

基于文本挖掘的我国电子商务政策演化研究

张千禧

江苏大学科技信息研究所, 江苏 镇江

收稿日期: 2025年5月12日; 录用日期: 2025年5月26日; 发布日期: 2025年6月23日

摘要

互联网产业的快速发展, 也给电子商务产业带来了新的发展机遇。政策作为引导电子商务产业发展的重要因素, 揭示了电子商务政策的演化路径对电商产业的发展具有重要意义。本文基于政策生命周期理论, 对从中国政府网、国家发改委和商务局官方网站上收集到的345份电子商务政策进行生命周期划分, 共划分了三个阶段, 利用LDA主题模型提取各个阶段的政策主题, 并结合Word2Vec词向量模型对其进行主题相似度计算, 构建了我国电子商务政策的演化路径。研究发现, 我国电子商务政策呈现从基础设施建设到体系优化, 再到产业融合的渐进发展路径, 且重点在于关注农村电子商务和农产品跨境电子商务两大产业领域。

关键词

电子商务, 政策演化, 文本挖掘

Research on the Evolution of China's E-Commerce Policy Based on Text Mining

Qianxi Zhang

Institute of Science and Technology Information, Jiangsu University, Zhenjiang Jiangsu

Received: May 12th, 2025; accepted: May 26th, 2025; published: Jun. 23rd, 2025

Abstract

The rapid development of the Internet industry has also brought new development opportunities to the e-commerce industry. Policy, as an important influencing factor guiding the development of the e-commerce industry, reveals the evolution path of e-commerce policies, which is of great significance to the development of the e-commerce industry. Based on the policy life cycle theory, this paper divides the life cycle of 345 e-commerce policies collected from the official websites of the Chinese government, the National Development and Reform Commission, and the Commerce Bureau into three stages. The

LDA topic model is used to extract the policy topics of each stage, and the Word2Vec word vector model is combined to calculate the topic similarity. The evolution path of e-commerce policies in China has been constructed. Research findings indicate that China's e-commerce policies follow a progressive development path from infrastructure construction to system optimization and then to industrial integration, with a focus on two major industrial fields: rural e-commerce and cross-border e-commerce for agricultural products.

Keywords

E-Commerce, Policy Evolution, Text Mining

Copyright © 2025 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

近年来,随着互联网等信息化技术的不断发展,电子商务产业也获得了迅猛的发展。互联网技术等新一代信息化技术的不断发展,给我国电子商务产业带来了新的发展动力,催生了跨国电商、农村电商等一系列电商新生态。电子商务产业涵盖了从在线购物、支付、物流到技术支持、营销服务等全链条的商业生态。电子商务产业主要是指通过互联网等数字化技术为基础,通过电子化手段进行商品或服务交易及相关活动的综合性产业[1]。政策通常是指国家权力机关在一定时间内为了实现一定的目标而制定的行为准则,对相关企业或互联网平台具有引导性的作用。电子商务政策主要是指政府针对电子商务活动进行指导、为了促进电子商务产业健康发展和产业结构升级而制定的行为规范[2]。

随着电子商务产业的壮大升级,各级政府都非常关注电子商务产业的发展,制定了一系列的产业政策和发展措施。因此,系统化地梳理我国电子商务政策的发展脉络,分析其发展历程和实践过程,对推动电子商务产业发展升级和发展壮大具有重要意义。

2. 理论基础与文献综述

2.1. 政策生命周期理论

政策生命周期理论是由美国政策学家琼斯首先提出来的[3]。政策生命周期理论主要是指政策从形成到衰退的动态过程。该理论认为政策遵循从形成、发展到衰亡的过程,即政策是一个具有阶段性特征的“生命过程”。政策生命周期理论不仅可以用于分析单一政策,还可对比不同国家或领域的政策演进规律,为政策优化提供参考。政策的发展是动态的、变化的,遵循着一定的发展规律,并且随着时间的变化也会呈现出不同的特征。电子商务政策同样也遵循着这种政策生命周期理论。

2.2. 电子商务政策研究

关于我国电子商务政策的研究主要集中在对农村电子商务政策的研究和对跨境电子商务政策的研究等方面。如钮钦[4]从政策工具和商业生态系统两个维度对中国农村电商的政策文本进行分析,并对农村电商发展政策提出了相关建议;肖开红等[5]基于政策文本计量分析,对我国在2001~2018年间国家层面发布的涉农电子商务政策进行了演进分析;姚遥[2]还对南陵县电子商务政策的执行问题和对策进行了研究;赵峰含等[6]基于跨境电子商务发展背景,从政策效力、参与主题、发展阶段、政策工具四个维度对

