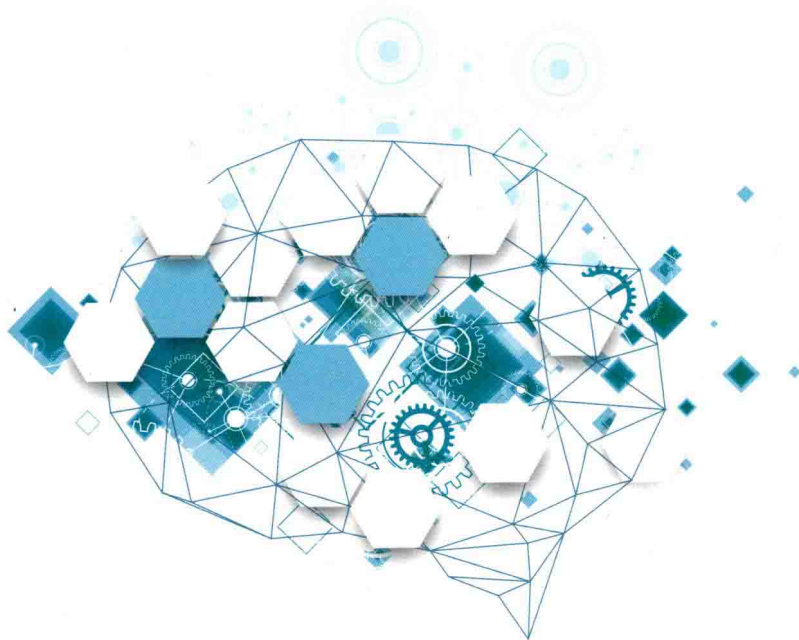




大数据征信背景下的 信息质量度量与提升研究

Information Quality Measurement and Improvement
in the Background of Credit Big Data

宋 媚 著



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

国家自然科学基金项目（71503108）和中国博士后科学基金特别资助项目
（2015T80165）研究专著

大数据征信背景下的 信息质量度量与提升研究

Information Quality Measurement and Improvement
in the Background of Credit Big Data

宋 媚 著



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书主要关注大数据背景下的信息质量度量方法,并有针对性地提出了信息质量管理提升策略。本书具体包括:大数据征信背景下的信息质量评价与管控;征信信息质量度量模型的构建;征信信息质量度量模型应用;征信系统的信息质量管理与价值提升策略分析。

图书在版编目(CIP)数据

大数据征信背景下的信息质量度量与提升研究 / 宋媚著.

—上海:上海交通大学出版社,2016

ISBN 978-7-313-16211-3

I. ①大… II. ①宋… III. ①信息管理-质量管理-研究

IV. ①G203

中国版本图书馆 CIP 数据核字(2016) 第 280915 号

大数据征信背景下的信息质量度量与提升研究

著 者: 宋 媚	地 址: 上海市番禺路 951 号
出版发行: 上海交通大学出版社	电 话: 021-64071208
邮政编码: 200030	
出 版 人: 郑益慧	
印 刷: 上海天地海设计印刷有限公司	经 销: 全国新华书店
开 本: 710mm×1000mm 1/16	印 张: 10.25
字 数: 167 千字	
版 次: 2016 年 12 月第 1 版	印 次: 2016 年 12 月第 1 次印刷
书 号: ISBN 978-7-313-16211-3/G	
定 价: 39.00 元	

版权所有 侵权必究
告 读 者: 如发现本书有印装质量问题请与印刷厂质量科联系
联系电话: 021-64366274

前 言

本书是国家自然科学基金项目(71503108)和中国博士后科学基金特别资助项目(2015T80165)的研究专著。本书依托我国央行的企业信用信息基础数据库,就信息质量的度量与提升问题,按照知识获取、模型构建与分析、提升策略分析的顺序逐步展开。首先,结合前期征信系统数据质量管理总结的经验和知识,基于美国麻省理工学院信息质量计划和全面数据质量管理负责人、信息质量国际会议主席 Richard Wang 的信息质量度量框架,构建信息质量度量模型,探求适用于大数据征信信息的质量度量统一指标体系和各信息质量维度的测量方式。其次,结合国内外信息质量管理经验,探索采数据与流程于一体的信息价值提升策略。

