

国内外数据新闻研究与实践

——基于文献计量和内容分析的比较研究

周谦豪 梁佳霖 许鑫

(华东师范大学经济与管理学部信息管理系, 上海 200062)

摘要: [目的/意义] 分析和探讨国内外数据新闻发展的差异、差距及其成因, 提出促进数据新闻本土发展的建议。[方法/过程] 从数据新闻研究论文数据和竞赛作品数据等多源国内外数据出发, 使用文献计量法、内容分析法和对比分析法, 从多个角度分析比较国内国际数据新闻的学界研究和业界实践。[结果/结论] 建议从数据新闻人才培养、学术交流平台建设、新技术的接纳运用方面入手, 促进数据新闻的本土发展。

关键词: 数据新闻 精确新闻 大数据 新闻报道

分类号: G210.7

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2021.02.04

0 引言

2009年,《卫报》上线“数据博客”;2010年,首届国际数据新闻圆桌会议举办;2011年,搜狐网建立“数字之道”专栏;2012年,《数据新闻手册》出版、“全球数据新闻奖”创立;2015年,首届中国数据新闻大赛举办……一系列事件无不标志着在新闻业,数据新闻已经成为一股热潮。为了展望这股热潮的未来,有两大问题值得探讨:一是国内国际在数据新闻发展上的差距几何;二是学界研究与业界实践在主题上有哪些差异。

1 相关研究

数据新闻作为一个较“年轻”的概念,目前未有关于其定义的定论^[1]。现有的主流观点包

[作者简介] 周谦豪 (ORCID: 0000-0001-9831-4304), 男, 硕士研究生, 研究方向为科技情报, Email: qianhao.zhou@foxmail.com; 梁佳霖 (ORCID: 0000-0002-4078-3690), 女, 本科生, 研究方向为数据新闻, Email: 10164800141@stu.ecnu.edu.cn; 许鑫 (ORCID: 0000-0001-7020-3135), 男, 博士生导师, 教授, 博士, 研究方向为情报技术、信息分析, Email: xxu@infor.ecnu.edu.cn (通讯作者)。



括从生产流程^[2]、与传统新闻的差异^[3]、呈现形态和行业发展^[4]及其它角度^[5]出发,对数据新闻进行定义。其中最被认可的是 Jonathan Gary 等在《数据新闻手册》中提出的“数据新闻是用数据做的新闻”^[3]这一简洁明了的定义。虽然国内国际学界和业界人士对数据新闻的理解不尽相同,但可以归纳出,数据新闻的核心在于分析挖掘出数据背后的有价值的新信息。

除对数据新闻的定义、概念和内涵的研究,学界也充分关注数据新闻自身的发展。数据新闻的前身是20世纪60年代的精确新闻。这是一种“将社会科学和行为科学的研究手法应用于新闻实践的报道”^[6],其特点是用精确的数据和概念来分析新闻,使新闻报道更为客观、科学。20世纪90年代,随着信息时代的到来,在精确新闻的基础上又演化出了计算机辅助新闻。本世纪以来,世界逐步迈入数据时代。在大数据兴起及开放数据运动铺开的背景下,衍生出了新的新闻生产方式,即数据新闻。相关研究指出,精确新闻、计算机辅助新闻和数据新闻三者是一脉相承的,呈递进发展的关系^[7]。在国内,郭晓科^[8]、史安斌等^[9]较早地尝试对这一领域的实践发展作出总结。蒋欣欣等^[10]通过梳理国内数据新闻演变发展的脉络,认为数据新闻在我国同样经过了从精确新闻、计算机辅助新闻到数据新闻的发展阶段,但起步即落后于国外,如上世纪60年代诞生的精确新闻,直到90年代才在国内出现。唐述权^[11]基于人民日报的实践,总结了数据新闻在该媒体机构的发展历程,并对数据新闻在中国的未来发展趋势、机遇与限制等作出判断。在研究综述方面,马玉霞^[1]重点关注了国内数据新闻研究的现状,并指出研究欠缺创造性和缺乏经典研究成果两大不足;秦静等^[12]总结国内数据新闻研究现状后预测其存在四大发展趋势:注重构建理论体系、注重本土化和普及、研究视角多元化、聚焦人才培养。

为了促进国内数据新闻研究和实践的发展,本研究提出基于相关研究论文题录数据及其引证数据、竞赛作品及其元数据等国内外多源数据,使用文献计量法、内容分析法和对比分析法对国内国际的学界业界在数据新闻领域的研究和实践进行比较分析。在方法上,本研究不同于现有研究多采用案例调查法和案例分析法,因而具有一定的创新性。

2 研究方案设计

本研究选择基于数据视角,多维度比较国内、国际在数据新闻领域的学界研究与业界实践,特设计了以下研究方案。

2.1 对象选择

本研究以文献计量法、内容分析法和对比分析法为研究方法,对国内、国际及学界、业界在数据新闻领域的研究探索情况进行比较分析。文献计量法的研究对象是数据新闻研究论文;内容分析法的研究对象是数据新闻竞赛作品;对比分析法的研究对象同时囊括研究论文和竞赛作品。

2.2 数据采集

本研究以表1所示的数据来源作为获取数据的源头。为了反映数据新闻的整体情况和最新发展,在研究数据新闻作品时,不同于现有研究聚焦于国内外领军数据新闻栏目^[13],本研究选择以国内外知名数据新闻竞赛官方网站为数据来源。



周谦豪, 梁佳霁, 许鑫. 国内外数据新闻研究与实践——基于文献计量和内容分析的比较研究 [J]. 文献与数据学报, 2021, 3 (2): 047-064.

表 1 研究数据来源情况

区域	数据类型	数据来源	备注
国内	研究论文数据	中国知网 (含中国学术期刊全文数据库、中国优秀博士学位论文全文数据库等)	截至 2020 年 3 月 18 日, 以主题含“数据新闻”为检索条件, 经去除重复, 检索并获取到 2197 条文献信息。其中, 618 条文献为发表在核心期刊上的期刊论文。核心作者数据源于中国引文数据库。
	竞赛作品数据	中国数据新闻大赛官方网站	截至 2019 年底, 该大赛分别于 2015 年、2016 年、2018 年和 2019 年举办了四届比赛。本研究以四届比赛中的入围作品为研究对象。
国际	研究论文数据	Web of Science 核心合集	截至 2020 年 3 月 18 日, 以主题含“data journalism”为检索条件, 经去除重复, 检索并获取到 1175 条文献。
	竞赛作品数据	全球数据新闻奖官方网站	为了配合与国内数据新闻竞赛进行比较, 本研究仅选取 2015 年、2016 年、2018 年和 2019 年比赛中的入围作品为研究对象。

2.3 研究方法

2.3.1 文献计量法

在对国内国际数据新闻研究论文进行比较分析的过程中, 本研究采用了文献计量法。该方法是利用数学和统计学方法定量研究文献外部特征的科学研究方法^[14]。本研究在利用文献计量法进行分析的过程中, 以 CiteSpace 为工具, 以年为时间单位, 以发文量、核心作者、核心期刊、合作网络为分析维度完成计量分析。借助工具所提供的关键词共现和关键词聚类功能, 本研究对国内国际学界在数据新闻方面的研究领域和热点进行了识别。

