

图书在版编目 (CIP) 数据

大数据导论 / 刘春燕, 司晓梅主编. — 武汉: 华中科技大学出版社, 2022.5

ISBN 978-7-5680-5767-7

I . ①大… II . ①刘… ②司… III . ①数据处理—高等学校—教材 IV . ① TP274

中国版本图书馆 CIP 数据核字 (2022) 第 077335 号

大数据导论

Dashuju Daolun

刘春燕 司晓梅 主编

策划编辑: 康 序

责任编辑: 刘姝甜

封面设计: 抱 子

责任监印: 朱 玟

出版发行: 华中科技大学出版社 (中国 · 武汉)

电话: (027)81321913

武汉市东湖新技术开发区华工科技园

邮编: 430223

录 排: 武汉创易图文工作室

印 刷: 武汉开心印印刷有限公司

开 本: 787 mm × 1092 mm 1/16

印 张: 15.5

字 数: 384 千字

版 次: 2022 年 5 月第 1 版第 1 次印刷

定 价: 58.00 元



本书若有印装质量问题, 请向出版社营销中心调换
全国免费服务热线: 400-6679-118 竭诚为您服务
版权所有 侵权必究

前言

PREFACE

随着移动互联网、物联网、云计算等新一代信息技术的应用和推广，大数据技术成为又一颠覆性的技术，备受人们关注。身处大数据时代，我们已经感受到大数据对人们的思维模式和生活方式的改变，大数据对人类的社会生产和生活必将产生重大而深远的影响。

本书定位为大数据技术入门教材，以大数据的基本技术路线为框架，通过基本理论和应用实例相结合的方式，介绍大数据技术，帮助读者形成对大数据知识体系及其应用领域的轮廓性认识，为读者在大数据领域的继续深造奠定基础。

本书旨在服务大数据初学者，为适应初学者学习特点，适当增加了广度而降低了深度，在数据挖掘部分尽可能少地使用数学知识，对于一些不可避免的部分，力求展现其中的精华，而在大数据实验部分，必须掌握的基础性编程语言也有涉及。本书主要以 Java 语言为基础。

本书第 1 章为大数据概述，介绍大数据的基本概念和应用领域，回顾大数据理念和技术的发展历程，阐述大数据的发展前景。

第 2~7 章介绍大数据采集与预处理、大数据计算平台、大数据管理、数据挖掘、大数据隐私与安全和人工智能 6 个大数据领域及其中的主要技术。第 2 章主要介绍大数据采集技术，包括大数据的来源、采集方法及数据预处理方法等，最后对大数据采集应用案例进行分析，帮助读者更好地理解大数据采集技术。第 3 章介绍大数据处理架构 Hadoop 以及与大数据技术密不可分的云计算技术及其应用。由于 Hadoop 已经成为应用最广泛的大数据技术，本书的大数据相关技术主要围绕 Hadoop 展开，包括 HDFS 和 MapReduce。第 4 章介绍大数据管理，包括分布式数据库（HBase）、常用的 NoSQL 数据库和云数据库。第 5 章从数据挖掘的概念入手，介绍数据挖掘的几种算法以及算法的应用。第 6 章提出大数据面临的安全隐患，介绍大数据安全的基本概念以及大数据安全与隐私保护的主要方法。第 7 章介绍人工智能的起源和基本概念，通过人工智能在生活中的应用案例引发读者对智能时代的思考。

第 8~11 章包含 4 个实验，对应数据采集技术、云计算技术和数据挖掘技术展开。

本书在重视理论的前提下，不忽视实际的可操作性，注重问题的解决，“大数据基础”与“大数据技术”部分每章均设有习题，以帮助读者巩固所学知识。

本书由武汉华夏理工学院刘春燕和司晓梅主编。在本书编写的过程中，编者参考了国内外大量大数据及云计算技术的文献资料，且书中部分案例来自网络，在此一并对相关作者表示感谢。

由于编者能力有限，书中难免存在不妥之处，恳请读者朋友提出宝贵意见，不胜感激。

为了方便教学，本书还配有电子课件等资料，可以登录“我们爱读书”网（www.ibook4us.com）浏览，任课教师可以发邮件至 hustpeiit@163.com 索取。

目录

CONTENTS

第1章 大数据概述 /001

- 1.1 大数据兴起之谜 /001
- 1.2 无处不在的大数据 /004
- 1.3 大数据的概念和特征 /013
- 1.4 大数据的关键技术 /016
- 1.5 大数据与云计算、物联网 /022
- 1.6 大数据的发展、应用及挑战 /027

第2章 大数据采集与预处理 /032

- 2.1 大数据的来源 /032
- 2.2 大数据的采集方法 /034
- 2.3 数据预处理 /044
- 2.4 大数据采集应用案例——互联网行业职场分析 /052

