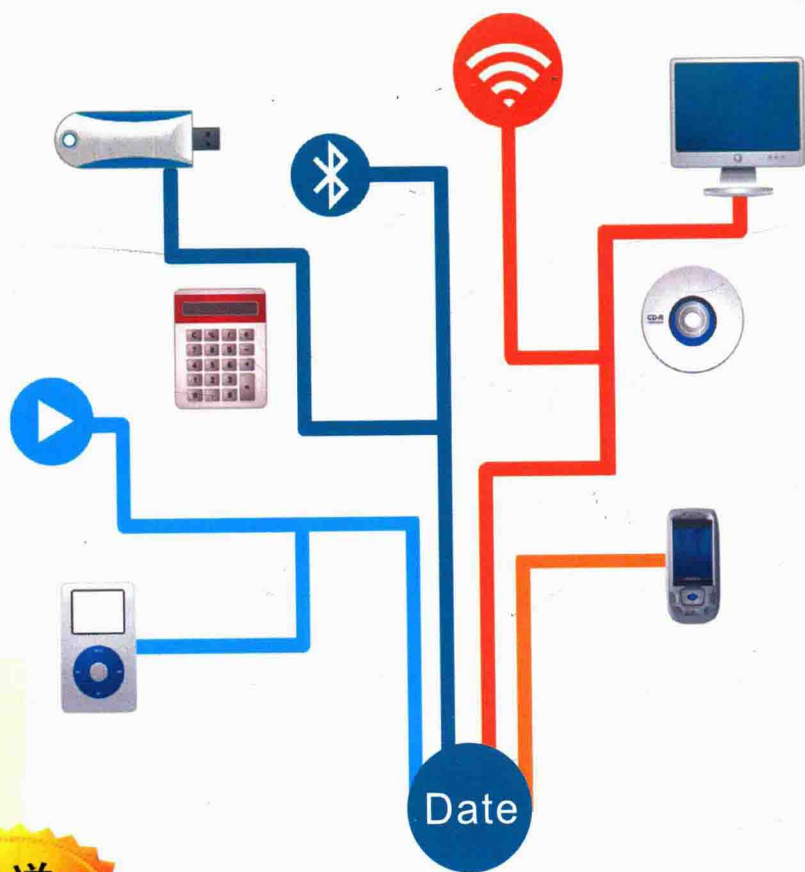


普通高等教育数据新闻系列教材

数据新闻入门教程

主编 陈积银 曹树林



赠送
电子课件

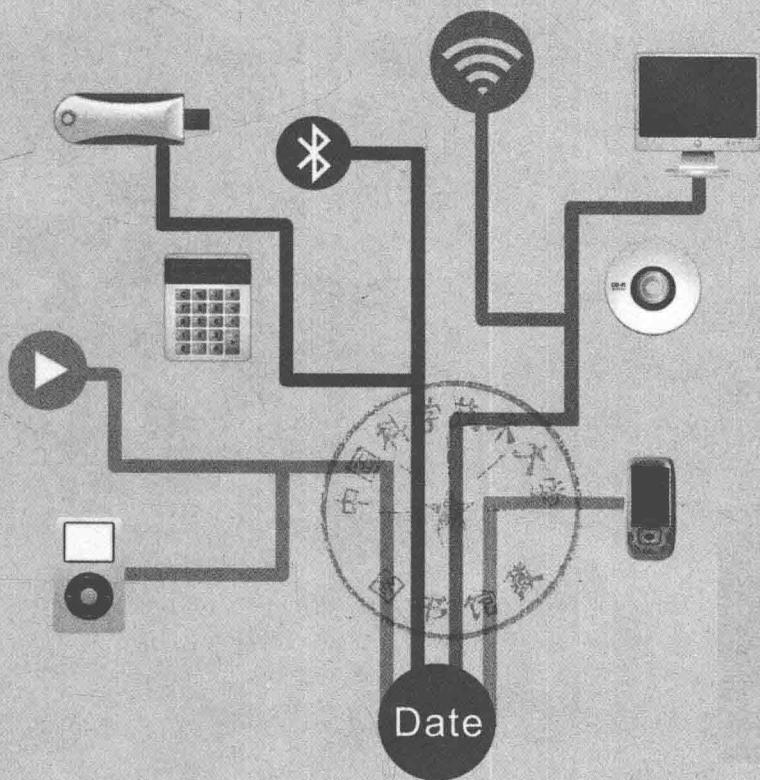


西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

普通高等教育数据新闻系列教材

数据新闻入门教程

主编 陈积银 曹树林



 西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

图书在版编目(CIP)数据

数据新闻入门教程/陈积银,曹树林主编. —西安:西安
交通大学出版社,2016.4
ISBN 978-7-5605-8432-4

I. ①数… II. ①陈 …②曹 … III. ①新闻学-教材
IV. ①G210

中国版本图书馆 CIP 数据核字(2016)第 072959 号

书 名 数据新闻入门教程
主 编 陈积银 曹树林
责任编辑 史菲菲

出版发行 西安交通大学出版社
(西安市兴庆南路 10 号 邮政编码 710049)

网 址 <http://www.xjtupress.com>
电 话 (029)82668357 82667874(发行中心)
(029)82668315(总编办)