跨境电子商务政策进行了量化分析；赵杨等[7]对我国跨境电子商务具体政策的实施效果进行了重点评价与分析。

2.3. 政策文本挖掘研究

政策文本是国家各级权利机关或行政部门以文件形式颁布的法律、法规、规章等官方文献，也包括政策制定过程中形成的研究、咨询、决议等公文文档[6]。政策文本挖掘主要是以政策文本为研究对象，通过对其进行文本解读或主题分析，进而反映政策特征。李柯等[8]基于文本词频统计、语义网络和聚类分析等方法对我国航空安全政策进行了研究，为完善航空安全政策提供了理论参考；孔令凯等[9]基于文本挖掘对我国共性技术政策进行主题识别、量化共性技术政策主题强度，从综合性、一致性及均衡性等方面探究了政策组合特征；也有学者利用文本挖掘等方法对科技成果转化政策[10]、现代物流产业政策[11]、电子信息产业[12]等进行了演进分析。

2.4. 现有文献述评

现有研究对电子商务政策的研究主要是从政策计量和基于政策工具等角度展开，主要是对电子商务政策的发文主体、发文数量等特征进行描述分析，并深入到电子商务政策的文本主题层面。且现有基于政策文本挖掘的研究也较为普遍，但也是基于文本词频统计、聚类等方法对政策主题进行挖掘。本文主要基于 LDA 和 Word2Vec 等方法对政策文本进行主题提取，并根据所提取出的政策主题利用 Word2Vec 进行相似度计算，以政策主题为分析对象，借助政策生命周期理论，进而揭示电子商务政策的演化路径。

LDA 模型是一种强大的文本挖掘工具[13]，能够处理大规模的非结构化的政策文本信息，通过概率分布和语义关联，可以实现特征词的语义聚类，高效率地揭示文本内容结构和主题内涵；相较于 LDA 模型，Word2Vec 是一种基于深度学习的神经网络模型，能够捕捉到文本上下文语义信息，能够将文本内容转化为词向量，从而更加深层次地揭示文本之间的关联关系。

3. 数据处理与政策生命周期划分

3.1. 政策文本来源与处理

对于电子商务政策文本的收集，通过在中国政府网、商务部网站和国家发改委官网等各官方网站检索政策文件，检索词为“电子商务”、“电商”、“电商产业”，时间跨度为 2005~2024 年，针对各个检索词在各官方网站进行检索，并对获得的政策文本进行阅读和筛选，剔除掉无关信息、活动通知、视频信息或相关度较低的文本，最终筛选出相关度较高的政策文本 345 份。

3.2. 政策生命周期划分及数据预处理

电子商务政策的逐年发布量如图 1 所示，横轴表示政策发布的年份，纵轴表示政策发布的数量。从图 1 可以看出，电子商务政策在 2005~2014 年发布的政策数量并不是很多，每年发布的数量在两份左右，且分布并不均匀，从 2015 年开始，政策数量开始显著增加，所以 2014 年是政策从零星发布转向快速增长的关键节点。随着电子商务产业的发展，关于电子商务产业政策的发布数量也随之增多，到了 2015 年，政策的发布数量开始呈现上升的趋势，并且在 2016 年达到最高峰，该年电子商务政策的发布数量为 74 份，但是到了 2017 年，政策的发布数量又开始出现下降的趋势，2018 年又开始增加，2019 年后政策发布量持续下降，从这一年开始，政策从波动上升到持续下降的转折，可见 2019 年是政策成熟期的关键节点，2015~2019 年这一阶段，电子商务政策呈现出波动上升的趋势。在 2020~2024 年这一阶段，电子商务

政策的发布数量又开始呈现下降的趋势。

我国电子商务政策发文量分布

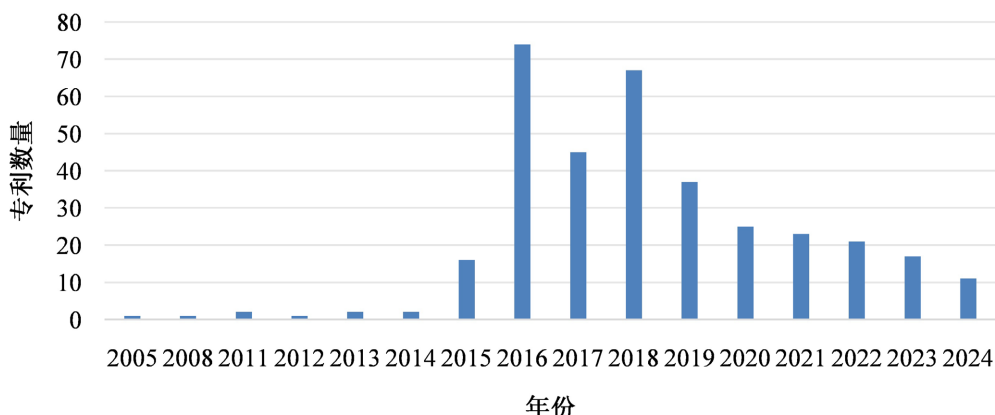


Figure 1. E-commerce policy release trends

图 1. 电子商务政策发布趋势

根据政策生命周期理论并结合我国电子商务政策在 2005~2024 年的发布趋势, 本文将 2005~2014 年划分为第一阶段, 在这一阶段, 政策发布数量极少, 电子商务产业处于探索期; 2015~2019 年为第二阶段, 政策发布数量的最高峰在这一阶段, 电子商务产业进入快速发展期; 2020~2024 年为第三阶段, 在这一阶段, 电子商务政策发布数量呈现逐年下降的趋势, 电子商务产业进入发展的成熟阶段, 对应的电子商务产业政策发布数量随之减少。