第3章 大数据计算平台 /058

- 3.1 Hadoop 平台 /058
- 3.2 HDFS /062
- 3.3 大数据计算模式 /067
- 3.4 MapReduce /069
- 3.5 Spark 平台 /078
- 3.6 流计算框架 Storm /087
- 3.7 云计算平台 /091
- 3.8 云计算的关键技术 /096

第4章 大数据管理——大数据的高效之道 /101

- 4.1 大数据管理之 NoSQL 数据库 /101
- 4.2 NoSQL 与关系型数据库的比较 /104
- 4.3 NoSQL 的四大类型 /107
- 4.4 NoSQL 的三大基石 /111
- 4.5 新兴数据库技术 /114
- 4.6 大数据应用案例——在“北上广”打拼是怎样一种体验 /120

第5章 数据挖掘——大数据的智慧之道 /126

- 5.1 数据挖掘概述 /126
- 5.2 大数据挖掘技术 /128
- 5.3 分类和预测 /129
- 5.4 聚类分析 /139
- 5.5 关联规则分析 /146
- 5.6 商业智能的分析预测 /154
- 5.7 社交大数据的成功密码 /158
- 5.8 大数据应用案例——大数据预测 /163

第6章 大数据隐私与安全 /168

- 6.1 安全与隐私问题 /168
- 6.2 大数据面临的问题 /169
- 6.3 大数据的安全防护策略 /171
- 6.4 如何解决隐私保护问题 /171

- 6.5 大数据应用案例——智慧城市中的安全防护 /175

第7章 人工智能——科幻到现实的蜕变 /182

- 7.1 人工智能的起源 /182
- 7.2 当人工智能遇上大数据 /190
- 7.3 人机大战：AI会挑战人类吗？ /192
- 7.4 AI会取代人类吗？ /194
- 7.5 AI时代的教育与个人发展 /199
- 7.6 大数据应用案例——神秘AI的魅力 /203

第8章 数据采集实验 /206

第9章 CloudSim 虚拟平台实验 /212

第10章 数据挖掘算法之 Apriori 算法实验 /215

第11章 数据挖掘算法之决策树算法实验 /216

附录 A /218

附录 B /219

附录 C /229

参考文献 /239

第 1 章 大数据概述

大数据作为继云计算、物联网之后 IT 领域又一种颠覆性技术，备受人们的关注。大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，对人类的社会生产和生活必将产生重大而深远的影响。

大数据时代悄然来临，带来了信息技术发展的巨大变革，并深刻影响着社会生产和人们生活的方方面面。世界各国均高度重视大数据技术的研究和发展，纷纷把大数据上升为国家战略加以重点推进。

本章主要概略介绍大数据的兴起、生活中的大数据、大数据的概念及特征，以及大数据的关键技术和大数据的发展。

1.1 大数据兴起之谜

◆ 1.1.1 大数据产生的背景

早在远古时代人们就已经在石头、树木上记载相应的数据了，再到后来，人们用竹筒、布帛等记载和传输数据，在这一阶段，数据的记录和传播都是非常有限的；到后来纸张出现，印刷术被发明，数据的记录和传播有了第一次长足的进步，但是此时的数据量仍旧相当小，传播速度也较为缓慢，传播范围相对狭窄，人们对数据的分析和使用十分有限；计算机和磁盘等存储介质出现后，人们记录数据和计算分析数据的能力有了质的飞跃，随着以博客、社交网络、基于位置服务为代表的新型信息发布方式的不断涌现，以及云计算、物联网等技术的兴起，数据以前所未有的速度在不断地增长和积累，至此，人们进入所谓的大数据时代。

大数据浪潮汹涌来袭，与互联网的发明一样，这绝不仅仅是信息技术领域的革命，更是在全球范围内启动透明政府建设、加速企业创新、引领社会变革的利器。现代管理学之父德鲁克曾经说过：“预测未来最好的方法，就是去创造未来。”“大数据战略”，则是当下领航全球的先机。

越来越多的政府、企业等机构开始意识到数据正在成为组织最重要的资产，数据分析能力正在成为组织的核心竞争力。大数据时代对政府管理转型来说是一个历史性机遇，对企业来说，海量数据的运用将成为未来竞争和增长的基础。同时，大数据已引起学术界的广泛研究兴趣。

