



高等院校计算机类规划教材
国家新闻出版改革发展项目库入库项目
数据科学与大数据技术专业教材丛书

大数据入门与实验

INTRODUCTION AND EXPERIMENT OF BIG DATA

吴斌 滕一阳◎编著



北京邮电大学出版社
www.buptpress.com



高等院校计算机类规划教材
国家新闻出版改革发展项目库入库项目
数据科学与大数据技术专业教材丛书

大数据入门与实验

吴 斌 滕一阳 编著



北京邮电大学出版社
www.buptpress.com

内 容 简 介

本书是一本大数据技术入门教材,先介绍了大数据技术入门知识,再重点介绍了一个大数据分析教学平台——BDAP,并列出了大量大数据分析入门实验教学案例,详细展示了运用BDAP进行实验的过程。本书主要内容包括大数据初识、大数据技术初识、大数据分析教学平台、大数据分析教学平台的实验解析和扩展实验案例以及如何入门基础上渐进实现个性化培养提升等内容。本书的目标是在讲解大数据相关基础知识的同时培养读者运用大数据分析教学平台进行大数据分析实验的能力。本书可作为计算机学科相关专业,特别是数据科学与大数据技术专业的基础教材,也可以作为大数据知识普及类的学习参考书。

图书在版编目(CIP)数据

大数据入门与实验 / 吴斌, 滕一阳编著. -- 北京: 北京邮电大学出版社, 2022. 4
ISBN 978-7-5635-6603-7

I. ①大… II. ①吴… ②滕… III. ①数据处理—教材 IV. ①TP274

中国版本图书馆CIP数据核字(2022)第006883号

策划编辑: 姚 顺 刘纳新 责任编辑: 王小莹 封面设计: 七星博纳

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路10号

邮政编码: 100876

发 行 部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 保定市中华美凯印刷有限公司

开 本: 787 mm×1 092 mm 1/16

印 张: 13.75

字 数: 322千字

版 次: 2022年4月第1版

印 次: 2022年4月第1次印刷

ISBN 978-7-5635-6603-7

定价: 36.00元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

大数据顾问委员会

宋俊德 王国胤 张云勇 郑 宇
段云峰 田世明 娄 瑜 孙少隣
王 柏

大数据专业教材编委会

总主编：吴 斌

编 委：宋美娜 欧中洪 鄂海红 双 锴
于艳华 周文安 林荣恒 李静林
袁燕妮 李 劼 皮人杰

总策划：姚 顺

秘书长：刘纳新

为顺应全球范围内运用大数据推动经济发展、完善社会治理体系、提高政府服务和监管能力的大趋势,2015年国务院印发了《促进大数据发展行动纲要》,此后我国大数据相关产业迎来了进一步的发展。据2021年7月中国互联网协会发布的《中国互联网发展报告(2021)》统计,在大数据领域2020年我国大数据产业规模达到718.7亿元,增幅领跑全球数据市场。大数据已渗透到社会生活的方方面面,政府部门,医疗、金融、交通、教育、传媒、电商、电信、服务等行业对大数据相关人才的需求量逐年递增,因此各大高校争相开设大数据相关专业以培养产业急需的专业人才。2015年,教育部公布新增数据科学与大数据技术专业。2016年,获批开设此专业的学校有北京大学、中南大学和对外经济贸易大学三所学校,截至2022年3月,开设此专业的学校已经累计增加到700多所。并且,截至2022年3月,开设大数据管理与应用专业的学校达到182所。同时全国共有上千所职业院校成功申报了大数据技术与应用专业,以培养大数据应用型、实战型人才。大数据相关人才培养进入快车道。此外,大数据技术与当前新兴产业紧密关联,在大学生创新创业活动中,已经成为重要的基本技术之一,因此,大数据技术相关知识的学习在计算机科学与技术等一些相关专业上逐渐向通识教育转变。而大数据技术还是一门实践性非常强的专业技术,需要加强学生的专业实践能力,本书就是在这样的背景下编写的,它不仅满足大数据基本知识教育,而且也紧密结合实践能力培养。

