

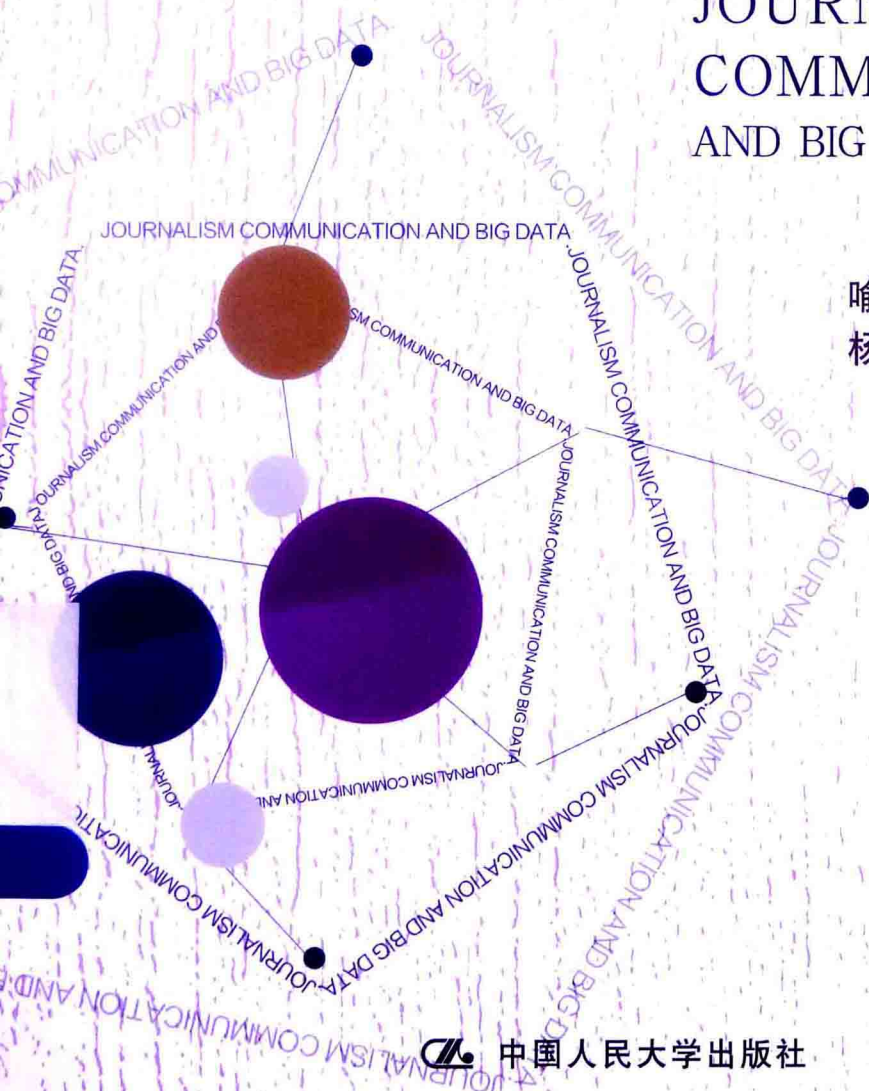
BIG DATA

新闻传播的 大数据时代

范式变革：大数据时代新闻传播的革命性创新

JOURNALISM
COMMUNICATION
AND BIG DATA

喻国明 李彪 著
杨雅 李慧娟



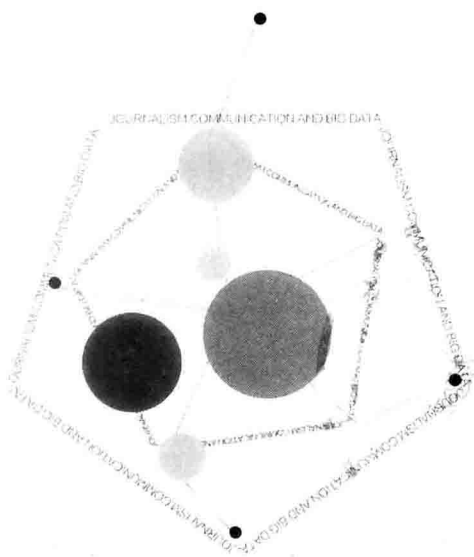
中国人民大学出版社

BIG DATA

新闻传播的 大数据时代

JOURNALISM
COMMUNICATION
AND BIG DATA

喻国明 李彪 著
杨雅 李慧娟



中国人民大学出版社

图书在版编目 (CIP) 数据

新闻传播的大数据时代/喻国明等著. —北京: 中国人民大学出版社, 2014. 10
ISBN 978-7-300-20084-2

I. ①新… II. ①喻… III. ①新闻学-传播学-研究 IV. ①G210

中国版本图书馆 CIP 数据核字 (2014) 第 226009 号



新闻传播的大数据时代

喻国明 李彪 杨雅 李慧娟 著

Xinwen Chuanbo de Dashuju Shidai

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

邮政编码 100080

电 话 010-62511242 (总编室)

010-62511770 (质管部)

010-82501766 (邮购部)

010-62514148 (门市部)

010-62515195 (发行公司)

010-62515275 (盗版举报)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com>(人大教研网)

经 销 新华书店

印 刷 北京鑫丰华彩印有限公司

规 格 170 mm×230 mm 16 开本

版 次 2014 年 10 月第 1 版

印 张 16.25 插页 1

印 次 2014 年 10 月第 1 次印刷

字 数 183 000

定 价 38.00 元

版权所有 侵权必究 印装差错 负责调换

大数据方法与新闻传播创新： 从理论定义到操作路线

喻国明^①

一、大数据方法与新闻传播创新

1. 何谓大数据