本研究实现了如下三个方面的创新:

(1) 提出信用信息应用价值提升的系统综合集成分析视角。

在前人的研究中,或者从宏观经济层面的信息质量管理应对效果上进行理论与实证分析,或者从微观管理层面进行业务流程再造分析,或者仅从技术实现角度着眼于算法本身的改进,缺乏面向应用需求与信息加工规律、融管理和技术为一体的系统性思考。本研究采用知识、信息、数据的倒序质量管理思路,基于 Richard Wang 信息质量度量框架和专家知识提取、前沿质量特征分量度量技术,构建符合大数据征信特征的综合集成分析体系。

(2) 加入相关性维度,提出了面向应用的双重赋权信息质量度量体系。

在前人的研究中,基于“真实世界”剖面,仅从基础数据的准确性、完整性、及时性开展信息质量度量研究。本研究通过引入面向外部报告使用、面向内部产品研发的应用视角,通过加入涉及任务数、使用频次、重要程度的相关性(Relevance)度量,构建了相关性、准确性、完整性、及时性“四性”相结合的双重赋权信息质量度量体系。

(3) 实现基本信息、信贷信息、非信贷信息质量管理上的适应性应用创新。

在前人的研究中,多数面向通用数据库的信息质量度量。本研究除了采用基本的信息质量分析框架外,还依据大数据征信的具体特征进行适应性地改变,如在信贷信息选取关键业务节点开展度量的思路、采用业务完整性代替结构完整性的信贷完整性度量方式等。本研究基于应用导向下的业务需求特征、层次性分析视角,开展信息质量度量方法的灵活应用,与现有的信息质量度量实践有效衔接,并积极梳理、应对目前数据质量治理中存在的问题,实现了信息质量管理上的应用创新。

目 录

第 1 章 绪论	1
1.1 研究背景	1
1.1.1 现实背景	1
1.1.2 理论背景	3
1.2 研究问题	5
1.3 研究意义	5
1.3.1 现实意义	5
1.3.2 理论意义	6
1.4 研究方案	7
1.4.1 研究范围	7
1.4.2 研究内容	11
1.4.3 技术路线	12
1.4.4 研究方法	14
1.5 创新点	15
1.6 本章小结	16
第 2 章 国内外研究综述	17
2.1 征信信息基本概念界定	17
2.1.1 信息的定义	17
2.1.2 征信信息的定义	18
2.1.3 征信信息质量的定义	19
2.2 征信市场中的信息共享文献综述	20

2.2.1	国内外征信市场的信息共享实践综述	20
2.2.2	国内外征信市场的信息共享理论综述	21
2.3	信息质量度量的文献综述	24
2.4	信息价值提升的文献综述	26
2.5	本章小结	27
第3章	大数据征信背景下的信息质量评价与管控	28
3.1	大数据征信信息质量评价与管控	28
3.1.1	量化考评,建立健全征信数据质量评价方法	28
3.1.2	源头控制,集中解决突出的数据质量问题	29
3.2	信息质量考核与管理的国际经验	30
3.3	信息质量提升的国际经验	32
3.4	征信系统信息质量度量与提升的关键点	34
3.4.1	企业征信系统信息共享的分类与汇聚	34
3.4.2	企业征信信息采集现状	35
3.4.3	企业征信系统信息质量度量的关键点	38
3.4.4	企业征信系统信息价值提升的关键点	40
3.5	征信信息质量评价结构分析	41
3.5.1	相关研究综述	42
3.5.2	研究模型的提出	43
3.5.3	研究假设的形成	44
3.5.4	研究设计	45
3.5.5	数据分析与结果	49
3.5.6	讨论	55
3.6	本章小结	56
第4章	征信信息质量度量模型构建	57
4.1	质量度量框架	57
4.1.1	基本理论视角:真实世界	57
4.1.2	基本维度突破:以用户为中心	57
4.1.3	选取可比的报告尺度:分维度测量	58