本书先介绍大数据的时代背景和大数据相关入门知识,包括大数据的基本概念、数据分析流程、大数据技术框架与生态等;然后介绍如何运用大数据基本知识与工具进行大数据预处理、分析及实验。本书重点介绍了一款大数据分析教学平台——BDAP,它是由北京邮电大学计算机学院数据科学与服务中心研发的大数据分析教学平台,包含了该中心在数据科学与大数据技术领域十多年积累的科研成果,已经在计算机学院本科生相关大数据实验课程教学中应用了四年多。上千名学生在此平台上进行学习与实验。而且,本书还介绍了大数据分析教学平台中的各类实验,从易到难地详细介绍了实验过程和要

求。为了进一步培养学生利用海量数据进行数据分析和挖掘的兴趣,本书由浅入深地介绍了许多与深度学习相关的概念、领域,并以实战为目标,对若干个流行的实验平台进行了简单的介绍。最后,本书总结结合 BDAP 的教学实践经验,展望科教融合下的课程建设。

读者可通过网址 <http://www.bupt-dssc.org.cn> 进入 BDAP。首先请刮开本书封底上的“北邮智信”验证涂层获得激活码,然后进行注册、登录。

本书内容源于作者四年多在北京邮电大学计算机学院创新创业实践课中大数据技术单元的教学讲义。北京邮电大学计算机学院开设的创新创业实验课已被列为国家级一流本科社会实践类。本书可以作为大数据相关专业课程的实验参考书,也可以用于计算机相关专业的大数据技术入门基础实验教学。

本书内容丰富,浅显易懂,适合对大数据技术感兴趣的零基础人员学习。同时,本书配备了专业学习实践平台,整理了大量学习实验案例,由浅入深,循序渐进地引导读者走进大数据技术的世界。

本书作为高等院校计算机类规划教材系列和数据科学与大数据技术专业教材丛书系列中的一本书,得到了北京邮电大学出版社的大力支持。同时,在本书的编写过程中,作者得了北京邮电大学计算机学院数据科学与服务中心的王柏老师、白婷老师,以及杨卉帆、曹晨雨、晏成昊、李芳涛、任浩箐、崔翔冲、胡益博、漆金晟、张静怡同学的支持,在此感谢为本书出版而付出辛劳的所有人员。

吴 斌
于北京邮电大学

第 1 章 大数据的时代背景	1
1.1 大数据的产生	2
1.2 大数据战略	3
1.2.1 大数据战略的内涵及意义	3
1.2.2 中国大数据战略的层次体系	4
1.3 大数据的应用	8
1.3.1 大数据在电子政务中的应用	8
1.3.2 大数据在网络通信行业中的应用	10
1.3.3 大数据在医疗行业中的应用	11
1.3.4 大数据在能源行业中的应用	12
1.3.5 大数据在零售行业中的应用	15
1.4 大数据人才的需求	15
1.4.1 大数据人才的能力要求	16
1.4.2 大数据人才的分类	17
1.5 数据科学与大数据技术专业	18
1.5.1 数据科学与大数据技术专业的定位	20
1.5.2 数据科学与大数据技术专业的培养目标	21
第 2 章 大数据初识	22
2.1 大数据的定义	22
2.2 大数据的特征	23
2.3 典型行业的大数据	24
2.3.1 金融行业的大数据	24
2.3.2 健康医疗行业的大数据	25