提到大数据的大，一般人认为指的是它数据规模的海量——人类在数据记录、获取及传输方面的技术革命，造成了数据获得的便捷与低成本，这便使原有的以高成本方式获得的描述人类态度或行为的、数据有限的小数据已然变成了一个巨大的、海量规模的数据包。这其实是一种不得要领、似是而非的认识。其实，前大数据时代也有海量的数据集，但由于其维度的单一，以及和人或社会有机活动状态的剥离，使其分析和认识真相的价值极为有限。大数据分析的价值和意义就在于，透过多维度多层次的数据，以及历时态的关联数据，找到问题的症结，直抵事实的真相。

2. 何谓大数据方法

大数据的数据跟过去传统的结构性的数据有很多不同。过去结构性的数据比较单一，成本、时间耗费比较多。而且从它的数据量以及解释的单

^① 喻国明：中国人民大学新闻学院副院长、国家二级教授，中国传媒经济与管理学会会长。

一化的层面上来说,都存在着缺陷,但是传统数据的优势在于结构性好、功能性强。大数据来自于生活的自然流露,包括在生活中的购物行为、搜索行为、表达行为等,这些都反映着人的真实生活状态。因此大数据的数据来源本身随着数字化记录、存储和传输技术的日臻完善而非常丰富,而且其数据的采集几乎可以与信息的发生同步,获得数据信息的成本又很低。至于大数据分析,现实的情况是,说的人很多,鼓吹其神奇价值的喧嚣声浪很高,却鲜见其实际运用得法的模式和方法。造成这种窘境的原因无外乎二:一是对于大数据分析的价值逻辑尚缺乏足够深刻的洞察;二是大数据分析中的某些重大要件或技术还不成熟。

3. 何谓大数据新闻传播

大数据新闻传播不同于传统新闻报道那样的简单数字交代,而是展示了一种从宏观与中观的层面对社会某一方面的趋势、动态和结构性的把握。大数据方法在新闻传播时间中的初级应用,是借助类似百度指数等各类数据采集和分析工具去挖掘散落在社会文本“碎片”中的具有新闻价值的资讯描述和意义表达。传统的新闻采集数据的方式更多的是通过线人、采访这种形态,而大数据方法为媒体工作者提供了一个全新的专业工具,帮助大家去挖掘新闻。

大数据方法视野下的新闻传播创新包含这样两个层次的内涵:首先,它是新闻形态的一种创新,包括可视化信息、人性化的嵌入。其次,它是一种全新意义上的内容创新,即通过碎片化的数据及文本的挖掘技术,实现了新形态的“减少和消除不确定性”的新闻内容。

二、大数据在新闻传播领域的实际应用

目前利用大数据资源的实际社会成效、有实际影响力的产品依然屈指可数。数据源的代表性和价值、良好的供给与需求的合作以及有广泛影响力的平台，是大数据应用获得成功的关键因素。