传 真 (029)82668280
印 刷 陕西新世纪印刷厂

开 本 787mm×1092mm 1/16 印张 15.625 字数 373 千字
版次印次 2016 年 8 月第 1 版 2016 年 8 月第 1 次印刷
书 号 ISBN 978-7-5605-8432-4/G·1399
定 价 34.80 元

读者购书、书店添货、如发现印装质量问题,请与本社发行中心联系、调换。

订购热线:(029)82665248 (029)82665249

投稿热线:(029)82668133

读者信箱:xj_rwjg@126.com

版权所有 侵权必究

内 容 提 要

本书是数据新闻课程的基础性教材。全书分为五个章节，分别是：第一章，数据新闻概论；第二章，数据新闻的选题；第三章，大数据语境下的信息提取：获取数据；第四章，数据分析；第五章，用数字讲故事：数据可视化。

本书图文并茂，读者可以从更加直观的角度来学习和了解数据新闻各个方面的知识。本书可以作为新闻专业学生的专业课教材，也可以作为新闻媒体从业人员的参考用书，还可以作为对数据新闻感兴趣人士的学习读本。

编委会成员(按姓氏拼音排名)

Pam Tobey Rick Dunham 白红义 蔡晓艺

曹树林 陈 杰 陈积银 高红波 郭亚娟

刘 鹏 刘文红 王 未 王建磊 王资佶

吴 恒 杨 廉 赵 洋 曾凡齐

近年来数据新闻在国内外发展得如火如荼,是大数据技术对新闻业全面渗透的必然结果。2013年被称为“大数据”元年。数据新闻正是在大数据技术背景下逐渐兴起的一种新型的新闻生产方式,其核心是数据的挖掘、分析和可视化呈现。数据新闻是大数据技术对新闻业全面渗透的必然结果。它在一定程度上改变了传统新闻的生产方式,是未来新闻发展的主要方向之一。目前,国内外的媒体都在尝试运用此方法生产新闻,部分国内外的一流院校如哥伦比亚大学、密苏里大学、清华大学、中国传媒大学、中山大学、上海大学、河北大学等已经开设了相关课程与专业。西北师范大学也在整合资源,开设了数据新闻的相关课程,并与哥伦比亚大学合作举办了“数据新闻工作坊”,进行了数据新闻生产的培训,邀请了全国近20家传媒机构参与。

甘肃省委依托西北师范大学成立了甘肃省融合媒体研训基地。基地成立一年来,邀请哥伦比亚大学的老师来甘肃为全国的莘莘学子培训数据挖掘和数据分析与可视化技巧。另外,基地还组织学员前往香港大学、复旦大学、上海大学等著名高校学习数据新闻的制作能力。

2015年4月,甘肃省融合媒体研训基地联合中国传媒经济与管理学会启动了首届中国数据新闻大赛。该活动历时3个月,在全国范围内征集了业界和学界的优秀数据作品88件,有70余家机构约300余人参与这项赛事。参赛的作品通过深度的数据分析、专题讨论或其他形式的研究来呈现其作品主题,内容涉及经济、环境、教育、时政、娱乐、文化等方面。此次大赛共收获了来自全球近10万人的投票参与。学界有哥伦比亚大学、纽约大学、密苏里大学、清华大学、北京大学、中国人民大学、中国传媒大学、西北师范大学等40多所高校,业界有人民日报社、上海报业集团、浙江日报集团、新京报、澎湃网、甘肃日报社等20多家媒体机构参与,受到了媒体从业人员、相关专业老师和高校学生的广泛关注。

2015年11月,甘肃省融合媒体研训基地联合北京师范大学新闻学院等机构在北京师范大学开启了第二届中国数据新闻大赛。经过近半年的组织,会务组共收到来自全国各地共计110余个队伍的报名信息,直接参与人数达600余人。此次活动的主题主要围绕新丝路经济带建设过程中的成就、机遇、挑战、传播与战略

合作问题。此次大赛共收获了来自全球 15 万余人的投票参与。业界包括：人民日报、光明日报等 10 余家业界机构；学界包括清华大学、中国人民大学、复旦大学、武汉大学、北京师范大学、中国传媒大学、上海大学、华东师范大学、西北师范大学等 70 多所高校。北京捷泰天域信息技术有限公司、兰州智慧传播有限责任公司、《新闻与传播研究》、《新闻记者》、《网络传播》、《传媒》杂志、《西北师范大学学报》等机构参与了协办。

然而，在和众多的从业者和高校师生们交流的过程中，我们发现，当前数据新闻的制作，还是一件不太容易的事。国内数据新闻教学的教材奇缺。因此，甘肃省融合媒体研训基地特别组织国内数据新闻精英以及基地本身的成员编写了《数据新闻入门教程》。