2.3.2 内容分析法

在对国内国际数据新闻竞赛入围作品进行比较分析的过程中, 本研究采用了内容分析法。该方法是根据预先设计的编码表读取文献内容并将之有序量化表达, 以研究文献内容特征的科学研究方法^[15]。本研究以数据新闻的议题类型、呈现形式、数据源、分析工具等维度构建数据新闻竞赛作品内容分析编码表。随后, 本研究依据该表对中国数据新闻大赛和全球数据新闻奖入围作品进行内容分析和基于图表的差异呈现。

2.3.3 对比分析法

在分别和综合分析国内国际数据新闻研究论文的计量分析结果和竞赛作品的内容分析结果的过程中, 本研究采用了对比分析法。对比分析法是通过比较客观事物以达到认识事物的本质和规律、做出正确评价的方法^[16]。在综合分析时, 本研究以区域(国内、国际)为横轴、以界别(学界、业界)为纵轴构建四象限, 按领域主题、工具方法和特色特点三个方面分析和归纳上述结果, 实施进一步的对比分析。

3 国内国际论文比较

3.1 发文量

从数据新闻研究论文的发文量维度出发, 以 2000 年至 2019 年逐年在国内发表的论文、在国内核心期刊发表的论文和收录在 Web of Science 核心合集集中的论文数量为依据, 绘制成如图 1 所示的国内国际数据新闻研究论文发文量逐年趋势图。

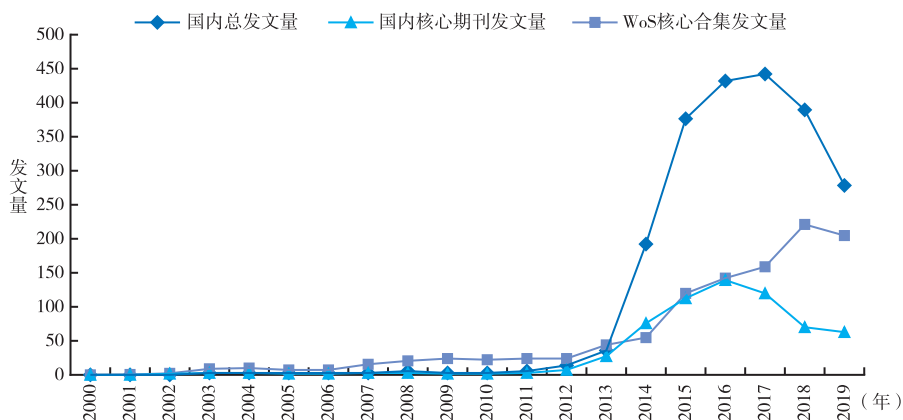


图1 国内国际数据新闻研究论文发文量逐年趋势

从图1可知,国际上数据新闻的研究先于国内起步。自2006年起,国际数据新闻研究发文量相比以往显著攀升并稳步递增,几乎同时期兴起的“开放数据运动”是推动数据新闻发展的重要动力。2012年及以后,随着全球数据新闻奖的设立和《数据新闻手册》的出版,在实践的反向推动下,国际数据新闻研究被不断地推向新的高潮,直至2019年热度有所回落。

同样从图1可知,国内数据新闻发文量自2013年起提升迅猛。这与同年《数据新闻手册》中译本出版不无关系。国内对于数据新闻的研究趋势在随后几年中快速到达过热阶段,甚至个别年份一年内核心期刊中的发文量可以匹敌当年Web of Science核心合集收录的同主题国际论文。过快的增长和过热的状态可能促使国内的数据新闻研究更早地进入到“退热”阶段,2017年发文量增速显著放缓,2018~2019年则进入回落阶段。同时,国内核心期刊发文量占国内总发文量的比重逐年降低,数据新闻研究文献越来越多地分散于非核心期刊上,这是国内数据新闻研究快速增长的必然结果。

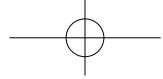
3.2 核心被引期刊

在分析数据新闻领域有影响力的期刊方面,本研究经过在中国知网中国引文数据库中进行检索,得到国内数据新闻领域的核心被引期刊数据。通过在CiteSpace工具中设置“Cited Journal”节点进行分析,得到国际数据新闻领域的核心被引期刊数据。经数据整理,国内国际排名前10的数据新闻领域核心被引期刊如表2所示。

分析表2可知,在数据新闻领域国内被引最频繁的期刊均为新闻传播学类的期刊。结合历年发文量可知,由于国内数据新闻研究起步晚于国外,在被引期刊方面暂未体现出数据新闻跨学科和学科交叉的特点;而在国际上,由于数据新闻研究历史更长,被引最频繁的期刊不仅有新闻传播学类的期刊,还有其它社会科学类期刊,具有鲜明的跨学科和学科交叉特点。从被引绝对次数来看,由于国内研究的发文量逐年快速增长、后来居上,因而国内核心被引期刊的被引频次显著高于国际核心被引期刊。

3.3 核心作者

在分析数据新闻领域有影响力的作者方面,本研究经过在中国知网中国引文数据库中进行检索,得到国内数据新闻领域的核心作者数据。通过在CiteSpace工具中设置“Cited Author”节点



周谦豪, 梁佳霖, 许鑫. 国内外数据新闻研究与实践——基于文献计量和内容分析的比较研究 [J]. 文献与数据学报, 2021, 3(2): 047-064.

进行分析, 得到国际数据新闻领域的核心作者数据。经数据整理, 国内国际排名前十位的数据新闻领域核心作者如表 3 所示。

表 2 数据新闻领域国内国际核心被引期刊排名

国内排名	国内核心被引期刊	被引频次	国际排名	国际核心被引期刊	被引频次
1	新闻记者	994	1	<i>JOURNALISM STUDIES</i>	447
2	编辑之友	768	2	<i>JOURNALISM</i>	435
3	国际新闻界	746	3	<i>JOURNAL PRACTICE</i>	371
4	新闻与写作	737	4	<i>NEW MEDIA & SOCIETY</i>	299
5	现代传播	704	5	<i>DIGITAL JOURNALISM</i>	298
6	青年记者	414	6	<i>JOURNAL OF COMMUNICATION</i>	246
7	新闻研究导刊	363	7	<i>MEDIA CULTURE & SOCIETY</i>	190
8	中国记者	363	8	<i>THESIS</i>	180
9	新闻界	279	9	<i>POLITICAL COMMUNICATION</i>	163
10	科技传播	263	10	<i>JOURNALISM & MASS COMMUNICATION QUARTERLY</i>	160

表 3 数据新闻领域国内国际核心作者

国内排名	国内核心作者	作者所在机构	被引频次	国际排名	国际核心作者	作者所在机构	被引频次
1	方洁	中国人民大学	680	1	Lewis SC	美国俄勒冈大学	160
2	喻国明	北京师范大学	493	2	Deuze M	荷兰阿姆斯特丹大学	133
3	颜冬	中国人民大学	355	3	Schudson M	美国哥伦比亚大学	125
4	文卫华	北京交通大学	349	4	Anderson CW	英国利兹大学	112
5	王斌	中国人民大学	337	5	Singer JB	英国伦敦大学	108
6	李冰	北京交通大学	321	6	Hermida A	美国哥伦比亚大学	98
7	刘义昆	中国地质大学	295	7	Hanitzsch T	德国慕尼黑大学	89
8	杨雅	北京师范大学	283	8	Carlson M	美国明尼苏达大学	87
9	章戈浩	澳门科技大学	283	9	Shoemaker PJ	美国雪城大学	84
10	彭兰	清华大学	281	10	Tuchman G	美国康涅狄格大学	83

