

Application and practice of big data in news reporting

Wurina

Alashan Right Banner Media Fusion Center, Alashan, Inner Mongolia, 737300, China

Abstract

With the rapid advancement of information technology, particularly the widespread application of big data, the news media industry is undergoing unprecedented transformation. Big data provides rich information sources, precise user profiles, and efficient content distribution mechanisms for news reporting, significantly enhancing the intelligence level and dissemination effectiveness of news production. This article, grounded in big data technology, systematically explores its specific application paths and practical outcomes in areas such as news selection, content editing, audience analysis, communication strategies, and public opinion monitoring. The study finds that through big data analysis, news organizations can transition from 'intuitive judgment' to 'data-driven' practices, improving the accuracy and timeliness of reports while also presenting new challenges to news ethics and content authenticity. The article aims to provide theoretical support and practical references for news media to promote digital transformation in the new era.

Keywords

big data; news reporting; intelligent collection and editing; public opinion analysis; media practice

大数据在新闻报道中的应用与实践

乌日娜

阿拉善右旗融媒体中心，中国·内蒙古阿拉善 737300

摘要

随着信息技术的迅猛发展，特别是大数据技术的广泛应用，新闻传播行业正经历前所未有的变革。大数据为新闻报道提供了丰富的信息来源、精准的用户画像和高效的内容分发机制，显著提升了新闻生产的智能化水平与传播效果。本文以大数据技术为研究基础，系统探讨其在新闻选题、内容采编、受众分析、传播策略与舆情监测等方面的具体应用路径与实际成效。研究发现，借助大数据分析，新闻机构能够实现由“感性判断”向“数据驱动”的转型，在提升报道精准度和时效性的同时，也为新闻伦理和内容真实性提出了新的挑战。文章旨在为新闻媒体在新时代背景下推动数字化转型提供理论支持与实践参考。

关键词

大数据；新闻报道；智能采编；舆情分析；传媒实践

1 引言

在数字化浪潮推动下，媒体行业正面临由传统模式向智能化、数据化转型的深刻变革。大数据作为信息时代的核心技术之一，不仅对商业、金融、医疗等行业产生深远影响，也正在重塑新闻传播的基本逻辑和运作模式。传统新闻采编依赖记者主观判断和有限线索来源，而大数据技术则通过对海量信息的采集、处理与挖掘，为新闻生产注入新的活力，构建起“数据驱动—精准传播—实时反馈”的新型传播机制。因此，有必要系统研究大数据在新闻报道中的实际应用路径，探讨其带来的结构性变革与应对机制，以期为传媒行业的转型发展提供理论依据和现实借鉴。

2 大数据技术的基本特征

大数据作为信息时代的基础资源，具备“体量大（Volume）、速度快（Velocity）、类型多（Variety）、真实性（Veracity）”等核心特征。这一“四V”特征不仅是其技术形态的本质体现，也是其在新闻传播领域展现赋能潜力的关键依据。第一，“体量大”意味着数据来源的广泛性与深层性。传统新闻工作依赖有限的线索、现场采访及资料检索，而大数据时代则可通过网络爬虫技术、开放数据库、用户终端行为日志、社交媒体等多元渠道，实时采集PB级甚至EB级的非结构化与半结构化数据。海量数据的可得性极大拓展了新闻线索的发现边界，为新闻议题设置提供坚实的数据基础。第二，“速度快”强调了数据获取与处理的时效性。在社交平台与移动终端高度普及的背景下，信息生成与传播的节奏呈指数级加快。借助流式数据处理架构（如Spark Streaming、Flink等），新闻机构可以实现秒级数据采集与分析，对突发事件做出快速响应。这种实时性提升了新闻报

【作者简介】乌日娜（1982-），女，蒙古族，中国内蒙古阿拉善人，本科，中级九级（记者），从事新闻研究。

道的时效优势，有助于第一时间占领信息传播高地 [1]。

第三，“类型多”体现出大数据在信息表现形式上的多样化。除了传统的文字资料外，图片、音频、视频、传感器数据、用户互动行为等非结构化信息日益成为新闻内容的重要组成部分。以短视频平台、直播平台为例，其产生的数据维度既包括内容信息，又涵盖用户评论、点赞、转发等多种交互行为，为内容创作与传播策略提供了丰富的素材与分析路径。第四，“真实性”作为大数据应用的关键保障，关系到新闻报道的公信力与社会影响力。数据来源的权威性、

采集过程的完整性、分析方法的科学性决定了大数据支持下的新闻内容是否具备可靠性与可验证性。在此基础上，引入数据溯源机制、交叉验证算法及人工干预环节，可进一步增强新闻报道在数据层面的可信度与专业性。为进一步系统呈现大数据在新闻报道中的应用逻辑，本文构建了一个覆盖选题策划、内容采编、受众分析与传播策略的全流程路径图。如图 1 所示，大数据作为驱动引擎，贯穿新闻传播全过程，构成了“数据采集—内容生成—用户理解—传播反馈”闭环机制，有力支撑了新闻生产的效率提升与智能转型。