本书共分为五章，第一章主旨在于厘清数据新闻的概念，讲述数据新闻的发展脉络，辨析数据新闻与统计新闻、精确新闻、算法新闻、计算机辅助新闻报道、数据可视化的概念。介绍在“信息爆炸”和“大数据”时代下，数据新闻的主要形态（包括数据地图、数据图表、数据漫画、专属新闻、数据驱动调查），并梳理国内外数据新闻的发展现状，包括当前业界的发展情况和学界数据新闻的教育情况。第二章从数据分析程度、选题的时效性两个维度梳理数据新闻可分为四类：展示型、分析型、周期型和突发型。并介绍数据新闻的选题类型、选题范围（包括互联网议题、体育议题、政治议题、体育议题）及发展现状，阐述了未来将呈现娱乐化、产品化、个性化的趋势。第三章主要基于大数据语境下的信息提取，介绍数据搜索的方法，数据来源渠道（包括国际官方机构、国内官方机构、自建数据库等），数据挖掘或转换数据用到的工具（包括火狐浏览器、八爪鱼、Document Cloud、Python、CiteSpace、ABBYY FineReader 等）。这些数据信息的提取方法可以帮助更好地完成数据新闻的数据收集工作。第四章以理论和实践操作结合案例进行讲解对获取的数据进行分析，提出数据分析的必备素养及思路（如 PEST 分析法、5W2H 分析法、逻辑树分析法），并介绍数据分析常用的工具（如 Excel、SPSS 等）。在未来的新闻业中，学会简单的数据分析是新闻人必备的技能之一。第五章主要对数据可视化的概念、优秀的数据可视化的标准以及数据可视化的常用工具（主要包括 Excel、iCharts、Tagxedo 和图悦、ARCGIS、Gephi、Tableau Dshtop、Echarts 和百度图说以及基于 HTML5 的数据可视化工具等）进行分析，通过案例对数据可视化有初步的了解和认识。同时，本书还对全国的数据新闻优秀团队一一进行了采访。

很明显,数据新闻在一定程度上改变了传统新闻的生产方式,是未来新闻发展的主要方向之一。当然也存在鱼龙混杂、泥沙俱下的情况。部分数据新闻开始忽略新闻真实含义,过分强调数据新闻的技术。其实最重要的是正确运用数据新闻其独特的叙述新闻方式,才能推动新闻更好地向前发展。

本书第一章由陈积银、王资信编写;第二章由吴恒编写;第三章由杨廉、蔡晓艺、郭亚娟编写;第四章由王未、杨廉编写;第五章由赵洋编写;附录1由曾凡齐编写,附录2由陈积银、刘文红整理。

感谢中国传媒经济与管理学会常务副会长吴信训教授,北京师范大学周星教授、张洪忠教授,南京大学王成军博士,中国社会科学院新闻研究所的朱鸿军博士,清华大学的 Rick Dunham、杭敏女士,他们的大力支持和鼓励是本书最终得以面世的保证。感谢中央电视台市场研究(CRT)总经理徐立军博士,复旦大学周葆华教授、徐笛博士,上海大学吴小坤博士,他们的鼓励与支持,使我们基地的团队在不断壮大。

尽管在编辑此书的过程中,我们就兢业业,不断研讨,搜集资料,唯恐出现错误。但是,由于水平有限,本书还难免存在一些问题。**本书中的图片仅具有示意作用,读者们如果要查看原图,请拨打 029-82668133 索取电子课件。**在此敬请读者批评指正!也恳请读者给我们及时反馈,以便我们修订完善。

编者

2016年4月10日

目录

Contents

第一章 数据新闻概论	(1)
第一节 数据新闻的发展历程	(1)
第二节 什么是数据新闻	(4)
第三节 数据新闻的主要形态	(10)
第四节 数据新闻的发展现状	(20)
第二章 数据新闻的选题	(26)
第一节 选题类型	(26)
第二节 选题范围	(37)
第三节 趋势	(50)
第三章 大数据语境下的信息提取:获取数据	(54)
第一节 获取数据总论	(54)
第二节 数据信息搜索	(56)
第三节 数据挖掘和转换数据工具	(81)
第四章 数据分析	(125)
第一节 什么是数据分析	(125)
第二节 数据分析开始前做什么	(127)
第三节 数据分析思路及方法	(133)
第四节 用工具来进行数据分析	(142)
第五章 用数字讲故事:数据可视化	(167)
第一节 什么是数据可视化	(167)
第二节 数据可视化:如何用数据讲故事	(174)
第三节 可视化用到的工具	(182)
第四节 优秀数据新闻可视化案例	(196)
附录	(199)
附录 1 基于 Python 语言的在线社区用户行为分析	(199)
附录 2 国内部分数据新闻团队介绍	(211)

第一章 数据新闻概论

数据新闻,简而言之,就是从数据中寻找故事,并利用数据来讲述故事的新闻样式。它是西方新闻业的发明,2009年,英国著名质报《卫报》开设了“数据博客”,专攻数据新闻。2014年,美国《纽约时报》设立了专注于数据新闻的网站——The Upshot。数据新闻正日益跻身新闻业主流报道样式。

本章主旨在于厘清数据新闻的概念。本章首先讲述数据新闻的发展脉络,继而对数据新闻的概念进行辨析,接着介绍当下数据新闻的主要形态,随后介绍国内外数据新闻的发展现状,最后展望数据新闻的未来。

第一节 数据新闻的发展历程

数据新闻并不是21世纪的新生事物,它在媒体上的最早现身可以追溯至1821年,甚至更遥远的年代。只是彼时,新闻业还未使用“数据新闻”这一概念。数据新闻可以说是计算机辅助报道(computer-assisted reporting, CAR)的后裔,它汲取了计算机辅助报道的部分内核,又吸收了21世纪迅猛发展的计算机技术的养分,在特定的社会、经济、科技和文化等要素下孕育而生。追溯数据新闻的历史印记,不仅有助于分析数据新闻热潮背后的历史动因,也有助于理解数据新闻在当下的发展态势。

