• 1.4 教育大数据的研究现状和国内外相关政策	15
1.4.1 教育大数据的研究现状	15
1.4.2 国内外教育大数据政策	17
• 1.5 教育大数据的意义	21
1.5.1 教育大数据驱动教育决策创新	21

1.5.2	教育大数据助力教育教学变革	22
1.5.3	教育大数据促进学习方式转变	22
• 1.6	教育大数据的关键领域	23
1.6.1	学习分析	23
1.6.2	教育数据挖掘	26
	参考文献	28
第2章	教育大数据的技术	31
• 2.1	教育大数据的处理流程与技术架构	32
2.1.1	教育大数据的处理流程	32
2.1.2	教育大数据的技术架构	33
2.1.3	教育大数据的技术架构案例	34
• 2.2	教育大数据的采集	36
• 2.3	教育大数据的预处理	40
2.3.1	数据清洗	41
2.3.2	数据转换	41
2.3.3	具体示例	41
• 2.4	教育大数据的存储	42
2.4.1	单机文件存储	42
2.4.2	关系型数据库存储	43
2.4.3	分布式存储	43
2.4.4	云存储	44
• 2.5	教育大数据的处理	45
2.5.1	常用的数据处理方法	45
2.5.2	处理架构	46
• 2.6	教育大数据的分析与挖掘	48
2.6.1	数据分析	48

2.6.2	数据挖掘	49
• 2.7	教育大数据的可视化	50
	参考文献	55
第3章	教育大数据技术标准	57
• 3.1	教育大数据技术标准国际情况	58
3.1.1	ISO/IEC JTC1(SC32、SC36; WG8、SG2; WG9)	59
3.1.2	IEEE BDGMM	61
3.1.3	IMS GLC	62
3.1.4	ADL(xAPI)	63
3.1.5	ITU - T(SG13、SG16、SG17、SG20)	64
3.1.6	NIST NBD - PWD(NBD - PWG)	65
3.1.7	NCES(CEDS)	65
3.1.8	国外具体标准	66
• 3.2	教育大数据技术标准国内情况	69
3.2.1	全国信标委大数据标准工作组	69
3.2.2	CELTSC	70
3.2.3	中国教育大数据标准体系框架构建	72
3.2.4	中国具体标准	75
	参考文献	81
第4章	教育大数据的治理与开放	83
• 4.1	教育大数据的治理	84
4.1.1	数据治理概述	84
4.1.2	教育大数据治理现状	88
4.1.3	教育大数据治理案例	89
4.1.4	教育大数据治理方法	93
• 4.2	教育大数据的开放	95

4.2.1	开放教育数据的概念和分类	95
4.2.2	教育大数据的开放和使用	96
4.2.3	全球教育大数据开放概况	97
4.2.4	教育大数据开放案例	98
4.2.5	教育大数据开放的问题和启示	102
	参考文献	104
第5章 教育大数据风险管理与法律政策		109
• 5.1	教育大数据风险管理	110
5.1.1	大数据安全风险概述	110
5.1.2	教育大数据风险问题	112
5.1.3	教育大数据风险机理	113
5.1.4	教育大数据风险特点	114
5.1.5	教育大数据重点风险领域	117
• 5.2	教育大数据法律政策	120
5.2.1	国内外教育大数据法律政策与风险防控实践	120
5.2.2	教育大数据法律政策体系建设与风险防控对策	126
	参考文献	129
第6章 教育大数据与课堂教学应用		131
• 6.1	教育大数据背景下的课堂教学	132
• 6.2	基于教育大数据的教学应用模式	133
6.2.1	个性化分组学习	133
6.2.2	交错式自主开放学习	133
6.2.3	游戏化翻转课堂教学	134
6.2.4	自适应测试应用	135
6.2.5	个性化学习诊断	135
• 6.3	案例 1：上海开放大学智慧学习分析与评估系统	137
6.3.1	简要介绍	137

6.3.2	功能层面	137
6.3.3	应用层面	138
<hr/>		
• 6.4	案例 2 : 上海电教馆的高中研究性学习智能支持系统	142
6.4.1	简要介绍	142
6.4.2	技术层面	143
6.4.3	功能层面	148
6.4.4	应用层面	152
	参考文献	154
<hr/>		
第7章	教育大数据与校园管理应用	157
<hr/>		
• 7.1	教育大数据背景下的校园管理	158
7.1.1	校园管理面临的挑战	158
7.1.2	教育大数据给校园管理带来的变革	160
<hr/>		
• 7.2	基于教育大数据背景下的校园管理应用模式	164
• 7.3	案例 1 : 上海大学基于大数据的决策支持系统	172
7.3.1	项目背景	172
7.3.2	技术架构	173
7.3.3	数据模型管理	175
7.3.4	功能模块	177
<hr/>		
• 7.4	案例 2 : 天津市和平区智慧教育数据中心建设	180
7.4.1	项目背景	180
7.4.2	设计思路	180
7.4.3	整体架构	183
7.4.4	大数据与智慧教育数据中心应用	184
	参考文献	190
<hr/>		
第8章	教育大数据与在线教学应用	191
<hr/>		
• 8.1	教育大数据背景下的在线教学	192

8.1.1	国内外在线教学的发展	192
8.1.2	在线教学面临的挑战	193
• 8.2	基于教育大数据在线教学应用模式	195
8.2.1	学习分析模式	195
8.2.2	教学分析模式	202
8.2.3	管理分析模式	205
• 8.3	案例 1: 可视化学习分析大数据应用	209
8.3.1	Open edX 学习分析	209
8.3.2	iForum 学习分析	213
• 8.4	案例 2: 上海学习网在线学习大数据应用	220
8.4.1	平台简介	220
8.4.2	学习功能	220
8.4.3	架构设计	221
8.4.4	平台应用	223
	参考文献	228
第9章	教育大数据与开放远程教育应用	233
• 9.1	教育大数据背景下的开放远程教育	234
9.1.1	开放远程教育发展面临的挑战	234
9.1.2	大数据带给开放远程教育领域的机遇	237
• 9.2	基于教育大数据的开放远程教育应用模式	241
9.2.1	构建全过程学习分析的开放学习平台	242
9.2.2	建立学习者体验支持下的开放学习资源	243
9.2.3	提供数据分析支持下的学习支持服务	243
9.2.4	支持数据驱动的精准管理与决策	245
• 9.3	案例 1: 上海开放大学个性化学习支持系统大数据应用	246
9.3.1	个性化学习支持系统设计目标	246