4.2 征信信息质量度量概述	58
4.2.1 面向企业征信信息的质量度量界定	59
4.2.2 面向大数据征信的信息质量度量体系及其扩展	61
4.3 本章小结	61
第 5 章 征信信息质量度量模型应用	63
5.1 企业征信相关性分析	63
5.1.1 信息相关性的概念	63
5.1.2 征信系统的信息相关性度量	63
5.1.3 企业征信系统信息相关性的测度结果	64
5.1.4 应用分析	67
5.2 基本信息质量分析	67
5.2.1 基本信息的完整性度量	67
5.2.2 基本信息的及时性度量	79
5.2.3 基本信息的准确性度量	84
5.2.4 企业征信基本信息的综合度量	97
5.3 信贷信息质量分析	99
5.3.1 信贷信息完整性度量	99
5.3.2 信贷信息及时性度量	102
5.3.3 信贷信息准确性度量	106
5.3.4 信贷信息各指标综合分析结果	110
5.3.5 企业征信数据质量度量方法的应用分析	111
5.4 非信贷信息质量分析	111
5.4.1 非信贷信用信息的完整性度量	111
5.4.2 非信贷信用信息的及时性度量	116
5.4.3 非信贷信用信息的准确性度量	118
5.4.4 企业非信贷信息的综合度量	125
5.5 本章小结	126

第 6 章 大数据征信背景下的信息质量管理与
价值提升策略分析 127

6.1 基于信息产品清单管理制度的信息质量管理策略 127

6.1.1 编制信息产品目录 127

6.1.2 确定影响突出的关键信息产品 130

6.1.3 创建关键信息产品的数据质量规范 132

6.1.4 记录关键信息产品的生产流程 133

6.1.5 信息产品定期核查 135

6.2 基于信息产品图和统一建模语言的信息价值提升策略分析 135

6.2.1 质量设计模型 136

6.2.2 质量改进计划与监控 136

6.3 本章小结 137

第 7 章 结论与展望 138

7.1 研究结论 138

7.2 下一步工作展望 142

参考文献 143

索引 149

后记 153

第 1 章

绪 论

“人而无信,不知其可也。大车无輗(ní 指大车车辕和横木衔接的活销),小车无軹(yuè 指古代车上置于辕前端与车横木衔接处的销钉),其何以行之哉!”取自《论语·为政》。可见,信用对于我们的的重要性,从古至今亘古不变。然而,如何通过对信用大数据的有效管理,提升决策的应用价值,是现今中国社会的又一重要课题。在如今大数据背景下,每天有不同“消息”“知识”碎片纷至沓来,我们并不缺乏数据,缺乏的是具可证伪性的知识和高品质的信息。因此,本研究致力于在信用大数据的质量度量与提高方面做出有益尝试,通过构建完备的知识框架,并采用实践应用分析作为案例实证,以期在这方面研究上先行先试,抛砖引玉。

本章首先针对大数据征信的信息质量度量与提升的现实与理论背景进行了阐述,提出全书的研究问题及其研究意义;而后从研究范围、研究内容、技术路线、研究方法四个模块来详尽介绍全书的研究方案;最后,提出本书的创新点。

1.1 研究背景

1.1.1 现实背景

服务银行与银行间以及政府和银行间信息共享的、承建国家设立的金融信用信息基础数据库的信息中介平台——中国人民银行征信中心建设维护的征信系统(以下简称央行征信系统),主要分为个人征信和企业征信两大部分,涵盖了基本信息、信贷信息、非信贷信息三大块(Jentzsch, 2008)。其中,基本信息又称为身份信息,用于信息主体身份识别;信贷信息用于了解信息主体借贷发生额、

余额、偿还情况等历史信用交易行为,用于发现贷款逾期、违约的风险;非信贷信息主要包括公共信息、商务信息,通称为非银行信息,目前金融信用信息基础数据库中以公共信息为主,用于了解信息主体在各社会政府部门、企事业单位遗留的“历史轨迹”——因后付费使信用交易触发的费用欠缴、行政备案、定期提交的材料、监管过程中的源头信息(司法判决、行政奖惩信息)等。这些“历史轨迹”可被划分为多种不同的采集目标、数据特征和使用方式。应用 Richard Wang 所提出的“面向应用”的信息质量管理思路,对以上三类信息进行梳理。

