

Python 视域下的“三农”政策文本量化与演变路径研究

——以 2015—2026 年中央一号文件为例

高旭

西安外国语大学英文学院

DOI:10.32629/as.v9i6.4070

[摘要] 中央一号文件是中国农业农村政策的风向标,其文本内容的历时性演变直观映射了国家“三农”战略治理逻辑的深刻转型。面对海量且非结构化的政策文本,传统的人工定性解读难以精准捕捉高频词汇背后的量化规律。因此本文基于计算语言学视角,以2015—2026年12份中央一号文件为研究对象,构建“语料库—文本挖掘”的分析框架开展研究。研究发现,近十二年间我国“三农”政策重心呈现阶段性演进,治理逻辑存在三方面转型:贫困治理从集中攻坚转向常态化防返贫监测,粮食安全底线地位持续强化,发展动能向县域产业全链条升级下沉。因此提出完善文本挖掘政策分析方法、强化政策阶段特征识别、优化“三农”治理逻辑适配性的举措,为客观认知我国农业政策演进、提升政策解读的精准性提供数据支撑。

[关键词] 计算语言学; 文本挖掘; 中央一号文件; Python; TF-IDF算法

中图分类号: H0 **文献标识码:** A

A Python-Based Quantitative Analysis of the Evolution of China's Rural Policy Texts: Evidence from the Central No. 1 Documents (2015–2026)

Xu Gao

School of English Studies, Xi'an International Studies University

[Abstract] The Central No. 1 Document serves as a key policy guide for China's agriculture, rural areas, and farmers. Its diachronic evolution provides valuable insights into shifts in the country's rural governance strategy. Given the large volume and unstructured nature of policy texts, traditional qualitative analysis is often insufficient for identifying underlying quantitative patterns. Drawing on computational linguistics, this study examines 12 Central No. 1 Documents issued between 2015 and 2026 and develops a corpus-based text mining framework for policy analysis. The findings indicate that China's rural policy priorities have evolved through distinct stages over the past decade. Three major transformations in governance logic are identified: the transition from targeted poverty alleviation to regularized anti-poverty monitoring, the continued strengthening of food security as a strategic priority, and the increasing emphasis on upgrading county-level industrial value chains. Based on these findings, the study proposes improvements to text-mining approaches for policy research, enhanced identification of stage-specific policy characteristics, and better alignment between governance mechanisms and evolving rural development objectives. The results provide empirical evidence for understanding the evolution of China's rural policies and contribute to more accurate policy interpretation.

[Key words] computational linguistics; text mining; Central No. 1 Document; Python; TF-IDF

“务农重本,国之大纲”。农业、农村、农民问题始终是关系国计民生的根本性问题。2004年以来,中共中央、国务院连续发布以“三农”为主题的中央一号文件,这不仅是中国政府重视农业发展的政治宣示,更是指导全年乃至更长时期农业农村工

作的行动指南^[1]。这一系列文件构成了一个庞大、连续且逻辑严密的政策文本体系,客观记录了中国从传统农业向现代农业转型、从城乡二元结构向城乡融合发展迈进的历史进程。政策文本是政府意志的直接体现,对其进行科学分析是政策科学研

究的基础。早期的政策文本分析主要依靠人工编码,如美国的“政策议程项目”,虽然准确度高,但耗时耗力且难以处理大规模数据。随着计算机技术的进步,文本挖掘技术逐渐被引入社会科学研究。郭晗潇等提出的TF-IDF(词频-逆文档频率)算法是信息检索领域的经典算法,它通过加权计算有效过滤了常见词,保留了具有区分度的特征词,被广泛应用于关键词提取^[2]。在中国,近年来已有不少学者尝试利用NLP技术对历年《政府工作报告》及五年规划等宏观政策文件进行文本分析与可视化,这些研究普遍证实了政策文本中的词频变化与国家宏观战略调整之间存在显著的正相关性^[3]。然而,目前针对特定领域(特别是农业领域)的政策文本挖掘研究相对较少,且现有研究多停留在简单的词频统计层面,缺乏针对农业公文特性的算法深度适配和优化。

