



数字治理

与政府治理改革

SHUZI ZHILI

YU ZHENG FU ZHILI GAIGE

王印红 著

数字治理

新华出版社



中国海洋大学一流大学建设专项经费资助项目

数字治理

与政府改革创新

SHUZI ZHILI

YU ZHENGFU GAIGE CHUANGXIN

王印红

新华出版社

图书在版编目 (CIP) 数据

数字治理与政府改革创新 / 王印红著.

北京: 新华出版社, 2019.1

ISBN 978-7-5166-4427-0

I. ①数… II. ①王… III. ①国家—行政管理—现代化管理—研究—中国

IV. ① D630.1

中国版本图书馆 CIP 数据核字 (2019) 第 003885 号

数字治理与政府改革创新

作 者: 王印红

责任编辑: 董朝合

封面设计: 王雪晨

出版发行: 新华出版社

地 址: 北京石景山区京原路 8 号

网 址: <http://www.xinhupub.com>

购书热线: 010—63077122

邮 编: 100040

经 销: 新华书店

中国新闻书店购书热线: 010—63072012

照 排: 北京钟书堂文化传播有限公司

印 刷: 北京市金星印务有限公司

成品尺寸: 170mm × 235mm 1/16

印 张: 16.75

版 次: 2019 年 5 月第一版

字 数: 257 千字

印 次: 2019 年 5 月第一次印刷

书 号: ISBN 978-7-5166-4427-0

定 价: 68.00 元

图书如有印装问题, 请与印刷厂联系调换 电话: 13020075142

序 言 / PREFACE

大数据时代毋庸置疑地来临了。计算机与信息系统大规模的运用，各种数字设备进入家庭，进入政府部门，为数据的大规模生产带来了空前的机遇和能力。

回首世纪之交，全球信息化浪潮风起云涌。1998年，美国提出“数字地球”的概念。随后，中国科学家展开针对性研究，并提出与之相对应的“数字中国”，数字浪潮风起云涌。2015年，在第二届世界互联网大会开幕式上，国家主席习近平面对多国政要及来自世界各国的互联网精英明确表示：要“推进‘数字中国’建设，发展分享经济，支持基于互联网的各类创新，提高发展质量和效益。”时光的指针进入新时代，中国的数字化进程已经扩展到政务、民生、实体经济等各个领域，数字化建设迎来了新纪元。

大数据时代的来临有其深刻的历史背景和发展必然性。

其一，信息技术特别是数字技术应用于普及的必然结果。信息技术、网络技术、数字技术为大数据的产生与发展提供了技术平台与支撑环境。从计算机的核心技术来说，微处理器的发展正如摩尔^①预测的那样“微处理器中的晶体管数每两年翻一番”，1971年之前，计算机是有一个房间那么

① 戈登·摩尔是intel公司的创始人，1965年他在准备一词计算机存贮会议的材料时发现：每一个新出的芯片上包含的晶体管数是上一代的两倍。每一代新芯片的生产周期是18—24个月。在1972年摩尔重新定义了他的说法：每两年芯片上的晶体管数翻一番，性能提高一倍。之后被各个理解者进行了不同的阐释，主要的版本有：1. 集成电路芯片上所集成的电路的数目，每隔18个月就翻一番；2. 微处理器的性能每隔18个月提高一倍，而价格下降一半；3. 用一个美元所能买到的电脑性能，每隔18个月翻两番。此类论述被称为信息时代第一定律，也称摩尔定律。

大的庞然大物。Intel 的 CPU4004 的产生改变了计算机发展的走势,计算机由原来的大型昂贵的科研设备,由机房走入家庭,由昂贵走向普遍。Intel 微处理器的出现改变了 IT 发展的走向,电脑由神秘走入普通人的生活。多核处理器已使得计算机的处理能力提升到惊人的水平,为信息处理提供了最基本的技术基础,计算机逐渐变革成更加个性化的、易于应用的、能够处理复杂事务的微型设备。正如尼葛洛庞帝所言“计算机已经与计算无关”^①。现在还有谁会说计算机的功能主要是计算呢?

