

政策文本分析文献综述

郑诗诗, 熊 蓓*

湖北汽车工业学院汽车商学院, 湖北 十堰

收稿日期: 2025年12月9日; 录用日期: 2026年1月14日; 发布日期: 2026年1月26日

摘 要

数字化时代, 政策文本作为社会价值观与政策意图的核心载体, 其分析对解构政策本质、支撑公共治理具有关键意义。但现有研究多聚焦单一模块, 缺乏“数据源-方法-应用”全链条系统性梳理。本文系统梳理领域研究进展以构建整合分析框架: 首先, 明确数据源涵盖国家及地方官方文件、专业数据库及政策舆情文本等多元类型; 其次, 解析文本预处理、内容分析法等基础技术, 纳入BERT、LDA、监督式机器学习等前沿方法应用场景; 聚焦政策主题演变、协同与冲突剖析、量化评估及多领域实践应用。研究构建技术赋能下的政策文本分析整合框架, 揭示机器学习技术对提升分析效率、深度及精准度的核心价值, 同时反思政策分析方法应用的挑战, 为未来数据质量优化、跨学科融合等方向提出参考。

关键词

政策文本分析, 内容分析法, 政策工具, 文本挖掘

A Literature Review of Policy Text Analysis

Shishi Zheng, Bei Xiong*

School of Automotive Business, Hubei University of Automotive Technology, Shiyan Hubei

Received: December 9, 2025; accepted: January 14, 2026; published: January 26, 2026

Abstract

In the digital era, policy texts serve as the core vehicle for social values and policy intentions. Analyzing them is crucial for deconstructing policy essence and supporting public governance. However, existing research often focuses on single modules, lacking systematic integration across the “data source-method-application” chain. This paper systematically reviews field research progress to establish an integrated analytical framework: First, it clarifies data sources encompassing diverse types such as national/local official documents, professional databases, and policy-related public opinion texts. Second, it analyzes foundational techniques like text preprocessing and content analysis,

*通讯作者。

while incorporating cutting-edge methods including BERT, LDA, and supervised machine learning. The study focuses on policy theme evolution, conflict analysis, quantitative evaluation, and multi-domain practical applications. By developing a technology-empowered policy text analysis framework, it reveals the core value of machine learning in enhancing analytical efficiency, depth, and precision. The research also reflects on challenges in policy analysis methodologies, providing references for future improvements in data quality optimization and interdisciplinary integration.

Keywords

Policy Text Analysis, Content Analysis Method, Policy Tools, Text Mining

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自然语言处理(NLP)、机器学习等数字技术创新推动社会科学研究进入“计算转向”新阶段,政策文本分析作为解析公共治理意图的核心范式,也随之迎来技术赋能契机。政策文本是政府及公共部门为实现治理目标制定的正式文件集合,其分析为政策制定、实施及评估提供支撑。在技术驱动下,政策文本分析从传统定性解读转向“定性 + 定量”融合研究,主题建模[1]、情感分析[2]等计算方法的引入,既提升了研究效率,又提供了政策剖析的宏微观新途径,数字技术与传统分析框架的融合还拓展了分析的广度与深度。虽然技术赋能使政策文本分析实践丰富,但现有研究仍存在局限,多聚焦单一技术或特定领域,缺乏对“技术演进 - 方法适配 - 应用落地”全链条整合,造成理论与实践需求衔接断层。在此背景下,明确文本分析在政策研究中的定位与价值、梳理领域进展意义重大,本文旨在梳理基于文本分析的政策研究领域,剖析分析方法应用现状与进展,涵盖多元技术适配逻辑、实践案例及研究挑战,最终明晰研究脉络与未来方向。

2. 政策文本分析的范围与概念

2.1. 政策文本的定义与范围

政策文本是一种复杂的政策过程产物,其规定了政治、经济、社会博弈的场域、主体与“游戏规则”[3],也是因政策活动而产生的记录文献,涵盖广泛的文献类型;其范围可划分为三个层级:核心层为各级权力及行政机关颁布的法律、法规、部门规章等官方文件;延伸层包括政策制定过程中形成的研究报告、咨询意见、听证记录等公文档案;关联层涵盖政策实施过程中产生的辩论纪要、舆论评论等政策舆情文本,其权威性与系统性使其成为政策研究的核心载体[4]。政策文本的核心价值蕴含于内容构成中,通常包含对拟解决问题和预期后果的声明[5],这种结构化表达明确了政策目标和方向,为实施提供理论基础和行动指南。因此,政策文本对于研究社会结构、政策过程、政策发展轨迹及批判性政策分析具有重要意义[6]。

2.2. 政策文本分析的核心概念

政策文本分析是一种系统的研究方法,指通过标准化方法解析政策文本的语义内涵与结构特征,既可以作为解读利益相关者在意见一致或分歧方面的重要手段[7]。宏观层面的政策分析则聚焦政策生命周期,通过整合多维度信息与理论视角,能够帮助研究者和实践者理解政策变革的原因和过程[8]。政策文

本计算是信息科学、政策科学与计算科学协同发展的实践成果[9]，其核心逻辑是通过对政策文件及政策评论的量化计算分析，使政策制定研究过程更加科学有效。

从核心分析框架来看，政策意图、政策目标与政策工具构成三维一体的解读体系，三者层层递进。政策意图反映了政策制定的初衷和方向；政策目标是政策意图的具体化，明确政策的具体成果；政策工具则是政府实施政策的首要手段，也是影响政策实施效果的关键因素[10]。该框架依托 20 世纪 60 年代兴起的政策工具理论，理论为政策文本分析提供了重要支撑[11] [12]。

2.3. 政策分析整合分析框架

结合政策文本语义场景依赖、方法局限等特征，政策文本分析依托“数据 - 技术 - 目标”三维框架展开，该框架既依托政策工具理论的经典支撑，又回应了当前政策文本分析中“语义场景依赖”“方法表面化”的现实痛点，使分析过程更具可追溯性与学术辩驳性。其具体逻辑架构见图 1：