分析表 3 可知, 在数据新闻领域国内核心作者主要集中于首都高校, 特别是中国人民大学。在被引频次位居前十位的作者中, 有三名作者来自该校; 其余作者中, 又有三名作者具有该校教育背景, 在一定程度上反映了中国人民大学的学者是国内学界研究数据新闻的主力军。中国人民大学对数据新闻领域的较早涉猎、相关领军学者的集中以及领先的新闻学学科排名相辅相成^[17], 共同促成了这一结果。在国际上, 鉴于数据新闻的前身源于并最先风靡于美国, 而《纽约时报》《华盛顿邮报》和《卫报》等当前最具代表性的数据新闻生产者大多属英美媒体^[18], 因此国际核心作者均为欧美学者, 以英、美学者为主力。此外, 与核心被引期刊的被引频次类似, 由于国内研究在量上后来居上, 国内核心作者的被引绝对次数显著高于国际核心作者。

3.4 作者关系

作者间的合作关系及其构成的合作网络同样是文献计量中作者分析所关注的部分。本研究通过在 CiteSpace 工具中设置“Author”节点并利用分析功能, 分别得到国内国际数据新闻研究论



利用可视化聚类功能, 分别得到如图 4、图 5 所示的国内国际数据新闻研究关键词聚类图。该工具以模块度 (Modularity Q) 和平均轮廓系数 (Mean Silhouette) 两项指标衡量聚类效果。模块度反映聚类的模块性, 模块度越接近于 1, 聚类效果越好。平均轮廓系数反映聚类集群的同质性, 平均轮廓系数越接近于 1, 聚类效果越好。对国内数据新闻研究关键词聚类的模块度为 0.8198, 平均轮廓系数为 0.8186; 对国际数据新闻研究关键词聚类的模块度为 0.7482, 平均轮廓系数为 0.7298。聚类效果均良好。



图 4 国内数据新闻研究论文关键词聚类

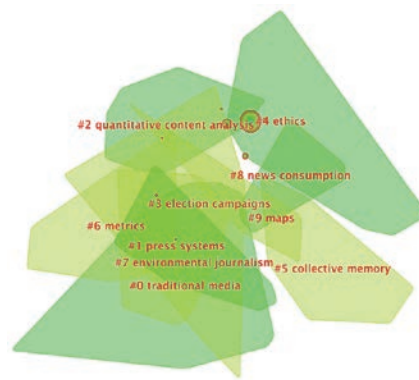


图 5 国际数据新闻研究论文关键词聚类

进一步分析图 4、图 5, 并对聚类结果按理论研究、应用研究、发展研究及人才研究进行归类, 结果如表 4 所示。由表 4 可知, 国内数据新闻研究主要集中在应用方面, 特别是对报道形式的研究, 而理论、发展、人才研究虽然存在, 但相对较少。相比之下, 国际数据新闻研究内容更全面。基于关键词聚类的具体比较结果如下:

①理论研究。国际研究在本方面的努力包括对数据新闻概念、原则和基本流程的探索。其中一项经典研究是通过 90 余项数据新闻学术研究和业界讨论的内容进行分析、归纳和总结, 得出对数据新闻尚无统一认识的结论^[19]。受国内研究起步晚、历史短的影响, 国内同类研究相对国际还处于概念引入与介绍层面, 以梳理归纳国际数据新闻研究为主, 创新性、权威性较欠缺, 各研究趋同性较强。

②应用研究。国际研究在本方面的努力主要为基于案例分析解读数据新闻生产的基本操作, 这与国外数据新闻已进入大量生产阶段, 且新闻媒体积累了丰富的相关经验密不可分。其中一项经典研究是前《卫报》员工 Simon Rogers 编写的《数据新闻大趋势》。基于《卫报》“数据博客”的实践, 其撰写了详细的案例, 分析解读数据新闻生产流程中的操作。起步相对较晚的国内研究针对国际经典数据新闻案例进行大量具体分析。其中一项经典研究是以《卫报》的数据可视化为例, 分析数据新闻的应用情况和给传统媒体带来的挑战与机遇^[20]。因国内数据新闻的实践相对而言不如国际成熟, 应用也相对欠缺, 国内研究偏好选取的对象以《卫报》等国外领军媒体为主, 存在研究对象同质性强的问题。

③发展研究。由于数据安全、数据伦理和隐私保护问题近年来日益凸显, 国际研究在本方面的努力主要为对数据新闻潜在问题 (如新闻伦理) 的反思与对其未来趋势的预测。鉴于上述问题

的全球性，国内在本方面的研究重点也囊括前者，此外包括对数据新闻冲击传统新闻的思考等。其中一项经典研究是针对我国数据新闻与公民隐私间的冲突、数据新闻生产与传统新闻的融合的思考和对策^[21]。

④人才研究。鉴于数据新闻率先发展于国外且人才储备比较充分，国际新闻界更应聚焦于数据新闻人才的运用和管理，国际研究在本方面的努力主要围绕新闻系统展开。国内研究主要针对从业人员素养和人才培养问题展开，主流观点是新闻从业人员必须调整自我、不断学习、提高素养以适应数据时代的新闻传播发展需要。如有国内研究指出，数据新闻编辑应当同时具备新闻素养、数据素养和职业精神^[22]；高校新闻专业应重视培养具备数据挖掘能力的复合型人才^[23]，这与国内数据新闻人才匮乏的现状是密不可分的。

表4 基于关键词聚类的国内国际数据新闻研究内容比较

内容	国内	国际
理论研究	大数据	quantitative content analysis、metric、environmental journalism
应用研究	“一带一路”、新闻报道、《新闻联播》、网站、调查性报道、媒体、可视化报道、视觉修辞、“两会”大数据	election campaigns、collective memory、maps
发展研究	/	ethics、news consumption
人才研究	/	press systems

4 国内国际竞赛比较

4.1 赛事规模

如图6所示，从举办时间和历届参赛作品数量来看，全球数据新闻奖诞生于国际数据新闻发展成熟之后，且作为全球性竞赛，从设立之初就受到关注，至今其赛事规模已显著扩大。中国数据新闻大赛设立之时，正处数据新闻引入国内后的快速发展阶段（从文献计量分析亦可发现），其赛事规模增长迅猛，参赛作品数量近乎成指数型增长（2017年因故未举办赛事）。最近一届赛事规模已可匹敌全球数据新闻奖。

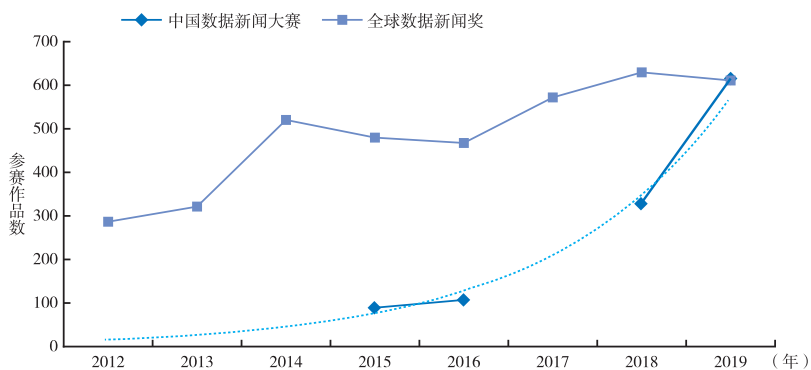


图6 全球数据新闻奖和中国数据新闻大赛的历届参赛作品数量比较



周谦豪, 梁佳霁, 许鑫. 国内外数据新闻研究与实践——基于文献计量和内容分析的比较研究 [J]. 文献与数据学报, 2021, 3(2): 047-064.