我们生活在一个模拟世界中,但数字技术已经成为主流技术。混合信号解决方案过去包含大量模拟数据,只需要少量数字信号处理,这种方案已经迁移到系统应用中,在系统中第一次产生了模数转换过程。模拟技术衰落的原因之一是建立在自身缺陷上的。摩尔定律适用于数字电路而不是模拟电路;晶体管可以而且必须做得更小,这有利于数字电路。但这对模拟晶体管的影响并不大,反而器件尺寸越小,模拟器件特性往往越差。器件的小型化一直是这个世界技术进步的关键,在这一点上模拟技术不能跟上时代,逐渐被淘汰。

原因之二是模拟技术在信号、时间和数值上都是连续变化的,这样的变化性质很难在传输的过程中被加密,因此,极易被窃取或者被监听,反观数字技术中采用离散数据(0—1)来表示传输信号,基本单元电路简单,对电路中各元件精度要求不很严格,允许元件参数有较大的分散性,只要能区分两种截然不同的状态即可。此外,数字技术具有抗干扰能力强、保密性好等特点,逐渐成为信息世界的主流技术。

因为数字技术的创新与应用,大数据时代启幕。顾名思义,大数据时代就是指有大量数据生成并利用这些数据的时代。IBM 估算,全球每天生成的数据量达到 2.5EB,而且其增长势头在不断加快^②。如果换算成家庭使用的普

① 尼古拉斯·尼葛洛庞帝(Nicholas Negroponte)出生于1943年,是一个美国计算机科学家,因为长期以来一直在倡导利用数字化技术促进社会生活的转型,被西方媒体推崇为电脑和传播科技领域最具影响力的大师之一。

② 人民网.2年生成人类史上9成数据,迅猛增长的大数据将改变未来[EB/OL].[2012—12—20].[2017—12—20]. <http://finance.people.com.cn/n/2012/1220/c348883-19959503.html>.

通蓝光光盘, 2.5EB 相当于 10 亿张容量为 25GB 的蓝光光盘。而且令人吃惊的是, 人类迄今为止生成的数据中, 有 90% 是在近两年内产生的。电子邮件、微信、微博、手机随手拍摄的图片、视频、各种传感器收集的数据以及网上购物的日志数据等, 都被存储起来, 成为大数据。

其二, 大数据改变生活、改革政府。计算机网络的出现, 将表面积 5.1 亿平方公里的地球连成一个小小的村庄, 人类向这个小村庄迁徙的过程称之为信息化。在信息化的过程中, 人类开拓了一个没有疆域的新型空间——网络空间 (Cyberspace)^①。这个虚拟空间却为各个国家、各个地区政府、组织和个人提供着实实在在的服务, 并彻底改变了人们获取信息、存储信息、传递信息的方式, 最重要的是改变了人类的生活方式, 人们在网上购买到商品与服务, 每个人都可能因为网络而结缘。

2018 年 1 月 30 日 CNNIC 在京发布第 41 次《中国互联网络发展状况统计报告》。报告称: “我国网民规模达 7.72 亿, 普及率达到 55.8%, 超过全球平均水平 (51.7%)4.1 个百分点, 超过亚洲平均水平 (46.7%)9.1 个百分点。我国网民规模继续保持平稳增长, 互联网模式不断创新、线上线下服务融合加速以及公共服务线上化步伐加快, 成为网民规模增长推动力。”^② 互联网的发展, 创造了一个新的群体——网民。网民可以称为新一代的公民, 他们的数量急剧增加, 他们的生活方式、工作方式也逐渐改变。

大数据改变生活。大数据已经渗透进人们的日常生活, 最明显的变化就是只需要一部手机就能走遍天下的梦想在我们身边已成为现实。早晨用手机叫车、中午掏出手机叫外卖、晚上在超市买完食材用手机扫码, 完成