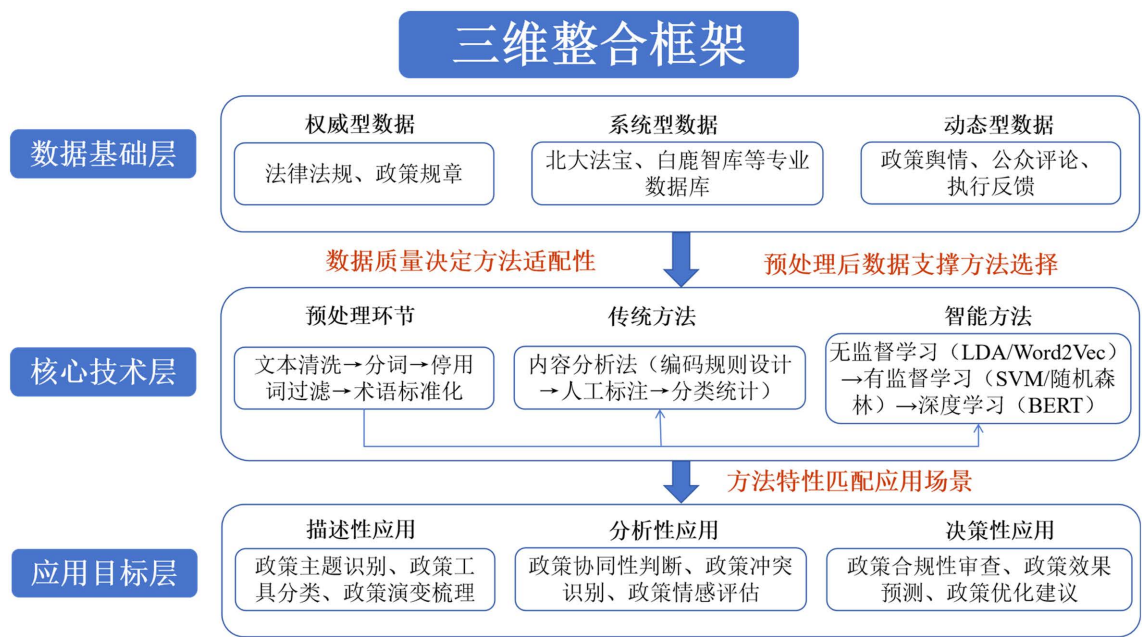


Figure 1. Three-dimensional integration analysis framework diagram of policy analysis
图 1. 政策分析三维整合分析框架图

数据基础层：三类数据互补形成矩阵——权威型数据保障合法性，系统型数据保障系统性，动态型数据保障针对性，数据特征直接限定方法选择。

方法技术层：预处理统一数据格式，传统方法与智能方法互补(前者深度解读、后者高效挖掘)，实现文本到信息的转化。

应用目标层：从“描述(是什么)”到“分析(怎么样)”再到“决策(怎么办)”，反向约束方法选择(如决策性应用需优先选择可解释性方法)。

3. 政策文本分析的方法

3.1. 文本预处理方法

数据是政策文本分析的核心基础，其可靠性与全面性直接影响研究结论的科学性。主要来源包括：国家层面的政策文本[13]，涵盖国家层面的法律法规、指导意见和规划纲要等。此外，专业数据库如北大

法宝数据库、白鹿智库政策大数据搜索平台及各个地方政府平台,也收录了大量政策文本,支持多维度检索,能为研究提供系统化数据支持。

文本预处理是政策文本计算的前置基础环节,出发点是对政策文本进行自然语言处理[14],该环节的核心步骤包括三部分:一是文本清洗(去除无关字符、标点、停用词、HTML 标签及特殊符号,同时进行大小写转换、繁简体标准化和拼写纠正);二是分词(将连续文本切分为独立的词汇单元,中文需借助 Jieba、JiebaNLP 或基于 Transformer 的模型,英文则按空格和标点分割,也可利用深度学习模型(如 LSTM)进行更精准的处理);三是词性标注(基于规则词典、序列标注模型(如 HMM/CRF)或预训练模型为分词结果标注语法类别,如名词、动词等)。此外,GPT 系列大语言模型(LLM)、RNN 等技术可适配更复杂的政策文本处理需求,进一步提升预处理精准度。

从既有研究实践来看,文本预处理的核心价值已得到充分验证。JIN Z 等[15]指出,自然语言处理(NLP)可实现文本分类、主题建模等后续分析,为政策制定提供信息提取支持;魏泽洋等[16]对生态环境准入清单政策文本进行分段、分词、去无效词等预处理,有效提取关键词;刘灵辉等[17]通过构建自定义词典、同义词表优化共同富裕政策文本分词效果;JIANG N 等[18]的研究中,对中美能源区块链政策文本实施收集、筛选、清洗等预处理,确保了后续分析的准确性。由此可见,文本预处理通过多步骤协同操作可实现数据质量提升,其处理效果直接影响后续量化分析、主题识别与趋势挖掘的可靠性,是政策文本计算中不可或缺的基础技术环节。

3.2. 内容分析法方法

政策文本多以长文本形态呈现,学者们常运用内容分析法对某一领域的政策语义和政策演化进行研究[19]。其本质是对文献内容开展定量与定性相结合的系统性语言分析[20],是一种对研究对象的内容进行深入分析,透过现象看本质的科学方法[21],最终揭示信息传递规律及背后的社会文化逻辑。在政策研究场景中,其聚焦语言特征、主题分布、政策工具使用[22]及变迁规律,核心环节为“编码-分类”,编码将文本信息转化为可分析数据,且因非介入性和可重复性降低了主观干扰,保障了研究信度[23]。

