

基于 LDA 模型的地方政府工作报告政策组合分析

曹权宽

(北京邮电大学经济管理学院; cqk0808@126.com)

摘要: 新质生产力的形成与发展离不开地方政府科学、有效的政策组合的支撑,为探究地方政府工作报告中政策组合关注重点及变化趋势,本文通过对全国 2010-2022 年期间地级市政府工作报告的文本分析,使用 LDA 模型及 JS 散度的计算方法,识别出地方政府工作报告中政策组合主题及其演变特征,总结政策组合主题各时期变化特点及背后可能的原因,并提出要重视地方政府政策组合的导向作用,形成科学有效的政策体系,助力我国经济高质量发展,加快形成新质生产力。

关键词: 政府工作报告; 政策组合; 文本分析

引言

以科技创新引领现代化产业体系建设,是聚焦经济建设这一中心工作和高质量发展这一首要任务的关键,习近平总书记指出:“科技创新能够催生新产业、新模式、新动能,是发展新质生产力得核心要素。”创新在我国现代化建设全局中居于核心地位,推动以科技创新为核心的全面创新,才能更好把握发展的时与势,在形成和发展新质生产力、促进经济高质量发展中实现“以进促稳”[1]。

在此过程中,政府要能够发挥积极的导向作用,地方政府工作报告是地方政府向人大及其常委会报告工作情况、总结过去工作、规划未来工作的重要文件,因此,对地方政府工作报告进行深入分析,能反映出地方政府在国家大力发展的重要领域的政策组合关注点和投入力度,把握地区经济发展的趋势和特点[2]。

1 相关文献综述

政策组合理论强调,在面对复杂的社会经济问题时,单一政策往往难以达到预期效果,需要通过一系列政策的组合和相互作用来实现政策目标。政策组合不仅涉及不同政策之间的协调,还包括政策在时间维度上的连续性和在空间维度上的协同效应。Flanagan 等认为政策组合概念的出现为我们提供了一个新的创新政策分析框架,以便更好地应对混乱复杂、多层次、多行为者的现实[3]。现有研究关于政策组合的概念还没有达成一致,早期的学者将政策组合定义为相同或不同领域政策工具之间的组合[4]。2010 年 OECD 将政策组合定义为政策之间的平衡和互动,而政策包括四个方面,战略工具、战略任务、支持干预的原理以及所聚焦的政策领域。总之,学者在研究政策组合时都各有侧重点,本文从政策主题的角度切入研究,识别地方政府工作报告中的政策组合,并分析在地方政府工作报告中发挥政策引领产业结构升级,实现高质量发展,形成新质生产力过程中所做出的努力[5]。

2 研究方法及数据说明

本文主要通过查阅各地级市政府官方网站,少部分缺失年份通过北大法宝或中国经济网进行查漏补缺,手动收集了全国范围内 2010-2022 年共 4057 份地级市政府工作报告,由于一些地级市政府工作报告存在部分缺失或内容不完整等问题,在兼顾样本区间和数量的情况下,最终选取有效样本为 272 个地级市在 2010-2022 年间的共 3536 份政府工作报告用于政策组合分析。

本文采用的 LDA 主题模型能够从大量非结构化文本中提取出潜在的主题，并通过计算文档和主题之间的分布来揭示文本的深层结构，并通过计算 JS 散度，分析不同年份间政策主题整体的变化，为政策组合演进的动态监测提供了依据[6]。

首先，本文收集 2010-2022 年间的全国范围内的地级市政府工作报告,对收集到的政府工作报告进行预处理，包括去除停用词、标点符号、数字和无意义的词汇，以便确保文本质量,并对文本进行了分词处理，对清洗后的文本进行依存句法分析，识别语句结构，有助于保留文本的语义关系。通过进一步文本分析识别确认，去除掉与本文所关注的区域发展关联性不强方面的关键词，识别并保留主要相关的关键词，这些关键词是构成政策主题的基础。利用 LDA 主题模型对预处理后的文本数据进行建模，将文档集转化为词语向量的集合。根据前人的研究和本文数据集的大小和复杂性，合理设置 LDA 模型参数。

