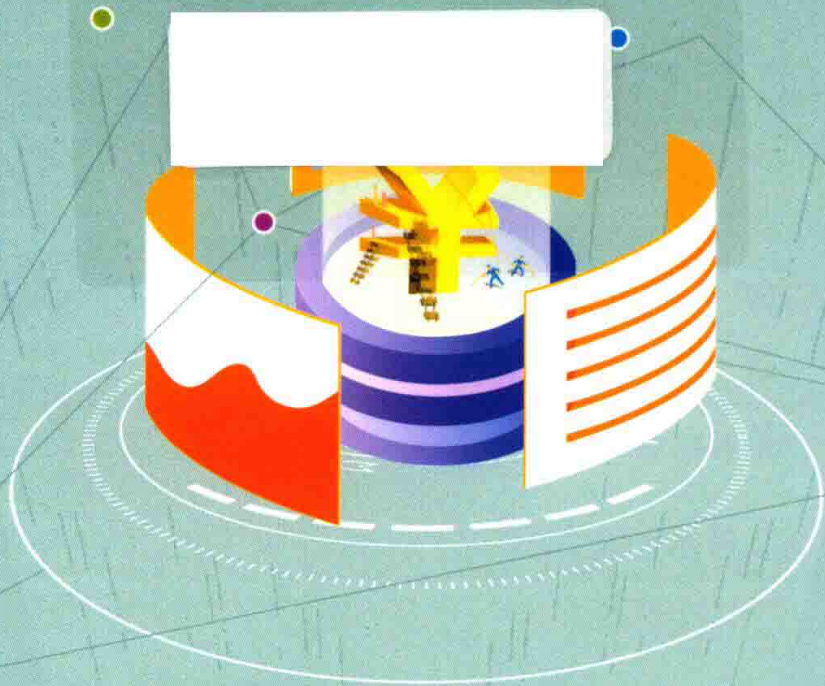


融 | 媒 | 体 | 前 | 沿 | 研 | 究 | 丛 | 书

大数据时代

高校财务数据分析与风险 防控之路

周星秀 连长嵩 潘 苗 著



中国传媒大学出版社

融 | 媒 | 体 | 前 | 沿 | 研 | 究 | 丛 | 书

大数据时代

高校财务数据分析与风险 防控之路

周星秀 连长嵩 潘 苗 著

中国传媒大学出版社
· 北京 ·

图书在版编目(CIP)数据

大数据时代:高校财务数据分析与风险防控之路 /周星秀,连长嵩,潘苗著.
-- 北京:中国传媒大学出版社,2019.12
ISBN 978-7-5657-2656-9

I. ①大… II. ①周… ②连… ③潘… III. ①高等学校—财务管理—
风险管理—研究—中国 IV. ①G647.5

中国版本图书馆 CIP 数据核字(2019)第 290548 号

大数据时代:高校财务数据分析与风险防控之路

DASHUJU SHIDAI:GAOXIAO CAIWU SHUJU FENXI YU FENGXIAN FANGKONG ZHI LU

著 者 周星秀 连长嵩 潘 苗

责任编辑 欧丽娜

封面设计  上古鸟设计

责任印制 李志鹏

出版发行 中国传媒大学出版社

社 址 北京市朝阳区定福庄东街1号 邮编:100024

电 话 86-10-65450528 65450532 传真:65779405

网 址 <http://cucp.cuc.edu.cn>

经 销 全国新华书店

印 刷 北京玺诚印务有限公司

开 本 710mm×1000mm 1/16

印 张 11

字 数 163千字

版 次 2019年12月第1版

印 次 2019年12月第1次印刷

书 号 ISBN 978-7-5657-2656-9/G·2656 定 价 58.00元

版权所有

翻印必究

印装错误

负责调换



序言

AI 取代会计?