(1) 基本信息用于信贷客户的主体身份识别,因此其要求在进行信息主体信贷信息认定时能够配套提供准确及时的身份信息。同时,基本信息的报送是由事件(或有事件)触发的,当发生新设或变更时,需要对应报送新增或变更的信息主体的基本信息;另外,每一信息主体的身份信息形成了一系列的“历史轨迹”,这些“历史轨迹”可被切分为多个不同的时间剖面,每个剖面都是信息主体当时身份状态在“真实世界”的映射,各个时间剖面间变更前后的信息并不包含明确的逻辑联系。

(2) 信贷信息是征信系统的主体信息,用于了解授信主体在授信机构(商业银行、财务公司、小额贷款公司、证券、保险等)的历史信用行为,因此其要求尽可能地真实反映当前信息主体的信用状态;同时,信贷信息的报送是业务状态变更触发的,当发生借据新增、还款、展期、欠息等信用状态变更时,需即时报送信贷类信用信息;另外,每一信息主体的业务状态会自动生成符合一定内部逻辑合理性的业务链条。以企业贷款为例,先签订授信合同(新增合同),而后发生放款(新增借据),之后贷款收回(还款);若到期借据项下能够全部收回则为结清,否则或转为展期(新增展期),或产生逾期、五级分类变更(借据变更),同时若有欠息(新增欠息)。

(3) 非信贷信息较为复杂,既包括提示或有风险,又包括了解信息主体除信贷信息之外的其他信用行为。提示或有风险主要有环保处罚、法院诉讼等,具有一定的不确定性;了解其他信用行为主要有公积金缴费信息、电信缴费信息等,其特征与信贷信息类似,但周期更有规律,多为按月缴纳、按月采集。

其实,每个节点下的信息主体征信构成了主体的单个剖面。每个剖面都是信息主体当前在“真实世界”的镜像。在各个剖面中,信息主体在不同政府部门的记录就像多传感器系统获取的目标量测值,即使花了很大的人力物力进行了搜集,但是若不能对各个“传感”视角的信息质量进行有效维护,不仅单个量测值

(单条记录)的价值相当有限,更有甚者会产生负面影响。例如,通过2014年7—8月间的企业信用报告应用情况调查工作,初步摸底发现,许多授信机构都反馈存有还款未及时录入、担保关系未能及时解除的问题。尽管信息主体通常也可采用异议通道进行权利救济,但在未核实解决前,授信机构为了防范信用风险,多会从审慎角度拒绝给其授信。处理的时滞及潜在损失造成了授信机构审贷工作的障碍和信息主体对于银行授信工作的不满。

目前,业界已建立的征信数据质量评价体系,主要针对信贷信息的关键点,因三类信息各有特点,无法有效涵盖所有征信信息可能产生异议和问题的各个风险点。因而,需要引入新的知识框架,能够有效涵盖三类信息的特征并进行适应性改造,对大数据征信背景下的信息质量开展连续度量与持续监督。通过对其开展全面而有针对性的分析,来减少信息主体的异议,提升面向授信应用的征信服务质量。因此,目前最基本、最重要的是立足实践需求,从征信数字产品的形成基础——信息质量管理出发,探索一套能够有效涵盖三类信息的科学信息质量度量体系与提升策略,提高信息服务与决策支持质量。

1.1.2 理论背景

从一定意义上说,在投入使用之前,数据是没有价值的,只有真正能够达到功能性目的的数据才成为信息。在为了达成目的,如用于回答 Who、What、When 和 How many 等问题时,数据与问题关联起来,解决相应问题,数据完成了从符号向有价值信息的转变。如果说信息是回答 What 等类的问题,而知识就是回答 Why 的问题。知识融合了上下文语境、价值观、经验、规则等,知识是随着时间推移多来源信息的综合体(Rowley, 2007)。因此,在 DIKW(Data-Information-Knowledge-Wisdom)层次论中,认为数据、信息、知识是递进的关系。数据是用于表征物体、事件和环境的符号,可被视为一系列不连贯的事实或观察,可通过分析、交叉引用、筛选、排序、汇总等处理转变为信息(Stonier, 1993; 1997);信息经过思维深加工赋予一定含义或关系释义就形成了知识(Zins, 2007)。传统的信息质量治理工作,也是遵循这种理念开展的。常规情况下,从数据层次出发,从数据中发现问题,发现后采用筛选、汇总等一系列操作归纳出问题点,而后探寻问题点产生的原因,并寻求解决方案,以期积累汇集成质量管理知识。