◆ 1.1.2 大数据的发展历程

大数据不是凭空产生的，它有自己的发展过程。大数据的发展大致分为三个阶段，如图 1.1 所示。

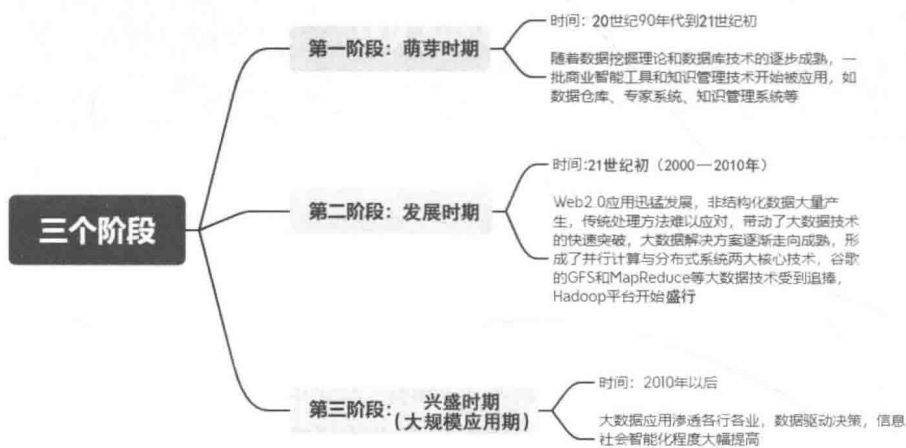


图 1.1 大数据发展的三个阶段

1. 萌芽时期（20 世纪 90 年代至 21 世纪初）

1997 年，美国国家航空航天局艾姆斯研究中心的大卫·埃尔斯沃思和迈克尔·考克斯在他们研究数据可视化时首次使用了“大数据”的概念。1998 年，《Science》杂志发表了一篇题为“大数据科学的可视化”的文章，“大数据”作为一个专业名词正式出现在公共期刊上。

在这一阶段，大数据只作为一个概念或假设，少数学者对其进行了研究和讨论，其意义仅限于表示数据量的巨大，对数据的收集、处理和存储没有进一步的探索。

2. 发展时期（2000—2010 年）

21 世纪刚开始的 10 年，互联网行业得到了快速发展。2001 年，美国 Gartner 公司率先开发了大型数据模型。同年，Doug Laney 提出了大数据的“3V”特性。2005 年，Hadoop 技术应运而生，成为数据分析的主要技术。2007 年，数据密集型科学出现，不仅为科学界提供了一种新的研究范式，而且为大数据的发展提供了科学依据。2008 年，《Science》杂志推出了一系列大数据专刊，详细讨论了一系列大数据的问题。2010 年，美国信息技术顾问委员会发布了一份题为“规划数字化未来”的报告，详细描述了政府工作中大数据的收集和使用。

在这一阶段，“大数据”作为一个新名词，开始受到理论界的关注，其概念和特点得到进一步丰富，相关的数据处理技术层出不穷，大数据开始显现出活力。

3. 兴盛时期（2011 年至今）

2011 年，IBM 公司开发了沃森超级计算机，通过每秒扫描和分析 4 TB 数据打破了世界纪录，大数据计算达到了一个新的高度。随后，MGI 发布了大数据前沿报告，详细介绍了大数据在各个领域的应用，以及大数据的技术框架。2012 年在瑞士举行的世界经济论

坛上学者们讨论了一系列与大数据有关的问题，发表了题为“大数据，大影响”的报告，并正式宣布了大数据时代的到来。

2011年之后大数据的发展可以说进入了全面兴盛的时期，越来越多的学者对大数据的研究从基本的概念、特性转到数据资产、思维变革等多个角度。大数据也渗透到各行各业之中，不断变革原有行业的技术，创造出新的技术，大数据的发展呈现出一片蓬勃之势。

2011年6月，麦肯锡咨询公司发布了《大数据：下一个创新、竞争和生产力的前沿领域》研究报告。研究报告中指出，数据正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。各行各业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，大数据时代已经降临。

2012年3月22日，美国政府宣布投资2亿美元发起大数据研究和发展倡议，致力于提高从大型、复杂数据中集中提取信息和知识的能力，并服务能源、健康、金融和信息技术等领域的高科技企业。

2012年4月，英国、美国、德国、芬兰和澳大利亚研究者联合推出“世界大数据周”活动，旨在促使政府制定战略性的大数据措施。

2012年5月，联合国发布了《大数据促发展：挑战与机遇》白皮书，指出大数据对于联合国和各国政府来说是一个历史性的机遇，人们如今可以使用极为丰富的数据资源来对社会经济进行前所未有的实时分析，帮助政府更好地响应社会和经济运行。

2012年7月，为挖掘大数据的价值，阿里巴巴集团在管理层设立“首席数据官”一职，“首席数据官”负责全面推进“数据分享平台”战略，并推出大型的数据分享平台——“聚石塔”，为天猫、淘宝平台上的电商及电商服务商等提供数据云服务。