关于中国“三农”政策的演变,国内学界已积累了丰富的定性研究成果。孔祥智系统梳理了中国共产党百年来的“三农”政策变迁,将其划分为土地革命、社会主义改造、改革开放以及新时代乡村振兴等不同历史阶段,深刻剖析了各阶段政策制定的历史逻辑与理论逻辑^[4]。罗必良则回顾了改革开放以来中国农业发展的历程,指出了农业政策重心从“制度变革”向“要素投入”再到“绿色发展”的动力转换机制,并探讨了未来农业现代化的实现路径^[5]。在此基础上,朱华东^[6]等以1982—2024年26个以“三农”为主题的中央一号文件为研究对象,运用质性研究和政策文本量化分析法,系统考察了农田建设政策的演进脉络,发现农田建设在“一号文件”中的词频数量经历了从低频次到高频次再到低频次的过程,并长期聚焦农田水利和高标准农田建设,政策发展历经群众自主、试点探索、规范完善和全面实施四个阶段,政策工具使用也从“直接供给、全能包办”向“上下联动、多元参与”转变,为农业基础设施政策的优化提供了经验启示。

综上所述,现有研究在方法上多偏重定性解读,在时间跨度上多止步于2022年或2023年,缺乏对最新政策走向的覆盖。本文的创新点主要体现在以下三个方面。首先,时间跨度新,本研究的周期延伸至2026年,涵盖了“十五五”开局的重要节点。其次,优化方法论,不直接使用通用停用词表,而是针对公文特点构建了包含“加强”、“推进”、“切实”等行政套话的专用停用词表,显著提高了TF-IDF算法提取关键词的准确度。最后,采用时序对比视角,将十二年划分为四个具有明确战略特征的阶段,通过横向对比与纵向追踪,立体地解构了政策变迁的动态过程。

1 资料与方法

1.1 数据来源

本文选取的数据源为2015年至2026年由中共中央、国务院发布的“中央一号文件”。所有文本均来源于中华人民共和国中央人民政府门户网站及农业农村部官方网站,确保了数据的权威性与准确性。为了更清晰地捕捉政策重心的阶段性转移特征,研究依据国家五年规划周期及重大战略任务的切换节点,将12份文件划分为四个独立的时间窗口。首先,第一阶段(2015-201

7),对应“十二五”收官与“十三五”开局。这一时期的核心背景是经济发展进入新常态,农业主要矛盾由总量不足转变为结构性矛盾,核心任务是“农业现代化起步”与“供给侧结构性改革”。其次,第二阶段(2018-2020),对应全面建成小康社会的决胜期。核心任务是“脱贫攻坚决战”,确保现行标准下农村贫困人口全部脱贫。以及,第三阶段(2021-2023),对应“十四五”关键期。核心任务是“巩固拓展脱贫攻坚成果同乡村振兴有效衔接”,以及在疫情与地缘政治冲突背景下保障“粮食安全”。最后,第四阶段(2024-2026),对应“十五五”前瞻期。核心任务是全面推广浙江“千万工程”经验,建设“宜居宜业和美乡村”,并向“农业强国”目标冲刺。

1.2 文本预处理与停用词表构建

获取原始文本后,首先利用正则表达式去除文件头尾的发布日期、文号及非正文信息,并将所有文本统一编码为UTF-8格式。由于中文文本缺乏像英文那样的自然空格分隔,本文采用Jieba分词组件进行处理。

在政策文本分析中,最大的干扰源来自于高频行政套话。诸如“加强”、“推进”、“全面”、“坚持”等词汇虽然使用频率极高,但在所有年份的文档中均无差异分布,无法反映具体的政策重心。若不加处理,这些词将占据词频统计的前列,掩盖实质性内容。因此,本文在通用停用词表的基础上,结合农业政策公文特点,构建了一份包含400余个词条的领域专用停用词表。

1.3 TF-IDF关键词提取算法

为了提取各阶段最具代表性的关键词,本文摒弃了简单的词频(Term Frequency, TF)统计,采用了TF-IDF(Term Frequency-Inverse Document Frequency)加权算法。该算法的核心思想是:如果一个词条在当前文档中出现频率很高(TF高),同时在语料库的其他文档中出现很少(IDF高),则认为该词条具有很好的类别区分能力,应赋予较高的权重。其计算公式(1)如下:

$$TF-IDF = TF \times \log\left(\frac{N}{n_i + 1}\right) \quad (1)$$

其中, N 为语料库文档总数, n_i 为包含该词的文档数。

2 结果与分析

实验结果表明,经过停用词过滤和TF-IDF加权处理后,提取出的关键词能够精准概括各阶段的政策核心。

2.1 供给侧结构性改革的起步与攻坚(2015-2017)

2015年至2017年为第一阶段,该阶段正值“十二五”收官与“十三五”开局,是中国经济进入“新常态”的磨合期。根据TF-IDF算法计算,该阶段排名前列的关键词包括“农产品”、“保护”、“生态”等。