然而,这种“自下而上”的信息质量管理策略往往效率并不高。最显著的表现,在浩如烟海的数据中找问题,发现问题是简单的,但是解决问题是困难的。

通常,发现乃至归纳问题点需要一到两个月的时间,但是解决一个问题可能需要一年甚至更长的时间。一年过后,解决了目前的主要问题,次要问题又上升为主要问题,然后开始下一周期的解决与应对。这种“解决了主要矛盾,次要矛盾又上升为主要矛盾”的工作方法,是符合常规哲学思维规律的,适用于数据质量管理工作未见规模的初期。伴随数据质量管理工作的受重视与不断推进,若大数据背景下的信息质量治理继续采用这种方法,会发现“随着系统增量开发,尽管人力、物力、财力不断增加,人财物的短缺却越发显得严重,问题越来越多,人手总是不足”的情况无法得到有效改善,甚至变得更为突出。

对此,美国麻省理工学院信息质量(MITI)计划和全面数据质量管理(TDQM)负责人、信息质量国际会议主席(ICIQ)Richard Wang 提倡采用知识、信息、数据的倒序层次管理思路:首先,对所需知识进行界定;而后再描述表达这一知识所需的信息;最后,关注转化为信息的原始数据的描述及其转化流程。基于这种思路,对数据和信息质量标准进行定义,基于前期积累的知识进行信息质量管理的科学规划与组织,确立更为全面、有效的信息产品质量标准,改变“头痛医头,脚痛医脚”的质量管理现状,提升信息质量管理工作的效率与效力。这些思路被国内外商务数据治理和医学数据管理所认可并采纳,但其应用对象多为同一家企业的进、销、存或是同一家医院的病历、诊疗、医药等领域,其关注的重点在于企业内部流程和系统的梳理、归因、整合、提升。截至目前,尚未见到将其系统应用于包含大量第三方数据来源的中介平台数据,这类平台数据多具有数据量大、来源多样、数据类型多样、数据综合后才能充分体现其价值等特点。中介平台的数据质量管理与传统意义上企业自身的核心业务数据管理有很大不同,前者面临着更大的机遇与挑战。大数据背景构成了发现数据治理新方法和适应性应用的新机遇,因为大数据、多来源的特点更便于发现并提取共性问题,归纳提炼出新知识的概率更高;但同时许多中介数据管理问题的复杂度和归因难度较传统数据质量管理更高,其应对和解决的多是传统数据管理中所不曾遇到过的挑战。因此,在新的大数据背景下,要有效应对这种机遇与挑战,关键是如何结合质量管理的相关知识,利用大数据分析优势探求信息质量度量的方法,根据度量结果进一步规范信息质量提升方案,而后构成新的质量管理知识,最终形成“知识归结——信息组织——数据治理——知识归结”螺旋式推进系统。

1.2 研究问题

大数据虽是一座宝藏,但需要有“大浪淘沙”的勇气和方法。本研究结合大数据征信实践,基于 Richard Wang 提倡的质量治理理念,探求大数据征信中介系统数字产品质量管理的研究关键点。针对大数据征信信息质量问题,哪些度量和提升的关键点为信用报告和信用增值产品研发所关注,如何有效地建模并表达、处理基本信息和信用信息,并有效整合提升,是理论界和实践界一直都没有明确回答的问题。本研究在这方面做出初步尝试,预期将在以下几个方面取得理论和应用价值。