本文基于政策生命周期, 将收集到的政策文本划分为 3 个阶段, 通过对三个阶段的政策文本进行分词、去除停用词等数据预处理工作。本文利用 Python 中的 Jieba 分词库进行分词, 停用词表来自哈工大停用词表, 并对其加以自定义修改与补充。通过数据预处理, 为数据下一步的分析做准备。

4. 我国电子商务政策演化研究

4.1. 政策主题提取

本文利用 LDA 对三个发展阶段的政策主题进行提取。LDA 是一种由 Blei 等[13]提出的一种非监督的机器学习技术。它是一个基于三层贝叶斯概率的模型, 通过生成文档 - 主题 - 词汇矩阵, 以文本 - 主题 - 词语的结构来识别文档中的主题, 将文档转化为词向量。本文利用 LDA 主题模型来识别电子商务政策中的主题信息。设置最优主题数目是 LDA 模型识别政策主题的重要部分, 本文通过 pyLDAvis 交互图谱确定最优主题数目, pyLDAvis 能够通过可视化的方式展示主题交互式图谱, 可以更加直观地展示主题之间的潜在联系, 当主题之间相互独立、互不交叉时, 表示对应的模型最优。

4.1.1. 第一阶段政策主题识别

通过优化最优主题数目, 将第一阶段的政策主题设置为 5, 其 pyLDAvis 主题交互式图谱如图 2 所示。将每个主题所包含的主题词的数量设置为 9, 并为每个主题进行编号, 如第一阶段的第一个主题编号为 T_{1-0} , 以此类推。第一阶段的政策主题 - 主题词分布如表 1 所示。

从表 1 可以看出, 第一阶段我国电子商务政策主题主要集中在电商交易平台、消费者权益保护、跨境电子商务物流、跨境贸易知识产权与监管机制等方面, 涵盖了消费者、经营者等主体, 还包括物流政策、贸易监管机制等方面。

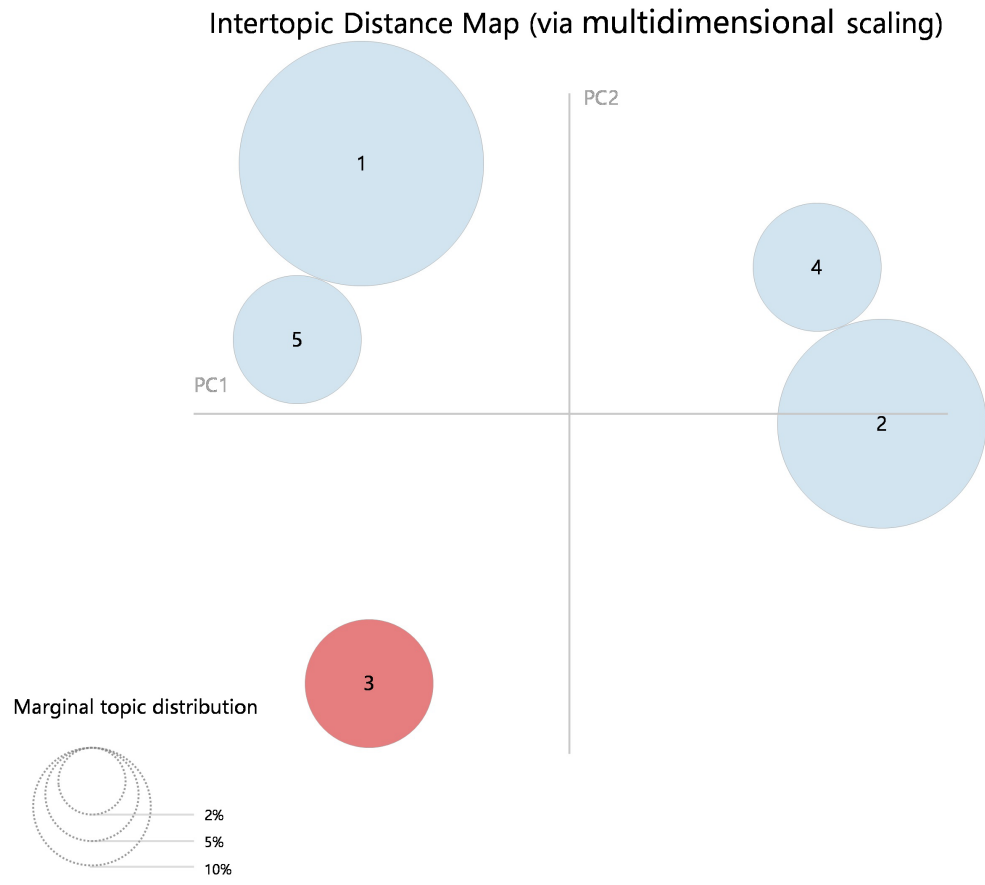


Figure 2. Interactive topic map of the first stage using pyLDAvis
图 2. 第一阶段 pyLDAvis 主题交互式图谱