2.3.3	电信行业的大数据	25
2.3.4	电子商务行业的大数据	26
第3章	大数据技术初识	28
3.1	数据分析流程	28
3.1.1	数据分析的基本流程	29
3.1.2	数据预处理	29
3.1.3	数据探查	30
3.1.4	数据建模与应用	30
3.1.5	数据可视化	31
3.2	数据分析技术	33
3.2.1	经典数据挖掘	35
3.2.2	机器学习简介	41
3.2.3	模型评估	48
3.3	大数据技术框架与生态	54
3.3.1	Hadoop	56
3.3.2	Spark	60
3.3.3	NoSQL	63
第4章	大数据分析教学平台——BDAP	66
4.1	BDAP 简介	66
4.1.1	大数据教学的现状与挑战	66
4.1.2	BDAP 总览	67
4.1.3	BDAP 的特点与优势	68
4.2	BDAP 的功能	68
4.2.1	文件管理功能	70
4.2.2	数据挖掘功能	74
4.2.3	可视化功能	79
4.2.4	深度学习探索功能	81
4.2.5	在线编程功能	82
4.2.6	作业管理功能	83
4.2.7	视频人物社交关系分析功能	85
4.3	BDAP 的使用方法	85
4.3.1	注册与登录	85

4.3.2	文件上传/下载	86
4.3.3	workflow 搭建	87
4.3.4	示例查看	89
第 5 章	大数据分析教学平台的实验解析	90
5.1	数据预处理实验	91
5.1.1	数据导入实验解析	91
5.1.2	数据清洗实验解析	93
5.1.3	数据生成与筛选实验解析	97
5.1.4	数据聚合与排序实验解析	100
5.2	基础实验	104
5.2.1	分类算法实验解析	104
5.2.2	聚类算法实验解析	112
5.2.3	关联规则实验解析	117
第 6 章	大数据分析教学平台的扩展实验案例	124
6.1	BDAP 的扩展算法实验	125
6.1.1	分类及聚类实验	125
6.1.2	文本分析实验	134
6.1.3	协同过滤实验	139
6.2	复杂网络分析实验	143
6.2.1	图基本属性实验	145
6.2.2	社团发现实验	147
6.2.3	最小生成树实验	149
6.3	在线编程实验	151
6.3.1	k -means 算法编程实验	152
6.3.2	手写数字识别实验	155
第 7 章	从初识走向熟练:个性化培养	159
7.1	能力测试	159
7.2	个性化培养目标及途径	162
7.3	技术渐进	162
7.3.1	深度学习	163
7.3.2	自然语言处理	167

7.3.3	计算机视觉	171
7.3.4	图计算	176
7.4	经典案例	178
7.4.1	Kaggle——以手写数字识别为例	179
7.4.2	阿里云天池——以新闻文本分类为例	185
7.4.3	Biendata	190
第8章	科教融合展望	194
8.1	BDAP的教与学	194
8.1.1	结合BDAP的教学方法	194
8.1.2	结合大数据特点的实验课程体系建设	195
8.1.3	教学实践经验总结	195
8.2	BDAP的科教融合	196
8.2.1	科教融合的内涵	197
8.2.2	大数据专业科教融合的特征	197
8.2.3	科教融合下的课程建设思路与实施步骤	198
8.2.4	实施科教融合的难度	199
参考文献	201

第 1 章

大数据的时代背景

2008 年 *Nature* 出版专刊 *Big Data*, 它从互联网技术、网络经济学、超级计算、环境科学、生物医药等多个方面介绍了海量数据带来的挑战^[1]。2011 年 *Science* 推出关于数据处理的专刊 *Dealing with Data*, 讨论了数据洪流 (data deluge) 带来的挑战, 特别指出, 倘若能够更有效地组织和使用这些数据, 人们将得到更多的机会, 从而发挥科学技术对社会发展的巨大推动作用^[2]。2011 年 5 月, 全球知名咨询公司麦肯锡发布报告称: “数据, 已经渗透到当今每一个行业和业务职能领域, 成为重要的生产因素。人们对于海量数据的挖掘和运用, 预示着新一波生产率增长和消费者盈余浪潮的到来, 标志着大数据时代的到来”。人、机、物三元世界的高度融合引发了数据规模的爆炸式增长和数据模式的高度复杂化, 世界已进入网络化的大数据 (big data) 时代。