当你推门走进一家线下实体银行网点,标准的语音向你问好:“你好,请问需要什么帮助?”就像很多年前,你通过几行简单的编程使得“hello, world!”滚动在电脑屏幕上。一个新的时代正在悄然进入你我之间。

2016年,德勤与 Kira Systems 联手,正式拉开了将人工智能引入财务工作的序幕。四大会计事务所的智能机器人中心均早早与多家企事业单位建立合作,提供财务自动化流程解决方案。

机器人投入到财务具体岗位之中,能够进行全流程的基础处理:比如去做发票的处理,开账单,记账和收款等工作。机器人为财务工作带来了效率的提升,帮助财务人员完成大量重复规则化的工作。其实,这里的机器人取代的是还仅仅停留在会计工作中那些重复性、缺乏创造性的程式化工作,属于用“自动化”取代了会计的传统作业。

如此“自动化”的初级应用足以敲响警钟,未来财务领域,不只是优秀的人和你我赛跑,更有一批不需要睡觉还会自动学习的真正 AI 机器人和我赛跑。真正的人工智能,还具备音像识别、计算机自然语言处理、推荐系统、专家系统等功能,能够做到及时处理票据、缩短信息传递时间、保障容错率、快速准确汇总数据、减少做假账的漏洞空间等。未来人工智能还可能进行更高一级的财务数据处理和决策。相对于人工而言,它是节约单位主体人力成本和时间资源的全时域虚拟劳动力。



我们要何以应对 AI? 唯有直面。新时代必造新英雄。财务人员要加强自身数字素养能力,转变思考模式,树立终身学习意识,终究人类与机器谁是最强大脑还很神秘。

回到最初的问题, AI 取代会计真正的危机,并非会计这个行业因人工智能技术而消失,真正的危机,是传统管理与作业模式受到冲击之下,那些仍固守单一思维不接受新数据技术的人被替代。

毁灭永远是创造的必经之路,请不要做灰飞烟灭的樯櫓,要当那个谈笑间利用新技术创造带来新一轮灰飞烟灭的舵手。



目 录

M U L U

| | |
|--------------------------------|------|
| 第一章 大数据正在改变世界 | / 1 |
| 第一节 大数据的发展历程 | / 1 |
| 第二节 大数据的技术框架 | / 12 |
| 第三节 大数据的应用分析 | / 16 |
| 第二章 大数据赋能高校财务发展 | / 21 |
| 第一节 高校财务数据 | / 22 |
| 第二节 大数据为高校财务带来机遇和挑战 | / 26 |
| 第三章 大数据时代高校财务管理困局 | / 29 |
| 第一节 高校财务管理的现状 | / 29 |
| 第二节 高校财务管理的困局 | / 32 |
| 第三节 高校财务管理的方向 | / 35 |
| 第四章 高校财务大数据平台建设 | / 40 |
| 第一节 平台设计目标 | / 40 |
| 第二节 平台功能 | / 43 |
| 第三节 平台框架设计 | / 46 |



| | |
|---------------------------------|-------|
| 第五章 应用子系统设计 | / 55 |
| 第一节 财务管理系统 | / 55 |
| 第二节 资产管理系统 | / 67 |
| 第三节 合同管理系统 | / 76 |
| 第四节 基建管理系统 | / 85 |
| 第五节 其他管理系统 | / 93 |
| | |
| 第六章 高校大数据平台财务数据链 | / 100 |
| 第一节 高校财务数据的来源 | / 100 |
| 第二节 高校财务数据的传递 | / 102 |
| 第三节 高校财务数据的存储 | / 112 |
| 第四节 高校财务数据的挖掘 | / 115 |
| 第五节 高校财务数据的安全 | / 118 |
| 第六节 实例分析 | / 124 |
| | |
| 第七章 高校大数据平台应用:风险防控 | / 128 |
| 第一节 高校风险防控 | / 129 |
| 第二节 高校风险防控框架模型 | / 133 |
| 第三节 高校风险防控的财务指标 | / 136 |
| 第四节 高校风险防控的数据分析策略 | / 141 |
| 第五节 基于数据分析的风险防控管理决策 | / 144 |
| | |
| 结语 未来:拥抱数据的高校财务 | / 148 |
| | |
| 附录 | / 150 |
| | |
| 参考文献 | / 166 |



第一章 大数据正在改变世界

第一节 大数据的发展历程

作为最早提出“大数据”时代到来的预言者,全球知名咨询公司麦肯锡称:“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。”如今,大数据的广泛应用越来越彰显其优势,其占领的领域也越来越大,电子商务、O2O、物流配送、财务管理等,各种利用大数据的领域正在协助企事业单位不断地发展新业务,创新运营模式。