Table 1. Theme-subject heading distribution of the first stage
表 1. 第一阶段主题 - 主题词分布

所处阶段	主题编号	政策主题	主题词
第一阶段	T ₁₋₀	电子商务交易平台与消费者权益保护	经营者，第三方，交易平台，消费者，电子商务交易平台，合法权益，工商行政，商品交易，交易规则
	T ₁₋₁	中小企业电商物流体系与信用服务	信息化，安全认证，国民经济，在线支付，中小企业，现代物流，法律法规，信用服务，国民经济运行
	T ₁₋₂	投融资与支付物流政策	投资人，中国证监会，销售业务，基础设施，在线支付，服务平台，服务提供商，物流配送，政策环境
	T ₁₋₃	农产品跨境电子商务的物流与政策环境	农产品，跨境电子商务，公共服务，物流配送，中西部，报关单，消费税，供应量，外贸企业
	T ₁₋₄	跨境贸易知识产权与监管机制	对外贸易，主管部门，知识产权，监督机制，国际贸易，售后服务，商品质量，信息安全，管理制度

4.1.2. 第二阶段政策主题识别

通过优化最优主题数目，第二阶段的最优政策主题数量同样是 5 个，其 pyLDAvis 主题交互式图谱如图 3 所示。将每个主题所包含的主题词的数量设置为 9，第二阶段的政策主题 - 主题词分布如表 2 所示。