一、追溯数据新闻的成长史

用图表展示数据,并用数据辅助报道,是新闻业由来已久的做法。1786年,苏格兰工程师、政治经济学家威廉·普莱菲(William Playfair)出版了《商业与政治图解》(The Commercial and Political Atlas)一书,书中他绘制了44幅图表来解释经济统计数据^①,此后由他创造的饼图、柱图等开始被广泛使用。图表登上报纸的版面至少可以追溯至1821年。根据英国《卫报》数据博客前主编西蒙·罗杰斯^②(Simon Rogers)的考察,1821年5月5日,《卫报》(当时《卫报》叫做《曼彻斯特卫报》)在其创刊号上刊登了一幅表格,列出了曼彻斯特和索尔福德地区每所学校的学生人数和平均学费,表格揭示了接受免费教育的学生人数,而这一人数与政府曾公布的数字相去甚远。表格的运用简洁明了地展示了事情的另外一面,有利于读者进一步了解真实情况。

到了19世纪后期,在新闻中运用数据已变得习以为常,商业新闻、竞技赛事都大量使用数

① 西蒙·罗杰斯.数据新闻大趋势:释放可视化报道的力量[M].岳跃,译.北京:中国人民大学出版社,2015:54.

② 西蒙·罗杰斯,为数据新闻作出了开拓性的贡献。2009年他参与创办了英国《卫报》数据博客,并任主编,2013年跳槽至Twitter担任数据编辑,2015年3月加入Google,担任趋势数据分析师。

据辅助报道。1884年,道琼斯公司开始在报纸上发布股市交易数据。数据正变得日益重要,它不仅为商业精英所用,更成为崛起中的中产阶级作各项决定的重要依据^①。

20世纪后半叶计算机技术的迅猛发展给新闻业注入了新的活力。1952年美国哥伦比亚广播公司率先使用一台大型计算机进行统计运算,预测了总统大选的结果^②。计算机也扩展了记者的工具包,美国记者菲利普·梅耶(Philip Meyer)将社会调查统计方法与计算机数据分析技术相结合,创造出了一种新的报道样式——精确新闻。简而言之,精确新闻是指运用社会科学或行为科学的研究方法来报道新闻事件^③。20世纪60年代,美国洛杉矶、纽瓦克和底特律等地区相继爆发种族骚乱,当时有两种流行观点解释骚乱成因:一是认为骚乱分子经济窘迫,又无力改变现状,借由骚乱来发泄绝望;二是认为南方黑人饱受种族歧视,当他们迁徙到北方后难以融入北方文化,骚乱成为他们长期被压制的发泄口。为验证这两种流行观点,菲利普·梅耶携手两位同事,通过随机抽样的方法对居住在底特律骚乱区的437名黑人进行了问卷调查,并用电脑对数据进行了统计分析。结果显示参加骚乱的人未必都来自社会底层,此外其中大部分在北方长大^④。梅耶据此写出了《十二街那边的人们》系列报道,并发表在《底特律自由报》(The Detroit Free Press)上^⑤,报道打破了人们的偏见,引起了轰动,并获得了普利策新闻奖,精确新闻这种报道样式也随之声名鹊起。

精确新闻在美国的流行并非偶然,它与20世纪美国社会科学的长足进展以及美国社会运转特点密切相关。社会科学中抽样与问卷调查的方法渐臻完善,这种方法也拓展到社会领域,民意调查成为体察社情民意的重要手段,它应用广泛,公共决策也大量依据民意调查的结果,以至于美国被称作是“调查国家”(survey country)。在20世纪,民意调查已经发展成为一项潜力巨大的产业。

二、数据新闻发展的社会背景

(一)“信息爆炸”和“大数据”时代到来

进入20世纪90年代后,信息量就以几何级别增长,而互联网的出现则促进了信息的传播,新网站的数目随之开始狂飙突进,每天更新的资讯不计其数,我们实在难以完全将其收入脑中。2011年4月7日的“存储网络世界”(Storage Networking World)年会上,美国加州大学圣地亚哥分校的研究人员给了我们一个数字:世界范围内服务器年处理为95700000000000000000字节,也就是9.57泽字节(1泽字节等于10的21次方)。这些数据到底有多少呢?具体来说,如果将地球的数据年处理量转换成书本格式,那么这些书本摞起来的厚度则长达90亿千米,这长度是地球与海王星距离的20倍^⑥!这种信息快速增长的现象就是“信息爆炸”。“信息爆

① Alexander Benjamin Howard. The art and science of data-driven journalism: when journalists combine new technology with narrative skills, they can deliver context, clarity, and a better understanding of the world around us[EB/OL]. <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>.

② Alexander Benjamin Howard. The art and science of data-driven journalism: when journalists combine new technology with narrative skills, they can deliver context, clarity, and a better understanding of the world around us[EB/OL]. <http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf>.

③ 菲利普·梅耶. 精确新闻报道:记者应掌握的社会科学研究方法[M]. 肖明,译. 北京:中国人民大学出版社,2015.