从理论框架构建来看,内容分析法的框架设计需兼顾利益相关方与政策生命周期阶段,通过适配不同场景的政策工具提升政策有效性、助力目标实现[24],而 Hellström [25]的研究为解析政策工具对政策行动的促进与限制作用提供了理论支撑。国内学者也逐步完善框架体系:杨正联[26]从“语句构成-有效性-系统性”三维度构建公共政策文本分析基础框架,为学理研究提供方法指引;范逢春[27]融合定量内容分析与质性批判话语分析,提出“价值-时间-内容”三维框架,为政策纵向比较提供了可复用路线。

在实证研究层面,内容分析法的应用价值也得到充分验证,杨志军等[28]构建“政策工具-价值链”二维框架,揭示政策工具使用偏好;黄先蓉等[29]解析网络文学治理政策工具应用逻辑;CHEN J 等[30]提出智能建造政策量化分析范式;此外,陈琤等[31]追溯 1980~2015 年 PPP 政策变迁,解释“政策热而实践冷”的制度根源;刘伟等[32]剖析 127 份科技成果评价政策,呈现科技成果评价政策的阶段特征。这些案例表明,内容分析法既能揭示政策工具与主题的使用特征,更能深度解构政策变迁背后的制度逻辑与实践挑战。

3.3. 文本内容挖掘方法

文本挖掘,也称作文本数据挖掘或文本知识发现,是指从大规模文本中提取隐藏的、先前未知的、潜在有价值的模式的过程[33]。作为政策文本分析的重要方法之一,文本挖掘方法能够通过技术手段揭示政策文本的结构特征、主题演化及潜在规律,为科学制定和优化政策提供数据支持与理论参考。

在具体研究中,文本挖掘方法在宏观长期政策分析中展现出显著适配性,张宝建等[34]借助 Rwordseg

技术提取关键词并构建关系矩阵, 通过 K-means 聚类实现政策主题分类, 从内容与性质维度多层解析主题演化过程; 祝鑫梅等[35]运用 R 语言开展词频统计与对应分析, 结合 Gephi 进行社会网络分析, 发现政策主题从基础规范向创新治理转变, 折射出政策与产业创新的协同强化趋势; 刘云等[36]则构建“创新制度 - 创新资源 - 创新主体”三维分类体系, 融合量化分析与文本挖掘方法划分四个发展阶段, 系统揭示政策演进特征并提出优化策略。上述研究均可印证文本挖掘方法在揭示长期政策演化与结构特征方面的优势。

3.4. 多方法对比

方法选择需以“研究目标 - 数据条件”的适配性为核心: 若聚焦小规模政策文本的深度解读(如地方专项政策的工具逻辑分析), 优先采用传统内容分析法; 若需处理大规模政策文本的主题挖掘、语义关联(如全国性政策的阶段演变), 可结合 LDA、Word2Vec [37]等智能方法; 而针对政策实体识别、合规性判断等精准性需求, 则适配 BERT + CRF [38]的组合模型。

同时, 单一方法的缺陷可通过“方法融合”规避: 例如用 LDA 挖掘政策主题后, 结合人工编码校验主题边界, 弥补 LDA 预设主题数的经验依赖[39]; 用 BERT 完成政策情感分类后, 通过传统内容分析法抽取典型语句, 解释情感倾向的核心依据, 破解模型的黑箱困境。

需警惕“唯技术论”的选择误区: 并非方法越前沿、模型越复杂效果越好[40]——用 BERT 分析十余篇地方政策文本, 既造成计算资源浪费, 也会因样本量不足导致过拟合; 而用传统内容分析法处理数万篇政策舆情, 则会因效率过低错失文本中的潜在关联。不同方法的具体特征、适用场景及短板, 详情见表 1。

Table 1. Comparison of main policy text analysis methods
表 1. 主流政策文本分析方法的特征对比表

| 分析方法 | 前提假设 | 核心优势 | 适用场景短板 | 数据要求 | 典型应用 |
|-------------------|-----------------------|----------------------------------|--|---------------------------------|--------------------------|
| 传统内容分析法 | 文本内容反映政策意图, 编码可标准化 | 低门槛、深度解读政策语境, 适配复杂文本 | 人工成本高(100 篇以上效率极低), 主观编码偏差难避免 | 100 篇以内政策文本, 需明确编码维度(如政策工具) | 生态环境政策工具分类; 共同富裕政策关键词提取 |
| LDA (潜在狄利克雷分配) | 文本是多个独立主题的概率混合 | 无监督学习, 无需标注数据, 主题可解释性强 n | 需预设主题数量(依赖经验), 对短文本/低频政策术语(如“双碳”新术语)敏感 | 1000+篇无标注政策文本(主题分布清晰) | 智能建造政策主题演化; 科技成果评价政策阶段特征 |
| BERT (预训练语言模型) | 语义依赖上下文双向编码 | 语义理解精度高, 适配短文本/歧义术语(如“减负”跨领域解读) | 计算成本高(需 GPU), 黑箱化严重(无法解释合规性判断依据) | 10,000+篇政策文本, 需标注样本(如情感倾向/政策类型) | APP 隐私政策合规性判断; 政策情感倾向分析 |
| Word2Vec (词向量模型) | 语义相近词汇的向量空间距离相近 | 训练高效、计算成本低, 适配大规模无标注数据 | 无法处理多义词(如“平台”在物流/科技政策中含义差异), 无法捕捉句子级语义 | 大规模无标注政策文本(术语出现频率 ≥ 5 次) | 绿色物流政策术语关联网络构建 |
| CRF (条件随机场) | 标签序列依赖上下文特征 | 序列标注精度高, 适配政策实体识别(如“政策工具”“实施主体”) | 特征工程复杂, 泛化性弱(跨领域需重新训练) | 中小规模标注政策文本(需标注实体标签) | 政策实施主体识别; 政策工具实体提取 |
| Jieba 分词(中文预处理工具) | 中文文本可通过词典/统计规则切分为词汇单元 | 轻量高效, 适配中文政策文本分词 | 对政策专业术语(如“放管服”)识别精度低, 需自定义词典优化 | 所有中文政策文本(需补充专业词典) | 共同富裕政策分词优化; 生态环境政策关键词提取 |