Big Data 专刊

大数据是现有产业升级与新产业诞生的重要推动力量。数据为王的大数据时代到来后, 产业界的需求与关注点发生了重大转变: 企业关注的重点转向数据; 计算机行业正在转变为真正的信息行业, 从追求计算速度转变为关注大数据处理能力; 软件行业从以编程为中心转变为以数据为中心。大数据处理的兴起也改变了云计算的发展方向, 使其进入以分析即服务 (AaaS) 为主要标志的 Cloud 2.0 时代。采用大数据处理方法, 生物制药、新材料研制生产的流程会发生革命性的变化, 可以通过数据处理能力极高的计算机并行处理, 同时进行大批量的仿真比较和筛选, 大大提高科研和生产效率, 甚至使整个行业迈入数字化与信息化的新阶段。数据已成为与矿物和化学元素一样的原始材料, 未来可能形成数据服务、数据探矿、数据化学、数据材料、数据制药等一系列战略性的新兴产业。



Dealing with Data

大数据还引起了科技界对科学研究方法论的重新审视, 正在引发科学研究思维与方法的一场革命。最早的科学研究只有实验科学, 随后出现了以研究各种定律和定理为特征的理论科学。由于理论分析方法在许多问题上过于复杂, 难以解决实际问题, 所以人们开始寻求模拟的方法, 导致计算科学的兴起。海量数据的出现催生了一种新的科研模式, 即面对海量数据, 科研人员只需从数据中直接查找或挖掘所需要的信息、知识, 甚至无须直接接触研究的对象。

1.1 大数据的产生

现在的社会是一个信息化、数字化的社会,互联网、物联网和云计算技术的迅猛发展,使得数据充斥着整个世界。与此同时,数据也成为一种新的自然资源,亟待人们对其进行合理、高效、充分地利用,使之能够给人们的生活工作带来更大的效益和价值。在这种背景下,不仅数据的数量以指数形式递增,而且数据的结构越来越趋于复杂,这就赋予了“大数据”比以往普通“数据”更加深层的内涵。

在科学研究(天文学、生物学、高能物理等)、计算机仿真、互联网应用、电子商务等领域,数据量呈现快速增长的趋势。美国互联网数据中心(IDC)指出,互联网上的数据每年将增长 50% 以上,每 2 年便将翻一番。数据并非单纯指人们在互联网上发布的信息,全世界的工业设备、汽车、电表上有着无数的数码传感器,随时感应和传递运动速度、温度、湿度乃至空气中化学物质的变化等,这也产生了海量的数据信息。

(1) 科学研究产生大数据。现在的科研工作比以往任何时候都依赖大量的数据信息交流处理,尤其是在各大科研实验室之间研究数据的远程传输的时候。例如,类似希格斯玻粒子的发现就需要每年 36 个国家的 150 多个计算中心之间进行约 26 PB 的数据交流。

(2) 物联网的应用产生大数据。物联网(the Internet of things)^[3]是新一代信息技术的重要组成部分,解决了物与物、人与物、人与人之间的互联。本质上,人与机器、机器与机器的交互大都是为了实现人与人之间的信息交互而产生的。在这种信息交互的过程中,催生了从信息传送到信息感知再到面向分析处理的应用。人们接收日常生活中的各种信息,将这些信息传送到数据中心,利用数据中心的智能分析决策得出信息处理结果,再通过互联网等信息通信网络将这些数据信息传递到四面八方。而在互联网终端的设备利用传感网等设施接收信息并进行有用的信息提取,得到自己想要的数据结果。

(3) 海量网络信息的产生催生大数据。互联网时代,数以亿计的机器、企业、个人随时随地都会获取和产生新的数据。互联网搜索巨头 Google 公司现在能够处理的网页数量在 10^{12} 以上,每月处理的数据超过 400 PB,并且呈继续高速增长的趋势;Youtube 中每天上传的总视频时长超过 7 万小时;截至 2021 年 3 月,淘宝年度活跃消费者达到 7.9 亿,月度活跃用户超 7.9 亿,淘宝上约有 8.8 亿件在线商品,每天超过数千万笔交易,单日产生的数据量超过 50 TB,存储的数据量达到 40 PB;新浪微博上每天有数十亿的外部网页和 API 接口访问需求,每分钟都会发出数万条微博。据 IDC 的研究结果显示,互联网每年产生的数字信息量在以 60% 的速度增长。截止到 2021 年,全球每年产生的数据信息已达到 59 ZB。所有的这些都是海量数据的呈现。