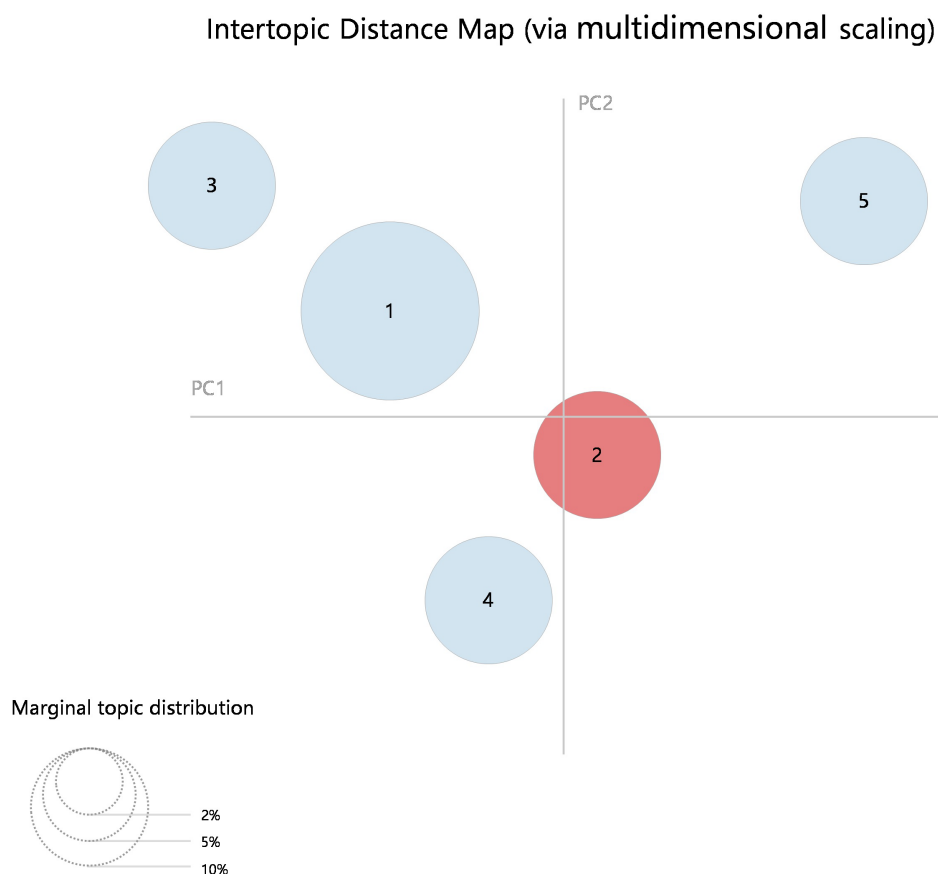


Figure 3. Interactive topic map of the second stage using pyLDAvis

图 3. 第二阶段 pyLDAvis 主题交互式图谱

Table 2. Theme-subject heading distribution of the second stage

表 2. 第二阶段主题 - 主题词分布

所处阶段	主题编号	政策主题	主题词
第二阶段	T ₂₋₀	农村电子商务发展与物流供应链优化	农村电子商务, 示范县, 农产品, 互联网, 贫困地区, 供应链, 因地制宜, 脱贫攻坚, 物流配送
	T ₂₋₁	跨境电子商务试验区管理机制与金融创新服务	互联网, 试验区, 产品质量, 进一步, 信息化, 跨境电子商务, 公共服务, 网络安全, 金融服务
	T ₂₋₂	农产品电商市场培育与政策支持	交易额, 万亿元, 农产品, 零售总额, 阿里巴巴, 跨境电子商务, 示范县, 省政府, 经济增长
	T ₂₋₃	跨境电商进口商品海关监管	跨境电子商务, 进口商品, 海关总署, 海关监管, 进一步, 税收政策, 工商登记, 数据安全, 服务商
	T ₂₋₄	电子商务市场规模与企业发展政策引导	试验区, 跨境电子商务, 产业园, 代理人, 进出口, 国际贸易, 农村电子商务, 数字签名, 综试区

从表 2 可以看出, 在第二阶段, 我国电子商务政策主要集中在农村电子商务产业发展和跨境电子商务发展两个方面, 其中包括对农产品的培育, 建立试验区、示范县等, 并且以当地的农产品为特色, 打造农产品电商产业链, 带动当地经济增长和推动农产品销售; 在跨境电子商务产业中, 也在积极打造试验区, 加强海关监管力度、优化税收政策等, 进一步促进电子商务产业的发展。

4.1.3. 第三阶段政策主题识别

通过优化最优主题数目，第三阶段的最优政策主题数量也是 5 个，其 pyLDAvis 主题交互式图谱如图 4 所示。将每个主题所包含的主题词的数量设置为 9，第三阶段的政策主题 - 主题词分布如表 3 所示。

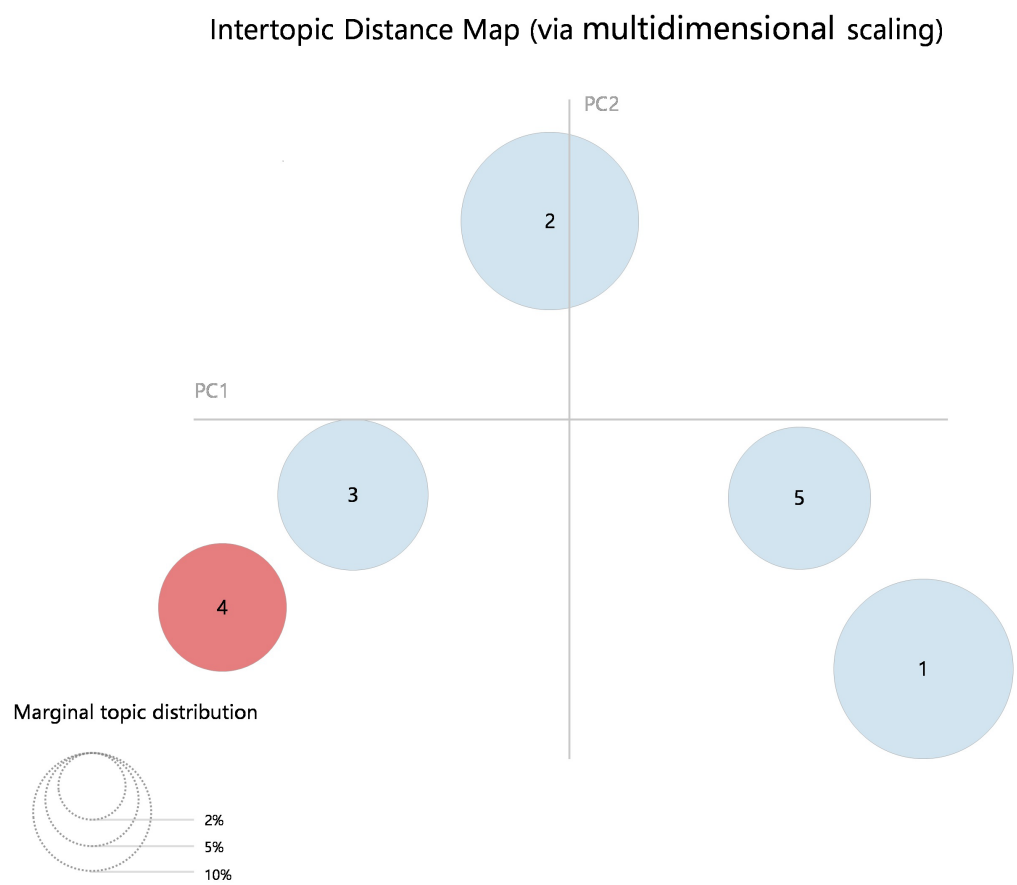


Figure 4. Interactive topic map of the third stage using pyLDAvis
图 4. 第三阶段 pyLDAvis 主题交互式图谱

Table 3. Theme-subject heading distribution of the third stage
表 3. 第三阶段主题 - 主题词分布