4.2 团队规模

如图 7 所示, 全球数据新闻奖的入围团队平均规模显著大于中国数据新闻大赛。后者规模最大的入围团队为 14 人, 而前者则有 370 人以上的团队参赛并入围。团队的规模在一定程度上决定了新闻调查和数据处理的能力, 较大规模的团队也能容纳更多领域的人才, 但组建大规模、跨领域数据新闻团队的前提是存在充足的人才储备, 且其他领域的相关人才愿意进入数据新闻领域。这正是国内数据新闻业界目前欠缺的条件, 也是国内赛事入围团队规模较小的主要原因。

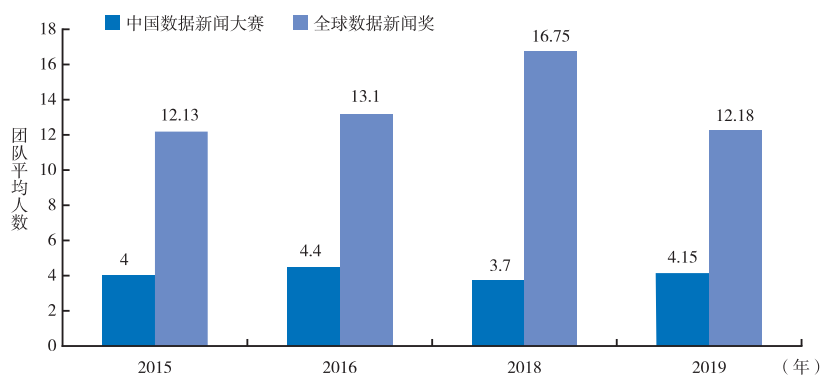


图 7 全球数据新闻奖和中国数据新闻大赛入围团队规模 (平均人数) 的比较

4.3 团队类型

如图 8 所示, 中国数据新闻大赛的入围作品主要来自学界团队, 少量来自业界团队; 全球数据新闻奖的入围作品则基本来自业界团队。究其原因, 是由于国外数据新闻发展早、应用早, 拥有数据新闻团队或具有数据新闻生产经验的媒体更多, 随着赛事规模、团队规模的增加, 支持基础和实践经验相对不足的学界团队被挤出入围行列; 而国内虽然已有部分媒体展开对数据新闻应用的探索, 但近八成媒体暂无数据新闻团队, 因而业界的实践不如国外普及, 参赛团队多来自于开设了数据新闻课程的高校等。可以说, 国内学界团队的参赛水平能够代表国内数据新闻实践的普遍水平。

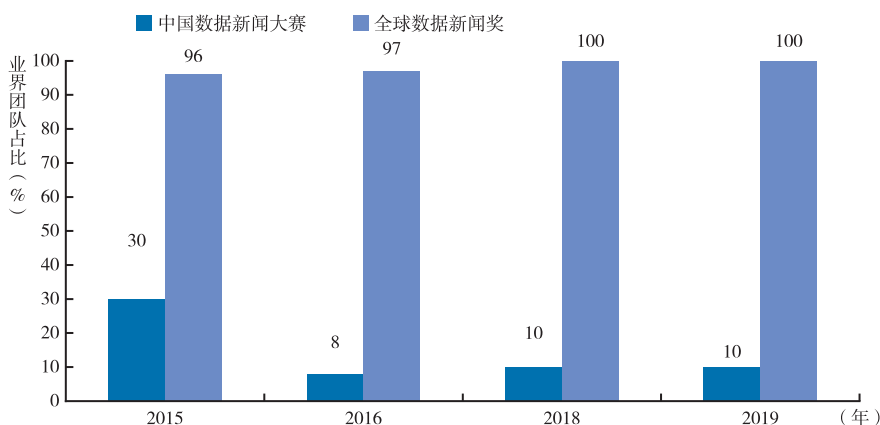


图 8 全球数据新闻奖和中国数据新闻大赛入围团队中业界团队占比的比较

4.4 作品可访问

不同于可通过纸张、广播、电视、网页中的一种或多种媒介传播的传统新闻，依赖动态、可交互的可视化功能的数据新闻一般都需要以网页为媒介。在瞬息万变的互联网上，网页的可访问是一个重点问题，依赖网页的数据新闻必须重视这个问题。如图9所示，早期中国数据新闻大赛的入围作品因距今时间较长、网站或网页缺乏维护等原因而失效甚至不复存在，这对数据新闻概念和内容的传播是一种阻碍。鉴于类似的原因，相同的问题也发生在全球数据新闻奖的入围作品上。后者在访问方面的问题还包括权限设置对于新闻获取的阻碍。

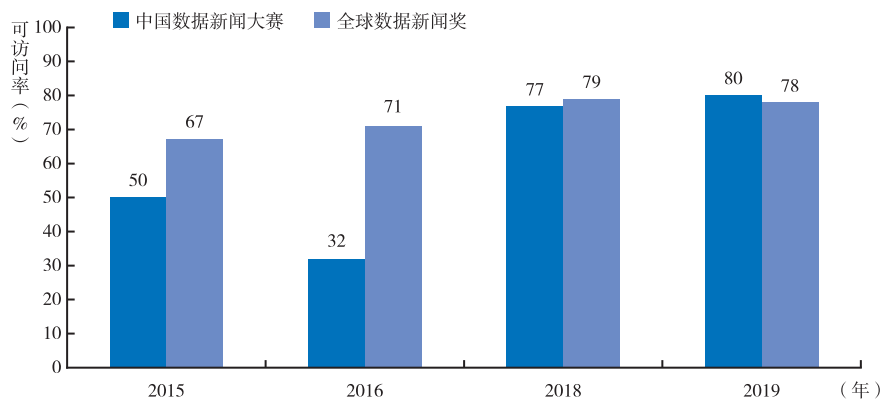


图9 全球数据新闻奖和中国数据新闻大赛入围作品可访问率的比较

4.5 议题分布

本研究根据两大赛事入围作品的内容特征，将其议题内容分为5大类进行编码：①政治经济；②文化；③社会；④教育；⑤环境。其中，文化议题涵盖历史、体育、娱乐等内容；社会议题涵盖社会保障、医疗健康、法律与犯罪等内容。

