



大数据分析及应用实践

总主编 马小东

主 编 杨和稳

副主编 任越美 张 旭 库 波 金 艺

高等教育出版社

BIG DATA

DATA MINING

ANALYSIS

VISUALIZATION

APPLICATION

大数据分析及应用实践

DASHUJU FENXI JI YINGYONG SHIJIAN

总主编 马小东

主 编 杨和稳

副主编 任越美 张 旭 库 波 金 艺

高等教育出版社·北京

内容提要

本书是大数据系列教材之一，是大数据分析理论与实践相结合的入门级教材。

本书主要内容包括认识大数据、大数据技术基础、大数据管理、大数据统计分析技术、数据挖掘、大数据可视化、大数据安全、大数据应用案例实操等。本书编写是以国云数据自主研发的大数据分析工具——“魔镜”为平台，理论结合实践，通过具体案例介绍如何进行大数据分析。读者可通过移动设备扫描书中的二维码观看彩色图片、操作视频等多媒体教学资源，还可通过书后提供的账号与密码登录大数据实验科研平台（www.labbigdata.com）进行实践操作。

本书可作为高等院校大数据入门课程的教材，亦可作为大数据技术相关培训教材，同时也可供大数据相关的从业人员参考。

图书在版编目(CIP)数据

大数据分析及应用实践 / 杨和稳主编. — 北京：高等教育出版社，2016.7
ISBN 978-7-04-044993-8

I. ①大… II. ①杨… III. ①数据处理—高等职业教育—教材 IV. ①TP274

中国版本图书馆CIP数据核字(2016)第150516号

策划编辑 孔全会 责任编辑 张朶琳 王 威 版式设计 高海钰 张文豪
封面设计 张文豪 责任印制 高忠富

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100120
印 刷 上海天地海设计印刷有限公司
开 本 787mm×1092mm 1/16
印 张 13.25
字 数 293千字
购书热线 021-56717287
010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
<http://www.hep.com.cn/shanghai>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版 次 2016年7月第1版
印 次 2016年7月第1次印刷
定 价 35.00元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换
版权所有 侵权必究
物 料 号 44993-00

目 录

第1章 认识大数据

1.1 大数据综述	1
1.1.1 大数据的产生	1
1.1.2 大数据的特征	2
1.2 大数据的基本概念	5
1.2.1 大数据的定义	5
1.2.2 大数据的特点	5
1.3 大数据思维	7
1.3.1 大数据思维的变革	7
1.3.2 大数据思维的关注点	9
1.4 大数据的处理过程	10
1.4.1 大数据采集	10
1.4.2 大数据导入与预处理	10
1.4.3 大数据统计与分析	11
1.4.4 大数据挖掘	11
1.5 大数据分析工具	12
1.5.1 InfoSphere BigInsights简介	12
1.5.2 BigQuery简介	13
1.5.3 “魔镜”简介	13
1.6 大数据的应用	14
1.6.1 大数据助石油公司智能营销、精准管理	14
1.6.2 大数据助乳业公司预测产奶量	15
实验1 认识大数据分析工具“魔镜”	16

第2章 大数据技术基础

2.1 基础架构支持	24
2.1.1 Hadoop	24
2.1.2 HBase	26
2.1.3 MapReduce	26

2.1.4	Hive.....	27
2.1.5	Pig.....	27
2.1.6	R语言.....	28
2.2	云计算.....	28
2.2.1	云计算的特点.....	29
2.2.2	云计算与大数据.....	30
2.3	数据采集.....	30
2.3.1	数据采集的意义.....	30
2.3.2	数据采集的方法.....	31
2.4	数据存储.....	32
2.4.1	数据存储的概念.....	32
2.4.2	数据的存储方式.....	32
2.4.3	常见数据源类型.....	33
	实验2 数据源接入与编辑.....	36