2012年12月，维克托·迈尔·舍恩伯格的《大数据时代》开始在国内风靡，推动了国内大数据的发展。

2013年被称为大数据元年。百度、阿里巴巴、腾讯各显身手，分别推出创新型大数据应用。同年12月，中国计算机学会发布《中国大数据技术与产业发展白皮书》，系统总结了大数据的核心科学与技术问题，推动了我国大数据学科的建设与发展，并为政府部门提供了战略性的意见与建议。

2014年4月，世界经济论坛以“大数据的回报与风险”为主题发布了《全球信息技术报告（第13版）》。报告认为，在未来几年，针对各种信息通信技术的政策会十分重要，接下来将对数据保密和网络管制等议题展开积极讨论。全球大数据产业的日趋活跃，技术演进和应用创新的加速发展，使各国政府逐渐认识到大数据在推动经济发展、改善公共服务、增进人民福祉乃至保障国家安全方面具有重大意义。

2015年8月，国务院印发《促进大数据发展行动纲要》，全面推进我国大数据发展和应用，加快建设数据强国。

2016年4月，在“2016大数据产业峰会”上工信部透露，我国将制定出台大数据产业“十三五”发展规划，有力推进我国大数据技术创新和产业发展。

2017年工业和信息化部印发大数据产业“十三五”发展规划。2017年5月，全国首个总体解决方案——《政府数据共享开发（贵阳）总体解决方案》通过评审，全国首部政

府数据共享开放地方性法规诞生。

2018年1月,国家发改委宣布了政务信息系统整合共享工作最新进展,已有71个部门、31个地方实现了国家共享交换平台的对接。

2019年5月,《2018全球大数据发展分析报告》显示,中国大数据技术创新能力有了显著的提升。

2019年9月,大数据产业生态联盟联合赛迪顾问公司发布《2019中国大数据产业发展白皮书》,指出2018年中国大数据产业规模为4384.5亿元,预计2021年将达8070.6亿元。



2020年至今,伴随着国家部委有关大数据行业应用政策的出台,国内的金融、政务、电信、物流等行业中大数据技术应用的价值不断凸显。同时,随着我国大力发展数字经济,推进数字中国建设,大数据产业将迎来高速发展期。

1.2 无处不在的大数据

从文明之初的“结绳记事”,到文字发明后的“文以载道”,再到近现代科学的“数据建模”,数据一直伴随着人类社会的发展变迁,承载了人类基于数据和信息认识世界的努力和取得的巨大进步。然而,直到以电子计算机为代表的现代信息技术出现,为数据处理提供了自动化方法和手段,人类掌握数据、处理数据的能力才实现了质的跃升。信息技术及其在经济社会发展的方方面面的应用(即信息化),推动数据(信息)成为继物质、能源之后的又一种重要战略资源。

我们生活在一个充满数据的时代,我们打电话、用微博、聊QQ、刷微信,我们阅读、购物、看病、旅游,都在不断产生新数据,“堆砌”着“数据大厦”。大数据已经与我们的工作生活息息相关、须臾难离。中国工程院院士高文说:“不管你是否认同,大数据时代已经来临,并将深刻地改变着我们的工作和生活。”

大数据就在你我身边,虽然我们看不到它、不在意它,但它却已经并将继续影响我们的生活。习近平主席及其夫人彭丽媛访问英国的帝国理工学院时,校方赠送给彭丽媛的羊绒披肩就有大数据的功劳——披肩尺寸是利用计算机图像分析技术得出的。校方还向习主席展示了如何运用大数据的方法分析国内人口迁移情况、“一带一路”倡议的国际影响力、个性化医疗的推广以及上海地铁的负载分布和应急办法等。

◆ 1.2.1 教育中的大数据

大数据将给教育带来革命性的变化。

在传统教育模式下,分数就是一切,一个班上的几十个人,使用同样的教材,同一个教师上课,课后布置同样的作业。学生是千差万别的,在这种模式下,不可能真正做到因材施教。在大数据的帮助下,个性化教育将真正实现,大数据将显示学生如何学、教师如何教,通过个性化的大数据积累及数据分析的结果给予反馈,并调整教学思路方法,实现因材施教。

又如“大数据情书”,即学校写给毕业生的“情书”,记录了毕业生在学校的点点滴滴,

这就是华中科技大学利用大数据制作的个性化毕业礼物——一封名叫“光阴的故事——致××”的电子信件和截图（见图1.2）在华中科技大学毕业生的微信、朋友圈流传。每一位即将离校的学子只要打开链接，输入自己的校园账号就能获取在校期间的学习、生活等方面数据和收获。