大数据可以实现一种在兴味盎然的“新闻游戏化”的参与中完成的传播读解和消费的过程。例如《华盛顿邮报》关于奥巴马就职典礼的报道，其网站中贴出了千兆像素的巨幅图片展示奥巴马宣誓就职场景，可以清晰地放大每一个局部，让每个参与者“找到”自己或自己的熟人。还有获得美国新闻奖的2013年的雪崩报道，以及阿拉伯之春立体四维报道西亚北非17国的情形，都是运用最新大数据的技术手段改革新闻报道的典型案例。

近期较为成功的大数据报道的案例，就是央视《新闻联播》在2014年春节期间播出的11集“据说”春运和春节。数据说春运和春节的成功有几大因素，其中包括新闻形式上的可视化的突破、新闻内容上的数据化和故事化的画面表达，呈现出“大数据小故事”。

大数据已经是一种客观存在。只不过相当长一段时间，人们缺少有效整合这些数据的技术和手段，并且人们对大数据的使用成本很高。互联网的OTT突破了原来的局限，解决了信息不对称的问题。其中的关键是大数据拥有方的合作与开放。“据说春运”节目的合作方的是百度公司，百度作为最大的中文搜索平台，每天要处理60亿次相关的搜索请求，其海量的数据能够生动翔实地反映中国网民具体的需求、兴趣点，搜索者本人的个人特点，等等。实际上，百度已经可以被视为中国最大的内容提供者。

三、现阶段大数据方法在新闻传播创新中的难点与关键

大数据方法在新闻传播创新中的难点与关键，现阶段主要集中在大数据方法与新闻传播价值逻辑之间的矛盾，以及大数据的数据源的开放等问题。

1. 大数据与新闻的价值逻辑之间的矛盾

大数据与新闻报道之间存在着几对矛盾。首先，事实之间的相关关系在新闻传播中却进行因果关系的解读所构成的矛盾。大数据的核心特色是强调伴随性指标的相关关系，大数据方法甚至于公开拒绝因果关系的认识逻辑。但新闻的传统解读却是具有强烈的因果逻辑的。如果在新闻报道的呈现中不把因果关系考虑进去，不但与人们的认识逻辑相悖，而且也容易滋长解读上的随意性和偶然性，这样便使数据对于新闻报道来说失去了核心意义。其次，大数据的内在逻辑与新闻表达的逻辑在某种程度上是相悖的。因为大数据强调的是信息结构化，抛开故事中心，“去故事化”，这就和传统报道中的故事化诉求产生了矛盾。如何将结构化的数据表现出人类生活的温度和质感，是大数据在应用于新闻传播过程中的一项极为重要的课题。另外，大数据方法与新闻传播所要求的精确性之间也是存在矛盾的，新闻要讲究精确性，而大数据方法却是以模糊性的呈现和把握为特点的。

2. 数据源的开放问题

在人人都在说大数据的时代，数据源的开放便非常重要了。互联网本

身是由开放精神主导的。如果我们无从得到权威的数据源的话，大数据方法就是一句空话。因此，大数据时代，Google也好，百度也好，必须要有开放的心态。很多媒体在前两年打造自己的微博，看起来是一个交流的平台，但是完全忽略了一个核心原则——只有开放和实现彼此连接，才是具有真正交流价值的平台。就现实而言，有质量的大数据源常常掌握在政府及大公司手中，如何开放这种大数据源的使用，事关社会的发展和人民生活的福祉，必须从制度和机制上给予保障。在这方面，美国政府的数据开放政策不但为政府开放数据源起到了一种很好的参照作用，对于大公司所掌握的数据源的开放也有着重要的借鉴意义。如果掌握着数据的公司或政府将数据源封闭在自己的圈子里，数据的巨大社会价值和商业价值就无法实现。实践表明，对于掌握着数据源的大公司和政府而言，数据放开会使政府和公司得到的比贡献出去的更多。