在确定主题数量时，本文采用了困惑度（Perplexity）这一指标来评估模型的性能。困惑度是一种衡量模型预测文本数据概率的指标，其值越低，表示模型的预测能力越强。在实验过程中，首先使用 LDA 主题模型对预处理后的文本数据进行了主题建模。然后分别计算了不同主题数量下的模型困惑度。通过对比发现当主题数量为 15 时，模型的困惑度达到最低。模型的预测能力最强，能够较好地捕捉文本数据中的潜在主题分布。考虑到关键词的代表性和区分度，对每个主题下保留 10 个关键词，反映出每个主题的特征和重点，形成主题关键词列表。

本文采用 JS 散度（Jensen-Shannon Divergence）来分析政策组合演化趋势。JS 散度是一种对称的、非负的量度,用于衡量两个或多个概率分布之间的差异。使用 JS 散度来计算每个城市在时间序列上政策组合的变化，由当年和前一年的政策主题强度分布差异情况计算得到，从而得到政策组合的一致性指标。

JS 散度计算公式如下：

$$JSD(P||Q) = \frac{1}{2}D(P||M) + \frac{1}{2}D(Q||M)$$

其中， $M = \frac{1}{2}(P + Q)$ ， P 、 Q 为两个概率分布，而 JS 散度是用于刻画概率分布 Q 拟合概率分布 P 的程度。JS 散度的取值范围为[0,1]，较低的 JS 散度意味着政策变化不大，或者变化趋势较为一致；而较高的 JS 散度则意味着政策变化较大，或者变化趋势不一致。

3 结果与分析

根据政策所涉及的具体领域，在经过 LDA 主题模型的分析后共取得 15 个主题与关键词对应表，详细的关键词汇如表 1 所示。

表 1 政策组合的主题与关键词列表

主题 ID	主题	占比	关键词
Topic 0	经济增长	15.13%	经济、增长、经济运行、GDP、经济结构、改革、计划、政策、发展、经济指标
Topic 1	产业升级	10.64%	产业、转型升级、产业结构、产业链、改革、园区、产业政策、优化、调整、供给侧
Topic 2	投资与消费	8.99%	投资、消费、消费市场、股权、消费结构、项目、消费升级、商品、投资环境、资产项目
Topic 3	金融改革	5.71%	金融、改革、市场、金融科技、服务、风险、创新、机构、监管、政策
Topic 4	创新驱动发展	13.25%	创新、发展、体系、能力、机制、创新网络、平台、资源、环境、成果
Topic 5	科技研发	8.47%	科技、知识产权、研发投入、技术、项目、企业、成果、技术创新、转化、政策

表 1（续）

Topic 6	信息化建设	3.93%	信息、数字化、产业、通讯、技术、通信技术、物联网、安全、改革、渠道
Topic 7	数字经济	5.13%	大数据、经济、人工智能、增长、技术、产业智能化、设备、服务、智能产业、普惠性
Topic 8	数字政府	1.50%	数字、公共服务、数据、民生、协同、数据平台、治理、优化、效率、改革
Topic 9	互联网经济	4.22%	互联网、融合、电商、企业、平台、需求、技术、便利、互联网服务、赋能
Topic 10	高新技术产业	6.68%	高新技术、新兴产业、研究、关键技术、园区、开发、成果转化、基地、培养、领域
Topic 11	工业转型	2.55%	工业、制造、工厂、企业、自动化、升级、智能生产、设备、整合、集成
Topic 12	创新创业集群	8.89%	创新、创业、环境、服务、经济研究、政策、项目、人才、设施、创业园区
Topic 13	智慧城市	1.12%	城市、智能、规划、服务、生态圈、社区、交通、治理、数据、信息化
Topic 14	绿色经济	3.78%	绿色、产业、效率、技术、能源、生态建设、环保、生产、可持续、资源节约

通过分析一组主题关键词，我们可以理解每个关键词背后的具体意义，进而揭示这些关键词所代表的主题的深层含义。2010-2022 年政策组合主题分布如上所示。根据主题的比重进行排序，主题依次为经济增长（15.13%）、创新驱动发展（13.25%）、产业升级（10.64%）、投资与消费（8.99%）、创新创业集群（8.89%）、科技研发（8.47%）、高新技术产业（6.68%）、金融改革（5.71%）、数字经济（5.13%）、互联网经济（4.22%）、信息化建设（3.93%）、绿色经济（3.78%）、工业转型（2.55%）、数字政府（1.50%）、智慧城市（1.12%），各政策主题分布占比如图 1 所示。