图 1.2 华中科技大学利用大数据制作的毕业礼物截图

“65 门必修，7 门公选，70 位老师……”“四年里，你偏爱集锦园食堂，消费金额位居榜首……”看到这些在华中科技大学留下的痕迹，毕业生纷纷表示感动和温暖。

此套针对毕业生个性化服务的大数据系统，由华中科技大学率先推出，并已在全国高校成功上线，作为送给 2016 届毕业生的一份特别礼物，为学子勾勒出在校期间生活的立体图景。

这些数据都是学生日常登录 HUB 系统或者使用校园卡时留下的，临近学生毕业，HUB 系统将数据整合，以毕业礼物的形式献给毕业生，从学生入学开始，到毕业离校，用数据描述了他们在校期间的经历，包括生源地、在校班级、是否转专业、学期注册、所修课程、授课教师、加权平均成绩、四六级英语成绩、奖助学金、科研成果、荣誉称号、访问华中科技大学教学信息服务平台、计算机等级考试成绩、校园卡消费、食堂、校内超市、校车、图书馆借书和门禁、党员发展历程等情况，用数据和场景故事逐页展示出来。该系统还根据学生注册积极程度，对部分学生授予了“注册神人”“注册牛人”“注册达人”的称号；根据成绩在专业的排名，对部分学生授予了“学圣”“学神”“学霸”的称号；根据在图书馆的借书数量排名，对部分学生授予了“读书达人”称号。

厦大图书馆设计的一个名为“囿·时光 (Tuan Time)”的网站，收集整理了毕业生大学时代的阅读记录、进馆次数等，被毕业生视为大学生涯的图书馆记忆。登录“囿·时光 (Tuan Time)”网页后，学生可以看到自己大学期间的图书馆印记，包括第一次到访图

书馆的时间，借阅的第一本书，最喜欢的座位，最常阅读的图书类别，以及一份书单。

南京理工大学的“大数据精准扶贫”，利用大数据分析为贫困生充饭卡。每个月在食堂吃饭超过 60 顿、一个月总消费不足 420 元 的学生，被列为受资助对象。学生无须填表申请，不用审核。这种“润物细无声”的善举——给贫困生的伙食补贴通过直接打入饭卡的方式进行，在确保学生尊严的基础上，给贫困学生带来了温暖。

电子科技大学曾有“寻找校园中最孤独的人”课题，从 3 万名在校生中，采集到了 2 亿多条行为数据，数据来自学生选课记录，进出图书馆、寝室，以及食堂用餐、超市购物等。通过对不同的校园一卡通前后刷卡的记录进行分析，可以发现一个学生在学校有多少亲密朋友，比如恋人、闺蜜。最后，通过这个课题找到了 800 多个校园中最孤独的人，他们平均在校两年半时间，一个知心朋友都没有。这些人中的 17% 可能产生了心理疾病，也可能用意志力暂时战胜了症状，但需要学校和家长重点予以关爱。

◆ 1.2.2 食品中的大数据

大家每天吃几餐？每天吃了多少米饭？吃了多少肉？相信这些问题一定难不倒大家。如果再问大家每天摄入了多少蛋白质、吸收了多少碳水化合物呢？大家想要清楚地回答这些问题可就不那么容易了。但是，有了大数据，这些问题都将不再是问题。我们每天吃了什么，吃了多少，该吃多少，这些大数据都会帮我们记录。

某穿戴设备用光谱扫描仪来扫描检测食物中的成分（通过对光子的波长进行排序，结合频谱来描述食物里面的成分），这样你将知道“吃了什么”；服务器将食物中的过敏原、化学成分和营养成分等相关信息发送到你的手机上，这样就告诉你“吃了多少”“该吃多少”，营养成分是否足够；它还能记录你的饮食信息，从而提示你是否已经达到了每天所需的摄取量，并根据具体情况给出合适的营养食谱。比如，你最近在减肥，不能吃太多，那就给你提供一份减肥食谱；如果你肠胃炎，只能清淡饮食，就会给你提供一份清粥小菜。这些智能的穿戴设备可能会成为我们私人定制的“营养师”，为我们提供服务，通过数据的采集及分析处理，告诉我们该吃什么，不该吃什么，该吃多少等。

如何吃得放心？你知道自己每天吃的肉是从哪里来的吗？大数据已经应用到食品安全领域的每个角落。建立食品追溯系统需要物联网技术的运用和普及，以期实现对食品生产、加工、运输、包装、储存等方面质量问题的监管，理论上实现对食品“从农田到餐桌”的全面监控。同时，利用大数据给食品安全分析过程中的风险评估、风险管理和风险交流提供相应的变化和动力，可以大数据工程为抓手，深入开展食品安全信息化建设，为我国食品安全长效机制的建立提供有效的工程技术保障。上海动物及动物产品产业链全程监管系统，对 17.5 万头能繁殖的母猪植入 RFID 式电子耳标，从猪刚出生一直到我们的餐桌，它都可以利用物联网技术进行跟踪，这样就可以告诉我们，今天餐桌上的这盘肉来自哪里，中间养了多长时间，有没有喂饲料，等等，通过智能识别和无线传感等技术，实现动物健康养殖、动物产品安全经营和质量安全溯源管理。动物及动物产品检疫监督管理系统示意图如图 1.3 所示。