4. 基于文本分析方法研究的政策领域

文本分析方法已广泛渗透至水资源管理、智能煤矿、区块链、互联网医疗、物流及危机应对等多领域政策研究, 其研究对象涵盖国家与地方层面政策文本, 核心围绕政策目标、工具、实施主体、主题及效力等维度展开。

在特定政策领域应用中, 文本分析方法为政策优化提供了有力实证支撑。CHENG Z 等[41]从三维度构建分析框架, 系统评估政策实施成效与存在问题; 类似地, WO 等[42]采用 TF-IDF 关键词提取、共现网络与中心性分析, 刻画了核心模块政策的演化趋势。王发明等[43]运用 CiteSpace 工具明确区块链领域研究热点与前沿趋势。YANG 等[44]通过内容分析法和共现网络分析, 深入剖析了主题的演进特征。物流行业的政策研究中, 甘俊伟等[45]基于“政策主题-政策工具-政策效力”三维框架, 量化分析绿色物流政策特征; 龚爱清等[46]借助频数分析与共词聚类方法, 划分西部物流业政策发展阶段并总结演进重点; YIJUN 等[47]结合智能物流政策文本与 PSM-DID 模型展开实证研究, 验证智能物流政策对企业绩效的积极影响。

此外, 文本分析方法在跨域整合与跨国比较研究中也展现出显著价值。NAM 等[48]运用文本挖掘与网络分析方法, 从关键词频率、TF-IDF、n-gram 及中心度等多维度揭示疫情危机管理核心议题, 为政策制定提供参考。刘裕等[49]识别 APP 隐私政策的信息安全风险并提出协同治理策略。

5. 基于文本分析的政策研究热点领域

5.1. 政策主题的演变研究

政策主题演变是政策文本分析的研究热点, 国内外学者运用多元量化方法已形成丰富成果。从方法脉络看, 早期研究以基础定量分析为主, 完成初级编码后会对主类目内容进一步划分子类目[50]。典型研究如曹玲静等[51]采用嵌入式主题模型(ETM)结合主题相似度计算, 可视化呈现美国科技政策主题演变路径; 郑定良等[52]定量分析 1978~2008 年中国高新技术政策文本的分布规律与演变趋势; 袁定欢等[53]运用主题分析法对政策文本进行编码与分类, 梳理政策主题分布及演变趋势。该研究为理解中国快速城市化背景下的政策适应性提供了典型案例支撑, 其主题分类框架和方法也为其他领域的政策文本分析提供了参考。

此外, 学界在方法上持续创新, HUANG 等[54]提出基于“政策目标-政策工具”的分析模式, 构建政策网络, 分析政策主题的演变规律。YAO 等[55]构建了“政策工具-政策主题-演进阶段”三维框架, 量化剖析政策工具结构与阶段演变特征。从方法应用与理论构建双维度丰富了政策主题演变的研究体系, 为深度解读政策文本内涵提供了多元化分析视角。

5.2. 政策协同与冲突分析

政策协同并非静态概念, 而是多元主体参与的动态协作过程, 核心是在不打破部门边界的前提下, 通过功能互补、资源整合实现跨部门高效合作, 最终达成公共服务无缝供给; 而政策冲突则是政策体系在目标、工具等维度适配不足, 会削弱政策实施成效。诸多学者借助文本分析方法针对不同领域政策开展实证研究, 为政策协同与冲突的量化分析提供了多元范例。郑琼鸽等[56]对 2013~2023 年产教融合政策文本量化分析, 构建政策协同度量模型, 为政策优化提供实证支持。聿凌云等[57]基于政策文本量化分析, 指出中国居民节能政策存在工具协同不足与冲突偏差。ZHAO 等[58]基于 PMC-Index 模型发现公共卫生应急政策协同欠缺。李江等[59]研究者提出的政策文献计量方法为政策协同与冲突研究提供了新研究工具; JABAL 等[60]通过 RBAC [61]和 XACML 模型[62]实现政策矛盾的精准识别, 拓展了政策冲突分析的技术路径。

5.3. 政策文本的量化分析

除主题演变与协同分析外,“AI 技术 + 政策场景”的深度融合下的量化技术创新,成为政策文本分析另一研究热点。随着人工智能与机器学习发展,政策文本定量分析已形成多维度框架,可通过计量模型、文本挖掘等解析政策工具特征、规律与效能。例如,廖燕珠[63]通过网络爬虫构建政策数据库,用监督机器学习算法自动编码,比较常态与危机时期政策系统要素调整,为政策文本量化分析提供实证依据与方法借鉴;JUHÁSZ [64]使用监督式机器学习对产业政策文本进行分类,实现高效准确的政策量化;KUANG 等[65]人结合人工智能与文本挖掘技术,构建 BERT 优化模型提升教育政策分析精度和 F1 值。在政策文本量化研究的多维图景中,朱侯等[66]用 BERT 模型实现自动分类,结合 L2 归一化算完整性得分;YANG [67]利用编码 - 频率统计 - 生命周期交叉矩阵,并以可视化方式呈现政策工具与治理阶段的动态耦合。龙春晓等[68]则借助 NVivo 编码与 Ucinet 网络,揭示供给型工具独大、需求型不足、主体协同薄弱的结构性失衡,为政策文本量化研究提供跨维度范式。在多学科融合推动了政策文本计算科学体系的成型,智能技术为语义挖掘提供支撑,既提升研究效率,也增强了决策科学性[69]。

