

算法赋能与数据洞察：Python 在新闻传播学中的核心应用研究

张秋水, 杨学永*

南京航空航天大学, 江苏 南京 211100

DOI:10.61369/SE.2025110006

摘要： 大数据时代的新闻传播已进入“计算传播学”新阶段，Python 算法以其高效的数据处理能力与灵活的分析特性，成为新闻传播学研究与实践的核心工具。本文基于2018–2023年《新闻与传播研究》《国际新闻界》等CSSCI核心期刊文献，系统梳理 Python 算法在新闻传播学中的应用脉络，重点剖析其在数据采集清洗、内容特征挖掘、受众画像构建、舆论演化建模四大场景的技术路径与实践价值。结合“媒体深度融合”战略要求，总结当前应用中存在的技术门槛、伦理风险等问题，提出“技术普及 – 方法创新 – 伦理规制”的三维发展路径，为 Python 算法在新闻传播学中的规范化应用提供理论与实践参考。

关键词： Python 算法；计算传播学；数据处理；受众分析；舆论监测

Algorithm Empowerment and Data Insights: A Core Application Study of Python in Journalism and Communication

Zhang Qiushui, Yang Xueyong*

Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211100

Abstract： News communication in the era of big data has entered a new phase of "computational communication," with Python algorithms becoming a core tool for research and practice in journalism and communication due to their efficient data processing capabilities and flexible analytical features. Based on literature from CSSCI core journals such as "Journalism & Communication Research" and "Chinese Journal of Journalism & Communication" from 2018 to 2023, this paper systematically reviews the application trajectory of Python algorithms in journalism and communication, focusing on analyzing their technical pathways and practical value in four key scenarios: data collection and cleaning, content feature mining, audience profiling construction, and public opinion evolution modeling. In light of the strategic requirements for "deep media integration," it summarizes issues such as technical barriers and ethical risks in current applications and proposes a three-dimensional development pathway of "technology popularization – method innovation – ethical regulation," providing theoretical and practical references for the standardized application of Python algorithms in journalism and communication.

Keywords： Python algorithms; computational communication; data processing; audience analysis; public opinion monitoring

引言（研究背景与意义）

《关于加快推进媒体深度融合发展的意见》明确提出“运用大数据、人工智能等技术提升内容生产效率”，这一要求推动新闻传播学从“定性描述”向“定量实证”转型。传统研究方法在面对社交媒体日志、新闻传播轨迹、用户行为数据等海量非结构化数据时，往往面临处理效率低、分析精度不足的困境。Python 作为开源编程语言，其丰富的库生态（如用于数据处理的 Pandas、文本挖掘的 NLTK、机器学习的 Scikit-learn）能够实现数据采集、清洗、分析、可视化的全流程覆盖，为新闻传播学研究提供了技术支撑。

基金项目：

教育部产学合作协同育人项目：241204554182113，无人机飞行教学训练虚拟仿真；

中央高校基本科研业务费专项资金：ND2025003，智能传播时代下媒体与受众互动模式的变革方向与模式。

通讯作者：杨学永，云南昭通人，助理研究员，研究方向：媒介技术、智能传播、媒介治理。

现有研究已零散涉及 Python 算法的应用，但缺乏系统性梳理，尤其对“数据处理 - 分析建模 - 实践落地”的逻辑链条阐释不足。本文以 2018-2023 年新闻传播学 CSSCI 核心期刊为样本，筛选以“Python”“计算传播”“数据挖掘”为核心议题的高影响力文献（被引频次 ≥ 15 次），系统构建 Python 算法在新闻传播学中的应用框架，这一研究既可为学者提供技术应用参考，也能为媒体机构的智能化转型提供思路。

一、Python 算法在新闻传播学中的应用脉络与核心场景

（一）应用脉络的三阶段演进

Python 算法在新闻传播学中的应用与技术迭代、数据爆发形成共振，呈现清晰的阶段性特征：2018-2019 年为“技术引入期”，研究聚焦 Python 基础功能，以数据采集（如 Requests 库爬取新闻评论）和简单统计为主，《国际新闻界》2018 年《基于 Python 的社交媒体数据采集与分析实践》是该阶段代表成果；2020-2021 年为“深度应用期”，算法应用从数据处理延伸至内容分析，如使用 NLTK 库进行新闻文本情感计算、LDA 模型实现主题挖掘，喻国明（2020，《新闻与传播研究》）提出“Python+机器学习”的传播效果分析框架；2022-2023 年为“融合创新期”，结合深度学习的 Python 应用成为热点，如基于 BERT 模型的新闻文本分类、LSTM 模型的舆论趋势预测，实现“数据 - 模型 - 洞察”的闭环。