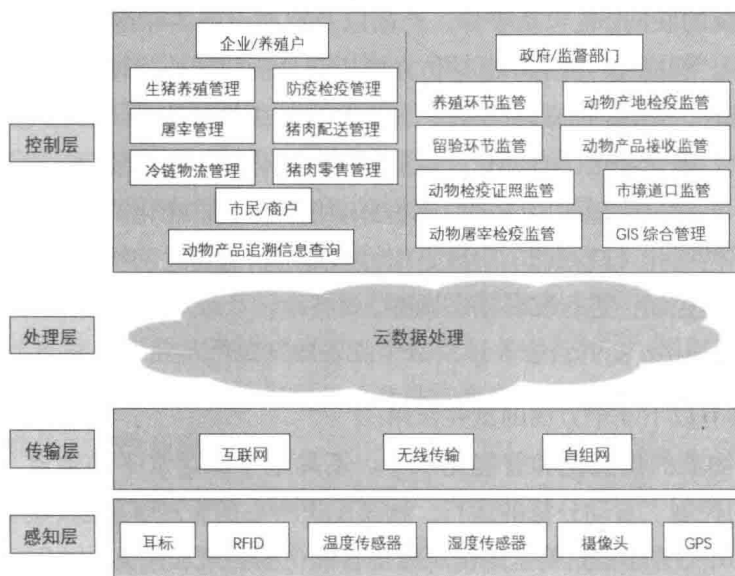


图 1.3 动物及动物产品检疫监督管理系统示意图

以一个人“吃了什么”“吃了多少”产生的数据来进行分析,然后告诉这个人“该吃多少”以及“如何放心地吃”已经是非常大的数据处理量了。然而,众口难调,习惯各异,涉及食物的做法、吃法、成分、营养价值、价格、食物来源等的数据看起来相当混乱(传统的数据库不能完全明确),如果要满足一个 1000 万人的“该吃多少”以及“如何放心地吃”的解答需求,就要用到大数据技术,通过收集、管理和分析每个人看似随意平常的饮食信息,就可以为每个人定制出满足其需求的个人食谱。

◆ 1.2.3 物流中的大数据

智能物流是大数据在物流领域的典型应用。智能物流融合了大数据、物联网和云计算等新兴 IT 技术,该物流系统能模仿人的智能,实现物流资源优化调度和有效配置以及物流系统效率提升。自从 IBM 在 2010 年最先提出“智能物流”概念以来,智能物流在全球范围内得到了快速发展。在我国,阿里巴巴集团联合多方力量共建“中国智能物流骨干网”,计划在 8~10 年的时间内建立一张能支撑日均 300 亿元(年度约 10 万亿元)网络零售额的智能物流骨干网络,支持数千万家新型企业成长发展,让全中国任何一个地区都能做到 24 h 内送货必达。大数据技术是智能物流发挥其重要作用的基础和核心,物流行业在货物流转、车辆追踪、仓储等各个环节中都会产生海量的数据,分析这些物流大数据,将有助于深刻认识物流活动背后隐藏的规律,优化物流过程,提升物流效率。

智能物流,又称智慧物流,是利用智能化技术,使物流系统能模仿人的智能,具有思维、感知、学习、推理判断和自行解决物流中某些问题的能力,从而实现物流资源优化调度和有效配置、物流系统效率提升的现代化物流管理模式。

智慧供应链具有先进化、互联化、智能化三大特点。“先进化”是指:数据多由感应设备、识别设备、定位设备产生,替代人为获取供应链动态并进行可视化自动管理,包括自动库存检查、自动报告存货位置错误等。“互联化”是指:整体供应链联网,不仅是客户、

供应商、IT系统的联网,也包括零件、产品以及智能设备的联网。联网赋予供应链整体计划决策能力。“智能化”是指:通过仿真模拟和分析,帮助管理者评估多种可能性选择的风险和约束条件,使供应链具有学习、预测和自动决策的能力,无须人为介入。

“智能物流”概念经历了自动化、信息化、网络化3个发展阶段。自动化阶段是指物流环节自动化,即物流管理按照既定的流程自动化操作;信息化阶段是指现场信息自动获取与判断选择;网络化(泛在化)阶段是指将采集的信息通过网络传输到数据中心,由数据中心做出判断与控制,进行实时动态调整。