④ 方洁. 数据新闻概论:操作理念与案例解析[M]. 北京:中国人民大学出版社,2015.

⑤ 菲利普·梅耶. 精确新闻报道:记者应掌握的社会科学研究方法[M]. 肖明,译. 北京:中国人民大学出版社,2015.

⑥ 果网. 全球信息大爆炸[EB/OL]. (2011-04-13). <http://www.guokr.com/article/20331/>.

炸”使得世界上的数据快速积聚,互联网中、服务器上都充斥着海量的数据,“大数据”时代随之而来。大数据的力量影响着全球的许多领域、行业。Google 公司利用自己所掌握的海量用户搜索数据便可预测美国哪些区域将有可能爆发流感;Farecast 软件通过分析海量数据便可预测当下的票价是否最便宜;亚马逊通过分析海量的顾客购买清单,便可预测顾客接下来的购买行为,进而有选择地为客户推荐相关产品。许多行业利用周围的海量数据而得到更有价值的信息,从而取得更大的成功。在“大数据”时代,人们不再为寻找数据而发愁,人们只需为如何开发、利用这些数据资源而绞尽脑汁。“大数据时代的预言家”舍恩伯格认为在“大数据”时代,就要拥有“大数据”思维。

首先,要摒弃随机抽样,而是使用全体数据。随机抽样是信息处理能力受限的时代产物,在大数据时代,技术的进步使得我们获取数据、存储数据、处理数据的成本低、易实现,人们有条件使用尽可能多的数据来分析问题,在大数据时代“样本=总体”。

其次,要放弃对精确的执著,学会接受混杂。所谓接受混杂,有两方面含义:其一是接受数据集中的错误信息,由于数据集中的数据量巨大,筛选错误信息非常浪费成本与时间,而且对于海量数据而言一些错误的信息基本可忽略不计;其二是接受数据的非结构化,因为只有 5% 的数据是结构化且能适用于传统数据库的,如果不接受非结构化,那么剩下的 95% 的数据都无法被利用。

最后,应关注相关关系,而不必过于追求因果关系。在大数据时代,我们要让数据自己“发声”,有些时候知道“是什么”就已足够,没有必要刨根问底。同时,大数据的相关关系分析法更准确、更快,且不易受偏见的影响。

在“大数据”思维的影响下,许多公司开发出了对“大数据”导入、整理、分析的工具、软件,统称为“大数据”技术。它为新闻实践提供了新的思路。一方面,许多媒体自身就掌握着丰富的数据资源,但缺乏利用的技术。另一方面,“大数据”技术可以发掘出大数据背后更有价值的信息,为记者提供了更多的新闻资源。总之,“大数据”时代的到来,成为数据新闻得以产生的一个重要驱动力。

(二)可视化技术进入新闻领域

“大数据”及相关技术虽然可以让记者们发现更多的新闻资源,但这种资源只具备新闻价值中的“时新性”,而可视化技术则增加这些信息的“趣味性”“可读性”,使这些资源成为真正的新闻作品。

可视化技术的概念界定有狭义和广义之分。狭义上的可视化技术就是指利用计算机技术和图形处理技术,将抽象的概念、文字、数据转化为易于理解且结构更为清晰、一目了然的图表信息;广义上的可视化技术,不只是将文字、数据转换成图表信息,还要包含人机交互性,人们可以与可视化后的信息终端进行互动活动。然而在可视化技术的应用上狭义与广义的概念并没有严格区分。

数据新闻工作者将可视化技术应用于数据新闻的制作过程中,一方面由于数据新闻使用了大量的数据,简单的数据罗列,会使读者产生反感,甚至找不到自己需要的信息,而可视化技术清晰、形象、生动地将大量数据以一些色彩明艳、生动有趣的画面形象将大量数据更有效地呈现出来,让读者理清复杂难懂的数据。另一方面,可视化技术中的交互处理技术应用于新闻呈现既实现了新闻与读者间互动,又满足了读者的个性化、碎片化阅读需求。

因此,将可视化技术应用于新闻实践,是数据新闻得以产生的另一个重要的驱动力。总

之,“大数据”及“大数据”技术,让记者能够有效利用数据,掌握更多的新闻资源。可视化技术,可以帮助读者更好地理解数据,读懂数据新闻,甚至喜欢上数据新闻的阅读方式。这两者都是数据新闻产生的重要驱动力。

第二节 什么是数据新闻

一、数据新闻辨析

(一)数据新闻与传统叙述式的新闻报道

从功能上讲,数据新闻与传统叙述式的新闻报道最根本的区别在于,前者为社交媒体时代的新闻记者赋予了一种新的核心竞争力,即同时拥有敏锐的新闻嗅觉和使用大规模数据处理信息的能力,从而完成更具有深度和专业性、更富于逻辑性和感染力的报道。简言之,当今的专业记者应当善于挖掘数据,将其转化为生动的故事和深邃的洞见,并且借助于新媒体使新闻报道可视化、纵深化、互动化,使新闻报道更精确、深入、直观。

数据新闻涵盖的往往是与公共事务和国计民生密切相关,但又不容易通过文字或图表等传统手段理解和阐明的领域。目前较为成功的数据新闻集中于政治选举活动、财经、能源、环境、体育等领域的相关选题。