大数据方法在未来发展中的“行动路线图”是怎样的呢？首先，大数据分析在方法论上需要解决的问题在于：如何透过多层次、多维度的数据集实现对于某一个人、某一件事或某一种社会状态的现实态势的聚焦，即真相再现；其中的难点就在于，我们需要洞察哪些维度是描述一个人、一件事以及一种社会状态的最为关键的维度，这些维度之间的关联方式又是怎样的，等等。其次，如何在时间序列上离散的、貌似各不相同的数据集合中，找到一种或多种与人的活动、事件的发展以及社会的运作有机联系的连续性数据的分析逻辑。其中的难点就在于，我们对于离散的、貌似各不相同数据如何进行属性标签化的分类。概言之，不同类属的数据集的功能聚合模型（用于特定的分析对象）以及数据的标签化技术，是大数据分析的技术关键。

其次，从表现角度来说，嵌入是关键词。我认为，大数据呈现的结果

和结论，与人的需求、人的行为、人的认识逻辑需要有一种相适应的嵌入。尤其是在大数据刚刚进入社会生活领域的时候，一定要顺势而为，跟人的需求相关，跟人的认识行为逻辑相关，这样人们就比较容易去解读它，然后去把握它，去消费它，去使用它。比如，与新闻媒介相结合的时候，新闻媒介自身的传播逻辑、传播规则、传播样式，新闻媒介在传播过程的需求点上可以嵌入当中的哪些大数据的服务。这要有更多的数据专家去解读。

最后，与现有的可视化技术发展相联系。日常生活中一些重大的新闻如果能够运用大数据来报道，其深度会大大强化，也能够给人以更强的动感和说服力，并且帮助人们比较准确地把握未来。在这个意义上，大数据与新闻的结合，将是新闻竞争的巨大技术驱动力。

概言之，大数据方法是需要社会规则创新和高智力投入的一项伟大的认识与实践范式。本书的价值仅仅在于梳理了其基本原理与操作逻辑以及迄今为止在新闻传播领域中的典型应用实例，作为有志于此的同道步入这一殿堂的阶引。本书参考了国内外诸多研究者的见解与研究成果，我们在行文中尽量一一标注其出处。但由于资料来源庞杂，难免有所遗漏，特向所有本领域的研究者和思考者致敬。



目 录

第一章 大数据时代：新闻传播学的新纪元 1

01 认识大数据 2

一、大数据的产生 2

二、大数据的内涵 3

三、大数据的特点 5

02 方法变革：大数据方法打开了新闻传播学的新境界 8

一、大数据为我们提供了观察世界的新工具 9

二、大数据为我们提供了建构世界的新范式 11

三、大数据为我们提供了预测世界的新技术 11

03 思维变革：大数据时代新闻传播的认识创新和逻辑创新 13

一、思维革命：从“为什么”到“是什么” 14

二、范式转换：从“随机样本”到“总体” 17

三、结果转换：从“具体结论”到“宏观趋势” 18

01 大数据新闻：从理论概念到操作路线图 22

一 大数据新闻的由来沿革 22

解读大数据新闻 22

02 大数据新闻：精确新闻报道的进阶形态 26

精确新闻学：大数据新闻的萌芽 26

大数据新闻与精确新闻：从黄州城湖塌潰的启示 27

大数据新闻：精确新闻的进阶形态 32

四 大数据新闻的未来趋势 33

03 大数据新闻的功能与价值 37

编辑：重构新闻的时空数据 37

大群：在格网上的事件集结 38

预测：把握未来的信息先机 39

五 “双源”基础信息的双维交叉 39

六 大数据新闻的融入价值：大数据新闻、数据新闻

04 大数据新闻对新闻业存在的挑战 42

一 个人信息隐私权面临的挑战 42

二 媒体经营权的挑战 44

三 用户经营权的挑战 47

05 大数据新闻的隐私权和版权争议 79

- 一、大数据新闻的隐私权保护 79
- 二、大数据新闻与新闻版权：由“今日头条”引出的版权之争 83

第三章 大数据时代的广告营销 87

01 大数据与消费者（用户）洞察 88

- 一、营销思维的变革：从传统数据库营销到大数据库营销 88
- 二、用户数据类型的变革：从属性数据到多维数据 92
- 三、数据营销工具的变革：从粗线条营销到精细营销 93
- 四、案例分析——大数据对于用户行为特征与行为的还原 94

02 大数据与媒介接触界面的情境洞察 100

- 一、媒介嬗变趋势分析：智能化界面与人性化趋势 100
- 二、媒介权力与传受分析：电子对话与关系传播 102
- 三、媒介接触界面的脑认知机制分析：对用户信息处理方式的洞察 103
- 四、案例分析——界面智能化的宜家 IKEA Catalog 107