随着社交网络的逐渐成熟,传统互联网转变为移动互联网,移动宽带速度的迅速提升,除了个人计算机、智能手机、平板电脑等常见的客户终端之外,更多更先进的传感设备、智能设备(如智能汽车、智能电视、工业设备和手持设备等)都将接入网络,由此产生的数据量比以往任何时期都多,数据增长速度也比以往任何时期都快,互联网上的数据量正在迅猛增长^[4]。

1.2 大数据战略

1.2.1 大数据战略的内涵及意义

(1) 大数据战略产生的背景

当前信息技术发展迅速,新型信息的发布方式不断涌现,引发了数据规模的爆炸式增长和数据模式的高度复杂化。除人力资源、自然资源外,数据正成为另一种重要的战略资源。作为新一代信息技术与服务业态,大数据以从海量数据集中发现新知识、创造新价值、提升新能力为主要特征,对国民经济、国家安全、科学研究、生活方式等产生重要影响。

许多国家在多个领域内积极支持大数据的发展和应用,制定大数据战略,力求抢占优势地位,各个国家间的竞争日趋激烈。美国高度重视大数据的研发与应用,是首个研究大数据并将其列为国家战略的国家,早在2012年就发布《大数据研究与发展计划》,将大数据作为加强国家安全的关键因素,并成立“大数据高级指导小组”,启动部署大数据国家战略;2014年发布《大数据:把握机遇,维护价值》,强调大数据的有效利用;2016年又发布了《联邦大数据研究与开发战略计划》,围绕大数据研究和开发的关键领域提出下一步发展战略,以维持其在数据科学领域强有力的领导地位。澳大利亚政府则于2013年发布《公共服务大数据战略》,侧重于大数据在公共服务中的应用,在大数据的管理、大数据与隐私保护、大数据整合分析、大数据开放共享4个方面给予特别关注。欧盟于2014年发布《数据驱动经济战略》,聚焦大数据价值链,倡导欧洲各国紧抓大数据发展机遇。此外,英国、日本等国家也高度重视本国的大数据发展,纷纷出台大数据战略。其中:英国为帮助该国经济从疫情中复苏,2020年发布《国家数据战略》;日本将大数据视为信息通信领域的战略重点,重视大数据的务实应用,先后推出《面向2020年的ICT综合战略》《创建最尖端IT国家宣言》《数据与竞争政策研究报告书》等,在数据开放流通、培养大数据人才等方面做出部署。

中国正处于从追随者、模仿者到开拓者、引领者的过渡时期。面对大数据领域日益激烈的竞争,实施国家大数据战略、统筹制定大数据中长期发展路线图、重构国家综合竞争优势、谋求未来竞争中的新制高点迫在眉睫。

(2) 大数据战略的内涵

“大数据”从被提出到获得普遍认可并成为全球热词,伴随的是数据在各行业领域的深层渗透与应用。作为复杂而庞大的数据集,大数据具备很大的分析与挖掘价值,是影响竞争和发展的重要因素。大数据战略是视大数据为一种战略资源的科学战略布局,而大数据国家战略则是将大数据提升到国家战略资源规划层面。其目的是关注如何利用大数据技术解决民生、国家治理、国家安全以及国际竞争力等问题。