通过反映两项赛事入围作品议题分布比较的图10可知，两项赛事的入围作品在议题分布上具有较强的相似性，表现在相同的议题占比排序（政治经济、社会、文化、环境、教育）。进一步分析发现，虽然两项赛事入围作品都青睐政治经济议题，但中国数据新闻大赛的此类作品以宏观报道为主，表现为围绕时政话题（如“一带一路”）对相关的统计数据整理归纳；全球数据新闻奖的同类作品则以微观报道为主，表现为围绕政治经济领域的某一具体事件深入挖掘。这一现象源于国内外在政经报道上的话题、角度选择不同。在综合比较部分，本研究进一步探讨了其中的成因。

4.6 数据源

本研究根据两大赛事入围作品的数据源性质，将其数据源分为5大类进行编码：①自主获取；②研究机构；③政府（或政府间机构）；④企业（或非政府组织）；⑤媒体。其中，自主获取包括自主调查、自主采访和自主抓取数据等；研究机构包括科研院所、高校、智库、独立学者等；政府（或政府间机构）包括政府及政府间组织机构（如世界贸易组织）；企业（或非政府组织）包括企业及非政府组织（如红十字会）。

如图11所示，每一类数据源的百分比代表在特定赛事所有入围作品中，有多少比例的作品



周谦豪, 梁佳霁, 许鑫. 国内外数据新闻研究与实践——基于文献计量和内容分析的比较研究 [J]. 文献与数据学报, 2021, 3(2): 047-064.

使用了该类数据。鉴于政府数据的权威性，政府（或政府间组织）是全球数据新闻奖和中国数据新闻大赛入围作品采用最多的数据源。鉴于国外政府数据申请公开的特点，全球数据新闻奖作品较多利用了向政府申请公开的数据。由于我国重视大数据发展，近年来稳步推动公共数据资源开放，建设政府数据资源共享开放工程^[24]，中国数据新闻大赛作品较多利用了由政府主动公开的数据。两项赛事差异较大的方面在于对研究机构和媒体提供的数据利用，在这方面全球数据新闻奖入围作品远远少于中国数据新闻大赛入围作品；而在自主获取数据方面，中国数据新闻大赛入围作品远远少于全球数据新闻奖入围作品。前者是由于国内研究机构和新闻媒体的可信度高，因此成为中国数据新闻大赛入围团队青睐的数据源；后者反映了中国数据新闻大赛入围团队对于自主采集数据的积极性不如全球数据新闻奖入围团队，这可能源于国内数据新闻团队在数据人才方面的欠缺，从而导致在数据采集技术上稍欠火候。

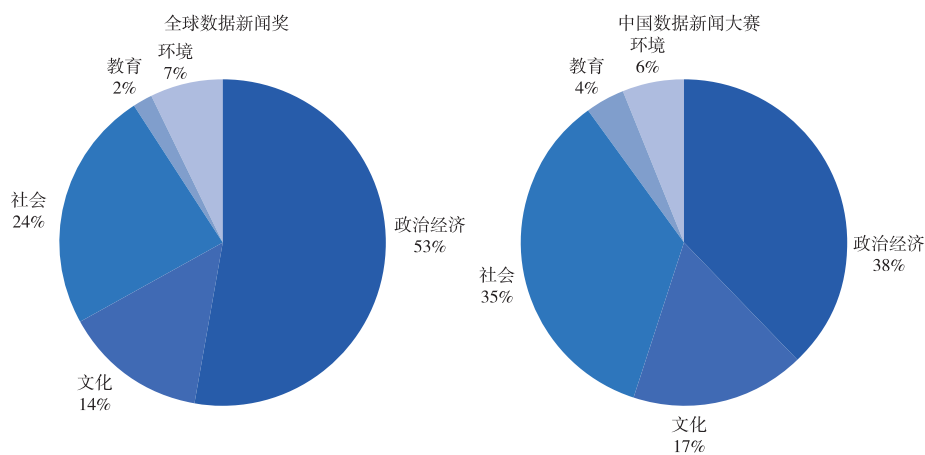


图 10 中国数据新闻大赛和全球数据新闻奖入围作品议题分布比较

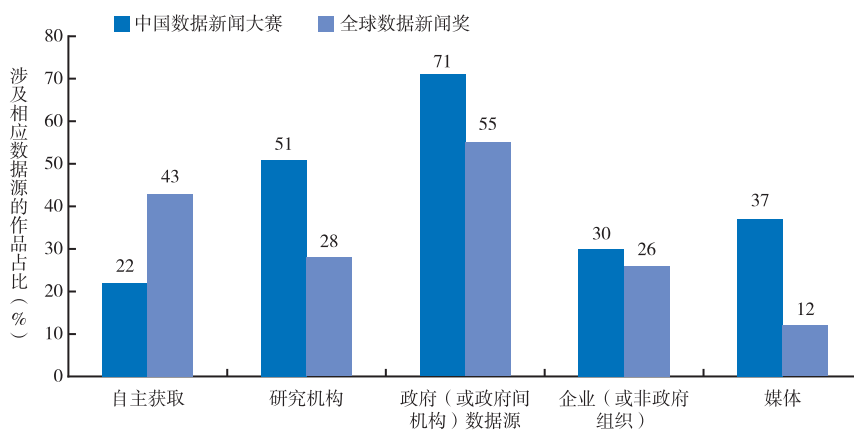


图 11 全球数据新闻奖和中国数据新闻大赛入围作品使用数据源的比较

如图 12 所示，中国数据新闻大赛入围作品所使用的数据源比全球数据新闻奖入围作品往往更多。进一步分析则发现，中国数据新闻大赛入围作品使用的数据源多但数据量小，全球数据新

闻奖入围作品使用的数据源少但数据量大。结合先前分析，其潜在原因在于中国数据新闻大赛入围团队偏好引用数据，为了确保数据新闻的全面性和准确性，势必要利用多源数据进行相互补充和交叉验证；而数据人才的匮乏、数据技术的欠缺则限制了入围团队所能利用数据量的上限。对于全球数据新闻奖入围团队而言，由于国外数据新闻人才供应更充足、团队背景构成更丰富，其更偏好自主获取一手数据并自行加工处理，能够深入挖掘更大量的数据。

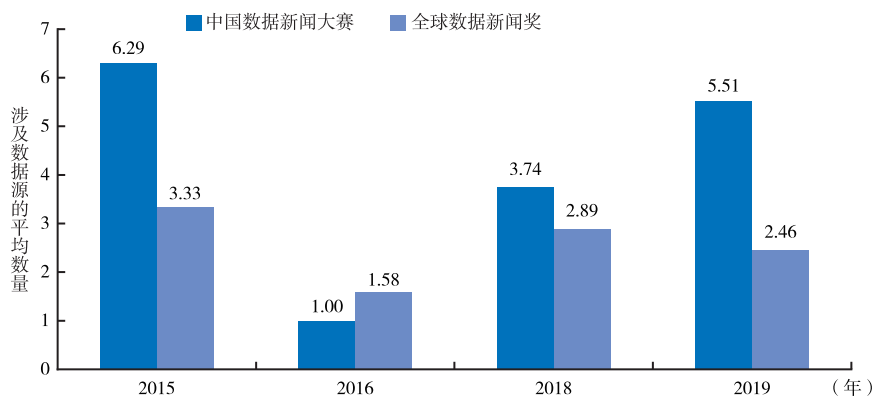


图 12 全球数据新闻奖和中国数据新闻大赛入围作品使用数据源的平均数量的比较

4.7 工具使用

数据处理是进行数据新闻报道的核心技能之一。为了对数据进行采集、清洗、分析、挖掘和可视化，必须熟练使用各类工具、软件。如图 13 所示，早期中国数据新闻大赛入围团队完成一件作品使用的工具、软件较为单一化。但是，最近一届赛事中入围团队平均使用的工具、软件数量有了飞跃式的提升，对数据处理和呈现效果的影响显著。其背后的原因离不开国内数据新闻的逐步成熟、人才队伍的逐渐健全和技术水平的不断提高。全球数据新闻奖入围团队虽然一直使用多样化的工具、软件，但近几年来平均完成每件作品使用的工具、软件数量略有下降，在工具选择方面变得更聚焦。工具在减少的同时，质量却在提高，反映出国际上对数据新闻从注重形式逐步转向注重内核，是从成熟走向更加成熟的标志。