(1) 对照数据质量考核与管理与数据质量提升的国际经验,基于大数据征信信息质量评价的特殊性,通过阶梯深度访谈法进行知识获取,识别并梳理信息质量度量和信息质量提升的关键点,获取征信信息质量度量与提升的“知识”,并进行决策相关性的界定与归结。回答数据质量度量与提升知识基础问题。

(2) 基于前期访谈“知识”,结合前期梳理的文献,构建信息质量评价结构模型;结合不同的评价维度,进行数据质量特征分量的选取。结合 Richard Wang 的信息质量度量框架,探索征信基本信息、信贷信息、非信贷信息的信息质量度量方法,发现并对目前征信信息进行产品划分与质量问题归结。回答信息质量度量问题。

(3) 基于专家知识和大数据分析,识别关键信息产品,创建关键产品的数据质量规范,确定定时检查的信息产品内容与流程;结合之前的信息产品目录与关键信息产品,建立基于信息产品图的质量设计模型,并制定质量改进计划与监控机制。回答信息价值提升问题。

1.3 研究意义

1.3.1 现实意义

在大数据背景下,需要面对类型更为多样、分散聚合的数据,需要化繁就简、全面整合。本研究是从大数据征信的异议和存在的问题切入,从管理策略和方法论层面为信息质量度量及价值提升提供了新的视角,开发了新的度量模型与工具,通过理论创新和应用拓展,为信息增值产品研发打好基础,也为提升信息

主体信用信息报告质量提供有效的信息支持和方法论指导,提高征信服务水平。具体表现为如下几个方面:

(1) 通过实证验证了大数据征信的信息度量维度,可为同类研究与分析提供科学依据。通过集成前期国内外学者信息质量评价的通用做法,并参考金融机构营运系统信息质量关注视角,进行了征信信用信息质量度量维度研究,确立了大数据征信信息质量科学全面的度量维度。

(2) 本研究通过引入面向外部报告使用、面向内部产品研发的应用视角,通过涉及任务数、使用频次、重要程度三个方面的讨论和度量,构建了相关性测度模型,基于度量结果构建了信息质量三级管理制度,为信息质量治理工作的分级管理奠定了坚实的基础。

(3) 结合征信实践,并基于相关理论,进行了大数据征信信息质量度量方法的开创性探索。完整性度量实现了从概念或需求层面覆盖程度、属性层面覆盖程度、记录层面覆盖程度开展的全方位度量;准确性度量依据平均距离类别划分为语义准确性和合理范围准确性,基于形式错误、逻辑错误、非主键错误等反向推导有效度量准确记录数占比;及时性度量根据流通时间与不变时间的相对状况,实现了从绝对及时性度量方式向相对及时性度量方式的扩展。以上为中介平台数据的完整性、准确性、及时性等科学度量提供方法论借鉴。

(4) 基于非信贷信息质量度量维度,构建了相关性、准确性、完整性、及时性的“四维度”度量模型。而后综合四维度,提出了面向应用的双重赋权信息质量度量体系。前人研究中,基于“真实世界”剖面,多从基础数据的准确性、完整性、及时性开展信息质量度量,本研究通过引入内外部征信非信贷信息应用情况,通过征信数据库的数据分析,构建了相关性、准确性、完整性、及时性相结合的双重赋权信息质量度量体系。

1.3.2 理论意义

本研究将信息质量度量理念和基本方法论拓展到了基于征信中介的信息质量管理领域,并在此基础上开展了适用于两类共享信息的数据质量管理策略研究。

信息共享是解决信息不对称的主要途径(Brown and Zehnder, 2010)。根据 Clarkson et al.(2007)的划分,信息共享可划分为横向信息不对称下的信息共享和纵向信息不对称下的信息共享。横向信息不对称下的信息共享,共享双方拥有共同知识,共享可以增加有效数据持有量,在征信上表现为银行与银行间的

信贷信息共享(有中介的 B2B 信息共享)。而纵向信息不对称下的信息共享,共享双方拥有不同信息,由有优势的一方提供给需要的另一方;这种信息共享实际上是一种信息服务,共享方向是单向的,在征信上表现为政府和银行间的非信贷信息共享(有中介的 G2B 信息共享)。前期国内外研究中多关注于征信市场上银行间的 B2B 信息共享,对于有中介的 G2B 异源信息共享关注不足。本研究立足于我国国情,结合我国有中介的 G2B 信息共享实践,探讨兼容 B2B 和 G2B 信息共享的价值度量和提升方式。