(二)数据新闻与统计新闻

统计界有名言:“统计是流动的历史,历史是流动的统计。”每项统计报告,都记载了刚刚过去的社会经济的数据历史;及时讲述这段历史,告诉人们这些刚产生的数字,此刻代表、说明、揭示、预示了什么^①。统计新闻是用经过抽象量化的统计数字来反映某个经济活动或经济现象的数量状态,让人们从这个数量形态上去认识这个经济活动或经济现象的意义。它不同于一般的经济新闻直接反映具体的、实在的、可感的经济活动和经济现象。如果说新闻是新近发生的事实的报道,那么可以说,统计新闻是新近发生的事实的间接性报道。

统计新闻除了描述新闻事实的数字材料外,大量的引证和分析数字的材料往往是背景材料。而数据新闻则是基于数据的抓取、挖掘、统计、分析和可视化呈现的新型新闻报道方式。数据新闻是在大数据的技术背景下产生的,是随着数据时代的到来出现的一种新型报道形态,是数据技术对新闻业全面渗透的必然结果。

(三)数据新闻与精确新闻

精确新闻由美国学者菲利普·迈耶在20世纪60年代提出并出版了专著《精确新闻学——一种社会科学报道的理论》,主张将社会调查研究方法应用到新闻实践中,科学地收集资料、查证事实,用数据来说话,从而提高新闻报道的准确性和客观性。

精确新闻更像是用社会科学的研究方法做新闻,而数据新闻是基于数据的现实,或者说是数据驱动,它们的历史关系不可否认。

数据新闻在一定程度上混杂性是通过大量数据更加促进了新闻事实的精确性,所以国内很多学者都把数据新闻作为精确新闻的补充。

^① 高颀伟. 统计新闻深度报道的采写探析[J]. 新闻世界, 2013(10): 108-109.

大数据时代下的数据新闻和精确新闻有着三个方面的不同:

1. 分析和处理的数据量不同

大数据时代产生的数据是巨量的,数据新闻在报道新闻事件时,分析和报道的数据量也是精确新闻无法比拟的。

2. 数据的地位和作用不同

在精确新闻中,更多的是以文字为主、数据为辅,或者两者相辅相成。而数据新闻则是数据驱动新闻,数据为主、文字为辅,新闻报道的思路和流程发生了改变。

3. 新闻报道的载体不同

精确新闻产生于传统媒体发达阶段,呈现载体是传统媒介。而数据新闻由于其分析的数据在空间和时间上的广阔,包含的数据量巨大,以及新的科技手法的出现,更适宜在数字平台发布。

而相对于精确新闻,数据新闻也有两大主要进步之处:

(1)数据获取方法和来源更多元,记者足不出户就可以借助互联网搜索、机器抓取获取数据;

(2)基于互联网的新闻数据可视化,表现力远远超出单调的统计数字和图表。

由于信息技术的进步,数据概念的扩展,数据新闻是在使命上继承了精确新闻的衣钵,而在方法论(数据获取与展示)上有了极大的丰富,更加回归新闻讲故事的本质^①。

(四)数据新闻与算法新闻

算法新闻,即基于算法的机器新闻写作,新闻内容生产和编辑、出版均自动化。最早产生于谷歌新闻开启的机器自动选编和推荐新闻的做法。

算法新闻是大数据环境下数据新闻的一种形式,但又与数据新闻有些许差别。首届“国际数据新闻”圆桌会议^②对数据新闻概念作出了如下界定:“‘数据新闻’是一种工作流程,包括下述基本步骤:通过反复抓取、筛选和重组来深度挖掘数据,聚焦专门信息以过滤数据,可视化地呈现数据并合成新闻故事。”^③算法新闻是基于数据新闻流程自动生成新闻的模式。它与数据新闻的区别在于:

①除前期技术开发外,人工参与过程并不是算法新闻生成的关键环节。

②算法新闻不同于早期纯粹的信息收集、抓取程序,而是在数据收集、处理的基础之上自动生成完整的新闻样式。算法新闻是在特定的计算机程序系统的基础上,对信息内容进行抓取、分析后自动形成完整的新闻报道的新的生产方式。

③算法新闻的成果是完整的新闻报道,算法新闻的最大特征是自动化的新闻生成,在技术研发阶段以后,人工的参与不对新闻生成产生决定性影响。

目前,“算法新闻”多运用于体育、经济、灾难等信息报道中。因技术的局限性,“算法新闻”的适用对象多为具有数据性、时效性特征的简短消息。

① <http://www.zhihu.com/question/23524711/answer/26510287>.

② 首届国际数据新闻圆桌会议(data-driven journalism Amsterdam round-table),由米尔科·劳伦兹主持,于2010年在阿姆斯特丹召开,由欧洲新闻中心(European Journalism Centre)和荷兰阿姆斯特丹大学(the University of Amsterdam)主办。

③ 方洁,颜冬.全球视野下的“数据新闻”:理念与实践[J].国际新闻界,2013(06).