第3章 大数据管理

3.1	大数据的清洗.....	45
3.1.1	数据质量.....	46
3.1.2	数据清洗的作用.....	48
3.1.3	数据清洗的方法和过程.....	49
3.1.4	数据清洗的实例.....	50
3.2	数据类型和数据转换.....	51
3.2.1	数据类型.....	52
3.2.2	数据转换.....	55
3.3	大数据的提取和加载.....	56
	实验3 数据处理.....	58

第4章 大数据统计分析技术

4.1	统计分析概述.....	69
4.1.1	统计分析的概念.....	69
4.1.2	统计分析的特点.....	71

4.1.3 统计分析的应用.....	73
4.2 统计分析的常见指标	75
4.2.1 统计指标概述.....	75
4.2.2 总量指标.....	77
4.2.3 相对指标.....	80
4.2.4 平均指标.....	83
4.2.5 变异指标.....	88
4.3 回归与预测	90
4.3.1 回归.....	90
4.3.2 预测.....	92
 实验4 数据分析.....	93

第5章 数据挖掘

5.1 大数据挖掘概念.....	100
5.1.1 数据挖掘的起源.....	100
5.1.2 数据挖掘的定义.....	102
5.2 数据挖掘任务	102
5.2.1 数据总结.....	102
5.2.2 分类.....	102
5.2.3 关联分析.....	103
5.2.4 聚类.....	103
5.3 数据挖掘流程.....	103
5.3.1 业务理解.....	104
5.3.2 数据理解.....	104
5.3.3 数据准备.....	105
5.3.4 建立模型.....	105
5.3.5 结果评价.....	106
5.4 数据挖掘的常用方法	106
5.4.1 决策树.....	106
5.4.2 遗传算法.....	109
5.4.3 神经网络.....	111

5.4.4 关联规则.....	112
5.4.5 粗糙集.....	113
5.4.6 判别分析.....	114
 实验5 数据挖掘.....	116

第6章 大数据可视化

6.1 数据可视化分类.....	122
6.1.1 结构可视化.....	122
6.1.2 功能可视化.....	123
6.1.3 关联关系可视化.....	123
6.1.4 趋势可视化.....	124
6.2 可视化表现形式.....	125
6.2.1 二维可视化形式.....	125
6.2.2 三维可视化形式.....	126
6.2.3 仪表盘.....	128
6.2.4 定制可视化形式.....	129
6.2.5 大数据可视化方式的选择.....	130
 实验6 数据可视化.....	132

第7章 大数据安全

7.1 大数据安全的内容.....	158
7.1.1 物理安全.....	159
7.1.2 网络安全.....	160
7.1.3 应用安全.....	160
7.1.4 数据隐私.....	162
7.2 数据保护技术.....	163
7.2.1 镜像技术.....	163
7.2.2 快照技术.....	164
7.2.3 持续数据保护技术.....	165
7.2.4 用户管理技术.....	166
 实验7 权限管理机制.....	170



第8章 大数据应用案例实操

8.1 学生成绩分析.....	174
8.1.1 背景分析.....	174
8.1.2 需求分析.....	175
8.1.3 大数据分析过程.....	175
8.1.4 分析结论.....	184
8.2 电商行业销售分析.....	184
8.2.1 背景分析.....	184
8.2.2 需求分析.....	184
8.2.3 大数据分析过程.....	185
8.2.4 分析结论.....	195
参考文献.....	196

第1章

认识大数据

近年来，大数据风云乍起，正在深刻影响着政治、经济、安全、健康、教育等各个方面，人们已意识到大数据是重要的数据资产，充分重视和运用大数据，能够产生更大的社会和经济价值。从数据到大数据，不仅是量的积累，更是质的飞跃。通过整合、分析不同来源、不同形式、包含不同信息的海量数据，使原本孤立的数据变得互相关联，让人们有新的发现，创造新的价值。