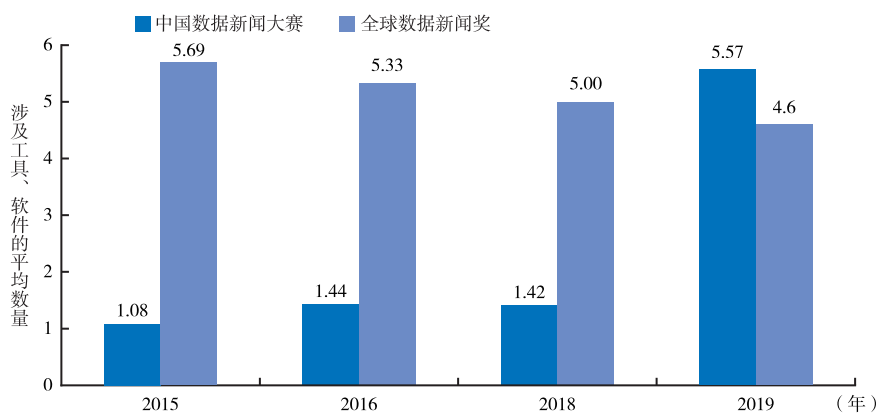


图 13 全球数据新闻奖和中国数据新闻大赛入围团队使用工具、软件的平均数量比较



周谦豪, 梁佳霁, 许鑫. 国内外数据新闻研究与实践——基于文献计量和内容分析的比较研究 [J]. 文献与数据学报, 2021, 3 (2): 047-064.

本研究对两项赛事所使用的工具、软件进一步分析、统计, 得到常用软件、工具比较结果, 如表 5 所示。其中, 中国数据新闻大赛入围团队最常使用的是日常最普及的电子表格软件 Excel, 其可满足数据处理、分析、可视化的一条龙流程。Web 前端语言 JavaScript 则成为全球数据新闻奖入围团队最常使用的工具, 服务于数据新闻的前端呈现。热门编程语言 Python 则是两项赛事中团队都常用于数据分析挖掘的语言, 但全球数据新闻奖入围团队还会选择有着统计学背景并具备强大可视化功能的 R 语言完成作品。鉴于国内数据新闻生产者的教育、工作背景多离不开属于人文社科领域的新闻学(业), 因此更倾向于选择通用、易获取、易上手、图形化的工具、软件。而在国外, 由于数据新闻团队背景构成更丰富, 具有数据专业背景的人才会促使团队倾向于选择专业性更强的工具、软件。

表 5 全球数据新闻奖和中国数据新闻大赛入围团队常用工具、软件的比较

排名	中国数据新闻大赛		全球数据新闻奖	
	工具或软件	使用频次	工具或软件	使用频次
1	Excel	62	JavaScript	71
2	镝数	22	D3.js	65
3	Python	17	HTML	48
4	HTML	10	Python	45
5	Tableau	9	Excel	40
6	Echarts.js	6	R	35
7	Adobe Photoshop	4	Adobe Illustrator	29
8	Adobe Illustrator	3	jQuery	25
9	Highcharts	2	CSS	24
10	D3.js	2	QGIS	23

4.8 呈现形式

本研究根据两大赛事入围作品的可视化形式, 将其呈现形式分为 4 大类进行编码: ①全文字; ②动态交互图表; ③静态图表; ④动画。

图 14 反映了两项赛事入围作品呈现形式的比较, 每一类呈现形式的百分比代表在特定赛事所有入围作品中, 有多少比例的作品采用了该类呈现形式。中国数据新闻大赛入围作品的呈现形式有多样化倾向, 即同一作品会采用多于一种可视化形式搭配呈现。但是, 其一般模式以静态信息图表为主, 少量动态交互图表和动画为辅, 动态、交互式设计普遍比较简单。全球数据新闻奖入围作品倾向于以动态交互图表搭配静态图表。其中, 以动态交互为主导、以静态图表为辅助, 交互式设计多丰富新颖。究其原因, 是数据新闻的呈现形式在一定程度上受制作时所选择的工具限制。结合工具使用情况, 不难看出以使用 Excel 为绝对主流的中国数据新闻大赛入围团队, 其作品主要呈现形式必然是以静态统计图为主的各类图表。而以编程语言作为主流工具的全球数据新闻奖入围团队, 其动态交互图表作品必然更为多见。这背后反映的则是国内国际数据新闻团队的技术实力差异。全文字的静态作品是全球数据新闻奖入围作品所独有的。此类作品不注重数据可视化, 而一概采用文字进行数据的解读和信息的传达。国内数据新闻实践相当重视数据可视化, 因此没有产生同类作品。

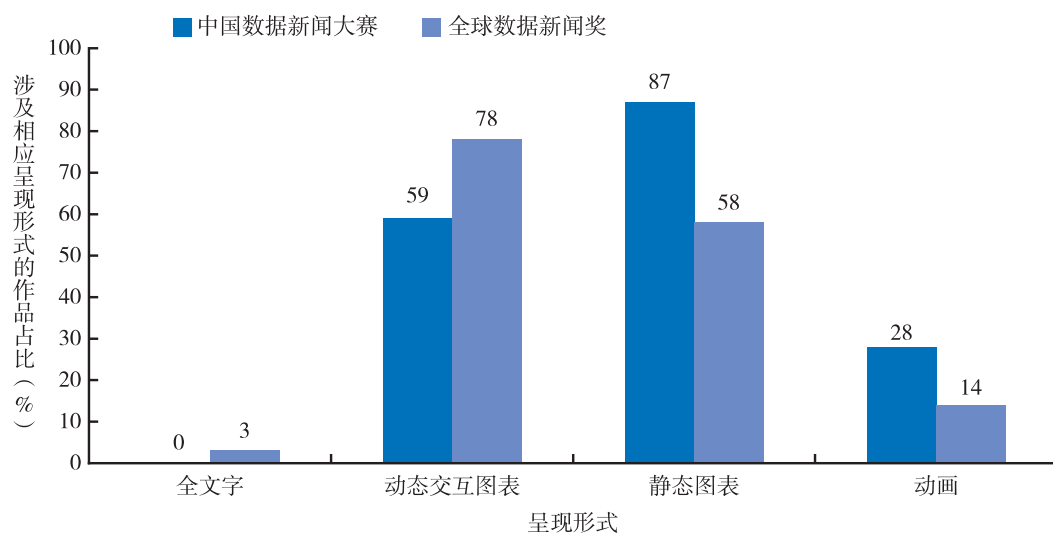


图 14 全球数据新闻奖和中国数据新闻大赛入围作品呈现形式的比较

5 研究与实践的综合比较

上述研究基于国内国际数据新闻研究论文和竞赛作品数据，结合利用对比分析法、文献计量法和内容分析法，分别得出了数据新闻领域研究与实践的国内国际初步比较结果。为了进一步探明数据新闻领域国内学界、业界与国际学界、业界的差距，本研究以国内—国际为横轴、以学界—业界为纵轴，建立国际学界、国内学界、国内业界、国际业界四个象限，归纳上述比较结果，对国内国际数据新闻研究与实践进行进一步比较。