1. 智能物流的作用

智能物流具有以下3个方面的重要作用。

(1) 提高物流的信息化和智能化水平。不局限于库存水平的确定、运输道路的选择、自动跟踪的控制、自动分拣的运行、物流配送中心的管理等问题,物品的信息也将存储在特定数据库中,并能根据特定的情况做出智能化的决策和建议。

(2) 降低物流成本,提高物流效率。由于交通运输、仓储设施、信息通信、货物包装和搬运等对信息交互和共享的要求较高,可以利用物联网技术对物流车辆进行集中调度,有效提高运输效率;利用超高频RFID标签读写器实现仓储进库管理,快速识别货物的进库情况;利用RFID标签读写器建立智能物流分拣系统,有效地提高生产效率并保证系统的可靠性。

(3) 提高物流活动的一体化。通过整合物联网相关技术,集成分布式仓储管理及流通渠道建设,可以实现物流中运输、存储、包装、装卸等环节全流程一体化管理模式,以高效地向客户提供满意的物流服务。

2. 智能物流的应用

智能物流有着广泛的应用。国内许多城市都在围绕智慧港口、多式联运、冷链物流、城市配送等方面,着力推进物联网在大型物流企业、大型物流园区的系统级应用。应用智能物流,可以将射频标签识别技术、定位技术、自动化技术以及相关的软件信息技术集成到生产及物流信息系统领域,探索和用物联网技术实现物流环节的全流程管理模式,开发面向物流行业的公共信息服务平台,优化物流系统的配送中心网络布局,集成分布式仓储管理及流通渠道建设,最大限度地减少物流环节、简化物流过程,提高物流系统的快速反应能力;此外,还可以进行跨领域信息资源整合,建设基于卫星定位、视频监控、数据分析等技术的大型综合性公共物流服务平台,发展供应链物流管理。

3. 大数据是智能物流的关键

在物流领域有两个著名的理论——“黑大陆”说和物流冰山说。著名的管理学权威P. F. 德鲁克提出了“黑大陆”说,认为在流通领域中物流活动的模糊性尤其突出,是流通领域中最具潜力的领域。提出物流冰山说的日本早稻田大学教授西泽修认为,物流就像一座冰山,其中沉在水面以下的是我们看不到的黑色区域,这部分就是“黑大陆”,而这正是物流尚待开发的领域,也是物流的潜力所在。这两个理论都旨在说明物流活动的模糊性和巨大潜力。对于如此模糊而又具有巨大潜力的领域,我们该如何去了解、掌控和开发呢?

答案就是借助于大数据技术。

发现隐藏在海量数据背后的有价值的信息,是大数据的重要商业价值。大数据是打开物流领域这块神秘的“黑大陆”的一把金钥匙。物流行业在货物流转、车辆追踪、仓储等各个环节中都会产生海量的数据,有了这些物流大数据,所谓的物流“黑大陆”将不复存在,我们可以通过数据充分了解物流运作背后的规律,借助于大数据技术,可以对各个物流环节的数据进行归纳、分类、整合、分析和提炼,为企业战略规划、运营管理和日常运作提供重要支持和指导,从而有效提升快递物流行业的整体服务水平。

大数据将推动物流行业从粗放式服务到个性化服务进行转变,颠覆整个物流行业的商业模式。通过对物流企业内部和外部相关信息进行收集、整理和分析,可以做到为每个客户量身定制个性化的产品和服务。

4. 中国智能物流骨干网——“菜鸟”

2013年5月28日,阿里巴巴集团、银泰集团联合复星集团、富春控股、顺丰集团、“三通一达”(申通、圆通、中通、韵达)、宅急送、汇通以及相关金融机构共同宣布,开始联手共建“中国智能物流骨干网”(China Smart Logistic Network, CSN),又名“菜鸟”。

“菜鸟”网络由物流仓储平台和物流信息系统构成。物流仓储平台将由8个左右大仓储节点、若干个重要节点和更多城市节点组成。大仓储节点将针对东北、华北、华东、华南、华中、西南和西北7个区域,选择中心位置进行仓储投资。物流信息系统整合了所有服务商的信息系统,实现了骨干网内部的信息统一,同时该系统将向所有的制造商、网商、快递公司、第三方物流公司完全开放,有利于物流生态系统内各参与方利用信息系统开展各种业务。