一 大数据的概念界定

业界对“大数据”的经典定义可以被归纳为“4个V”(简称“多快好省”):

(一)海量的数据规模(Volume)

指代大型数据集,一般在10TB规模左右。但在实际应用中,很多企业用户把多个数据集放在一起,形成了PB级的数据量,并且这样的数据集很难再用传统数据库工具对其内容进行抓取、管理和处理。



(二)多样的数据类型(Variety)

数据类别多种多样。数据来自多种数据源,数据种类和格式日渐丰富,已冲破了以前所限定的结构化数据范畴,囊括了半结构化和非结构化数据。

(三)快速的数据流转和动态的数据体系(Velocity)

数据处理速度快。在数据量非常庞大的情况下,也能够做到数据的实时处理,从各类型的数据中快速获得高价值的信息。

(四)巨大的数据价值(Value)

大数据的价值性高。随着社交数据、企业内容、交易与应用数据等新数据源的兴起,传统数据源的局限被打破,企业越发需要有效的信息资源以确保其价值性。由于数据产生量巨大且速度非常快,必然形成各种有效数据和无效数据错杂的状态,因此数据价值的密度低。

以上是最被业内人士接纳的大数据的概念特点,也是本书后续进行数据平台搭建、数据分析挖掘的基石。

此外,百度将大数据的概念界定为巨量资料,指的是大数据所涉及的资料量规模巨大,以至于无法通过目前的主流软件工具,在合理时间内达到摄取、管理、处理并整理成为帮助企业经营决策的资讯。

《互联网周刊》认为,大数据的概念远不止大量的数据和处理大量数据的技术,或者所谓的“4个V”之类的简单概念,而是涵盖了人们在大规模数据的基础上可以做的任何事情,而这些事情在小规模数据的基础上是无法实现的。换句话说,大数据让我们以一种前所未有的方式,通过对海量数据进行分析,获得有巨大价值的产品和服务,或深刻的洞见,最终形成变革之力。

高德纳咨询公司提出,大数据是依靠新的处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长率和多样化的信息资产。



从数据的类别上看,大数据指的是无法使用传统流程或工具处理或分析的信息,其定义是指那些超出正常处理范围和大小、迫使用户采用非传统处理方法的数据集。

尽管对“大数据”的概念界定说法不一,但基本能达成的共识是,大数据是由数量巨大、结构复杂、类型众多的数据结构构成的数据集合,在合理时间内可以通过对该数据集合的管理与处理,最终使其成为能帮助企事业单位管理与决策的有用信息。它的数量如此巨大,以至于分析它需要多个工作负载。当现有的技术达到极限时,也就是数据的极限。面对大数据浪潮的侵袭,重要的不是如何定义“大数据”,而是如何使用“大数据”,最大的挑战在于,哪些技术能更好地使用和挖掘数据以及大数据的应用情况如何。

二 大数据的发展历程

由于大数据技术的特点和重要性,目前国内外已经出现了“数据科学”的概念,即数据处理技术将成为一个与计算机科学并列的新的科学领域。关于大数据的发展历程,主要可分为以下几个阶段^①:

(一)大数据萌芽阶段(1980—2005年)

1980年,著名未来学家阿尔文·托夫勒的《第三次浪潮》一书将“大数据”称为“第三次浪潮的华彩乐章”。1997年,美国宇航局研究员迈克尔·考克斯和大卫·埃尔斯沃斯首次使用“大数据”这一术语来描述不能被处理和不可可视化的模拟飞机周围的气流,将“大数据问题”归结为数据集大到超出了主存储器、本地磁盘,甚至远程磁盘的承载能力。经过“9·11”袭击后,美国政府自2002年起已经涉足大数据挖掘以阻止恐怖主义,其组建了一个用

^① 开放大数据需上升到国家战略层面[J]. IT时代周刊,2015(5).