大数据战略落脚于战略,实质上是关于大数据发展应用的全局策划和指导。大数据并不简单指大规模的数据集合,它是海量数据与处理这些数据的技术的结合,是对数据集

合的有效应用,通过采集、加工、处理和分析庞杂多样的数据以发现新知识并创造新价值。所以,对大数据含义的界定实际上是一种观念变革,大数据战略的真正内涵也正在于此。

(3) 大数据战略的意义与作用

大数据是众多关键行业关注的问题。在信息化发展的新阶段,大数据对经济发展、社会秩序、国家治理、人民生活都会产生重大影响。扩张发展大数据与其他行业的结合会催生新业态。一方面,数据的挖掘如同开采新型能源,可以发现新知识、创造新价值,成为驱动发展的重要推动力;另一方面,大数据作为提升国家治理能力和重塑国家竞争优势的新手段,制定大数据战略,提高数据治理能力有利于实现政府的科学决策、社会的精准治理和公共服务的高效有序。最后,当今世界综合国力的竞争不再局限于传统实体资源的竞争,数字主权与数据安全也成为影响国际竞争的重要因素,“谁掌握了数据,谁就掌握了主动权”成为全球共识,将大数据发展提升至国家战略高度已经成为普遍共识。全球实践也证明,大数据战略不仅可以提升社会生产力、创造新的社会价值,而且可以提高政府管理效率、提升服务水平、加快创新能力建设^[5]。因此,在挖掘大数据价值和促进大数据发展方面,制定和实施国家大数据战略势在必行。

现阶段,尚未有任何国家对大数据领域形成垄断。中国作为重要的大数据资源集散地,有着世界上最多的互联网用户,这是成为数据强国的历史性机遇。面对新形势下的机遇与挑战,推动实施国家大数据战略,深刻了解其科学内涵,是加快数字中国建设的必然选择。

1.2.2 中国大数据战略的层次体系

(1) 国家层次的大数据战略

① 综合型大数据战略

近年来,我国推出了一系列大数据政策,以指导大数据技术应用、产业及技术的发展。党中央、国务院有关政策的部分统计见表 1.1。

2015 年,中共十八届五中全会提出“国家大数据战略”之后,国务院发布了《促进大数据发展行动纲要》,从国家大数据发展战略的全局高度出发,明确提出要促进数据资源的开放共享、构建数据强国、推动大数据治国,这是第一份指导我国大数据发展的权威性、系统性纲领文件,为我国大数据应用、产业和技术的发展提供了行动指南。2016 年 7 月,党中央和国务院立足于中国信息化建设进程和新形势,出台了《国家信息化发展战略纲要》,它是规范和指导未来 10 年中国信息化发展的纲领性文件,是国家战略体系的重要组成部分,是信息化领域规划、政策制定的重要依据。2016 年 10 月,中共中央政治局就实施网络强国战略进行集体学习,习近平总书记提出要建设全国一体化的大数据中心,加快落实国家大数据战略。2016 年 12 月,国务院发布了《“十三五”国家信息化规划》,明确提出要建立统一开放的大数据体系,加强数据资源规划建设,建立国家互联网大数据平台和国家治理大数据中心,深化大数据应用,强化数据资源管理,注重数据安全保护。2017 年 10 月,习近平总书记在党的十九大报告中指出,要推动大数据和实体经济深度融合,建设网

络强国、数字中国和智慧社会。同年12月,习近平在中共中央政治局对国家大数据战略实施进行第二次集体学习时指出,“大数据发展日新月异,我们应审时度势、精心谋划、超前布局、力争主动”,并深入分析大数据发展的现状和趋势,强调根据中国实际部署实施国家大数据战略,加快建设数字中国,让大数据在各项工作中发挥更大作用,服务社会与民众。2018年1月,为进一步加强和规范科学数据管理,保障科学数据安全,提高开放共享水平,更好地支撑国家科技创新、经济社会发展和国家安全保障,国务院办公厅印发《科学数据管理办法》。2020年5月,《中共中央 国务院关于新时代加快完善社会主义市场经济体制的意见》发布实施,进一步提出加快培育发展数据要素市场。