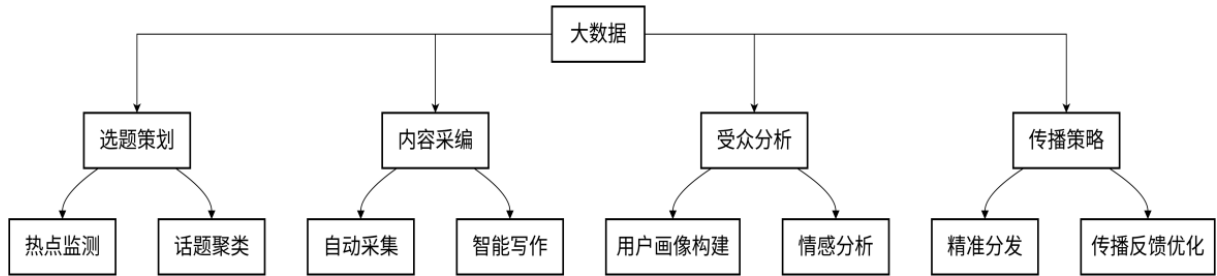


图 1 大数据在新闻报道环节中的集成应用流程图

3 大数据在新闻采编与选题策划中的应用

3.1 数据驱动的选题机制

在传统新闻报道机制中，选题策划往往依赖编辑、记者的主观经验、现场感知与直觉判断，虽然具有人文敏锐性与社会关照力，但也存在滞后性、盲目性与同质化等问题。而大数据技术的引入则为选题策划提供了量化依据与科学支撑，使选题逻辑从“人本驱动”逐步向“数据驱动”转变，实现了新闻议题的前瞻性布局与精准化选择。当前，新闻媒体普遍借助数据挖掘与社会计算工具开展多维度的信息采集与热点监测。同时，基于历史数据建模，还可以预测某一类事件的传播演化路径、公众关注生命周期，从而制定符合用户认知节奏的报道计划 [3]。

3.2 智能采编与内容生成

在新闻内容生产过程中，大数据与人工智能的融合正在改变记者的工作方式与采编流程。以往依靠人工搜集资料、编辑撰写与校对发布的多环节模式，正逐步被数据分析辅助、自动化写作与人机协同的新型机制所替代，实现了内容生产效率与质量的同步提升。在素材收集与资料分析阶段，媒体可运用信息抓取引擎（如 Octoparse、Scrapy 等）

对网络中的文本、图片、视频、政策文件等多源异构数据进行自动抓取与结构化处理，极大提升了新闻初步素材的整理效率。在此基础上，通过文本分类、实体识别、情感分析等 NLP 技术，可对内容进行语义理解与主题标注，辅助编辑准确定位新闻核心要素与事实关系。

自动化新闻写作技术已日趋成熟，特别在财经、体育、天气、股市、赛事等结构化数据密集的领域表现尤为突出。腾讯的“Dreamwriter”则已实现体育赛事报道的实时更新，从比赛数据生成完整的报道稿件，在数秒内完成内容生产与发布，大大提升了新闻即时性与覆盖能力。尽管智能化写作与采编系统为新闻机构带来显著效益，但也引发了对“记者职业去技能化”与“信息真实性削弱”的担忧。因此，未来的智能采编不应以完全替代人为目标，而应强调人机协同，借助 AI 辅助记者在数据处理、事实验证、表达策略等方面做出更为精准与高效的判断，从而保障新闻的深度、温度与价值。为更直观地展示大数据在新闻生产不同环节中的典型应用方式与技术依托，本文对选题、采编、分发、分析与舆情等核心流程进行了归纳总结。表 1 列出了各环节所匹配的应用路径与具体技术平台，体现出大数据从数据源到平台工具的全链条支撑能力。

表 1 大数据在新闻报道环节中的典型应用路径与技术示例

新闻环节	应用方式	技术手段或平台示例
选题策划	热点话题识别、趋势预测	微博热搜、百度指数、情感分析模型
内容采编	自动采集、智能撰稿	Scrapy 爬虫、Dreamwriter、NLP 写作系统
受众分析	用户画像构建、行为数据建模	推荐系统日志、深度学习点击模型
内容分发	个性化推荐、多端同步推送	今日头条推荐引擎、DSSM 算法
舆情监测	舆情波动监控、意见领袖识别	清博舆情、蚁坊软件、SNA 社交网络分析
危机应对	谣言识别与辟谣、信息溯源	辟谣宝平台、区块链新闻溯源机制