总体而言,政策文本的量化分析已形成较为成熟的方法体系,并在不同政策领域展现出较强的解释力。未来研究可进一步结合机器学习等新技术,提升文本分析的深度与广度,以更精准地指导政策优化与实践创新。

6. 政策分析方法应用的挑战与反思

政策文本分析在技术赋能下实现了效率与精度的提升,但实践中仍面临多重核心难题,需理性审视其局限与边界。

6.1. 可解释性不足制约决策适配

以 BERT 为代表的深度学习模型虽语义捕捉能力突出,但“黑箱化”特征明显。政策分析的核心价值是为决策提供可追溯、可辩驳的依据[70],而这类模型仅能输出分类、关联等结论,无法清晰阐释判断逻辑(如政策工具定性、合规性判定的核心依据)。这既降低了决策者对分析结果的信任度,也难以满足学术研究“可重复检验”的要求,限制了其在关键政策场景中的深度应用。

6.2. 数据与算法的客观性短板

数据层面,现有政策文本多依赖官方文件、专业数据库,缺乏基层执行主体、弱势群体的视角表达,存在“精英化偏向”[71];政策舆情数据又受“沉默的螺旋”影响,难以覆盖全体利益相关者诉求。算法层面,训练语料与政策公文语境的适配鸿沟、特征工程中的隐性主观选择,进一步放大偏差,导致分析结论可能偏离政策实践的真实需求。

6.3. 寓意解读易陷失真风险

政策文本的术语语义具有强烈的场景依赖性(如“减负”在教育、企业政策中内涵迥异),但主流分析方法多聚焦文本表面特征[72]。LDA 模型忽视政策历史背景与实施场景,Word2Vec 难以处理多义词的语境差异,即使是 BERT 也缺乏对政策领域专业知识的深度融合,易导致关键词提取、主题划分等基础环节失真,进而影响后续分析的准确性。

7. 结论与展望

7.1. 结论

本文系统梳理了基于文本分析的政策研究最新进展,涵盖核心概念、研究方法、热点领域及未来挑

战。近年来,从数据处理、内容分析到文本挖掘,方法的不断演进推动了对政策主题演变、协同与冲突等复杂问题的深入理解,凸显了文本分析在政策研究中的重要地位。研究表明,政策文本分析不仅是理解政策意图、评估政策效果和优化政策设计的重要工具,更在推动政策研究的科学化、系统化和数据化方面发挥了关键作用。随着自然语言处理、机器学习等技术的引入,大规模政策文本的量化分析成为可能,不仅提高了研究效率和精度,也为揭示政策运行机制、预测政策趋势以及跨领域比较提供了新的路径;同时,本文也对政策分析方法应用中的现实瓶颈与潜在风险展开反思,为技术工具的合理适配与规范应用提供参考。未来,伴随数据规范化、方法可解释性和跨学科融合的持续推进,政策文本分析将在服务政策制定与公共治理现代化中展现更广阔的前景。

7.2. 未来研究展望

7.2.1. 数据质量与可获取性

政策文本数据的质量管控与可获取性提升是未来研究面临的首要挑战。一方面,随着政策文本数据规模扩大和结构复杂化,数据的准确性、完整性和一致性成为制约分析效度的核心要素。这就需要明确如何设计适配多源异构政策文本(官方文件、舆情数据、执行记录)的智能清洗算法,实现噪声数据的自动识别与修复,在此基础上开发更高效的数据清洗与预处理算法,通过数据融合、交叉校验等技术手段整合多源数据、填补数据缺口并验证数据真实性,为高质量文本分析夯实数据基础。另一方面,政策文本数据的分散性与访问限制严重影响研究开展,当前大量政策文本散落于不同数据库与平台,部分数据存在访问权限壁垒,核心难题在于如何构建“安全-开放”平衡的政策文本数据共享机制,破解涉密数据与公共研究需求的矛盾。对此,应推动政策文本数据的开放共享,构建统一规范的数据平台,整合分散资源并提供标准化数据格式与访问接口,为研究者获取高质量数据、开展可靠分析创造条件。

7.2.2. 跨学科研究的融合

公共政策作为一个跨学科领域,位于传统学科与政府政策研究需求的交叉点。尽管公共政策的实践可以追溯到古代文明,但作为一门独立学科,公共政策在20世纪后半叶才逐渐成型,其领域文献呈现分散化特征,尚未形成广泛认可的经典文献体系,而核心文献的整合是凝聚关键概念、构建连贯理论框架的基础[73]。

鉴于政策研究横跨公共管理、政治学、经济学、计算机科学等多学科,未来的研究应强化跨学科深度融合,而非单纯的方法借用,聚焦具体可执行的研究议程推动理论与技术耦合。例如,开发面向政策因果推断的文本分析模型、构建融合法律知识图谱的政策合规性自动审查方法等,可联合公共管理与计算机团队攻关因果推断模型,融合法学与数据科学搭建合规审查方法;政治学家与经济学家也可借助文本挖掘技术,从多元视角剖析政策综合效应。此外,搭建跨学科研究平台、举办联合研讨会,将成为促进知识共享与方法创新的有效途径。

7.2.3. 新兴领域的研究

文本分析技术为探索新兴政策领域提供了有力工具。例如,在颠覆性创新政策[74]、科技人才政策协同[75]、公共数据开放[76]等领域,文本分析已展现出巨大潜力。展望未来,可持续发展政策与健康政策的研究将借助LDA (Latent Dirichlet Allocation)模型[77]和TF-IDF (Term Frequency-Inverse Document Frequency)模型[78]等先进文本分析技术,深入剖析政策文本的内涵、实施效果及其影响因素,核心需解决如何通过文本挖掘技术揭示可持续发展政策中经济、环境、社会维度的主题协同与冲突机制,以及如何构建时序文本分析模型,实现健康政策实施效果的动态预测与优化方向识别这两大问题。在全球可持续发展目标的引领下,LDA模型可挖掘政策文本中的潜在主题,清晰界定政策关注的核心议题与重点方向;