① 赛博空间是美国科幻作家威廉·吉布森 (William Gibson) 在 1984 年写下了一个长篇的离奇故事, 书名叫《神经漫游者》(Neuromancer), 小说出版后, 好评如潮, 并且获得多项大奖。故事描写了反叛者兼网络独行侠凯斯, 受雇于某跨国公司, 被派往全球电脑网络构成的空间里, 去执行一项极具冒险性的任务。进入这个巨大的空间, 凯斯并不需要乘坐飞船或火箭, 只需在大脑神经中植入插座, 然后接通电极, 电脑网络便被感知。当网络与人的思想意识合为一体后, 即可遨游其中。在这个广袤的空间里, 看不到高山荒野, 也看不到城镇乡村, 只有庞大的三维信息库和各种信息在高速流动。吉布森把这个空间取名为“赛博空间”(Cyberspace), 也就是现在所说的网络空间。

② CNNIC. 第 41 次中国互联网络发展状况统计报告 [EB/OL].[2018-3-5],[2018-4-6].
http://www.cnnic.cn/hlwzfzyj/hlwzxbg/hlwjtjbg/201803/t20180305_70249.htm.

支付。其实，不只是这些。你现在甚至都不用再带着各种会员卡、各种打折卡，因为这些实体卡如今都能够通过虚拟形式存储在手机里。你也不用再拿着一大堆保险资料和表格填个半天，因为通过手机不仅能够直接了解详细内容，而且还能够直接完成购买的每一步。还有更多，比如直接通过手机进行股票买卖、银行转账及业务办理；通过手机挂号、预约专家，甚至是看病；通过手机直接遥控电视、空调、冰箱、热水器、电饭煲；通过手机共享位置，聚会不用再四处寻找。这样的例子，数不胜数。

毫无疑问，这些变化都是大数据的功劳。曾有专家对此表示，21世纪是数据信息大发展的时代，包括互联网、移动互联网、物联网、车联网、GPS、医学影像、安全监控、金融、电信等都在疯狂地产生着数据。对于这些信息数据的利用也就完全改变了我们的生活方式。

大数据改变政府。在政府行政体制改革、公共服务领域和日常管理中，政府已把大数据作为基础性战略资源逐步加以发展和运用，并充分落实创新发展的理念，加快政府职能的转变，提升政府治理能力。2015年4月3日，国务院在政府信息公开工作要点中表示要积极推进政府数据公开，把运用大数据等现代信息技术作为促进政府职能转变、简政放权、放管结合和优化服务的有效手段。大数据也助力于政府转型，首先是帮助政府从管理型政府向服务型政府转型，利用大数据技术和方法提升公共服务能力。其次，原来的政府管理属于粗放型管理，数据可以支撑政府管理走向精细化。第三，以前的政府管理是零散孤立的，社保就是社保的事，公安就是公安的事，医疗就是医疗的事，联结各部门数据可助推政府向协作共享型转变。第四，就是从柜台式向自助式转型，利用互联网以后可以不用到窗口办事。最后，是大数据推动政府从被动响应，向主动预见型转型，即通过大数据分析预测未来的发展，主动采取预见性行动，这在军事领域尤其重要。

网民生活方式、工作方式逐渐发生明显的变化，网民们已经具备或即将具备通过网络同政府打交道的能力。当网民在逐步了解网络和利用网络，在非公共领域有了一些良好的服务体验后，将不愿再忍受公共领域存在的低劣的、非个性化的服务。如果通过网络可以查询自己的通话记录、详细清单，人们也希望通过网络查询医疗记录，查询户籍信息；如果通过网络可以

交易股票，人们也希望采取类似的方式进行交税。在信息时代，人们希望无论是在家中，还是在办公室里，都能得到全天候的及时的公共服务。作为管理国家和社会事务并且为所有社会成员提供服务的政府，当前面临巨大的压力。面对这种情况，各国政府唯一的选择就是顺应信息时代的发展，运用网络和信息技木，改进工作流程、优化服务手段，进行体制创新和管理创新，提升管理水平和服务质量。

政府顺应时代需求，运用网络技术、信息技术来改造政府的过程，就是互联网建设和发展的过程。互联网作为一个新生事物，天生就与网络结合在一起，它承袭了网络世界中突出体现出来的先进理念，如效率、平等、共享等。同时，互联网又诞生于民权呼声日益高涨的时代，作为一项政府治理工具，又吸取了当代政府善治的理念，这些理念是互联网建设的根基。因为有根基互联网才茁壮成长，生机勃勃。因为秉持这些理念，笔者才对互联网背景下大数据的未来感到欣喜，并充满信心。