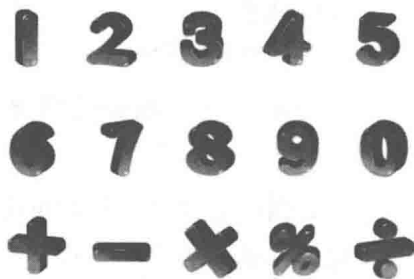
1.1 大数据综述

1.1.1 大数据的产生

在古埃及，人类已经学会了通过数据来计量财富和安排日常的生活。在文艺复兴后的欧洲，人们开始用数据来刻画自然乃至人文规律。随着近现代信息技术的发展和数字化进程的日益加快，数据逐渐脱离了仅仅作为刻度表征的特性，成为世界万物的量化映射。世界可

以通过数据来表示，人类也可以通过数字化的信息对客观世界进行再认识。数据成为描述客观世界的有力工具，文字、图像、视频、音频等均可以用数据形式进行描述。总之，世界的一切关系皆可用数据来表征。

通过数据来研究规律、发现规律，贯穿了人类社会发展的始终。人类



科学发展史上的不少进步都和数据采集、分析直接相关。1854年，伦敦发生了大规模的霍乱，很长时间没有办法控制。一位医师用标注地图的方法研究了当地水井分布和霍乱患者分布之间的关系，发现有一口水井周围霍乱患病率明显较高，借此找到了霍乱暴发的原因：一口被污染的水井。关闭这口水井之后，霍乱的发病率明显下降。这种方法，充分展示了数据的力量。

近现代以来，随着面临的问题越来越复杂，通过演绎的方式来研究这些问题常常变得很困难，这就使得数据归纳的方法变得越来越重要，数据的重要性也越发凸显出来。

随着互联网逐渐过渡到物联网，网络中互联的主体不仅仅是人，也可以是智能终端、传感器，乃至可穿戴设备等。人们的各方面信息正通过无处不在的物联网被采集、汇总和辨析，人类的生活最终进入虚拟化。数据的爆发式增长产生了各种各样的大数据。大数据又称巨量数据，指的是海量、高增长率和多样化的信息资产。大数据革命正在对世界产生巨大的系统性影响和深远意义。早在2012年美国就发布了《大数据研究和发展计划》，并成立了“大数据高级指导小组”。欧盟也正在力推《数据价值链战略计划》，英国发布了《英国数据能力发展战略规划》，日本的《创建最尖端IT国家宣言》和韩国的“大数据中心战略”也陆续出台。我国在2016年发布了“互联网+”行动计划，推进信息化与工业化深度融合，开放、共享和智能的大数据时代已经来临。

1.1.2 大数据的特征

与传统数据的产生方式相比，大数据具有三个明显的特征——数据量大、非结构性和实时性。

数据量大是大数据的明显特征，一般的计量单位都是PB、EB，甚至ZB。

大数据既包含结构化数据也包含非结构化数据，而一般传统数据都是结构化的数据，更容易被理解。非结构化数据的收集、存储及使用是非常复杂的，人们需要通过特定的大数据技术从大量非结构化数据中提取有用的信息。

数据是永远在线的、可以随时调用和计算的，这是大数据相较于传统数据最大的不同。在互联网高速发展的背景下，大数据不仅仅数量巨大，实时性、动态性也是大数据的重要特征。

中国古代治国就已经有重数据的思想，如商鞅提出：“强国知十三数……欲强国，不知国十三数，地虽利，民虽众，国愈弱至削”。大数据时代，循“数”治国将更加有效。传统数据时代，政府做决策更多依凭经验和局部数据，难免头痛医头、脚痛医脚，例如，交通堵塞就多修路。大数据时代，政府做决策能够从粗放型转向集约型。路堵了，利用大数据分析，可以得知哪一时间、哪一地段最容易堵，或在这一地段附近多修路，或提前预警引导居民合理安排出行，实现对交通资源的最佳配置和控制，改善交通。

大数据时代，每个人都会“自发地”提供数据。人们的各种行为，如点击网页、使用手机、刷卡消费、观看电视、坐地铁出行、驾驶汽车，都会生成数据并被记录下来，人们的性别、职业、喜好、消费能力等信息，都可以被商家从中挖掘出来，提供商机。