而 TF-IDF 模型则通过衡量关键词的重要性, 帮助研究者精准把握政策文本的语义结构和核心诉求。在健康政策研究中, LDA 模型可用于追踪政策主题的演变, 评估政策的适应性与前瞻性; TF-IDF 模型可通过关键词频率与权重分析, 识别政策实施的关键抓手与优化空间。上述技术的应用将为政策制定者提供科学依据, 助力实现可持续发展目标, 同时为社会的稳定与发展提供有力支撑。

基金项目

湖北汽车工业学院博士科研启动基金项目(BK202444)。

参考文献

- [1] 闫盛枫. 融合词向量语义增强和 dtm 模型的公共政策文本时序建模与演化分析——以“大数据领域”为例[J]. 情报科学, 2021, 39(9): 146-154.
- [2] Liu, H., Chen, X. and Liu, X. (2022) A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis. *IEEE Access*, **10**, 32280-32289. <https://doi.org/10.1109/access.2022.3160172>
- [3] 窦玉鹏. 文本分析在政策分析中的应用[M]. 青岛: 中国海洋大学出版社, 2024.
- [4] 裴雷, 孙建军, 周兆韬. 政策文本计算: 一种新的政策文本解读方式[J]. 图书与情报, 2016(6): 47-55.
- [5] Vogel, B. and Henstra, D. (2015) Studying Local Climate Adaptation: A Heuristic Research Framework for Comparative Policy Analysis. *Global Environmental Change*, **31**, 110-120. <https://doi.org/10.1016/j.gloenvcha.2015.01.001>
- [6] 涂端午. 中国高等教育政策制定的宏观图景——基于 1979-1998 年高等教育政策文本的定量分析[J]. 北京大学教育评论, 2007(4): 53-65, 185.
- [7] 杨慧. 社会科学研究中的政策文本分析: 方法论与方法[J]. 社会科学, 2023(12): 5-15.
- [8] Browne, J., Coffey, B., Cook, K., Meiklejohn, S. and Palermo, C. (2018) A Guide to Policy Analysis as a Research Method. *Health Promotion International*, **34**, 1032-1044. <https://doi.org/10.1093/heapro/day052>
- [9] 张涛, 蔡庆平, 马海群. 一种基于政策文本计算的政策内容分析方法实证研究——以互联网租赁自行车为例[J]. 信息资源管理学报, 2019, 9(1): 66-76.
- [10] Huang, X., Gao, J. and Zhang, H. (2022) Low-Carbon Eco-City Development Policy in China: A Bibliometric Analysis of Policy Documents. *Kybernetes*, **52**, 601-624. <https://doi.org/10.1108/k-02-2022-0230>
- [11] 王旭, 李雨晴. 我国个人信息保护政策文本量化研究及启示——基于“工具-效力-主题”三维框架的探析[J]. 情报科学, 2023, 41(1): 126-133+142.
- [12] 曾刚, 赵雪芹, 杨一凡. 政策工具视角下我国个人金融信息保护政策文本分析与对策研究[J]. 现代情报, 2021, 41(9): 84-93.
- [13] 黄如花, 温芳芳. 我国政府数据开放共享的政策框架与内容: 国家层面政策文本的内容分析[J]. 图书情报工作, 2017, 61(20): 12-25.
- [14] 张涛, 马海群, 易扬. 文本相似度视角下我国大数据政策比较研究[J]. 图书情报工作, 2020, 64(12): 26-37.
- [15] Jin, Z. and Mihalcea, R. (2022) Natural Language Processing for Policymaking. In: Bertoni, E., Ed., *Handbook of Computational Social Science for Policy*, Springer International Publishing, 141-162. https://doi.org/10.1007/978-3-031-16624-2_7
- [16] 魏泽洋, 汪自书, 宫曼莉, 等. 基于自然语言处理(NLP)的生态环境准入清单政策内容分析[J]. 环境工程技术学报, 2025, 15(1): 1-10.
- [17] 刘灵辉, 程丹, 柯子怡. 中国共同富裕政策的结构特征与演变历程——基于中央层面政策文本的分析[J]. 西南大学学报(社会科学版), 2024, 50(6): 64-80.
- [18] Jiang, N., Han, Q. and Zhu, G. (2023) A Three-Dimensional Analytical Framework: Textual Analysis and Comparison of Chinese and US Energy Blockchain Policies. *Sustainability*, **15**, Article No. 5192. <https://doi.org/10.3390/su15065192>
- [19] 汪大钊, 化柏林. 政策文本量化研究综述[J]. 科技情报研究, 2023, 5(1): 92-105.
- [20] 李钢, 蓝石. 公共政策内容分析方法: 理论与应用[M]. 重庆: 重庆大学出版社, 2007.
- [21] 邱均平, 邹菲. 关于内容分析法的研究[J]. 中国图书馆学报, 2004, 30(2): 12-17.
- [22] 张锐, 牛建华. “冷资源”变“热经济”中的政策工具组合运用: 基于国家层面冰雪产业政策文本分析[J]. 北京体育