表 1.1 我国大数据政策部分汇总

时间	出处	文件名称/讲话
2015.08	国务院	《促进大数据发展行动纲要》
2015.10	中共中央	《中国共产党第十八届中央委员会第五次全体会议公报》
2016.01	发展和 改革委员会	《国家发展改革委办公厅关于组织实施促进大数据发展重大工程的通知》
2016.03	国务院	《中华人民共和国国民经济和社会发展第十三个五年规划纲要》
2016.07	国务院	《国家信息化发展战略纲要》
2016.09	国务院	《政务信息资源共享管理暂行办法》
2016.10	习近平	“建设全国一体化的国家大数据中心”
2016.12	国务院	《“十三五”国家信息化规划》
2017.05	国务院	《政务信息系统整合共享实施方案》
2017.05	习近平	“一带一路”国际合作高峰论坛上的讲话
2017.10	中共中央/ 习近平	《决胜全面建成小康社会 夺取新时代中国特色社会主义伟大胜利》
2017.12	习近平	“实施国家大数据战略,加快建设数字中国”
2018.01	国务院	《科学数据管理办法》
2020.04	国务院	《关于构建更加完善的要素市场化配置体制机制的意见》
2020.05	国务院	《中共中央 国务院关于新时代加快完善社会主义市场经济体制的意见》

② 各行业的大数据战略

围绕宏观大数据国家战略,各部委和相关行业也出台了一系列政策来推动大数据的应用发展,具体政策统计见表1.2,包含医疗、政务信息资源、农业、经济、生态环境、国土资源、林业、交通运输、水利、气象、教育、交通旅游服务等方面。

从总体来看,相关部门及行业从多方位响应了国家大数据战略,尤其是关系国民经济建设与发展的重点行业领域,且不断深化战略部署,引导大数据应用向专业化、个性化纵深发展。不同于国家的宏观战略,各部委根据各自的职能分工进行相应部署,发挥“互联网+”的应用潜力,挖掘信息资源价值,推动大数据与各行业领域的深度融合。

表 1.2 部分行业领域大数据政策汇总

时间	出处	文件名称
2015.12	农业农村部	《农业部推进农业农村大数据发展的实施意见》
2016.01	发展和改革委员会	《国家发展改革委办公厅关于组织实施促进大数据发展重大工程的通知》
2016.02	国务院	《中医药发展战略规划纲要(2016—2030年)》
2016.03	生态环境部	《生态环境大数据建设总体方案》
2016.06	国务院	《国务院办公厅关于促进和规范健康医疗大数据应用发展的指导意见》
2016.07	自然资源部	《关于促进国土资源大数据应用发展的实施意见》
2016.07	国家林业和草原局	《关于加快中国林业大数据发展的指导意见》
2016.07	中国煤炭工业协会、 中国煤炭运销协会	《推进煤炭大数据发展的指导意见》
2016.08	交通运输部	《关于推进交通运输行业数据资源开放共享的实施意见》
2016.08	发展和改革委员会	《国家发展改革委办公厅关于请组织申报大数据领域创新能力建设专项的通知》
2016.10	农业农村部	《农业农村大数据试点方案》
2016.10	国务院	《“健康中国2030”规划纲要》
2016.11	国家中医药管理局	《中医药信息化发展“十三五”规划》
2017.01	工业和信息化部	《大数据产业发展规划(2016—2020年)》
2017.01	卫计委	《“十三五”全国人口健康信息化发展规划》
2017.05	水利部	《关于推进水利大数据发展的指导意见》
2017.09	气象局	《气象大数据行动计划(2017—2020年)》
2017.09	公安部	《关于深入开展“大数据+网上督察”工作的意见》
2018.01	教育部	《教育部机关及直属事业单位教育数据管理办法》
2018.03	交通运输部/ 文化和旅游部	《关于加快推进交通旅游服务大数据应用试点工作的通知》
2018.04	国务院	《国务院办公厅关于促进“互联网+医疗健康”发展的意见》
2019.02	工业和信息化部/ 国家机关管理局/ 国家能源局	《关于加强绿色数据中心建设的指导意见》
2019.07	工业和信息化部	《电信和互联网行业提升网络数据安全保护能力专项行动方案》
2020.02	工业和信息化部	《工业数据分类分级指南(试行)》
2020.04	工业和信息化部	《关于公布支撑疫情防控和复工复产复课大数据产品和解决方案》
2020.05	工业和信息化部	《关于工业大数据发展的指导意见》