大数据也将使个人受益。从生物学、医学上讲，以前生物学家只是通过对单个或几个基因的操控来观察其对生物体的影响，很难发现整体的关联。现在由于技术的发展，可以分析遗传信息、全体基因的表达信息、蛋白质族谱信息、全基因组甲基化信息、表观遗传信息等更多的信息，同时还有个人健康指标、病历、药物反应等数据。如果能达成生物学上多维多向数据的有机融合，就能够把一个人完整地描述出来，从而实现精准医疗的目的。

大数据时代，审核数据的真实性也有了更有效的手段。大数据的多样性，使得不同来源、不同维度的数据之间存在一定的关联度，可以交叉验证。例如，某地的工业产值虚报了一倍，但用电量和能耗却没有达到相应的规模。这就是数据异常，很容易被系统识别出来。发现异常后，相关部门再进行复核，就能更有针对性地防止、打击数据造假。

数据是一种资源，但数据又跟煤、石油等物质性资源不一样。物质性资源不可再生，因而很难共享。而数据可以重复使用，不断产生新的价值。大数据资源的使用是非恶性竞争的，在共享的前提下，更能够制造多赢。

大数据由于本身附带或隐含的价值，被类比为新时代的石油、黄金，甚至被视为一种与资本和劳动力并列的新经济元素，即大数据不仅对生产过程中形成产品和产生价值起着重要的作用，其本身也可看作为生产力要素，是产品生产中不可或缺的元素，是产品中不可分割的一部分。其价值主要体现在以下三个方面。

1. 大数据为新一代信息技术产业提供核心支撑

大数据问题的爆发以及大数据概念在全球的普及,是现代信息技术发展的必然。互联网以及移动网络的飞速发展使得网络基础设施无处不在,网络带宽也在不断拓展,最新的移动 4G LTE 网络支持 166Mbps 的峰值下行速度,下载一部蓝光电影只需几分钟,这使得人们能够随时随地访问数据。而云计算、物联网、社交网、购物网等的兴起和发展,使得数据以史无前例的速度每时每刻都在产生。大数据是信息技术和社会发展的产物,而大数据问题的解决又会促进云计算、物联网等新兴信息技术的发展和应用,大数据正成为未来新一代信息技术融合应用的核心,为云计算、物联网、移动网络等各项信息技术相关的应用提供坚实的基础。

2. 大数据正成为社会发展和经济增长的高速引擎

大数据蕴含着巨大的社会、经济和商业价值。大数据市场的井喷会催生一大批面向大数据市场的新模式、新技术、新产品和新服务,进而促进信息产业的加速增长。大数据对于全球经济、国计民生、政策法规等方面都至关重要。大数据在我国正在大力推行建设的“智慧城市”中将是不可或缺的。智慧城市的本质是将各行各业的数据关联打通,从中分析挖掘,从而形成城市的智慧联动。而其中从数据的采集到数据的分析挖掘,以及形成智能决策的每个过程,都离不开大数据的支撑。随着智慧城市建设的推进,大数据将有力地促进政务及社会化管理水平的提升,改进民生,发展生产,形成一系列有地方特色的、有清晰运营模式的新一代智能行业应用。

大数据将助力企业深度挖掘和利用数据价值,完成智能决策,在企业运营中提高效率、节约成本;在市场竞争中制订正确的市场战略,把握市场先机,规避市场风险;在市场营销中全面掌握用户需求,进行精准营销和个性化服务。企业的决策正从“应用驱动”到“数据驱动”,能够有效地利用大数据并将其转化为生产力的企业,将具备核心竞争力,成为行业领导者。同时,大数据已经深入与人们生活息息相关的各个领域。在休闲娱乐、教育、旅游、健康等各个领域,都能见到大数据的应用。

3. 大数据将成为科技创新的新动力

传统行业的信息化建设思路和技术落后,导致了大量数据被分离、闲置在各类彼此隔离的系统之中,同时各行各业、政府等也面临实时数据处理、应用方面的巨大挑战,依托传统的信息处理方式已无法取得质的突破。因此,如何以新的数据技术整合数据、存储数据、处理数据、应