5.1 领域、主题

从领域、主题方面对国内国际数据新闻的研究与实践建立四象限进行归纳分析，比较结果如图 15 所示。

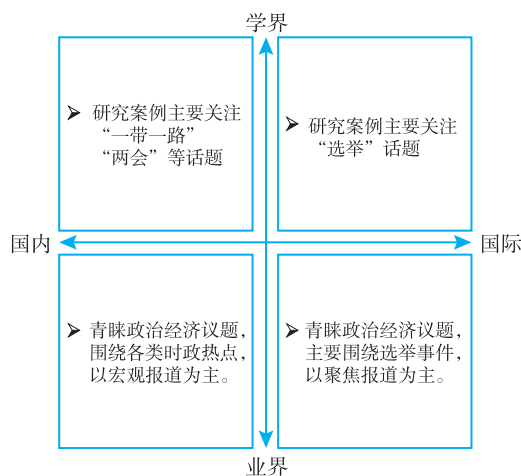


图 15 领域、主题方面国内国际数据新闻研究与实践的四象限比较



由图 15 可知, 国内国际数据新闻无论是学界研究还是业界实践, 最为关注的都是政治经济领域, 但在话题选择、报道角度上则存在不同, 产生这样的差别主要可归为三方面原因:

首先是国内外政治结构、制度体系不同。选举成为国外数据新闻报道的首选话题, 与西式选举制度“在一定意义上被等同于西式民主政治”^[25]是分不开的。西方国家的选举特别是其国家元首的选举进程往往漫长而复杂, 背后利益牵扯错综复杂, 给媒体以深挖曝光的机会。相比近年来低效日益凸显的西方代议选举制度, 以选举民主和协商民主二者有机结合的中国特色社会主义民主政治制度规避了西式选举的种种乱象和弊端^[26-27], 由此国内媒体更加着眼于国家发展和改革动态及人类命运共同体构建等话题。

其次是虽然国内近年在开放数据建设, 特别是政府数据开放方面有了长足进步, 但是开放数据中普遍存在数据粒度大、格式不便处理等问题。而企业的具体数据往往涉及其核心经营状况, 且企业一般将数据视为财产, 要使企业真正开放数据乃至基于其生产新闻报道困难重重。因此在目前的开放数据环境下, 只能催生更多的政治经济领域宏观数据新闻报道, 难以具体聚焦。

最后是国内在数据新闻全面型、复合型人才方面的欠缺。新闻人才数据素养不足, 数据人才极少参与新闻生产。由此, 在数据新闻生产过程中对数据的挖掘普遍停留在浅层次。虽然产出一些宏观报道较为容易, 但可能导致数据新闻泛泛而谈、价值缺失。

5.2 工具、方法

从工具、方法方面对国内国际数据新闻的研究与实践建立四象限进行归纳分析, 比较结果如图 16 所示。

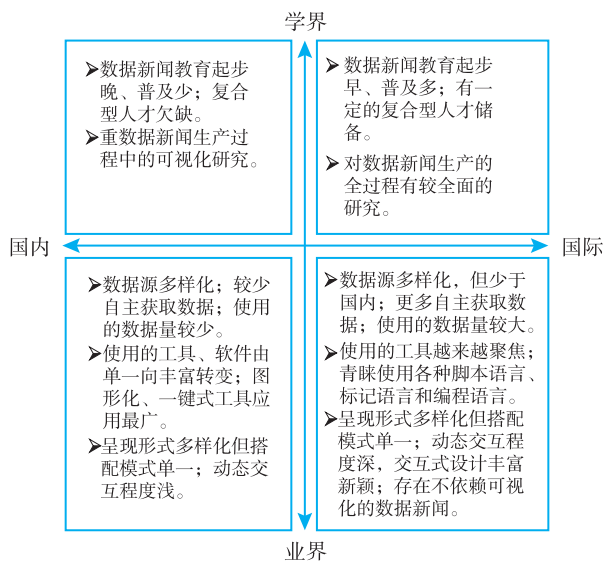
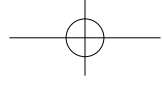


图 16 工具、方法方面国内国际数据新闻研究与实践的四象限比较

由图 16 可知, 由于国内在数据新闻人才培养、全过程研究方面的暂时落后, 数据新闻制作中对相关工具、方法的应用与国际存在显著差距。此外, 值得注意的是, 国内数据新闻可视化应



避免走入重表现、轻内容和重效果、轻交互的怪圈。

5.3 特色、特点

从特色、特点方面对国内国际数据新闻的研究与实践建立四象限进行归纳分析，比较结果如图 17 所示。

由图 17 可知，由于国内在数据新闻研究和实践的起步上晚于国际，因此无论是从研究体系的完备性还是从报道产出的成熟性方面考量，目前国内的水平尚不能与国际相提并论。前者的主要问题是研究层次浅和研究主题集中，如：理论研究以梳理和引用国际研究为主；应用研究在案例选择上具有同质性，创新性有待加强；研究热点集中但不同研究之间联系松散、缺乏紧密的逻辑支撑；研究的指导意义还有进一步提升的空间。与此同时，国际数据新闻研究已经形成较完整的体系。对于后者而言，国内数据新闻实践本质上多为多源数据融合下的多媒体化精确新闻；国际实践则多为基于海量数据挖掘、分析和可视化下的数据新闻。

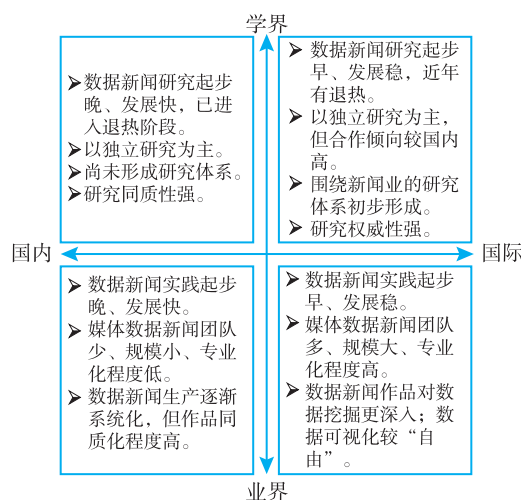


图 17 特色、特点方面国内国际数据新闻研究与实践的四象限比较

6 结语

我国历来重视新闻媒体的作用，进入数据时代后，又引领全球趋势重视大数据技术，这是数据新闻概念虽引入国内较晚但发展迅速的主要原因。在本土化数据新闻蓬勃生长的同时，必须注意到国内数据新闻领域受数据资源、人才培养、技术水平等方面的限制，依然与国际水平存在比较明显差距。为了通过数据新闻助力讲好中国故事，提升我国的国际新闻话语权，加快构建和完善中国特色新闻传播学学科体系、学术体系及话语体系，本研究提出以下三方面建议：

一是建设专业化的数据新闻人才队伍。建设专业化的人才队伍有两种途径：一是依托高校新闻传播类和数据科学类、计算机类、信息科技类专业间的合作，直接向社会输出全面型、复合型人才，实现一人多能；二是向新闻业引入数据类人才，建设融合多种类不同方面人才的数据新闻团队，实现人尽其能的合作。



周谦豪, 梁佳霁, 许鑫. 国内外数据新闻研究与实践——基于文献计量和内容分析的比较研究 [J]. 文献与数据学报, 2021, 3(2): 047-064.