- 大学学报, 2023, 46(3): 10-24.
- [23] 姜雅婷, 柴国荣. 安全生产问责制度的发展脉络与演进逻辑——基于 169 份政策文本的内容分析(2001-2015) [J]. 中国行政管理, 2017(5): 126-133.
- [24] Si, L., Liu, L. and He, Y. (2023) Scientific Data Management Policy in China: A Quantitative Content Analysis Based on Policy Text. *Aslib Journal of Information Management*, **76**, 269-292. <https://doi.org/10.1108/ajim-05-2022-0257>
- [25] Hellström, T. and Jacob, M. (2017) Policy Instrument Affordances: A Framework for Analysis. *Policy Studies*, **38**, 604-621. <https://doi.org/10.1080/01442872.2017.1386442>
- [26] 杨正联. 公共政策文本分析: 一个理论框架[J]. 理论与改革, 2006(1): 24-26.
- [27] 范逢春. 建国以来基本公共服务均等化政策的回顾与反思: 基于文本分析的视角[J]. 上海行政学院学报, 2016, 17(1): 46-57.
- [28] 杨志军, 耿旭, 王若雪. 环境治理政策的工具偏好与路径优化——基于 43 个政策文本的内容分析[J]. 东北大学学报(社会科学版), 2017, 19(3): 276-283.
- [29] 黄先蓉, 贺敏. 政策工具视角下我国网络文学治理政策文本分析[J]. 出版发行研究, 2021(5): 43-49.
- [30] Chen, J., Huang, M. and Liu, R. (2025) Textual Analysis of Intelligent Construction Policies from the Perspective of Policy Instruments in Fujian Province, China. *Buildings*, **15**, Article No. 1306. <https://doi.org/10.3390/buildings15081306>
- [31] 陈琤, 李丹. Ppp 政策变迁与政策学习模式: 1980 至 2015 年 ppp 中央政策文本分析[J]. 中国行政管理, 2017(2): 102-107.
- [32] 刘伟, 范旭. 基于中国创新能力演变的科技成果评价政策研究——对改革开放以来 127 份政策文本分析[J]. 科技管理研究, 2021, 41(12): 26-34.
- [33] 谌志群, 张国焯. 文本挖掘与中文文本挖掘模型研究[J]. 情报科学, 2007(7): 1046-1051.
- [34] 张宝建, 李鹏利, 陈劲, 等. 国家科技创新政策的主题分析与演化过程——基于文本挖掘的视角[J]. 科学学与科学技术管理, 2019, 40(11): 15-31.
- [35] 祝鑫梅, 余晓, 卢宏宇. 中国标准化政策演进研究: 基于文本量化分析[J]. 科研管理, 2019, 40(7): 12-21.
- [36] 刘云, 叶选挺, 杨芳娟, 等. 中国国家创新体系国际化政策概念、分类及演进特征——基于政策文本的量化分析[J]. 管理世界, 2014(12): 62-69, 78.
- [37] 马海群, 崔文波, 张涛. 我国数据安全政策文本主题挖掘及其演化分析[J]. 现代情报, 2024, 44(8): 28-38.
- [38] 谢腾, 杨俊安, 刘辉. 基于 BERT-BiLSTM-CRF 模型的中文实体识别[J]. 计算机系统应用, 2020, 29(7): 48-55.
- [39] 杨慧, 杨建林. 融合 LDA 模型的政策文本量化分析——基于国际气候领域的实证[J]. 现代情报, 2016, 36(5): 71-81.
- [40] Grimmer, J. and Stewart, B.M. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, **21**, 267-297. <https://doi.org/10.1093/pan/mps028>
- [41] Cheng, Z., Wang, N., Ouyang, R., Wang, H. and Song, Z. (2022) A Content Analysis of the Strictest Water Resources Management Policy in China. *Water Economics and Policy*, **8**, Article ID: 2250011. <https://doi.org/10.1142/s2382624x22500114>
- [42] Wo, X., Li, G., Sun, Y., Li, J., Yang, S. and Hao, H. (2022) The Changing Tendency and Association Analysis of Intelligent Coal Mining in China: A Policy Text Mining Study. *Sustainability*, **14**, Article No. 11650. <https://doi.org/10.3390/su141811650>
- [43] 王发明, 朱美娟. 国内区块链研究热点的文献计量分析[J]. 情报杂志, 2017, 36(12): 69-74, 28.
- [44] Yang, F., Shu, H. and Zhang, X. (2021) Understanding “Internet plus Healthcare” in China: Policy Text Analysis. *Journal of Medical Internet Research*, **23**, e23779. <https://doi.org/10.2196/23779>
- [45] 甘俊伟, 李欣芮, 贾璐瑜, 等. 我国绿色物流政策文本量化研究——基于“政策主题-政策工具-政策效力”三维分析框架[J]. 北京交通大学学报(社会科学版), 2025, 24(3): 83-96.
- [46] 龚爱清, 罗柳平. 我国西部地区物流业政策变迁及特征——基于政策文本量化的实证研究[J]. 商业经济研究, 2020(6): 99-102.
- [47] Liu, Y., Kim, S. and Sun, J. (2024) The Implications of Smart Logistics Policy on Corporate Performance: Evidence from Listed Companies in China. *Heliyon*, **10**, e36623. <https://doi.org/10.1016/j.heliyon.2024.e36623>
- [48] Nam, H. and Nam, T. (2021) Exploring Strategic Directions of Pandemic Crisis Management: A Text Analysis of World Economic Forum COVID-19 Reports. *Sustainability*, **13**, Article No. 4123. <https://doi.org/10.3390/su13084123>