用数据，解决业务系统实时性问题、并发性问题、海量数据存储计算问题、数据价值挖掘及应用问题是传统行业迫切的需求。各行各业需要科技创新，大数据成为科技创新的新动力。例如国内某家大型家电制造商，在生产线上安装传感器，记录大量一线的生产信息，通过分析、处理这些信息，来提高产品的优品率。再以国家电网为例，智能电表及电能量采集系统收集数据的目的首先是为收费，实际上这些数据能够发挥更大的作用。首先，能够判断房屋是否闲置，可以用于指导一个城市房价、地价的调控；其次，电网的监控数据可以用来判断整个区域的制造业、商业在不同区域的发展状况。可以预见，在政府、电力、金融、石油、交通、社保、公安、医疗等数据高度集中的行业中，大数据将成为各企业、部门、机构提高核心竞争力、抢占市场先机的关键，成为企业从“业务驱动”向“数据驱动”转变的重要推力，为企业带来自主技术与产品研发的新契机。



1.2 大数据的基本概念

1.2.1 大数据的定义

麦肯锡作为美国首屈一指的咨询公司，是研究大数据的先驱。在其报告“Big data: The next frontier for innovation, competition, and productivity”中给出的大数据定义是：大数据指的是大小超出常规数据库工具获取、存储、管理和分析能力的数据集，即大数据是现有数据库管理工具和传统数据处理手段很难处理的大型、复杂的数据集，涉及采集、存储、搜索、共享、传输和可视化等方面。大数据的“大”是一个动态的概念，以前的GB级数据是天文数字，但现在，在地球、物理、基因、空间科学等领域，TB级的数据已经很普遍，有的达到PB（1 024 TB）级或EB（1 024 PB）级。数据的类型多种多样，包括结构化数据、半结构化数据和非结构化数据。巨大的数据量和种类繁多的数据类型给大数据系统的存储和计算带来了很大挑战。

1.2.2 大数据的特点

如图 1-1 所示，大数据的特点可归纳为“4V”：Volume（容量），即海量的数据规模；Variety（种类），即多样的数据类型；Velocity（速度），即快速的数据流转和动态的数据体系；Value（价值），即巨大的数据价值。

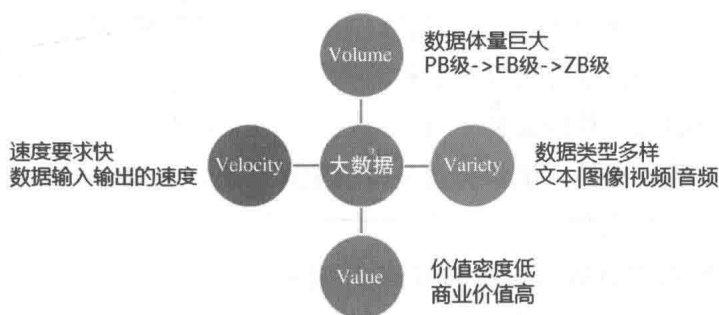


图 1-1
大数据的特点

1. Volume，数据量巨大

数据量的大小决定数据的价值及其潜在的信息量。伴随着各种可穿戴设备、物联网和云计算、云存储等技术的发展，人和物的所有轨迹都可以被记录，数据因此被大量生产出来。微博、照片、录像、自动化传感器、生产监测、环境监测、刷卡机等大量自动或人工产生的数据通过互联网聚集到特定地点，例如政府、银行、企业等机构，形成了海量的大数据。

2. Variety，数据类型繁多

因为数据源不同，非结构化数据越来越多，需要进行数据清洗、整理、筛选等操作，将其变为结构化数据。这意味着要在海量、种类繁多的数据间发现其内在关联。互联网时代，各种设备通过网络连成了一个整体。进入以互动为特征的时代，用户不仅可以通过网络获取信息，还成为了信息的制造者和传播者。这个阶段，不仅是数据量开始了爆炸式增长，数据种类也变得繁多。这必然促使人们对海量数据进行分析、处理和集成，找出原本看似毫无关系的那些数据之间的“关联性”，把似乎没有用的数据变成有用的信息，以帮助人们做出判断。