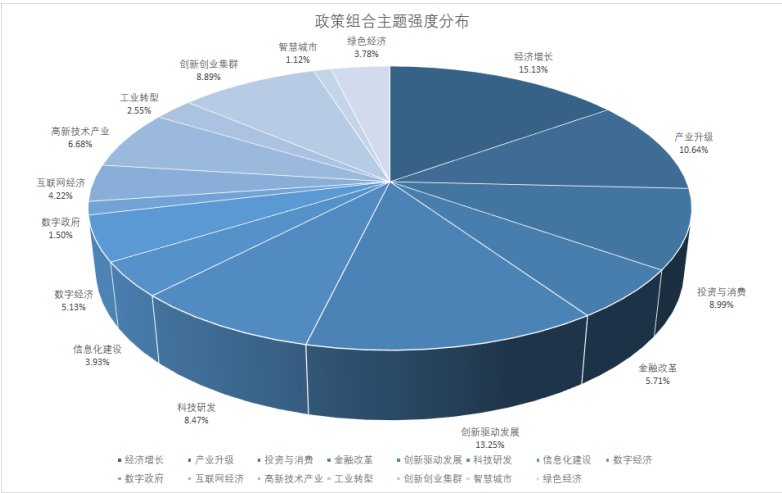


图 1 政策组合主题强度分布

依据前文所述，政策组合演化趋势分析是由判断 JS 散度指标数值而来，它反映了地方政府工作报告中当年和前一年的政策组合主题强度分布的变化，总体来看，2010-2022 年间 JS 散度的值在 0.1 到 0.5 之间波动，整体上数值均相对较低，政策组合一致性是较高的，政策组合关注方向趋势趋向一致。

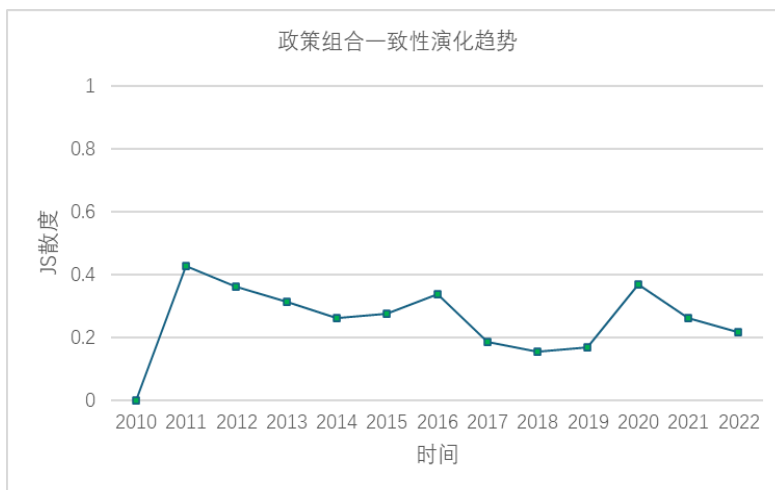


图 2 政策组合演化趋势分析

深入分析政策组合的演化趋势，根据 JS 散度数值大小，可以将其分为三个阶段，分别是政策组合波动期（2011-2015 年），政策组合稳定期（2016-2019 年），政策组合波动期（2020-2022 年），如图 2 所示。

在政策组合波动期（2011-2015 年），JS 散度数值在 0.2-0.5 之间，且有逐渐趋于平缓的趋势，在这一阶段，2011 年的 JS 散度达到了峰值，反映到地方政府工作报告中的政策组合方面，“产业升级”、“工业转型”、“科技研发”几个政策组合主题强度的增强，是构成这种波动的主要动因。2011 年是我国“十二五”规划起始之年，在这一时期国家发展的指导方针是坚持把科学发展作为主题，坚持把加快转变经济发展方式作为主线，提出五大方针：把经济结构战略性调整作为主攻方向，把科技进步和创新作为重要支撑，把保障和改善民生作为根本出发点，把建设资源节约型、环境友好型社会作为重要着力点，把改革开放作为强大动力。国家发展规划的大政方针在地方政府工作报告中得到了贯彻落实，中央领导层面的意见指导在地方政府政策组合的实施中形成了有力抓手。