二是设立平台, 促进数据新闻学术交流。促进学术交流有两种途径: 一是在相关期刊中设立有关数据新闻的专门栏目集中刊登研究成果, 或在有条件的情况下创办以数据新闻为主题的专门学术期刊; 二是举办数据新闻学术会议或研讨会, 以更好地适应数据新闻与信息技术密切相关、研究迭代速度较快的特点。

三是紧密拥抱 5G、人工智能时代, 加强新技术的接纳和落地。5G 作为下一代信息通信技术, 可以为数据新闻带来海量、实时、一手的素材, 又可以将形式多样的数据新闻报道实时传递出去, 新闻传播的时间延迟、空间限制再次缩小。人工智能依赖数据支撑, 又反过来服务于大数据, 因此人工智能的发展对数据新闻是天然利好。用好人工智能, 首先可以免去数据新闻工作者的一些重复初级劳动, 其次可以辅助新闻选题、采编和报送, 还可以避免报道和宣传产生偏差或错误^[28]。鉴于我国在 5G、人工智能领域具有先发和先天优势, 随着 5G+ 数据新闻、AI+ 数据新闻等的发展, 我国数据新闻的明天是光明的。

【参考文献】

- [1] 马玉霞. 数据新闻的兴起与发展文献综述 [J]. 新闻世界, 2015(06): 210-212.
- [2] EUROPEAN JOURNALISM CENTER. Data-driven journalism: what is there to learn? A paper on the data-driven journalism roundtable held in Amsterdam on 24 August 2010 [C]. Amsterdam: European Journalism Center, 2010.
- [3] GARY J, BOUNEGRU L, CHAMBERS L. The data journalism handbook [M]. Sebastopol, CA: O'Reilly Media, Inc., 2012.
- [4] 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, 35(06): 73-83.
- [5] 申琦. 数据新闻报道的伦理困境与出路 [J]. 青年记者, 2018(28): 16-17.
- [6] 喻国明. 从精确新闻到大数据新闻——关于大数据新闻的前世今生 [J]. 青年记者, 2014(36): 43-44.
- [7] 祝建华. 数据新闻的前世今生 [EB/OL]. (2014-07-10) [2020-05-10]. <http://media.people.com.cn/n/2014/0710/c386639-25265252.html>.
- [8] 郭晓科. 数据新闻学的发展现状与功能 [J]. 编辑之友, 2013(08): 87-89.
- [9] 史安斌, 廖蝶尔. “数据新闻学”的发展路径与前景 [J]. 新闻与写作, 2014(02): 17-20.
- [10] 蒋欣欣, 张超. 数据新闻在中国的发展阶段及其特点研究 [J]. 传播与版权, 2014(09): 11-13.
- [11] 唐述权. 让数据说话, 提升新闻价值——人民网数据新闻发展综述 [J]. 传媒, 2016(14): 9-13.
- [12] 秦静, 潘玉. 数据新闻国内研究综述 [J]. 新媒体与社会, 2015(03): 11-28.
- [13] 张雪婷. 国内外财经类数据新闻平台对比研究——以财新“数字说”和彭博“graphics”栏目为例 [J]. 文化与传播, 2019, 8(05): 97-105.
- [14] 陈忠海, 董一超. 统计分析法与文献计量法在档案学研究中的应用分析 [J]. 北京档案, 2017(03): 12-15.
- [15] 陈维军. 文献计量法与内容分析法的比较研究 [J]. 情报科学, 2001(08): 884-886.
- [16] 查先进, 杨凤. 基于对比分析法的专利情报分析实证研究 [J]. 图书馆论坛, 2008, 28(06): 193-197+204.
- [17] 教育部学位与研究生教育发展研究中心. 全国第四轮学科评估结果公布 [EB/OL]. (2017-12-28) [2021-05-31]. <http://www.cdgdc.edu.cn/xwyyjsjyxx/xkpgjg/>.
- [18] 陈力丹, 李熠祺, 娜佳. 大数据与新闻报道 [J]. 新闻记者, 2015(02): 49-55.
- [19] CODDINGTON M. Clarifying journalism's quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting [J]. Digital Journalism, 2015, 3(3): 331-348.
- [20] 章戈浩. 作为开放新闻的数据新闻——英国《卫报》的数据新闻实践 [J]. 新闻记者, 2013(06): 7-13.



- [21] 张玉. 对数据新闻的冷思考 [J]. 青年记者, 2016(33): 38-39.
- [22] 孟笛. 论数据新闻编辑素养 [J]. 中国出版, 2018(02): 56-60.
- [23] 刘方方. 我国数据新闻的发展现状与困境 [J]. 视听, 2019(12): 217-218.
- [24] 国务院. 促进大数据发展行动纲要 [EB/OL]. (2015-08-31) [2021-05-31]. http://www.gov.cn/gongbao/content/2015/content_2929345.htm.
- [25] 佚名. 西方选举制度真相 [EB/OL]. (2010-04-25) [2020-08-16]. https://www.guancha.cn/Rural/2010_04_25_60989.shtml.
- [26] 杨青芝. 中国特色社会主义民主政治制度优势探析——与西方资本主义民主模式比较 [J]. 社科纵横 (新理论版), 2012, 27(03): 1-2.
- [27] 张振波. 中国特色社会主义民主政治的优势 [EB/OL]. (2018-03-21) [2020-08-16]. <http://www.rmlt.com.cn/2018/0321/514384.shtml>.
- [28] 曹为鹏. 人工智能科技在新闻业的应用探究 [J]. 科技传播, 2017, 9(12): 6-7.

Research and Practice of Data Journalism in China and Abroad: A Comparative Study Based on Bibliometrics and Content Analysis

ZHOU Qianhao LIANG Jiaji XU Xin

(Department of Information Management, Faculty of Economics and Management,
East China Normal University, Shanghai 200062, China)

Abstract: [Purpose/significance] This research analyzed and explored the differences, gaps and causes of the development of data journalism at home and abroad. Suggestions were given to promote the development of data journalism in China. [Method/process] Based on multi-source data of research articles and competition works, bibliometrics, content analysis, and comparative analysis were used to compare domestic and international research and practices of data journalism from multiple perspectives. [Result/conclusion] The cultivation of data journalism talents, the construction of academic platforms, and the application of new technologies were proposed to promote the development of data journalism in China.

Keywords: Data journalism; Precision journalism; Big data; News report

(本文责编: 孔青青)