所处阶段	主题编号	政策主题	主题词
第三阶段	T ₃₋₀	农村电子商务与农产品流通	农产品，数字化，示范县，高质量，主管部门，农村电子商务，物流配送，公共服务，高度重视
	T ₃₋₁	数字化转型与高技术经济体制发展	数字化，高技术，一次性，交易额，消费者，制造业，贯彻落实，实体经济，进口药
	T ₃₋₂	农产品电子商务产业链监管与供应链优化	农产品，市场监管，便利化，网络覆盖，产业链，供应链，国家级，进一步，市场主体
	T ₃₋₃	进口商品信息化和实验基地建设	试验区，示范基地，综合评价，信息化，数字化，增值税，试点经验，产业园，进口商品
	T ₃₋₄	外贸与消费市场电商政策支持体系	零售额，分工负责，消费品，外贸企业，外商投资，招商引资，产业链，进出口，金融服务

从表 3 可以看出, 第三阶段的政策主题也主要关注农村电子商务产业的发展, 更加聚焦于农产品的流通、农产品的产业链和供应链优化等相关内容; 同时在第三阶段还出现了对进口商品信息化建设和贸易示范基地建设的关注。

4.2. 政策主题相似度计算

通过对电子商务政策主题的相似度计算, 从而构建电子商务政策的演化路径。本文利用 Word2Vec 对电子商务政策主题进行相似度计算。Word2Vec 是 Google 于 2013 年开源的一种工具, 其原理是基于深度学习算法, 是一种基于神经网络的词向量模型, 能够通过捕捉词语上下文信息, 将词语表示为语义向量 [14], 该方法可以较好地弥补 LDA 主题模型忽略上下文语义信息的缺陷通过模型训练与数据处理, 可以把文本内容转换为 K 维向量空间中的向量运算, 通过计算向量空间上的相似度进而表征文本语义上的相似度。本文通过余弦相似度计算主题之间的相似度关系, 计算公式如式(1)所示:

$$\cos(\theta) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (1)$$

其中, n 为文档集中的词向量个数, x_i 和 y_i 分别为第 i 个词向量的横纵坐标值。根据各阶段的主题相似度大小, 本文将主题之间的相似度阈值设置为 0.1, 即认为相似度大于 0.1 的主题之间存在演化关系。

本文利用热力图表示第一阶段与第二阶段、第二阶段和第三阶段的政策主题相似度。第一阶段和第二阶段的政策主题相似度如图 5 所示, 第二阶段和第三阶段的政策主题相似度如图 6 所示, 主题之间的相似度越高, 颜色越深。

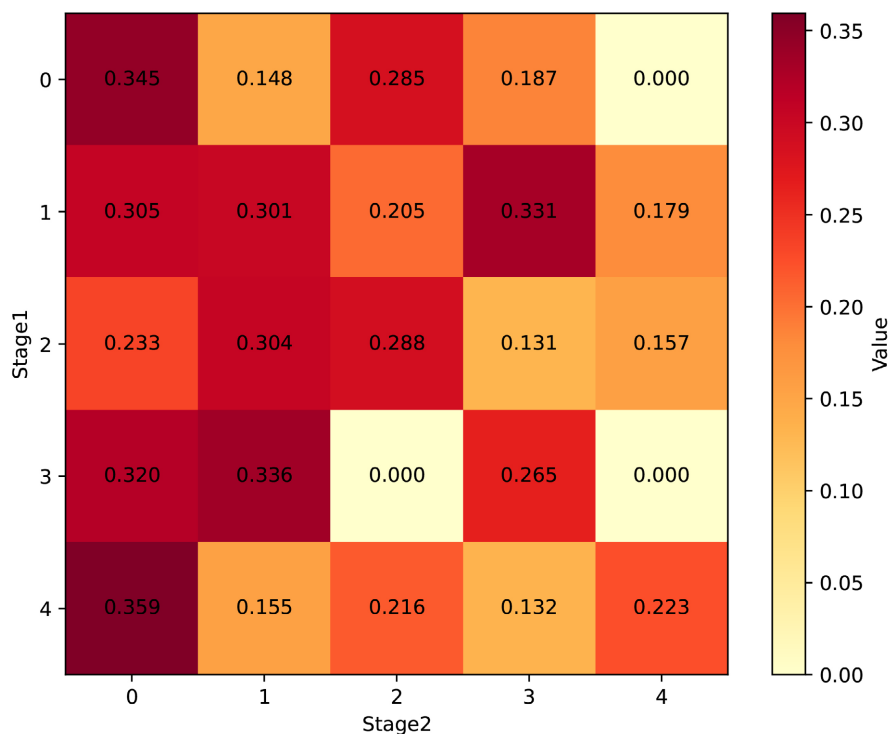


Figure 5. Heat map of the similarity of policy themes between the first stage and the second stage