当试着结束本书的导言部分时，我心思潮涌，感慨万分，原因是我们有幸生活在一个信息时代，我们拥有一个强有力的工具去改变自己，变革政府，甚至改变世界。在这个网络时代，每一个个体都能在网络上自由发出自己的声音，我们不仅是信息的接受者，而且是传播者；不仅仅是被管理者，而且是管理者；不仅仅是被服务者，而且是服务者；不仅仅是围观者，而且是参与者、建设者。每个草根都可能是明星，每个草根都可能是风云人物，每一个草根都可能成为影响世界的人。

contents

目录

序 言

第一章 数字时代的基本特征	1
第一节 大数据的特征与治理挑战	2
第二节 数字空间与乌托邦	10
第三节 “证”为何物，“办证”难在何处	15
第四节 行政审批改革与办证难的解决思路	22
第五节 信息主义社会	29
第二章 数字公平与信息扶贫	35
第一节 互联网的设计理念：平等通信权	36
第二节 政府数字化：高效政府治理基础	41
第三节 数字公平：新时代公平的新内涵	50
第四节 大数据时代的信息公平	56
第五节 数字鸿沟与信息扶贫	60
第三章 数字民主与网上议政	67
第一节 民主其实是为了自己	68
第二节 数字民主预示民主春天的到来	77
第三节 网上议事厅：有序参与的重要途径	90
第四节 电子投票系统：让大规模直接投票成为可能	95
第五节 博客问政：更广泛、更深度的参与式民主	98