（五）数据新闻区别于计算机辅助新闻报道

计算机辅助新闻报道发端于 20 世纪 50 年代,指的是用计算机来辅助收集和处理信息的新闻报道的方法。伴随着计算机技术的发展经历了使用大型机处理政府数据库、发现和调查新闻事实的早期阶段;再到 20 世纪 70—80 年代通过 PC 普及、商业和政府数据库的进一步开放形成的“以新闻报道为目的,对任何计算机化的信息来源的处理和使用行为”的第二阶段;发展到 20 世纪 90 年代中期以后利用互联网进行新闻采集、分析和制作的第三阶段^①。与数据新闻相比较,计算机辅助新闻报道更像是一种辅助工具,并不是一种独立的新闻报道方式,它只是强调计算机对于新闻报道是一种方法,而数据新闻则是大数据时代下新闻发展的一种形态,一种报道样式,计算机辅助新闻报道可以作为新闻生产中的一环,两者的内涵和外延均不同。

（六）数据新闻区别于数据可视化

提到数据新闻就一定会提到数据可视化,可视化(visualization)指的是利用计算机图形学和图像处理技术,将数据转换成图形或图像在屏幕上显示出来^②。随着数据新闻的发展,数据可视化成为数据新闻重要的组成部分,用以展现新闻事件背后多个维度的数据信息,使人们在理解这些信息及它们之间的关系时能够更直观,并发现其背后包含的信息。但是,数据新闻并不等同于数据可视化。

首先,可视化并不是新闻学的一个独立专业,数据可视化也并不只运用于新闻业。可视化作为一门涉及计算机图形、图像处理、计算机视觉、人机交互等多个领域的综合学科,不但广泛应用于如医学、生物、地理等领域的科学计算,而且在如金融、通信、网络等行业中信息处理方面的应用更是广泛。例如,医学领域可以用可视化技术呈现不同类型疾病在不同地域、不同年龄阶段的数据,从而得到某一地区对于哪种多发疾病针对哪些人群进行重点防治。

其次,数据可视化虽然被认为是数据新闻呈现数据、叙事的重要方式,但是并不是所有的数据新闻都使用可视化的形式呈现。西蒙·罗杰斯认为“数据新闻不是图形或可视化效果,而是用最好的方式去讲述故事,只是有时故事是用可视化效果或地图来讲述。数据也并不拘泥于可视化,它也可以以一个新闻故事的形式呈现,甚至仅仅展示数据”^③。

二、数据新闻的定义

“数据新闻”(data journalism),也被称为“数据驱动新闻”(data-driven journalism),这一概念近年来发端于新闻实践领域。西方主流大报和一些独立新闻机构设立了专门团队来设计一些新型的新闻应用,即运用各种技术软件来抓取、处理、分析和形象化呈现数据,数据呈现方式包括可视化数据图、互动图表和网络在线演示等,开启了数据新闻实践领域的第一页。如今包括中国在内的世界各国的传统媒体、新兴网站和独立新闻机构正在逐步接受数据新闻的理念,并进行相应的实践尝试。数据新闻研究的奠基之作《数据新闻手册》认为,数据新闻简单地说就是用数据做新闻,它与其他新闻的不同之处在于,将传统的新闻敏感与数据及数据技术结合后,为新闻的出现提供新的可能,这种可能可以出现在新闻生产流程的任何阶段,它可能出

① 方洁,颜冬.全球视野下的“数据新闻”:理念与实践[J].国际新闻界,2013(06).

② 好搜百科.可视化[EB/OL].<http://baike.so.com/doc/6488288.html>.

③ 章戈浩.作为开放新闻的数据新闻——英文《卫报》的数据新闻实践[J].新闻记者,2013(06).

现在人们使用软件来自动搜集或汇总来自政府、公安局和其他公民机构信息资源的过程中,也可能出现在人们使用软件发现海量数据间相关关系的过程中。在数据新闻中,记者通过简单的图表就可以表达复杂的事件,记者通过数据新闻也可以让读者更懂得新闻事件与自身的相关程度。数据既是数据新闻的资源,也是数据新闻表达的工具^①。

劳伦兹在第一届国际数据新闻圆桌会议中将数据新闻定义为一种工作流程,它主要包括以下步骤:通过反复抓取、筛选和重组来深度挖掘数据,聚焦专门信息以过滤数据,可视化地呈现数据并合成新闻故事,他认为数据新闻可被视为一个不断提炼信息的过程,在这一过程中,原始数据转换成有意义的信息,当把复杂的事实组织成条理清晰、易于理解和记忆的故事时,公众才能获取更多益处^②。

清华大学李希光教授则将数据新闻定义为在多学科技术手段条件下,把庞大的数据集中在不同变量的复杂关系及其与整个社会发展的关系,用视觉语言向公众展示,以这种更客观更友好的报道方式激发公众对公共事务的探讨与参与^③。

所谓数据新闻,即将数据及数据技术应用于新闻生产流程中,以可视化技术来呈现新闻,在数据新闻的世界里,记者不仅是信息的传达者,更是意义构建者。

三、数据新闻的功能

斯坦福大学杰夫·麦吉(Geoff McGhee)教授曾长期担任《纽约时报》等媒体记者,他于2009年至2010年间开始研究数据新闻。他指出,现在的新闻越来越多地和数据有关,媒体的责任是如何向公众解释复杂难懂的数据。数据的爆炸式增长使我们需用工具来进行分析,数据可视化专家正在开发工具帮助我们更好地理解和使用数据,记者的工作是运用数据使新闻报道更加有说服力^④。