图 5. 第一阶段和第二阶段政策主题相似度热力图

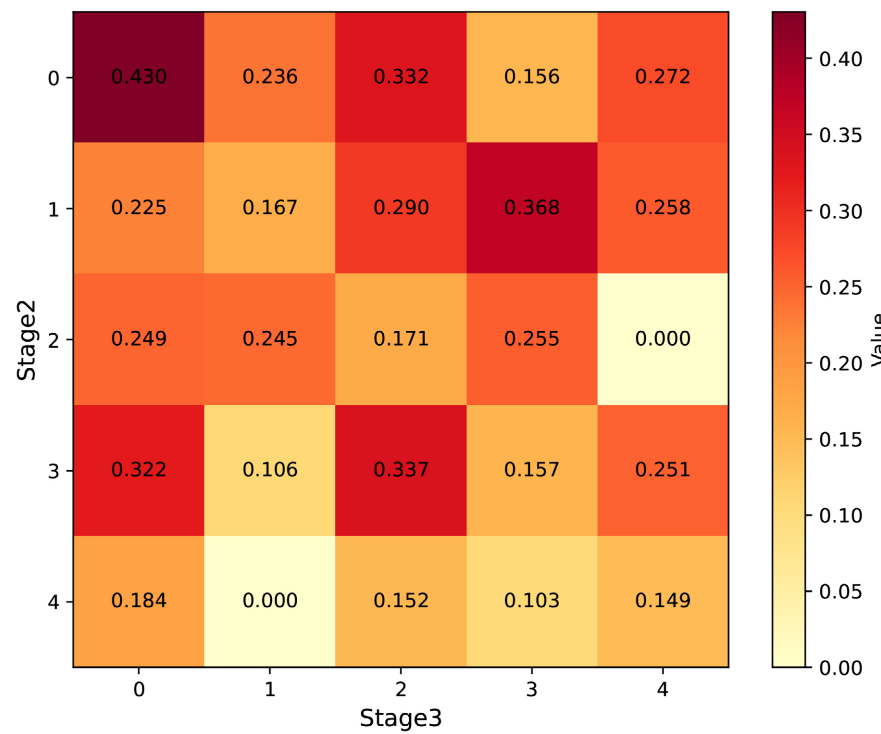


Figure 6. Heat map of the similarity of policy themes between the second stage and the third stage
图 6. 第二阶段和第三阶段政策主题相似度热力图

4.3. 政策主题演化路径

根据图 5 和图 6 中政策主题在不同阶段的相似度，绘制政策主题演化图谱，并以桑基图的形式加以呈现，如图 7 所示。

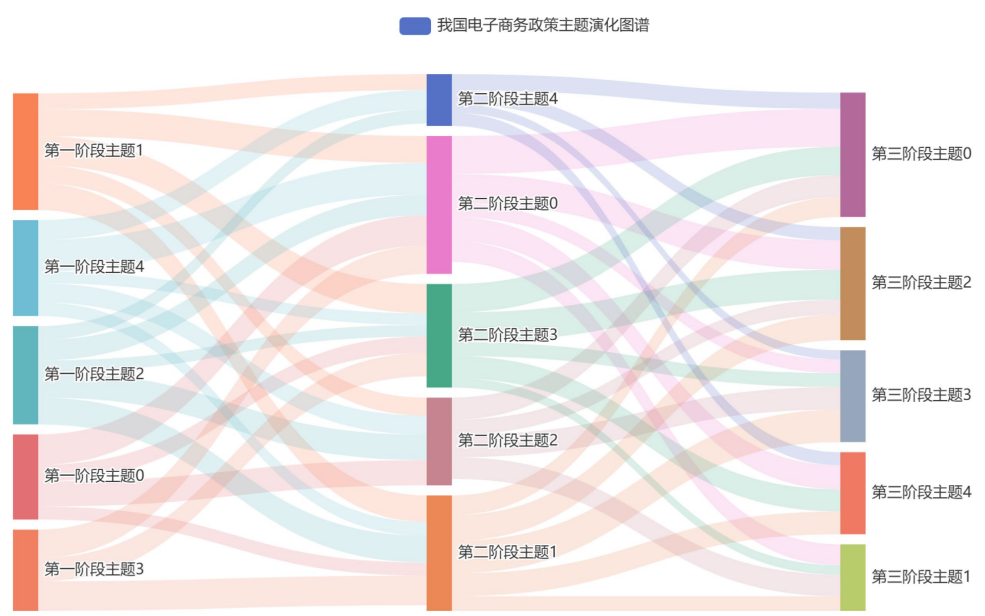


Figure 7. Three-stage evolution map of e-commerce policy themes
图 7. 电子商务政策主题三阶段演化图谱

图 7 中每个元素块都表示一个政策主题，元素块的大小与该主题所包含的主题文档数量有关，文档数量越多，元素块越大。连接各个元素块的线条表示各个主题之间的相似度，相似度越大，线条越粗。