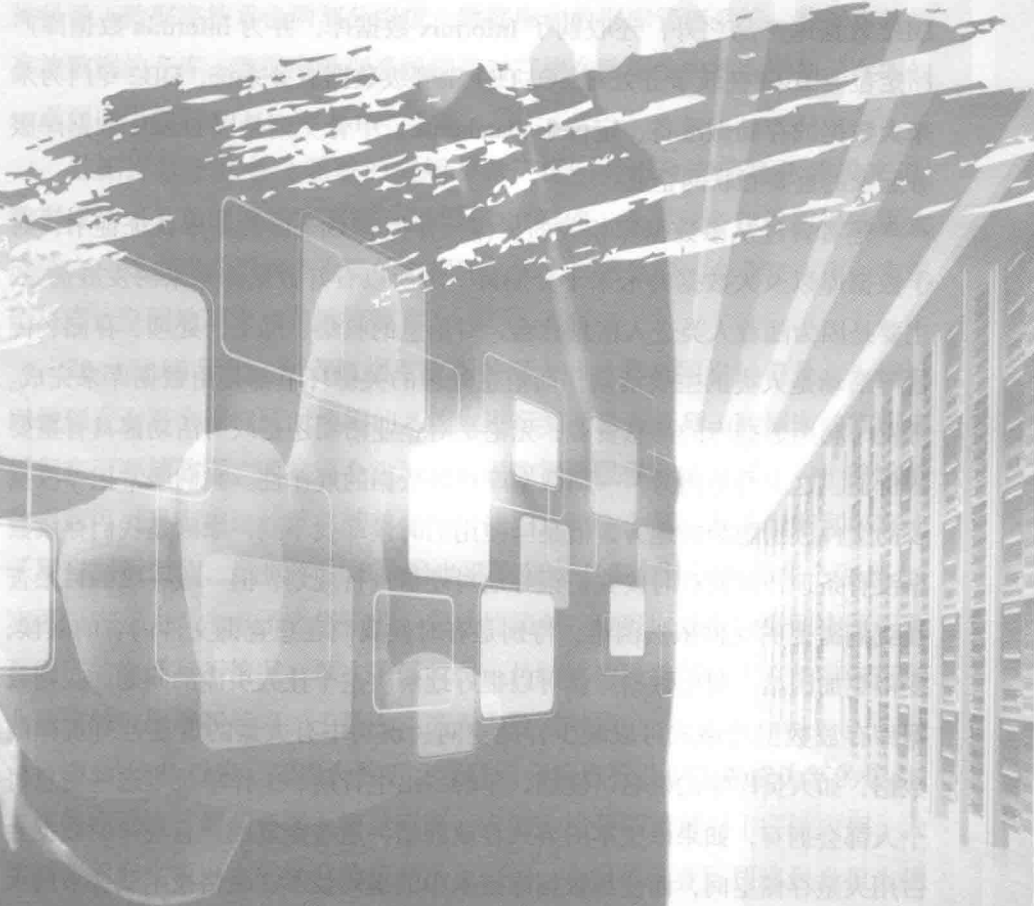


第四章 大数据时代的隐私与数据保护	107
第一节 数据安全五要素	108
第二节 看不见的数据安全问题	111
第三节 网络战争：战争的新变种	115
第四节 大数据时代隐私保护可能是个伪命题	121
第五节 大数据时代的网络实名制	126
第六节 更高的数据安全，需要更高的成本	132
第五章 互联网与数字治理立法	139
第一节 神治、人治到法制：数字空间需要法治	140
第二节 网络管辖权争议	149
第三节 人工智能及其规制	153
第四节 从《电子签名法》看大数据时代立法过程之难	158
第五节 《政府信息公开条例》颁布实施 10 周年回望	163
第六章 政府治理改革创新	171
第一节 政府改革价值取向的重新审视	172
第二节 强政府还是强社会	183
第三节 政府改革转型：提供信息服务	190
第四节 政府改革：服务还是服务	195
第五节 数字化精品与工匠精神	204
第六节 数字产品的产权与保护	209
第七节 “数字城市”到“智慧城市”	213
第八节 政府改革创新与流程再造	217
第九节 “互联网+”的产业政策的必要性	223
第十节 推行地方政府 CIO 制度	228
后 记：什么是更好的政府	234
参考文献	243

量级数据空间。定制数据库可以从最小的数据单位解决其中“定制”问题。对于一个大文件来说，它可以有很多人可以同时访问，就像允许一个人为这个文件进行编辑等。每个人对一个文件同时访问的延迟会变大，数据库则能够通过表来定制数据库，定制数据库的规模大小，在一秒钟内。

第一章

数字时代的基本特征



第一节 大数据的特征与治理挑战

大数据 (Bigdata) 是指信息爆炸时代产生的, 无法在可接受的时间范围内, 用常规的软件工具进行捕捉、管理、处理的数据集合。大数据是相比较而言的, 原来的数据是可以存储在一台计算机的存储器中的, 但是数据的产生速度以及人们对于大数据的需求超出了计算机存储器的发展速度。随着人们对于快速存取数据的要求越来越高, 计算机科学家发展了数据库技术, 业界的大佬 IBM、HP, 微软均都有自己旗下的数据库产品, IBM 除了自己的 DB2 数据库产品之外, 还收购了 Informix 数据库, 并为 Informix 数据库产品定位为 IBM 在线事务处理 (OLTP) 旗舰级数据服务系统, DB2 专门为未来大数据的存储做准备, 定位于 Datahouse。甲骨文则是以 Oracle 数据库服务起家的重要互联网企业。

笔者曾经从事数据库相关的工作, 并拥有 Oracle 数据库认证证书, 对于数据库以及大数据的未来十分期待。之所以看好数据库未来的发展前景, 主要是因为随着人类进入信息社会, 对信息的收集、加工、处理、存储、传输等活动是人类的主要活动, 而信息处理的关键环节都是由数据库来完成。开发与利用数据库的信息资源, 无论是对企业活动还是政府活动都具有重要的价值。这主要是因为: 一则数据库存放数据的规律性。我们经常把一些重要的文件放到隐蔽的地方, 但是当使用的时候却找不到, 原因是我们存放数据没有条理, 即使有时候我们很好地对硬盘进行规划, 但一级一级的目录查找我们需要的文件依然困难, 特别是有时候我们还想查询文件内容的时候, 变得更加困难, 利用数据库就可以很好地解决这个让人头痛的问题。二则数据库存放数据的格式可以减少存储空间。现实中有大量的重复且有规律的数据, 如人员库中的姓名、性别、年龄、出生日期、工作单位等信息每个人都会拥有, 如果以文本的方式存放的话, 这些重复的、有规律的数据会占用大量存储空间, 而使用数据库技术中的编码技术、表格技术就会节约大