“农产品”以绝对高权重位居榜首,这一数值远超第二名,成为该阶段无可争议的核心议题。紧随其后的是“保护”、“力度”和“生态”。因此,政策高频强调“农产品”的有效供给与质量提升。同时,“生态”的高权重排名(第4位)标志着农业发展逻辑的重大转折,即从透支环境资源向保护绿水青山转变。此

外,“培育”(第9位)、“主体”(第8位)的出现,说明国家开始着力培育家庭农场等新型经营主体,以解决“谁来种地”的难题。

2.2 脱贫攻坚决战与乡村振兴战略确立(2018–2020)

进入2018年至2020年,政策关键词发生了根本性的语义偏移。这一数据结果印证了当时举国体制抓扶贫的历史事实。进入2018年,随着全面建成小康社会进入决胜期,政策话语体系发生了结构性巨变。“振兴”一词以极高权重首次跃居第一,这直接呼应了党的十九大正式提出的“乡村振兴战略”。紧随其后的是“脱贫”和“扶贫”,两者权重的叠加效应构成了这一阶段的最强音。数据客观记录了中国举国体制打赢脱贫攻坚战的历史洪流。国家试图通过“三权分置”改革,激活农村沉睡的土地资产,为乡村振兴和脱贫攻坚提供关键的要素支撑和资金来源。

2.3 乡村振兴全面推进与粮食安全底线强化(2021–2023)

随着脱贫攻坚战取得全面胜利,“三农”工作重心历史性地转移到全面推进乡村振兴上来。在本阶段,“振兴”的权重进一步攀升至0.1102,达到全周期内的峰值,表明乡村振兴战略已从规划层面走向全面落地。与上一阶段不同的是,“帮扶”首次进入前三,与排名第二的“脱贫”并存。这精准对应了“巩固拓展脱贫攻坚成果同乡村振兴有效衔接”的政策导向,即政策重心从集中性的“攻坚”转向了常态化的“帮扶”。另一个显著变化是“耕地”的权重较前两个阶段显著跃升,排位第四,结合“粮食”的回升,这反映了在疫情冲击和全球地缘政治动荡的背景下,国家将“耕地红线”和“粮食安全”视为压舱石,政策容忍度降低,保护力度空前加大。此外,“县域”(第5位)的出现,标志着城乡融合发展的落脚点开始下沉,县城成为连接城市与乡村的关键载体。

2.4 常态化帮扶机制与农业强国建设(2024–2026)

进入2024年后,中国农业农村发展进入了建设农业强国的新阶段。本阶段以“十五五”开局为战略节点,政策话语体系由脱贫攻坚、全面振兴向农业强国建设、高质量发展、常态化治理深度转型,政策重心呈现三大转向:一是防返贫治理从集中攻坚转向常态化帮扶与动态监测,“帮扶”取代“脱贫”成为核心关键词,标志减贫治理完成制度化、长效化转型;二是粮食安全底线地位持续强化,耕地保护、产能提升、种业振兴等议题权重显著提升,以“良田、良种、良机、良法”协同发力夯实粮食安全根基;三是发展动能持续向县域产业全链条下沉,以县域为城乡融合关键载体,推动乡村产业提质增效、农村一二三产业深度融合,激活乡村内生发展动力。

3 核心议题的纵向演变趋势分析

为了透视“三农”政策变迁的深层动态逻辑,本研究选取了涵盖民生、安全、经济、环境、创新与社会治理六大维度的核心关键词“脱贫、粮食、产业、生态、科技、治理”,提取其在四个历史阶段的TF-IDF权重。

最为显著的趋势变化体现在“脱贫”与“帮扶”议题的生命周期演替上。在第一阶段(2015–2017)权重较低,处于政策酝酿期;进入第二阶段(2018–2020),随着脱贫攻坚战进入决胜时

刻,该词权重瞬间飙升至0.07以上,并在第三阶段(2021–2023)的过渡期内继续维持在高位;而到了第四阶段(2024–2026),随着过渡期结束,“脱贫”一词的权重断崖式下跌归零。这种数据的“归零”并不意味着相关工作的终止,而是治理形态的升维。

4 讨论

本文通过对长周期政策文本的深度挖掘,不仅从量化层面验证了国家宏观战略叙事的历史转折,更通过关键词权重的微观变动,揭示了隐藏在文本表象之下的治理逻辑变革。这种基于计算语言学的实证研究,为理解中国公共政策的制定与演变提供了区别于传统定性分析的独特视角。