(2) 地方层次的大数据战略

国家大数据战略的实施离不开地方政府的协同配合与贯彻落实。作为大数据战略的前沿阵地,很多地方政府很早就预见到大数据在未来经济社会领域的发展前景与重要意

义,并进行了积极探索。2012年,广东省政府办公厅率先发布《广东省实施大数据战略工作方案》,成功试水大数据战略。2012年,上海市政府数据开放平台启动,上海市科学技术委员会颁布了《上海推进大数据研究与发展三年行动计划(2013—2015年)》。2013年6月,重庆发布了《重庆市大数据产业发展规划》。此外,北京、天津、武汉等地也出台了相关文件。地方政府的探索无疑加快了中央政府部署大数据战略的步伐,2015年,国务院印发的《促进大数据发展行动纲要》从国家大数据发展战略全局的高度,提出了我国大数据发展的顶层设计和统筹布局,支持地方开展大数据产业发展和应用试点,谋划当地的大数据发展规划。之后,各级地方政府更是积极响应,各地大数据发展政策密集出台,地方大数据管理机构也陆续成立。

与国家宏观规划不同,地方政府的战略充分体现的是该地区的资源优势与发展特色。例如,2020年广东省工业和工信厅发布的《广东省5G基站和数据中心总体布局规划(2021—2025年)》,提出加快以5G基站和数据中心为重点的新一代信息基础设施建设,加强顶层设计和规划引导,加快新型基础设施与数字产业融合发展,为广东省经济社会高质量发展提供有力支撑。从地区分布看,各地区均有大数据战略相关文件发布,其中发布文件最多的是贵州省,这一方面在于国家的扶持政策,如2015年《促进大数据发展行动纲要》把发展贵州大数据产业正式上升为国家战略,支持贵州大数据产业集聚区创建工作;另一方面在于贵州省自身的前瞻性、战略性探索。

总体来说,目前地方大数据规划主要有两类:一类是与地区发展现状紧密结合的引领性规划,如各省市制定的大数据发展意见、发展规划,这类规划从政策机制上对大数据进行战略部署,提出发展目标和行动方向;二是具备可操作性的具体措施落实型规划,这类规划往往提出某些具体领域的实施方案或在专项工程中的具体部署,而在具体部署中,则重点关注了信息基础设施建设、人才培养、法规标准制定等方面。值得一提的是,贵州省作为抓住发展大数据机遇、实现“弯道超车”的典型,走出了一条不同于东西部省份的发展新路。无论是《贵州省大数据产业发展应用规划纲要(2014—2020年)》《贵州省信息基础设施条例》《关于加快大数据产业人才队伍建设的实施意见》等发展政策的制订,还是中国首部大数据地方法规《贵州省大数据发展应用促进条例》、首部政府数据共享开放地方法规《贵阳市政府数据共享开放条例》、首部大数据安全管理地方性法规《贵阳市大数据安全管理条例》等法规的出台施行,都印证了贵州省在大数据发展及创新应用方面取得显著成效,展示了其在探索战略行动路径中积累了丰富经验,也为其他地区提供了良好借鉴。

(3) 企业层次的大数据战略

企业利用大数据可以更敏锐地感知市场变化,洞察客户、消费者以及合作伙伴与竞争对手的行为;更精准地优化企业运营,和商业伙伴一起开展协同创新。企业要顺应大数据时代,利用大数据和大数据分析来增强洞察力,释放企业潜力,实现转型和发展。对此,需要加强对大数据的持续投入和研发,注重合作和行业应用,制定与之契合的大数据战略。

根据2016年贝恩公司对大数据行业的调研可知,对于拥有优秀大数据能力的企业,其财务业绩排在行业前25位的可能性是没有该能力的企业的2倍;做出正确决策的可能性是没有该能力的企业的3倍;决策速度是没有该能力的企业的5倍。大数据对于企业