量的存储空间。三则数据库可以从最小的数据单位解决共享“写”问题。对于一个文件来说,可以有很多人可以同时读,但是只允许一个人对这个文件进行写操作,多个人对一个文件同时写的时候就会发生共享冲突。数据库则能够通过表锁、行锁来解决同时写入的冲突问题,从而实现更多的人共享数据。四则,最为重要的是,数据库高效灵活的存取技术,如索引、视图等数据库独特的技术可以保证数据存取高效灵活。

数据库技术是信息处理系统的一项核心技术,是一种计算机辅助管理数据的方法,它研究如何组织和存储数据,如何高效地获取和处理数据。简单地说,数据库技术是通过研究数据库的结构、存储、设计、管理以及应用的基本理论和实现方法,并利用这些理论来实现对数据库中的数据进行处理、分析和理解的技术,即数据库技术是研究、管理和应用数据库的一门软件科学。数据库技术由两部分组成,数据库和数据库管理系统。数据库就是存放数据的仓库,数据库管理系统就是对存放在数据库中的数据如何存取的一系列应用程序。

数据库技术在各个领域的应用非常广泛,对于政府信息化建设来说,数据库技术是重要的技术支撑。20世纪90年代,电子政务建设取得的重大成就之一就是建设了四个重要的数据库,即人口信息库、企业法人库、宏观经济库、空间地理库和自然资源库。

20世纪80年代,随着数据库应用的深入和计算机网络的普及,人们可以做到在上百个用户同时使用的系统中进行实时事务处理,数据库的作用不仅仅在于存储数据,通常还会为用户检索数据提供便利的工具。一个简单的数据库可以只有本地计算机的单个用户使用,而一个复杂数据库则可以有几千用户同时使用,数据库中的数据也可能分布在多台计算机和多个硬盘上,它可以表现为某个电子表格里的一份地址清单,也可以是电信公司中每天数百万次的电话接听情况、单次通话收费、月结话费账单等日志文件。各行各业这样的数据越积越多,于是人类进入大数据时代。

自2012年以来,随着大数据在互联网和信息行业的发展成为新的热点,并在医疗卫生、商业智能、智慧政府、安防监控等领域得到了广泛应用,大数据背后产生了巨大的社会价值和产业空间,全世界商业巨贾将目光投入到

大数据的应用之中。

大数据的价值为大家所公认，但大数据的定义并没有学界的共识，一般而言，业界常常用4V来概括大数据的四个典型特征。第一，Volume（数据体量巨大）。数据集合已由最初的MB、GB数量级，进入TB级别，一些数据库已经跃入PB乃至ZB^①级别行列。其容量和规模远远超出了对于传统数据的认识。第二，形式多样（Variety）。传统数据或者数据库一般具有相同的结构性，或者由结构性相同的表连接成为关系数据库。但是大数据具有不同的结构、不同的形式、不同的来源、没有进行规划的设计，这样的数据对于数据存储，快速地查询提出了更高的挑战。第三，产生高速（Velocity）。数据生成的速度基本呈指数增长，数据流在高速生产时，需要快速地、持续地的实时分析与处理。比如当你使用信用卡刷卡消费的时候，刷卡交易的信息和购物情况已经进入了银行和商场的数据库；当你拿起手机拨打电话的时候，话单的信息已经进入了电信运营商的数据库；当你在医院挂号就医的时候，门诊和处方的信息已经进入了医院的数据库；当你使用公交卡乘车的时候，购买车票的信息也已经进入了公交公司的数据库。只要有人的行动，就会伴随数据的产生，每时每刻都有大量新的信息在产生，如同上下班高峰期的人流，拥挤而又嘈杂。第四，极具价值型（Value）。越多的数据越具有价值，这是毫无疑问的。大数据最重要的不是“大”和“数据”，而是在于大数据背后隐藏的价值。因此，大数据绝非仅仅停留在海量数据汇聚的层面，对大数据进行汇聚和挖掘才是大数据应用的终极意义。李克强总理曾说“数据深藏闺中是极大的浪费。”但大数据的价值一个典型的特点是其价值隐藏。其价值隐藏在众多的数据背后，价值的密度与数据总量的大小成反比，应该说，大数据价值密度比较低。