(一)讲故事的新工具和新方法

数据新闻最重要的一项功能是使用数据可视化软件,通过统计大量的数据,帮助记者使用数据图表讲述错综复杂的故事,而这种讲故事的方式必须依赖于对大数据的分析和可视化处理,由于数据量巨大,按照传统的新闻生产方式是几乎不可能实现的。数据新闻最佳的阅读载体是交互性强的电子媒介而非传统的平面媒介。

2015年上半年,叙利亚、利比亚等中东、北非地区战乱不断、持续动荡,加上“伊斯兰国”极端组织的猖獗活动,使得大批难民外涌,源源不断的难民涌入欧洲,成为了欧洲难民危机的导火索。澎湃新闻的作品《一路向北:15天4000公里,全程记录难民欧洲逃亡路》(<http://image.thepaper.cn/html/zt/2015/refugee/index.html>)通过使用数据、地图、文字叙述、图片描述、动画等方式综合讲述了这一过程。如图1-1所示,左侧地图以动画形式,随着右侧文字的滚动,地图也不断地向前滑动。这种讲故事的新方式将错综复杂、跨越欧亚版图、历时长久的事件描述得极为清楚。

① Lucy Chambers. The Data Journalism Handbook[M]. O'Reilly Media, 2012.

② 方洁,颜冬. 全球视野下的“数据新闻”:理念与实践[J]. 国际新闻界, 2013(06).

③ 李希光,张小娅. 大数据时代的新闻学[J]. 新闻传播, 2013(01).

④ Geoff McGhee. Journalism in the Age of Data[EB/OL]. (2012-07-18). <http://datajournalism.stanford.edu/>.

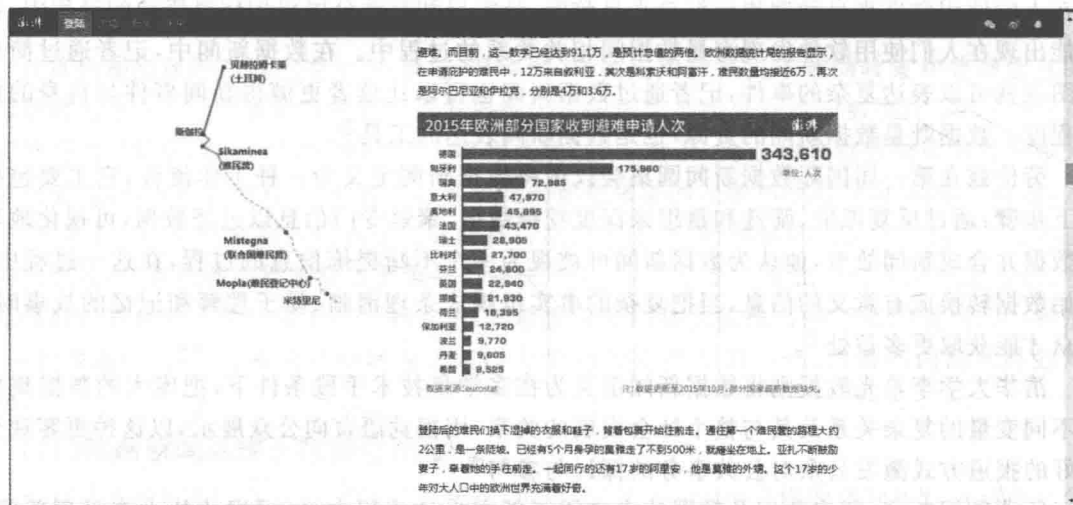


图 1-1 欧洲难民逃亡路

(二)解释宏大新闻事件与个人的关联

数据新闻可以帮助记者解释宏大背景下的新闻事件和个人之间的关联，新闻学对于报道公共事务的要求，是能够通过记者的报道来帮助读者认识到一项公共政策的实施或修订对个人造成的影响。数据新闻让读者们在阅读报道后对自己的生活提出问题，诸如，我的资金会受到什么影响等。总之，数据新闻的工作希望读者能在数据和新闻事件中找到与自身的关联，从而达到对现实政策和新闻事件的关注。

美联社的数据新闻《where do you stand》通过提出一系列问题让读者回答，根据答案判断哪个执政党适合这位读者的政治观点。

英国广播公司(BBC)和毕马威会计师事务所联合制作的《预算计算器:2012年的财政预算将如何影响你?》(Budget calculator: How will the Budget 2012 affect you?)能够帮助读者理解新的财政预算(税收计划)给个人生活带来的影响,用户只需要在界面上输入一些个人信息,它就能够自动计算出你需要为新的政府财政预算增加支付多少税,你的生活会变得更好还是更糟(见图 1-2)。

(三)记者角色的转变

通过数据的使用,记者工作的重点从“第一个报道者”转化成新闻事件的影响的阐释者。数据新闻报道的议题范围十分宽广,记者更有意义的工作是为读者提供经过定量分析的洞见,使用数据能够将抽象概念转化为普通人容易理解的事物,帮助记者讲述抽象的社会问题。记者还可以分析复杂形势中各种变量的动态关系,能够为读者预见下一次事件,指出主政机构的失误,甚至还能帮助人们寻找复杂问题的解决方案。