3. Velocity，处理速度快

数据的实时性需求越来越清晰。对普通人而言，开车去吃饭，会先用移动终端中的地图查询餐厅的位置，预计行车路线的拥堵情况，了解停车场信息甚至是其他用户对餐厅的评论。吃饭时，会用手机拍摄食物的照片，编辑简短评论发布到微博或者微信上……如今，通过各种有线和无线网络，人和人、人和机器、机器和机器之间产生无处不在的连接，这些连接不可避免地带来数据交换。而数据交换的关键是降低延迟，以近乎实时的方式传送给用户。

4. Value，大数据的最终意义——获得洞察力和价值

大数据的应用，正是在人工智能、机器学习和大数据挖掘等技术的迅

速发展驱动下,呈现这样一个过程:将信号转化为数据,将大数据分析为信息,将信息提炼为知识,以知识促成决策和行动。相关专家认为,就大数据的价值而言,就像沙里淘金,大数据规模越大,真正有价值的数据相对越少。

大数据价值密度低,由于数据采集的不及时,数据样本不全面,数据可能不连续等,数据可能会失真,但当数据量达到一定规模,便可以从更多的数据中提取有价值的信息,得到真实全面的反馈。所以行之有效的的大数据系统,就是把采集到的 ZB、PB 级数据,最终变成一个 bit,也就是最后的决策,这才是最关键的。



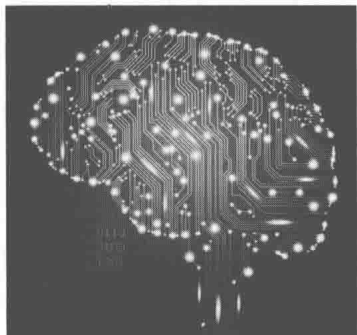
1.3 大数据思维

1.3.1 大数据思维的变革

近年来,大数据技术的快速发展深刻地改变了人们的生活、工作和思维方式。人们对大数据的认识发生了深刻的变革。

1. 从样本思维到总体思维的变革

社会科学研究社会现象的总体特征。以往,采样一直是主要的数据获取手段,这是人类在无法获得总体数据信息条件下的无奈选择。在大数据时代,随着数据收集、存储、分析技术的突破性发展,人们可以更加方便、快捷、动态地获得研究对象的所有数据,而不再因诸多限制不得不采用样本研究方法。相应地,对数据的思维方式也应该从样本思维转向总体思维,从而能够更加全面、立体、系统地把握全局。



2. 从精确思维到容错思维的变革

在小数据时代,由于收集的样本信息量比较少,所以必须确保记录下来的数据尽量结构化、精确化。否则,分析得出的结论在推及总体上就会“南辕北辙”,因此必须十分注重精确思维。然而,在大数据时代,得益于大数据技术的突破,大量的非结构化、异构化的数据能够得到储存和分析,这一方面提升了人们从数据中获取知识和洞见的能力,另一方面也对传统的精确思维形成了挑战。

在大数据时代,思维方式要从精确思维转向容错思维,当拥有海量即

时数据时，绝对的精准不再是主要目标，适当忽略微观层面上的精确度，容许一定程度的错误与混杂，反而可以在宏观层面拥有更好的洞察力。

3. 从关注因果关系到关注相关关系的变革

在对小数据进行分析时，人们往往执着于现象背后的因果关系，试图通过有限样本数据来剖析其中的内在机理。小数据的另一个缺陷就是有限的样本数据无法反映出事物之间普遍性的相关关系。而在大数据时代，人们可以通过大数据技术挖掘出事物之间隐蔽的相关关系，获得更多的认知与洞见。运用这些认知与洞见可以帮助人们捕捉现在和预测未来，而建立在相关关系分析基础上的预测正是大数据的核心议题。