密度比较低，那就不容易发现，因此数据挖掘就成为大数据时代重要的发现价值的技术手段。数据挖掘（DM，Data Mining）已渐渐成为管理科学与工程专业的课程内容，DM的意思就是从大量数据中，抽取潜在

① 数据的衡量单位。1Byte = 8 bit；1 KB = 1,024 Bytes；1 MB = 1,024 KB；1 GB = 1,024 MB；1 TB = 1,024 GB；1 PB = 1,024 TB；1 EB = 1,024 PB；1 ZB = 1,024 EB。为了更好地理解数据的大小，目前流行的U盘的容量为16GB或者32GB。

的、有价值的知识(模型或规则)的过程。DM 包含的内容丰富,如我们用统计方法研究相关关系,研究学生的数学成绩和物理成绩有没有相关性,研究人体的脂肪含量与人的寿命之间的关系。这些简单的知识发现,都可以称之为 DM。当然 DM 不仅包含这些,它更多倾向研究多元变量之间的关系,包含分类、估值、聚集、关联等内容。

人类社会进入信息时代已经几十年,我们拥有了大量丰富的数字信息,并已经存入大型甚至是超大型的数据仓库中。《纽约时报》由上世纪 60 年代的 10~20 版扩张至现在的 100~200 版,最高曾达 1572 版;《北京青年报》也已是 16~40 版;市场营销报已达 100 版。然而在现实社会中,人均日阅读时间通常为 30~45 分钟,只能浏览一份 24 版的报纸。如果我国已经实现了人口信息化,那么全国 13 亿人口就是一个超大型的数据库;银行系统已经建设了庞大的银行用户以及消费信息的超大型数据库;税务系统业已建设了全国的纳税人和税票数据库;电信部门则拥有庞大的手机用户群和用户通话记录。大量信息在给人们带来方便的同时也带来了一大堆问题:一是信息过量;二是信息真假难以辨识;三是信息安全难以保证;四是信息形式不一致,难以统一处理;五是信息很多,但知识很少,规律更少。于是人们开始提出一个新的口号:“要学会抛弃信息”。人们开始考虑:“如何才能不被信息淹没,而是从中及时发现有用的知识、提高信息利用率?”基于此,计算机专家们发展了数据挖掘技术。

数据挖掘(data mining)是指从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的但又是潜在有价值的信息和知识的过程。随着数据库技术的迅速发展以及数据库管理系统的广泛应用,人们积累的数据越来越多,激增的数据背后隐藏着许多重要信息,人们希望能够对其进行更高层次的分析,以便更好地利用这些数据。目前的数据库系统可以高效地实现数据的录入、查询、统计等功能,但无法发现数据中存在的关系和规则,无法根据现有的数据预测未来的发展趋势。由于缺乏挖掘数据背后隐藏的知识的手段,便导致了“数据爆炸但知识贫乏”的现象。

数据挖掘并不像定义的那样那么难以理解,简单地说就是从大量的数据中发现有价值的信息。最著名的例子是沃尔玛通过数据挖掘发现,啤酒的

销售与尿布的销售存在一定程度的相关性。一般看来,啤酒和尿布是属于不同顾客群的完全不相关的商品。但是沃尔玛通过收集到的一年的销售数据进行挖掘后发现:在某些居民区连锁店中尿布卖得好的店面啤酒也卖得很好。原因其实很简单,一般太太让先生下楼买尿布的时候,先生们一般都会犒劳自己两罐啤酒,因此啤酒和尿布在一起购买的机会是很多的。这是数据挖掘技术发现的众多秘密中的一个。要预测一只股票的走势几乎是不可能的,但是通过数据挖掘后,人们可以找出一只股票的走势与另一只股票走势的内在规律,如一数据挖掘专家得到过这个结论:“如果微软的股票下跌4%,那么IBM的股票将在两周内下跌5%”。