03 大数据与品牌软实力的构建 111

- 一、节点：大数据助力品牌接触点管理 111
- 二、链条：大数据助力品牌信息流导向 113
- 三、界面：大数据助力品牌认知网络整合 116

四、未来：大数据助力品牌营销个性化发展 119

五、案例分析——电视栏目品牌认知整合 121

04

大数据与市场预测 127

一、历时性数据分析与市场需求预测 127

二、数据库建设与潜在关系挖掘 131

三、基于数据的网络广告投放策略 134

四、基于眼动实验的广告视觉注意模型 138

五、基于数据的移动互联原生广告 143

05

大数据营销的操作流程图 147

一、关联 147

二、开放 148

第四章 大数据时代的社会舆情分析与把握 151

01

大数据对社会舆情研究的影响 152

一、大数据成为人类社会运行的“仪表盘”更有利于
舆情研究 152

二、大数据成为社会舆情研究的“利器” 153

三、大数据是社会关系的“沉淀池”，改变了传统舆情只
见“内容”不见“关系”的境况 153

四、大数据的预测性可以解决目前舆情研究的尴尬 155

02

大数据视域下社会舆情研究的若干转向 156

- 一、研究视角的转向：从单向度的内容研究转向“内容+关系”的多维度研究 157
- 二、研究方法的转向：由舆情信息采集转向数据加工、可视化等 157
- 三、数据库支持的转向：由简单的、有限的数据库转向非结构化的大数据库 158
- 四、舆情研究主体的转向：由小作坊式的单打独斗、面面俱到的舆情监控转向分工明确、高度聚合集约的舆情分析 159
- 五、舆情研究重点的转向：由舆情监测转向舆情预警乃至预测，从单向度的危机应对、品牌营销转向各领域的综合信息服务 159

03

大数据时代社会舆情研究领域的趋势 160

- 一、舆情产品从舆情软件开发到云平台搭建 161
- 二、大数据成为舆情研究关键技术的支撑和核心概念 161
- 三、舆情研究从信息传播到情绪传导机制研究 162
- 四、舆情研究从信息流研究转向社会关系网研究 162
- 五、舆情研究上升到新媒体与社会运动的研究层面 163
- 六、舆情预警成为大数据时代舆情研究的主要命题 163

04 大数据视域下舆情研究的若干做点 164

- 一、大数据收集：尽量扩大信息源 165
- 二、大数据存储：注意集约化 165
- 三、大数据处理：注意提纯和结构化 165
- 四、大数据结果呈现：低门槛和可视化 166

05 大数据社会舆情研究案例探索示例 167

- 一、案例 1：中国人在微博上都在说什么 168
- 二、案例 2：社会热点事件在微博中的传播机制 184
- 三、案例 3：不同网络世代的社会关注焦点 197
- 四、案例 4：社会各类运行指标分析 203

第五章 大数据传播：未来的发展与预测 221

01 观念转型：技术的高度决定思维的高度 222

02 范式转型：从数据范式到大数据范式 224

03 对象转型：复杂性社会及其现象的洞察 226

04 技术转型：可视化及可视化技术 228

05 大数据素养教育：警惕大数据的误区及陷阱 230

附：大数据分析资源和可视化工具 232

参考文献 243

第一章

大数据时代：新闻传播学的新纪元

Xinwen Chuanbo
de Dashuju
Shidai

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。



01

认识大数据

一、大数据的产生

(一) 物理基础：数字化时代无处不在的数据微处理器、传感器及互联网

大数据之所以产生，正如美国咨询大师托马斯·H·达文波特(Thomas·H·Davenport)所言，是因为今天无处不在的数字化的传感器和微处理器，比如物联网、云计算、移动互联网、车联网、手机、平板电脑、PC以及遍布地球各个角落的各种各样的传感器，无一不是数据生成、承载和传输的方式。我们正在迈进数字化时代。其实，人类的所有实践活动经过形式各异的机器或电子设备的作用都可以留下数据痕迹，这些痕迹记录了它的特征、位置或状态。这些设备和使用它的人们，通过网络之间交流和链接，“嫁接”出来了一个个庞大的数据源。这些品性各异的数据集在互联网的连接下构成了对于人类社会动态发展中个体与群体最为丰富、充分和翔实的数据描述。