“菜鸟”是阿里巴巴创始人马云整合各方力量实施的“天网+地网”计划的重要组成部分。所谓“地网”,是指“中国智能物流骨干网”,最终将建设成为一个全国性的超级物流网,这个网络能在24h内将货物运抵国内任何地区,能支撑日均300亿元(年度约10万亿元)的巨量网络零售额。所谓“天网”,是指以阿里巴巴集团旗下多个电商平台(淘宝、天猫等)为核心的大数据平台。由于阿里巴巴集团的电商业务在中国占据绝对优势地位,在这个平台上聚集了众多的商家、用户、物流企业,每天都会产生大量的在线交易,因此这个平台掌握了网络购物物流需求数据、电商货源数据、货流量与分布数据以及消费者长期购买习惯数据,物流公司可以对这些数据进行大数据分析,优化仓储选址、干线物流基础设施建设以及物流体系建设,并根据商品需求分析结果提前把货物配送到需求较为集中的区域,做到“买家没有下单,货就已经在路上”,最终实现“以‘天网’数据优化‘地网’效率”的目标。有了“天网”数据的支撑,阿里巴巴可以充分利用大数据技术,为用户提供个性化的电子商务和物流服务。用户从“时效最快”“成本最低”“最安全”“服务最好”等选项中选择快递组合类型后,阿里巴巴会根据以往的快递公司的服务情况、各个分段的报价情况、即时运力资源情况、该流向的即时件量等信息,甚至可以融合天气预测、交通预测等数据,进行相关的大数据分析,从而得到满足用户需求的最优线路方案供用户选择,并最终把相关数据分发给各个物流公司去完成物流配送。

“菜鸟”计划的关键在于信息整合,而不是资金和技术的整合。阿里巴巴的“天网”和“地网”,必须具备把供应商、电商企业、物流公司、金融企业、消费者的各种数据全方位、透明化地加以整合、分析、判断,并转化为电子商务和物流系统的行动方案的能力。

一年一度的“双 11 购物狂欢节”是中国网民的一大盛事,也是对智能物流网络的一大考验。在每年的“双 11”活动中,阿里巴巴都会结合历史数据,对进入“双 11”的商家名单、备货量等信息进行分析,并提前对“双 11”订单量做出预测,精确到每个区域、网点的收发量,所有信息与快递公司共享,这样,快递公司运力布局的调整就能更加精准。

“菜鸟”网络还将数据向电商开放,如果某个区域的快递压力明显增大,“菜鸟”网络就会通知电商错峰发货,或是提早与消费者沟通,快递公司可及时调配运力。2014 年度天猫“双 11 购物狂欢节”数据显示,大数据已经开始全面发力,阿里巴巴搭建的规模庞大的 IT 基础设施已经可以很好地支撑该活动当天 571 亿元的惊人交易量,6 h 处理 100 PB 的数据,每秒处理 7 万单交易,同时,以大数据为驱动,借助智能物流体系——“菜鸟”网络,天猫已经实现预发货,即“买家没有下单,货就已经在路上”。

◆ 1.2.4 能源中的大数据

各种数据显示,人类正面临着能源危机。以我国为例,根据目前能源使用情况,我国可利用的煤炭资源仅能维持 30 年,由于天然铀资源的短缺,核能的利用仅能维持 50 座标准核电站连续运转 40 年,而石油的开采仅能维持 20 年。

在能源危机面前,人类开始积极寻求可以用来替代化石能源的新能源,风能、太阳能和生物能等可再生能源逐渐被纳入电能转换的供应源。但是,新能源与传统的化石能源相比,具有一些明显的缺陷。传统的化石能源出力稳定,布局相对集中,而新能源则出力不稳定,地理位置也比较分散,比如风力发电机一般分布在比较分散的沿海或者草原、荒漠地区,风量小时发电量就多,风量小时发电量就少,设备故障检修期间就不发电,无法产生稳定可靠的电能。传统电网主要是为稳定出力的能源而设计的,无法有效吸纳、处理不稳定的新能源。

“智能电网”就是人们认识到传统电网的结构模式无法大规模适应新能源的消纳需求而提出的,人们认为必须将传统电网在使用中进行升级,既要完成传统电源模式的供用电,又要逐渐适应未来分布式能源的消纳需求。概括地说,智能电网就是电网的智能化,是建立在集成的、高速双向通信网络的基础上,通过应用先进的传感和测量技术、先进的设备技术、先进的控制方法以及先进的决策支持系统技术,实现电网可靠、安全、经济、高效、环境友好和使用安全的目标,其主要特征包括自愈、抵御攻击、提供满足 21 世纪用户需求的电能质量、容许各种不同发电形式的接入、启动电力市场以及资产优化高效运行等。

智能电网的发展离不开大数据技术的发展和运用,大数据技术是组成整个智能电网的技术基石,全面影响到电网规划、技术变革、设备升级、电网改造以及设计规范、技术标准、运行规程乃至市场营销政策的统一等方面。电网全景实时数据采集、传输和存储,以及累积的海量多源数据快速分析等大数据技术,都是支撑智能电网安全、自愈、环保及可靠运行的基础技术。随着智能电网中大量智能电表及智能终端的安装部署,电力公司可