在一个信息化的社会,可以肯定的是我们每个人都曾被挖掘过,如某个月发短信比较多,电信部门就会自动发一个短信过来,“如果本月再发100条,将会得到U盘一个”,我们发短信的记录全部存在电信商的数据库中,此种情况,就认为你是一个短信客户,电信部门会定期使用数据挖掘软件对众多的用户进行挖掘以发现他们需要的信息。

毫无疑问,由于技术的两面性,大数据产生之日,既是大数据价值利用之时,也是饱受大数据侵扰之时,新的安全威胁存在于大数据时代“潘多拉魔盒”中的魔鬼可能会随时出现。截至目前,可以毫无疑问地说,我们每个人都曾被挖掘过,聊天记录、上网偏好、搜索记录、打字习惯、消费记录,已经成为大数据的一部分,我们会被针对性地推送广告、商品。这些并不足惧,令人担忧的是这些偏好数据、隐私数据如何更好地得到保护,如何不被别有用心的人或企业不恰当地利用。

为实现对大数据效率、公平和法制的治理,人类社会仍然需要抖擞精神,凝聚智慧,接受大数据来临的治理挑战。

挑战一:大数据的巨大体量使得信息管理成本显著增加。4个“V”中的第一个“V”(Volume),描述了大数据之大,这些巨大、海量数据的管理问题是对每一个大数据运营者的最大挑战。在网络空间,大数据是更容易被“发现”的显著目标,大数据成为网络攻击的第一选择。一方面,大量数据的集中存储增加了泄露风险,黑客的一次成功攻击能获得比以往更多的数据量,无形中降低了黑客的进攻成本,增加了“攻击收益”;另一方面,大

数据意味着海量数据的汇集，这里面蕴藏着更复杂、更敏感、价值巨大的数据，这些数据会引来更多的潜在攻击者。

挑战二：大数据的繁多类型使得信息有效性验证工作大大增加。4个“V”中的第二个“V”(Variety)，描述了数据类型之多。大数据时代，由于不再拘泥于特定的数据收集模式，使得数据来自于多维空间，各种非结构化的数据与结构化的数据混杂在一起。未来面临的挑战将会是从数据中提取需要的数据，很多组织将不得不接受的现实是，太多无用的信息造成的信息不足或信息不匹配。我们可以考虑这样的逻辑：依托于大数据进行算法处理得出预测，但是如果这些收集上来的数据本身有问题又该如何呢？也许大数据的数据规模可以使得我们无视一些偶然非人为的错误，但是如果有个敌手故意放出干扰数据呢？现在非常需要研究相关的算法来确保数据来源的有效性，尤其是比较强调数据有效性的大数据领域。正是因为这个原因，对于正在收集和储存大量客户数据的公司来说，最显而易见的威胁就是在过去的几年里，存放于企业数据库中数以TB计，不断增加的客户数据是否真实可靠，依然有效。众所周知，海量数据本身就蕴藏着价值，但是如何将有用的数据与没有价值的数据进行区分看起来是一个棘手的问题，甚至引发越来越多的安全问题。

挑战三：大数据的低密度价值分布使得安全防御边界有所扩展。4个“V”中的第三个“V”(Value)，描述了大数据单位数据的低价值。这种广种薄收似的价值量度，使得信息效能被摊薄了，大数据的安全预防与攻击事件的分析过程更加复杂，相当于安全管理范围被放大了。大数据时代的安全与传统信息安全相比，变得更加复杂，具体体现在三个方面：一方面，大量的数据汇集，包括大量的企业运营数据、客户信息、个人的隐私和各种行为的细节记录，这些数据的集中存储增加了数据泄露风险；另一方面，因为一些敏感数据的所有权和使用权并没有被明确界定，很多基于大数据的分析都未考虑到其中涉及的个体隐私问题。另外，大数据对数据完整性、可用性和秘密性带来挑战，在防止数据丢失、被盗取、被滥用和被破坏上存在一定的技术难度，传统的安全工具不再像以前那么有用。

挑战四：大数据的快速处理要求使得独立决策的比例显著降低。4个