同时, Richard Wang 的知识、信息、数据的倒序信息质量管理理念和信息质量度量框架等均停留在信息质量度量与管理的理论层面。要真正开发成为可操作的中介平台数据质量度量方法,还需要进行方法论基础上的创造性设计和规划,并开展征信业务和征信信息的实地分析与扎实的实证研究,从而真正实现知识引导下的数据质量度量规则的提取与细化,并制定适宜的数据质量管理计划与改进方案。

1.4 研究方案

1.4.1 研究范围

1) 大数据

自 2008 年以来,大数据成为业界和专家们讨论的焦点,从 3V 到 4V,再到 5V,涵盖了大量(Volume)、高速(Velocity)、多样(Variety)、低密度(Value)、真实性(Veracity)。

(1) 大量(静态描述):通常大数据的体量比较大,往往达到 TB、PB 级,甚至达到 ZB 级。

(2) 高速(动态描述):大数据通常具有时间敏感性,需要支持实时分析的工具。

(3) 多样(数据类型):大数据的数据类型更为多样,通常包括结构化数据、半结构化数据及非结构化数据。

(4) 低密度(经济价值):单条信息通常价值较低,综合利用大数据,才能真正提炼、提升数据价值。

(5) 真实性(数据质量):数据质量的保证是大数据分析的基础。数据的内涵与外延、数据源的真实和全面等,是大数据科学权威性的重要保障。

大数据的特征逐渐被认识,其价值也得到越来越多的肯定。许多国家将大数据上升到国家战略层面,美国于2012年提出了改进知识获取能力的《大数据研究和发展计划》,澳大利亚于2013年发布了《公共服务大数据战略》,我国也于同年出台了将大数据发展纳入国家产业政策的相关指导意见。

2) 征信

征求、验证信用。征信是指专业化的、独立的第三方所从事的对企业、事业单位等组织的信用信息和个人的信用信息进行采集、整理、保存、加工,并向信息使用者提供的活动。征信体系是一个国家金融基础设施的重要组成部分。通过构建征信体系,旨在解决信贷市场的根本问题——借贷双方之间的信息不对称。

3) 大数据征信

大数据征信有狭义和广义之分。狭义上,大数据征信主要指诞生于互联网金融发展的过程中,来源于互联网各个平台的、与风险相关的信用数据。典型代表有阿里的“芝麻信用”、腾讯的社交数据以及91征信等。狭义层面的大数据征信在数据来源、权威性、完整性上区别于传统征信方式。广义上,大数据征信指“大数据时代的征信”,是指综合利用互联网线上线下信息建立的征信,既包括互联网金融发展过程中产生的征信数据,也包括传统征信。互联网时代能够记录借款人之前不可记录的信息,同时降低了信息获取的成本,为征信业务的发展和进步提供了有效支撑。广义层面的大数据征信改变了数据采集、整理、保存、加工、提供的方式和手段,但其征信活动的实质和征信业务开展的原则并不发生根本性的变化,也并非与传统征信相对立。本书中的“大数据征信”指的是广义上的大数据征信。

4) 信息质量

信息质量指信息可作为特定应用的可信资料来源的程度。因此,信息质量首先是面向应用的,是针对具体决策问题而言的。一般能够在合适的时间、合适的地点将合适的信息提供给合适的人员用于相应决策问题的解决,就是高质量的信息。而同时信息质量又有领域依赖,关于各个领域内劣质数据的相关问题和应对方案在过去近几十年得到许多学科领域的关注和探讨,有图书馆管理学领域在线数据库方面的,有会计领域财务数据方面的,有计算机编程领域数据处理方面的。总之,关于数据和信息质量问题的研究可以追溯到20世纪80年代甚至更早。同时,关于信息质量的定义也从简单的概念建构到多维层次知识体系的建立,关于信息质量度量与提升的研究也逐渐发展成为一门跨理论与实践