通过关注线性的相关关系，以及复杂的非线性相关关系，可以帮助人们看到很多以前不曾注意到的联系，还可以掌握以前无法理解的复杂技术和社会动态，相关关系甚至可以超越因果关系，成为人们了解这个世界的更好视角。在大数据时代，思维方式要从因果思维转向相关思维，努力颠覆千百年来人类形成的传统思维模式和固有偏见，才能更好地分享大数据带来的深刻洞见。

4. 从自然思维到智能思维的变革

大数据思维最关键的转变在于从自然思维转向智能思维，使得大数据具有生命力，获得类似于“人脑”的智能，甚至智慧。计算机的出现极大地推动了自动控制、人工智能和机器学习等新技术的发展，“机器人”研发也取得了突飞猛进的成果并开始实际应用。应该说，自进入到信息社会以来，人类社会的自动化、智能化水平已得到明显提升，但始终面临瓶颈而无法取得突破性进展，机器的思维方式仍属于线性、简单、物理的自然思维，智能水平仍不尽如人意。

大数据时代的到来，为提升机器智能带来契机，因为大数据将有效推进机器思维方式由自然思维转向智能思维，这才是大数据思维转变的关键所在和核心内容。众所周知，人脑之所以具有智能、智慧，就在于它能够对数据信息进行全面收集、逻辑判断和归纳总结，获得有关事物或现象的认识与见解。在大数据时代，随着物联网、云计算、社会计算、可视技术等突破发展，大数据系统也能够自动地搜索所有相关的数据信息，进而类似“人脑”一样主动、立体、逻辑地分析数据、做出判断、提供洞见，从而具有类似人类的智能思维能力和预测未来的能力。

总之，大数据时代将带来深刻的思维转变，大数据不仅将改变每个人的日常生活和工作方式，还将改变商业组织和社会组织的运行方式，而且

将从根本上奠定国家和社会治理的基础数据，使得国家和社会治理更加透明、有效和智慧。

1.3.2 大数据思维的关注点

大数据思维是客观存在的，大数据思维是新的思维方式。用大数据思维方式思考问题、解决问题是当下的潮流。大数据思维开启了一次重大的时代转型，人们对数据的关注点也发生了很大变化。

1. 数据思维的核心是利用数据解决问题

利用数据解决问题的核心是要深度了解需求，了解真正要解决什么样的问题，解决问题背后的真实目的是什么。在解决问题的过程中使用数据的方法，通常可以叫量化的方法。

2. 大数据关注“有用”

用数据价值思维方式思考问题、解决问题。信息总量的变化导致了质变，最先经历信息爆炸的学科，如天文学和基因学，引出了“大数据”这个概念。如今，这个概念几乎应用到了所有自然科学和社会科学领域中。从“功能为价值”转变为“数据为价值”，说明数据的价值在扩大，“数据为王”的时代出现了。

3. 由关注精确度转变为关注效率

大数据标志着人类在寻求量化和认识世界的道路上前进了一大步，过去不可计量、存储、分析和共享的很多东西都被数据化了，拥有大量的数据和更多不那么精确的数据为人们理解世界打开了一扇新的大门。大数据能提高生产效率和销售效率，原因是大数据能够让人们知道市场的消费需要。大数据让企业的决策更科学，由关注精确度转变为关注效率的提高，大数据分析能提高企业的效率，例如，企业产品迭代的速度在加快，三星、小米手机制造商半年就推出一款新智能手机。在利用互联网、大数据提高企业效率的形势下，快速就是效率，预测就是效率，预见就是效率，变革就是效率，创新就是效率，应用就是效率。

4. 关注定制产品

由企业生产产品转变为由客户定制产品。用定制产品思维方式思考问题、解决问题。大数据时代让企业找到了定制产品、订单生产、精准销售的新路子。企业下一波的改革是大规模定制，为大量客户定制产品和服务，成本低又兼具个性化。要真正做到个性化产品和服务，就必须对客户需求有很好的了解，这背后就需要依靠大数据技术。