于筛选通信、犯罪、教育、金融、医疗和旅行等记录来识别可疑人的大数据库。2004年,美国“9·11”委员会呼吁反恐机构组建“一个基于网络的信息共享系统”,以便能更快速地处理应接不暇的数据。

(二)大数据兴起阶段(2006—2011年)

“大数据”在云计算出现之后才凸显其真正价值。谷歌在2006年首先提出云计算的概念。2007—2008年,随着社交网络的激增,技术博客和专业人士为“大数据”概念注入新的生机,一些顶尖计算机科学家声称,“当前世界范围内已有的一些其他工具将被大量数据和应用算法所取代”,“应该深入参与大数据计算的开发和部署工作,因为它将直接有利于许多任务的实现”。2008年9月,《自然》杂志推出了名为“大数据”的封面专栏。2009—2010年,“大数据”成为互联网技术行业中的热门词汇。2009年,印度建立了用于身份识别管理的生物识别数据库;同年,联合国全球脉冲项目研究了如何利用手机和社交网站的数据源来分析预测从螺旋价格到疾病爆发之类的问题;美国政府通过启动Data.gov网站的方式进一步开放了数据的大门,该网站的超过4.45万量数据集被用于保证一些网站和智能手机应用程序跟踪信息,这一行动激发了从肯尼亚到英国范围内的政府们相继推出类似举措;欧洲一些领先的研究型图书馆和科技信息研究机构建立了伙伴关系,致力于改善在互联网上获取科学数据的简易性。2010年,肯尼斯库克尔发表了大数据专题报告《数据,无所不在的数据》。2011年6月,麦肯锡发布了关于“大数据”的报告,正式定义了“大数据”的概念。随后,这一概念逐渐受到了各行各业的关注,标志着“大数据”时代已经到来。2011年12月,在工信部发布的物联网十二五规划上,信息处理技术作为4项关键技术创新工程之一被提出来,其中包括海量数据存储、数据挖掘、图像视频智能分析,这些都是大数据的重要组成部分。



(三) 大数据发展阶段(2012—2016年)

2012年,“大数据”一词越来越多地被提及,人们用它来描述和定义信息爆炸时代产生的海量数据,并命名与之相关的技术发展与创新。数据正在迅速膨胀并变大,它决定着未来的发展,带来了思维变革、商业变革和管理变革。伴随着互联网的浪潮,“大数据”在各行各业中扮演了举足轻重的角色。2012年1月,在瑞士达沃斯召开的世界经济论坛上,大数据是主题之一,会上发布的报告《大数据,大影响》宣称,数据已经成为一种新的经济资产类别。2012年,美国奥巴马政府在白宫网站发布了《大数据研究和发展倡议》,这一倡议标志着大数据已经成为重要的时代特征。同年3月22日,奥巴马政府宣布将2亿美元投资大数据领域,这成为大数据技术从商业行为上升到国家科技战略的分水岭。

2012年,美国颁布了《大数据的研究和发展计划》;英国发布了《英国数据能力发展战略规划》;日本发布了《创建最尖端IT国家宣言》;韩国提出了“大数据中心战略”;其他国家也相继制定了相应的战略和规划。2012年7月,联合国在纽约发布了一份关于大数据政务的白皮书《大数据促发展,挑战与机遇》,总结了各国政府如何利用大数据更好地服务和保护人民。2013年,英国政府宣布注资6亿英镑发展8类高新技术,其中,1.89亿英镑用来发展大数据技术,旨在开放欧盟公共管理部门的所有信息。大数据掀起的变革,正在对现有的生产力和生产关系构成重要影响。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)》;同年5月,美国白宫发布了2014年全球“大数据”白皮书的研究报告《大数据:抓住机遇、守护价值》。截至2014年,数据开放运动已覆盖全球44个国家。

我国大数据业务相对起步较晚,在2013年全球开放数据的70个国家和地区中,中国仅位列第35位。阿里巴巴是最早提出通过数据进行企业数据化运营的企业,从2013年1月1日开始重塑平台、金融和数据三大业务。



2014年3月，贵州在北京宣布大数据产业起航；同年，“大数据”首次出现在当年的《政府工作报告》中。报告指出，要设立新兴产业创业创新平台，在大数据等方面赶超先进，引领未来产业发展。2014年，国务院通过了《企业信息公示暂行条例(草案)》，要求在企业部门间建立互联共享信息平台，运用大数据等手段提升监管水平，“大数据”成为国内热议词汇。