- [49] 刘裕, 周毅, 农颜清. 网络信息服务平台用户个人信息安全风险及其治理——基于 117 个 app 隐私政策文本的内容分析[J]. 图书情报工作, 2022, 66(5): 33-43.
- [50] Zhou, W., Dai, L., Zhang, Y. and Wen, C. (2021) Personal Information Management on Social Media from the Perspective of Platform Support: A Text Analysis Based on the Chinese Social Media Platform Policy. *Online Information Review*, **46**, 1-21. <https://doi.org/10.1108/oir-06-2020-0249>
- [51] 曹玲静, 张志强. 二十一世纪以来美国科技政策主题分析及发展态势研判[J]. 情报学报, 2024, 43(5): 616-632.
- [52] 郑代良, 钟书华. 1978-2008: 中国高新技术政策文本的定量分析[J]. 科学学与科学技术管理, 2010, 31(4): 176-181.
- [53] 袁定欢, 黄小琦, 鲍海君, 等. 深圳市城中村改造政策主题及演变过程分析——基于 2009-2019 年政策文本分析[J]. 城市规划, 2021, 45(2): 92-98.
- [54] Huang, C., Yang, C. and Su, J. (2018) Policy Change Analysis Based on “Policy Target-Policy Instrument” Patterns: A Case Study of China’s Nuclear Energy Policy. *Scientometrics*, **117**, 1081-1114. <https://doi.org/10.1007/s11192-018-2899-z>
- [55] Yao, X., Hu, Y., Gong, H. and Chen, D. (2021) Characteristics and Evolution of China’s Industry-University-Research Collaboration to Promote the Sustainable Development: Based on Policy Text Analysis. *Sustainability*, **13**, Article No. 13105. <https://doi.org/10.3390/su132313105>
- [56] 郑琼鸽, 徐思, 许世建. 我国职业教育产教融合政策协同的度量模型和现实图景——基于 2013-2023 年国家层面政策文本的量化分析[J]. 职教论坛, 2024, 40(12): 30-37.
- [57] 毕凌云, 杨洁. 中国居民生活节能引导政策的效力与效果评估——基于中国 1996-2015 年政策文本的量化分析[J]. 资源科学, 2017, 39(4): 651-663.
- [58] Zhao, Y. and Wu, L. (2022) Research on Emergency Response Policy for Public Health Emergencies in China—Based on Content Analysis of Policy Text and PMC-Index Model. *International Journal of Environmental Research and Public Health*, **19**, Article No. 12909. <https://doi.org/10.3390/ijerph191912909>
- [59] 李江, 刘源浩, 黄萃, 等. 用文献计量研究重塑政策文本数据分析——政策文献计量的起源、迁移与方法创新[J]. 公共管理学报, 2015, 12(2): 138-144, 159.
- [60] Jabal, A.A., Davari, M., Bertino, E., Makaya, C., Calo, S., Verma, D., *et al.* (2019) Methods and Tools for Policy Analysis. *ACM Computing Surveys*, **51**, 1-35. <https://doi.org/10.1145/3295749>
- [61] Cruz, J.P., Kaji, Y. and Yanai, N. (2018) RBAC-SC: Role-Based Access Control Using Smart Contract. *IEEE Access*, **6**, 12240-12251. <https://doi.org/10.1109/access.2018.2812844>
- [62] Kuang, T.P., Ibrahim, H., Sidi, F., Udzir, N.I. and Alwan, A.A. (2021) An Effective Naming Heterogeneity Resolution for XACML Policy Evaluation in a Distributed Environment. *Symmetry*, **13**, Article No. 2394. <https://doi.org/10.3390/sym13122394>
- [63] 廖燕珠, 莫桂芳. 变与稳: 危机事件触发政策系统的差异化调适——基于机器学习的省级环境政策文本分析[J]. 公共管理与政策评论, 2025, 14(2): 39-54.
- [64] Juhász, R., Lane, N.J., Oehlsen, E. and Perez, V.C. (2025) Measuring Industrial Policy: A Text-Based Approach. NBER Working Paper No. 33895. National Bureau of Economic Research.
- [65] Kuang, H., Tian, P. and Liang, X. (2024) Policy Analysis Combining Artificial Intelligence and Text Mining Technology in the Perspective of Educational Informatization. *Humanities and Social Sciences Communications*, **11**, Article No. 1517. <https://doi.org/10.1057/s41599-024-04076-0>
- [66] 朱侯, 吴子帅, 韦秉东. 基于 BERT 文本分类模型的 APP 隐私政策完整性评价研究[J]. 现代情报, 2023, 43(3): 123-134.
- [67] Yang, J. and Ding, X. (2024) Textual Analysis of China’s Environmental Policies from the Perspective of Policy Instruments. *Sustainability*, **16**, Article No. 9787. <https://doi.org/10.3390/su16229787>
- [68] 龙春晓, 李承璐, 范阳东, 等. 我国慢性病防治政策文本的量化分析: 基于政策工具和政策演进及政策主体的三维框架[J]. 中国全科医学, 2025, 28(20): 2457-2463, 2500.
- [69] 胡吉明. 政策文本研究: 从内容计算到功能理解[J]. 图书情报知识, 2023, 40(4): 145-152.
- [70] Weimer, D.L. and Vining, A.R. (2017) Policy Analysis: Concepts and Practice. 6th Edition, Routledge.
- [71] 周雪光. 基层政府间的“共谋现象”——一个政府行为的制度逻辑[J]. 社会学研究, 2008(6): 1-21+243.
- [72] 单飞跃, 张玮. 经济法中的政策——基于法律文本的实证分析[J]. 社会科学, 2012(4): 9.
- [73] Fan, L.L. (2013) Canonic Texts in Public Policy Studies: A Quantitative Analysis. *Journal of Public Affairs Education*,

- 19, 681-704. <https://doi.org/10.1080/15236803.2013.12001759>
- [74] 张瑶, 张光宇. 双重视角下颠覆性创新的政策文本分析[J]. 技术经济与管理研究, 2021(11): 35-40.
- [75] 苗宏慧, 全情爽, 舒心. 基于 BERTopic 的科技人才政策文本主题识别与量化分析——以东北三省为例[J]. 现代情报, 2025, 45(4): 110-121.
- [76] 白清礼, 吕仁杰, 贺亚萍, 等. 我国公共数据开放范式研究——基于 21 份政策文本的扎根分析[J]. 图书馆, 2025(11): 17-27.
- [77] 陈彦丽, 林陶玉, 李圆圆, 等. 基于 LDA 主题模型的我国医用耗材管理政策文本的量化分析[J]. 中国卫生事业管理, 2025, 42(1): 46-50.
- [78] 邓云峰, 冯永康, 王双燕. 应急决策文本的多维语义挖掘方法——基于 TF-IDF 和 PMI 的技术框架[J]. 中国安全生产科学技术, 2025, 21(5): 36-45.