在政策组合稳定期（2016-2019 年），JS 散度数值在 0.1-0.4 之间，多年数值均在 0.2 以下，政策组合一致性比较稳定，2016 年数值波动最为明显，其波动变化主要在于“创新驱动发展”、“高新技术产业”、“创新创业集群”几个政策组合主题强度的增强。2016-2020 年是“十三五”规划时期，这也是实现全面建设小康的规划，是实现第一个“一百年目标”的规划。创新、协调、绿色、开放、共享的新发展理念是“十三五”乃至更长时期我国发展思路、发展方向、发展着力点的集中体现，贯穿于“十三五”经济社会发展的各领域各环节。这一时期地方政府工作中也更加强调规模发展模式要让位于创新发展模式，在这一时期创新驱动还在产业向高端化和智能化转型升级中得到充分体现。

在政策组合波动期（2020-2022 年），JS 散度数值在 0.2-0.4 之间，在 2020 年数值最高，政策组合主题强度变化最大的是“信息化建设”、“绿色经济”、“数字政府”几类主题，2020 年是“十三五”规划的收官之年，是我国全面建成小康社会的决胜之年，是为“十四五”规划夯实基础的一年，也是抗疫攻艰的关键一年。“十三五”规划以来，在习近平新时代中国特色社会主义思想的科学指引下，我国深入推进供给侧结构性改革，着力构建现代化经济体系，经济实力大幅跃升，经济结构持续优化，深化改革成效显著，潜在增长动力强劲。在众多历史目标交汇点上，我国不仅继续明确对经济社会发展战略部署，加强自主创新，推动新兴产业发展，并且在增强政府治理能力，优化公共服务配置，提升地方政府数字化建设水平等方面也投入了更多的注意力。

4 结论

本文以 2010-2022 年地方政府工作报告为样本，借助文本分析的技术手段分析了地方政府工作报告中的政策组合及演化趋势，为更好地了解地方政府工作重点和政策组合关注方向提供了新的研究视角，为地方政府在

实践中更为灵活地应对问题，及时调整政策组合，评估政策实施效果提供了基础。结果表明我国地方政府紧跟国家大政方针，在地方政府工作报告中实施了能够有效服务创新发展、高质量发展的政策组合。

新时代中国的新质生产力的形成与发展要求地方政府继续发挥引领导向作用，出台一系列促进科技创新发展与经济高质量发展的政策组合，积极响应国家重大发展战略，着力推进地方政府的支撑作用，在完善高新技术人才培养，促进科技成果转化，加快产业结构转型升级方面继续创新服务职能，加快形成一整套可持续性的政策体系，为新质生产力的发展提供有力保障[7]。

参考文献

- [1] 刘伟. 科学认识与切实发展新质生产力[J]. 经济研究, 2024, 59(03): 4-11.
- [2] 卢盛峰, 李晓淳, 卢洪友. 地方政府财政治理特征研究：来自中国政府工作报告文本识别的经验证据[J]. 财政研究, 2020, (04): 99-114. DOI:10.19477/j.cnki.11-1077/f.2020.04.010.
- [3] Flanagan K, Uyerra E, Laranja M. Reconceptualising the ‘policy mix’ for innovation[J]. Research Policy, 2011, 40(5): 702-713.
- [4] Rogge S K, Reichardt K. Policy mixes for sustainability transitions: An extended concept and framework for analysis[J]. Research Policy, 2016, 45(8): 1620-1635.
- [5] 徐喆, 李春艳. 我国科技政策组合特征及其对产业创新的影响研究[J]. 科学学研究, 2017, 35(01): 45-53. DOI:10.16192/j.cnki.1003-2053.2017.01.006.
- [6] 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(05): 71-81.
- [7] 洪银兴. 新质生产力及其培育和发展[J]. 经济学动态, 2024, (01): 3-11.