2015年2月，工信部将贵阳正式列为全国唯一的大数据产业发展试点示范区。同年4月，国内首个大数据交易所在贵阳挂牌成立。2015年，大数据上升到国家战略层面，我国政府于当年8月通过了《关于促进大数据发展的行动纲要》。2015年10月26日至29日，党的十八届五中全会召开，开启了大数据建设的新篇章。大会公开提出要实施“国家大数据战略”，这是大数据第一次被写入党的全会决议，标志着大数据战略正式上升为国家战略。2016年12月18日，工业和信息化部《大数据产业发展规划(2016—2020年)》正式印发。《中国大数据发展调查报告(2017年)》称，2016年中国大数据市场规模为168.0亿元，增速达到45%，预计2017—2020年增速保持在30%以上。

(四)大数据爆发阶段(2017年至今)

2017年，大数据已经渗透到人们生活的方方面面，我国大数据产业的发展也进入爆发期，在政策、法规、技术、应用等多重因素的推动下，基本形成了跨部门数据共享共用的格局。京、津、沪、渝、冀、辽、贵、晋等省市政府相继出台了大数据研究与发展行动计划，整合数据资源，实现区域数据中心资源汇集与集中建设。至今，全国至少已有13个省成立了21家大数据管理机构，有35所本科院校获批开设“数据科学与大数据技术”本科专业，62所专科院校开设“大数据技术与应用”专科专业，申报数据科学与大数据技术本科专业的学校达到293所。

2017年11月22日，贵阳市委、市政府出台了《关于加快建成“中国数谷”的实施意见》。预计到2020年，“中国数谷”将形成1个EB以上的海量



数据存储能力,初步完成5个国家级试点示范任务,培育10个以上具有核心竞争力和商业影响力的大数据品牌,推出100个以上大数据应用领域(场景),形成1 000亿元以上主营业务收入,聚集10 000家以上大数据市场主体。贵阳将成为全国大数据创新策源地。同年12月8日,中共中央政治局就实施国家大数据战略进行第二次集体学习,中共中央总书记习近平在主持学习时强调,实施国家大数据战略,加快建设数字中国。《中国大数据发展调查报告(2018年)》指出,2017年中国大数据产业总体规模为4 700亿元人民币,同比增长30%;大数据核心产业规模为236亿元人民币,增速达到40.5%,预计2018—2020年增速将保持在30%以上。

2018年,达沃斯世界经济论坛等全球性重要会议都把“大数据”作为重要议题,进行讨论和展望。在2018年拉斯维加斯消费电子展(CES)上,美国消费技术协会总裁兼首席执行官加里·夏皮罗、英特尔首席执行官布莱恩·克尔扎尼奇等都表示,大数据将对人类生活产生深远影响,大数据是未来科技浪潮发展不容忽视的巨大推动力量,许多国家政府都对大数据产业发展抱有高度的热情。2018年,美国希望利用大数据技术实现在多个领域的突破,包括科研教学、环境保护、工程技术、国土安全、生物医药等。其中,具体的研发计划涉及了美国国家科学基金会、国家卫生研究院、国防部、能源部、国防部高级研究局、地质勘探局等6个联邦部门和机构。2018年,欧盟在大数据方面的活动主要涉及四方面内容:研究数据价值链战略因素;资助“大数据”和“开放数据”领域的研究和创新活动;实施开放数据政策;促进公共资助科研实验成果和数据的使用及再利用。

大数据发展浪潮席卷全球。全球各经济社会系统采集、处理、积累的数据增长迅猛,大数据全产业链市场规模逐步提升。2018年,大数据产业化发展呈现出以下七大趋势。第一,开源大数据商业化进一步深化。第二,打包的大数据行业分析应用开拓新市场。第三,大数据细分市场规模进一步增大。第四,大数据推动公司并购的规模和数量进一步提升。第五,大数据分析的革命性方法出现。第六,大数据与云计算将深度融合。第七,大数据一体机



将陆续发布^①。

2017年,全球的数据总量为21.6ZB(1个ZB等于十万亿亿字节)。目前,全球数据的增长速度在每年40%左右。日本著名的矢野经济研究所预测,2020年,日本大数据市场规模有望超过1兆日元。全球大数据市场规模年均实现15.37%的增长,预计到2020年,全球的数据总量将达到40ZB,预计到2022年,全球大数据市场规模将达到800亿美元。