（二）核心应用场景：数据处理与分析的技术实践

解决“数据获取难、质量低”的痛点。在数据采集方面，Python 的 Requests 库可实现对新闻网站、微博、抖音等平台的结构化数据爬取，如《新闻大学》2021 年研究利用 Selenium 库模拟浏览器操作，获取短视频平台的用户互动数据（点赞、评论、转发）；对于 API 开放平台（如微信公众号 API），则通过 Python 构建请求接口实现数据批量获取。数据清洗环节，Pandas 库的缺失值处理（dropna() 函数）、重复值剔除（drop_duplicates() 函数）及异常值检测（箱线图分析）功能，成为提升数据质量的核心工具，如澎湃新闻数据团队使用 Pandas 清洗 2022 年两会相关新闻评论数据，将有效数据率从 68% 提升至 92%（《新闻战线》2022）。

新闻内容特征挖掘实现“文本量化 - 主题聚焦”的突破。Python 的自然语言处理（NLP）库为内容分析提供技术支撑：一是文本分词与特征提取，使用 Jieba 库对新闻标题和正文进行分词，通过 TF-IDF 算法提取核心关键词，如《现代传播》2023 年研究提取主流媒体“乡村振兴”主题报道的高频词，发现“产业”“帮扶”“基层”等核心议题；二是情感倾向分析，借助 SnowNLP、TextBlob 库计算新闻文本的情感值，如《当代传播》2022 年研究对 10 万条疫情相关新闻评论进行情感分析，发现主流媒体报道的正面情感引导作用显著；三是主题建模，通过 LDA（潜在狄利克雷分配）模型挖掘新闻内容的潜在主题，如《国际新闻界》2021 年研究使用 Gensim 库构建 LDA 模型，识别出科技新闻的“技术突破”“伦理争议”“产业应用”三大主题聚类。

受众画像与行为分析能够精准匹配“内容供给 - 用户需求”。

Python 算法通过整合用户多维度数据（基本信息、浏览记录、互动行为），构建精细化受众画像：在人口统计学特征分析上，利用 Pandas 进行描述性统计，明确受众年龄、性别、地域分布；在行为偏好挖掘上，通过 K-Means 聚类算法对用户互动数据进行分类，如《新闻与传播研究》2023 年研究将短视频平台用户分为“资讯获取型”“娱乐消费型”“社交分享型”三类，并针对性提出内容推送策略；在需求预测上，使用 Scikit-learn 库的逻辑回归模型，基于用户历史行为预测其对特定新闻主题的兴趣度，如人民日报客户端利用该模型实现新闻内容的个性化推荐，用户点击量提升 35%（《中国记者》2023）。

构建“实时监测 - 动态预测”体系展示舆论演化与风险预警。Python 算法在舆论研究中的应用集中于三大环节：一是舆论监测，通过 Scrapy 框架构建爬虫系统，实时抓取多平台舆论数据，结合正则表达式提取舆论核心事件；二是传播路径分析，利用 NetworkX 库构建舆论传播网络图，识别意见领袖与传播关键节点，如《新闻大学》2022 年研究通过该方法发现某公共事件中，媒体官微与领域专家是舆论引导的核心力量；三是风险预警，基于 LSTM（长短期记忆网络）模型对舆论情感趋势进行预测，当负面情感值超过阈值时自动触发预警，如人民网舆情数据中心的“舆论预警系统”，将风险响应时间从 24 小时缩短至 4 小时（《新闻与写作》2023）。

二、Python 算法应用的研究方法特征与学科融合价值

（一）研究方法特征：定量主导与混合研究兴起

基于核心期刊文献统计，Python 相关研究中定量研究占比达 72%，形成“数据采集 - 算法建模 - 结果验证”的标准化流程。具体方法上，机器学习法（占比 38%）、内容分析法（占比 25%）、社会网络分析法（占比 19%）为主要类型，且呈现“多方法融合”趋势，如将 Python 文本挖掘与问卷调查结合，既实现大规模文本的量化分析，又通过定性数据补充研究深度。研究主体上，高校与媒体机构的合作研究增多，如中国人民大学新闻学院与字节跳动合作开发的“短视频舆论分析系统”，体现“学界 - 业界”的技术协同。

（二）学科融合价值：推动新闻传播学方法论革新

Python 算法的应用打破了新闻传播学与计算机科学、统计学的学科壁垒，推动形成“计算传播学”交叉学科方向。在理论层面，为议程设置、沉默的螺旋等经典理论提供量化验证工具，如通过 Python 分析新闻报道与公众讨论的相关性，实证检验议程设

置理论的适用性；在方法层面，弥补传统定性研究样本量小、主观性强的不足，实现“宏观趋势 + 微观行为”的双重分析；在实践层面，为媒体融合、舆论治理提供技术解决方案，契合“智媒时代”的行业发展需求。

三、Python 算法应用的挑战与伦理争议

（一）技术与研究层面的挑战

1. 技术门槛与资源不均：Python 算法应用需要研究者具备编程能力与数学基础，核心期刊数据显示，仅 35% 的新闻传播学研究者能独立完成算法建模，部分高校因缺乏数据与算力支持，难以开展大规模研究，导致“技术鸿沟”问题。