首先,针对贫困治理议题的词频演变,深刻折射了中国减贫机制从“运动式攻坚”向“制度化治理”的现代化转型。在“脱贫攻坚”阶段,政策文本体现出极强的动员性和时限性,具有“集中力量办大事”的特征;随着绝对贫困的消除,政策重心迅速转向建立防止返贫的动态监测和常态化帮扶机制。“帮扶”的高频出现,揭示了政策制定者致力于将减贫经验制度化、长效化,试图通过更精细的社会保障和倾斜政策解决相对贫困问题。这表明中国乡村治理正在从非常规的突击状态回归常规的行政状态,治理的韧性与可持续性得到了显著增强。

其次,关于“粮食”与“产业”两类议题权重的同步走高,揭示了新时期农业政策“底线思维”与“发展导向”的辩证统一。关于粮食安全与耕地保护的关注度在经历初期的去库存调整后,于后期呈现出螺旋式上升的态势。这种回归深刻反映了在逆全球化思潮抬头和地缘政治冲突加剧的外部环境下,国家安全逻辑对农业政策的深度重塑。“中国人的饭碗任何时候都要牢牢端在自己手中”已内化为政策文本中不可动摇的硬约束。同时,“产业”与“县域”权重的显著提升,则补齐了发展的拼图。这说明政策制定者深刻认识到,单纯依靠财政输血无法维系乡村的长期繁荣,必须依托“县域”这一城乡融合的关键载体,通过“产业”升级来激活乡村的内生造血能力。这种“一手抓安全底线,一手抓产业发展”的双轮驱动模式,构成了当前及未来一段时期中国农业政策的稳定内核。

最后,本研究的实证结果也充分验证了改进型文本挖掘方法在政策分析中的适用性。通过引入领域专用停用词表并优化TF-IDF模型,算法成功剥离了“加强”、“推进”等高频行政套话的干扰,敏锐捕捉到了实质性政策信号。结果证明,计算社会科学方法能够有效弥补传统定性研究难以精确度量关注度权重的缺陷,为客观认知国家战略意图提供了新的方法论支撑。当然,基于词袋模型的算法在处理复杂语义逻辑时仍有局限,未来研究可进一步引入深度学习模型,以期在更微观的粒子上解析政策意图的传导机制。

5 结论

本文立足于计算语言学视角,利用Python数据挖掘技术对2015—2026年中央一号文件进行了全景式、分阶段的量化分析。通过TF-IDF算法提取关键词并结合时序演变特征,本研究得出以下主要结论:

第一,中国“三农”政策具有鲜明的阶段性演进特征。从“十三五”时期的供给侧结构性改革,到“全面小康”时期的脱贫攻坚决战,再到“十四五”及“十五五”时期的乡村全面振兴,政策重心随着国家战略目标和主要矛盾的变化而动态调整,展现出极强的时代回应性与战略定力。

第二,“安全底线”与“内生发展”构成了新时期政策的双轮驱动。实证数据表明,粮食安全与耕地保护的战略地位在后期得到空前强化,构成了农业政策的稳定内核;与此同时,发展方式正加速转型,以“县域”为载体、以“产业”为核心的内生发展模式,已取代传统的要素投入模式,成为建设农业强国的核心动能。

第三,文本挖掘方法有效揭示了政策治理的微观逻辑。本研究证实,通过构建专用停用词表并运用TF-IDF算法,能够有效剔除行政噪声,精准捕捉到如“帮扶”机制建立、“县域”载体崛起等实质性政策信号。这证明了计算社会科学方法弥补了传统定性研究难以精确度量关注度权重的缺陷,为客观认知国家战略意图提供了新的方法论支撑。

未来研究可进一步引入深度学习模型(如BERT)进行语义网络分析,探究关键词之间的共现关系(如“产业”与“就业”

的关联),从而更立体地解构中国农业政策的传导机制与执行效能。

[参考文献]

[1]唐丽.基于内容分析法的我国“三农”政策文本量化研究[D].陕西师范大学,2021.

[2]郭晗潇.国家发展战略视角下的“三农”问题——以《人民日报》涉农社论为文本的分析[J].学习与探索,2022,(1):46-56.

[3]蔡颖.基于Python的文本数据处理研究[J].软件,2023,44(05):179-183.

[4]孔祥智.贯彻落实好新时代党的“三农”政策[J].红旗文稿,2022,(9):37-39.

[5]罗必良.中国农业现代化:时代背景、目标定位与策略选择[J].国家现代化建设研究,2023,2(1):65-78.

[6]朱华东,张蚌蚌.中国农田建设政策的演进脉络与深化改革方向——基于以“三农”为主题的中央“一号文件”文本分析[J].农业经济问题,2025,(9):38-57.

作者简介:

高旭(2000—),女,汉族,山西省临县人,硕士,西安外国语大学英文学院,研究方向:应用语言学。