图 7 中存在两条较为明显的政策主题演化路径，分别是第一阶段主题 1 (中小企业电商物流体系与信用服务)→第二阶段主题 0 (农村电子商务发展与物流供应链优化)→第三阶段主题 0 (农村电子商务与农产品流通)，在我国电子商务政策发展的第一阶段，主要突出对现代物流体系、信用服务、安全认证与数字信息化等内容的关注，同时还将电子商务与国民经济相结合，保障国民经济平稳运行，同时也为中小企业参与电商活动提供了基础支撑；在政策发展的第二阶段，仍然在对物流和供应链体系进行优化，在中小企业物流体系建立之后，开始将电商产业发展的重点转向农村电商，对电子商务产业的服务场景进行逐步扩展，为农产品拓展销售渠道和销售路径；在电子商务政策发展的第三阶段，依然结合农村电子商务进行拓展，加强了农产品与电商产业的深度融合，利用电子商务平台，为不同产品打开销路，加快农产品的快速流通。同时，将电商产业与农产品深度融合，还能促进乡村振兴，利用电子商务、现代化物流体系等一系列信息化、现代化设备带领乡村致富、农业振兴。

图 7 中另外一条政策演化路径为：第一阶段主题 3 (农产品跨境电子商务的物流与政策环境)→第二阶段主题 1 (跨境电子商务试验区管理机制与金融创新服务)→第三阶段主题 3 (进口商品信息化和实验基地建设)，在这条政策演化路径中，更加强调了农产品跨境物流环境的搭建与跨境电子商务试验区的管理机制和金融服务，在政策发展的第三阶段，又开始建设进口商品实验基地，并对进口商品进行信息化建设。

总而言之，我国电子商务政策的演化路径反映了从基础建设到体系优化，再到产业融合的渐进发展路径。从中小企业电商运营环境的搭建到农产品现代化物流体系的构建，最后到跨境电商试验区的建设，电子商务的服务边界不断向外拓展。随着农村电子商务物流体系的不断完善，我国农村电子商务物流体系也迎来了新的发展模式和产业创新，应当鼓励电商平台与现代农业、乡村旅游结合，探索“电商 + 文旅 + 农产品”模式。同时，支持人工智能、区块链技术在物流和信用服务中的应用。此外，由于区域发展不平衡，也应该针对欠发达地区，制定差异化的电商扶持政策，如税收优惠、人才培养等。

5. 结语

本文通过收集我国电子商务产业领域中的政策文本，利用政策生命周期理论对收集到的 345 份政策文本进行生命周期划分，将其划分为三个发展阶段，并利用 LDA 对其进行主题提取，通过 Word2Vec 词向量模型对政策主题进行相似度计算，并根据相似度计算结果构建政策主题之间的演化关系，发现农业电子商务和农产品跨境电子商务在政策演化过程中占据重要地位。

基金项目

本文系江苏大学第二十二批学生科研课题立项资助项目“基于专利信息分析的产业技术扩散演进路径研究”(编号: 22C302)的研究成果之一。

参考文献

- [1] 万文静. 电子商务发展对我国区域碳排放的影响及作用机制[J]. 商业经济研究, 2024(23): 189-192.
- [2] 姚遥. 南陵县电子商务政策执行问题与对策研究[D]: [硕士学位论文]. 芜湖: 安徽工程大学, 2023.
- [3] 刘江, 梁浩. 公共政策分析[M]. 沈阳: 辽宁大学出版社, 2013: 209.
- [4] 钮钦. 中国农村电子商务政策文本计量研究——基于政策工具和商业生态系统的内容分析[J]. 经济体制改革, 2016(4): 25-31.
- [5] 肖开红, 雷兵, 钟镇. 中国涉农电子商务政策的演进——基于 2001-2018 年国家层面政策文本的计量分析[J]. 电

- 子政务, 2019(11): 91-103.
- [6] 赵峰含, 潘勇. 我国跨境电子商务政策分析: 2012-2020 [J]. 中国流通经济, 2021, 35(1): 47-59.
 - [7] 赵杨, 陈雨涵, 陈亚文. 基于 PMC 指数模型的跨境电子商务政策评价研究[J]. 国际商务(对外经济贸易大学学报), 2018(6): 114-126.
 - [8] 李柯, 张世豪, 罗帆. 基于文本挖掘的我国航空安全政策研究[J]. 中国安全生产科学技术, 2025, 21(3): 195-200.
 - [9] 孔令凯, 杨晓萌, 林超然, 等. 基于文本挖掘的共性技术政策文本量化及政策组合特征研究[J]. 科技管理研究, 2023, 43(15): 35-45.
 - [10] 唐雪倩. 基于文本挖掘的我国科技成果转化政策演进研究[D]: [硕士学位论文]. 秦皇岛: 燕山大学, 2024.
 - [11] 俞太平. 基于文本挖掘的我国现代物流产业政策主题演化研究[D]: [硕士学位论文]. 南昌: 江西财经大学, 2024.
 - [12] 徐文. 基于文本挖掘的电子信息产业政策演化研究[D]: [硕士学位论文]. 杭州: 浙江工商大学, 2024.
 - [13] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, **3**, 993-1022.
 - [14] 胡泽文, 韩雅蓉, 王梦雅. 基于 LDA-Word2vec 的图书情报领域机器学习研究主题演化与热点主题识别[J]. 现代情报, 2024, 44(4): 154-167.