2. 方法应用的表层化：部分研究存在“算法崇拜”倾向，盲目使用复杂模型却忽视研究问题本身，如在小样本数据中强行应用深度学习模型，导致结果过拟合；

3. 数据质量与合法性问题：网络数据存在虚假信息、隐私信息混杂等问题，部分研究因数据采集未获得授权，触及法律风险，违背《个人信息保护法》相关要求。

（二）伦理争议：技术赋能背后的价值困境

1. 算法偏见风险：Python 算法的训练数据若存在偏向性（如某类群体数据缺失），会导致分析结果失真，如性别相关新闻的情感分析中，算法可能因训练数据的性别刻板印象，产生对女性议题的情感误判；

2. 隐私侵犯问题：在数据采集集中，易过度获取用户个人信息，如抓取社交媒体用户的私人评论与地域信息，侵犯公民隐私权；

3. 媒介伦理冲突：个性化推荐算法可能导致“信息茧房”，加剧受众认知固化，与媒体的公共服务职能存在张力，这一问题在核心期刊的伦理讨论中被频繁提及（彭兰，2022，《国际新闻界》）。

四、Python 算法在新闻传播学中的应用展望与规范路径

（一）应用展望：技术迭代与场景拓展

随着 ChatGPT 等技术发展，Python 将实现“算法建模 + 自

然语言交互”的融合应用，如通过 LangChain 库连接大语言模型，自动生成新闻数据分析报告，降低技术应用门槛；针对图片、视频等非文本数据，Python 的 OpenCV 库、PyTorch 框架将实现多模态数据的统一分析，拓展新闻内容研究的边界；对于跨平台协同分析，多平台数据融合分析打破平台数据壁垒，实现对舆论的全景式监测。

（二）规范路径：技术、伦理与政策的协同治理

1. 技术普及与能力建设：高校应开设“Python 与新闻数据分析处理”相关课程，媒体机构可开展技术培训，通过“开源社区 + 案例库”建设，降低算法应用门槛；

2. 伦理准则与行业规范：建立“数据采集 - 算法应用 - 结果发布”的伦理审查机制，明确数据使用的合法性边界，如采用数据脱敏技术保护用户隐私；

3. 政策引导与学科建设：相关部门可出台计算传播学研究规范，推动“新闻传播学 + 计算机科学”的交叉学科建设，鼓励学界与业界合作，实现技术创新与价值引领的平衡。

五、结语

Python 算法不仅是新闻传播学研究的技术工具，更是推动学科方法论革新与行业智能化转型的核心动力。其在数据处理、内容分析、舆论监测等场景的应用，实现了新闻传播学从“经验判断”向“数据驱动”的转型，为媒体融合与舆论治理提供了新路径。但同时，技术门槛、数据质量、伦理风险等问题也亟待解决。

未来，新闻传播学研究者需秉持“技术赋能而非技术主导”的理念，在提升 Python 应用能力的同时，强化伦理意识与人文关怀。通过“学界 - 业界 - 政策”的三方协同，推动 Python 算法在新闻传播学中的规范化应用，使其既发挥数据洞察价值，又坚守新闻传播的公共属性，为构建智媒时代的良好传播生态提供支撑。

参考文献

- [1] 喻国明, 兰美娜. 计算传播学: 智能时代新闻传播研究的新范式 [J]. 新闻与传播研究, 2020, 27(3): 24-46.
- [2] 彭兰. 算法社会的媒介逻辑与伦理风险 [J]. 国际新闻界, 2022, 44(5): 6-25.
- [3] 张明新, 李红涛. 基于 Python 的社交媒体数据采集与分析: 方法与实践 [J]. 国际新闻界, 2018, 40(12): 115-134.
- [4] 刘海龙, 方师师. 计算传播学的方法论反思: 基于核心期刊论文的分析 [J]. 新闻大学, 2022, (6): 32-48.
- [5] 王辰瑶, 黄典林. 算法驱动的新闻个性化推荐: 效果与伦理 [J]. 现代传播 (中国传媒大学学报), 2023, 45(2): 112-119.
- [6] 中华人民共和国全国人民代表大会常务委员会. 中华人民共和国个人信息保护法 [Z]. 2021-08-20.
- [7] 中共中央宣传部, 国家广播电视总局. 关于加快推进媒体深度融合发展的意见 [Z]. 2020-09-26.
- [8] 人民日报客户端数据团队. 个性化推荐算法在主流媒体的应用实践 [J]. 中国记者, 2023, (4): 78-83.
- [9] 人民网舆情数据中心. 基于 Python 的舆论预警系统构建与应用 [J]. 新闻与写作, 2023, (7): 56-63.
- [10] McKinney W. Python for Data Analysis[M]. 2nd ed. Beijing: O' Reilly Media, 2018.