4 数据化传播策略与分发机制优化

4.1 个性化推荐与精准传播策略

在传统媒体时代，新闻内容的传播模式以“编辑决定—大众接收”为核心，内容生产与受众需求之间存在信息错配问题。而大数据技术的发展，使媒体得以通过数据挖掘和算法模型实现传播路径的精细化与个性化，从而有效提高新闻内容的触达效率与传播效果。

个性化推荐的核心在于构建“用户画像”系统，即基于用户在平台上的历史行为数据（浏览记录、搜索关键词、点击偏好、停留时长、转发频率等），结合用户的基本属性（年龄、性别、地域、终端设备等）及其社交关系网络，对用户的兴趣点进行动态建模与实时更新。这一画像不仅揭示了用户对内容的偏好模式，也为新闻分发策略提供了数据基础。在技术层面，协同过滤算法（Collaborative Filtering）、内容推荐算法（Content-Based Filtering）及基于深度学习的推荐系统（如深度神经网络推荐模型、DSSM、Wide & Deep等）被广泛应用于主流新闻客户端和门户网站中。例如，今日头条基于用户点击行为与文本相似性构建个性化内容池，实现了“去中心化”的新闻传播逻辑，即每位用户所接收到的新闻推送结果均不相同，最大程度提高用户阅读兴趣与平台粘性。

4.2 算法分发的困境与治理路径

虽然大数据驱动下的算法分发极大提升了传播效率，但其所带来的负面效应亦逐渐显现，成为当前媒体伦理与信息治理领域的核心议题之一。最为突出的问题包括“信息茧房”“回音室效应”“算法操控”与“虚假信息泛滥”等，直接挑战了新闻传播的客观性、公正性与公共性。个性化推荐虽然提升了用户满意度，却容易使用户长期处于“信息舒适区”，只接收到符合其既有认知与兴趣的信息。这种“信息茧房”效应限制了用户的信息广度，削弱了媒体作为社会公共信息平台的知识启蒙功能。同时，算法推荐也可能强化群体极化与意见分裂，加剧网络空间的对立情绪，影响社会认知结构的稳定性。算法黑箱问题导致内容推送过程缺乏透明度。用户无法理解为何会接收到某类信息，媒体也难以有效监督平台是否存在倾向性推荐或商业操控行为。一些商业平台更可能基于“点击最大化”原则推送低俗、煽情、标题党等内容，损害了主流新闻的公信力与内容生态的可持

续性。

面对以上困境，亟需从制度设计、算法优化与媒体责任等多维度构建有效的治理体系：一是推动平台算法治理的透明化。建议平台披露推荐算法的基本逻辑与可解释路径，引入“算法知情权”与“用户选择权”机制，让用户有权选择关闭推荐或手动调整偏好，提升算法系统的可控性与信任度。二是强化内容推荐的多样性机制。在推荐模型中引入“多元曝光机制”或“多样性约束函数”，确保推荐结果包含不同观点、立场与题材的内容，打破单一信息流动路径，增强受众的信息宽容度。三是提升编辑人工干预能力，建立“人机协同推荐机制”。编辑应在算法推荐结果上进行质量把控与方向校正，防止内容被完全技术化操控。同时，舆情敏感类事件应由人工主导内容审核与推送节奏，保障公共事件的传播规范性与信息权威性。四是加强算法伦理建设与新闻职业规范融合。在大数据推荐系统中嵌入伦理判断框架，引导推荐系统优先推送具备公共价值、社会责任与专业深度的内容，防止算法沦为单纯“流量驱动”的商业工具。总之，传播机制的优化不仅应追求技术效率，也需兼顾社会价值与伦理边界。唯有在提升技术赋能的同时加强治理与引导，方能实现新闻传播在数据驱动时代的“善治”。

5 结论

大数据作为信息社会的重要引擎，正以前所未有的深度介入新闻传播全过程，重塑新闻生产逻辑与传播生态。其在选题策划、内容生成、传播分发与舆情监测中的深度应用，不仅提升了新闻机构的运营效率与传播效果，也推动了媒体治理体系的现代化转型。然而，技术红利背后亦隐藏着伦理、隐私、算法等方面的复杂挑战。未来，唯有在坚持新闻专业价值基础上，实现技术与价值的融合创新，方能推动新闻传播事业在智能时代实现高质量发展。

参考文献

- [1] 钟宝坤.大数据时代新闻报道创新路径探究[J].新闻研究导刊,2025,16(12):90-93.
- [2] 纪辰.数据新闻视角下新闻媒体的发展路径[J].新闻文化建设,2025(04):100-102.
- [3] 曾琢.基于大数据分析的国际新闻报道舆论引导研究[J].新闻文化建设,2025(01):84-86.