三 大数据的特征分析

大数据可以实时地为组织主体摄取、管理、处理、整理数据,生成为各方所需要的数据资料,同时大数据也蕴含着很高的商业价值,被称为“数字生产力”。所以越来越多的组织主体开始重视大数据建设^②。

(一)大数据 = 海量数据 + 复杂类型的数据

大数据包含了海量数据的含义,而且在内容上超越了海量数据,简而言之,大数据=海量数据+复杂类型的数据。大数据包括交易和交互数据集在内的所有数据集,其规模或复杂程度超出了常用技术按照合理的成本和时限捕捉、管理及处理这些数据集的能力。大数据由三项主要技术趋势汇聚组成:

1. 海量交易数据

在从ERP应用程序到数据仓库应用程序的在线交易处理(OLTP)与分析系统中,传统的关系数据以及非结构化和半结构化信息仍在继续增长。随着企业将更多的数据和业务流程移向公共和私有云,这一局面将变得更加复杂。

① 数据来源:《2018全球大数据产业将呈七大发展趋势》,新华网客户端,2018年2月5日。

② 涂兰敬. 专家观点:“大数据”与“海量数据”的区别[J],网络与信息,2011(12).



2. 海量交互数据

这一新生力量由源于 Facebook、Twitter、LinkedIn 及其他来源的社交媒体数据构成。它包括呼叫详细记录(CDR)、设备和传感器信息、GPS 和地理定位映射数据、通过管理文件传输(Manage File Transfer)协议传送的海量图像文件、Web 文本和点击流数据、科学信息、电子邮件等等。

3. 海量数据处理

大数据的涌现已经催生出了设计用于数据密集型处理的架构,例如具有开放源码、在商品硬件群中运行的 ApacheHadoop。对于企业来说,难题在于以具备成本效益的方式快速可靠地从 Hadoop 中存取数据。

(二)大数据包括 A、B、C 三个要素

大数据意味着通过更快获取信息以使做事情的方式变得与众不同,并因此实现突破。大数据被定义为大量数据(通常是非结构化的),它要求我们重新思考如何存储、管理和恢复数据。那么,多大才算大呢?考虑这个问题的一种方式就是,它是如此之大,以至于我们今天所使用的任何工具都无法处理它。因此,如何消化数据并把它转化成有价值的洞见和信息,这其中的关键就是转变。基于从客户那里了解的工作负载要求,大数据包括 A、B、C 三个要素:大分析(Big Analytic)、高带宽(Big Bandwidth)和大内容(Big Content)。

1. 大分析

帮助获得洞见,指的是对巨大数据集进行实时分析的要求。它能带来新的业务模式,提供更好的客户服务,并实现更好的结果。

2. 高带宽

帮助走得更快,指的是处理极端高速的关键数据的要求。它支持快速有效地消化和处理大型数据集。



3. 大内容

不丢失任何信息,指的是对于安全性要求极高的高可扩展的数据存储,并能够轻松实现恢复。它支持可管理的信息内容存储库而不只是存放过久的数据,并且能够跨越不同的大陆板块。

大数据是一股突破性的经济和技术力量,它为 IT 支持引入了新的基础架构。大数据解决方案消除了传统的计算和存储的局限。借助于不断增长的私密和公开数据,一种划时代的新商业模式正在兴起,它有望为大数据客户带来新的实质性的收入增长点以及富于竞争力的优势。

(三) 高速性与易变性

1. 高速性

高速性描述的是数据被创建和移动的速度。在高速网络时代,通过基于实现软件性能优化的高速电脑处理器和服务器,创建实时数据流已成为流行趋势。企业不仅需要了解如何快速创建数据,还必须知道如何快速处理、分析并返回给用户,以满足他们的实时需求。

2. 易变性

大数据具有多层结构,这意味着大数据会呈现出多变的形式和类型。相较传统的业务数据,大数据存在不规则和模糊不清的特性,造成很难甚至无法使用传统的应用软件进行分析的弊端。传统业务数据随时间演变已拥有标准的格式,能够被标准的商务智能软件识别。目前,企业面临的挑战是处理并从各种形式呈现的复杂数据中挖掘价值。

(四) 多样性与计算速度快

数据多样性的增加主要是由于新型多结构数据的出现,以及包括网络日志、社交媒体、互联网搜索、手机通话记录及传感器网络等数据类型所造