

基于BERTopic的电子商务技术主题挖掘与时间演化分析

朱勤尤

贵州大学公共大数据国家重点实验室, 贵州 贵阳

收稿日期: 2025年12月18日; 录用日期: 2025年12月31日; 发布日期: 2026年1月28日

摘要

本研究以2015~2024年间中国国家知识产权局收录的5750项电子商务专利为分析对象, 运用BERTopic深度主题模型对该领域技术主题进行挖掘与演化分析。研究识别出10个技术主题, 根据其时间演化特征归纳为四大特征集群: “生命周期型”硬件自动化、“达峰缓降型”核心平台、“稳态关注型”商品信息化及“低频成熟型”综合支撑。时序分析显示, 电子商务技术呈现出从基础建设到智能化发展的阶段性特征。基于上述发现, 本文从企业差异化技术投资与产业生态培育两个层面提出了优化路径建议。研究为电子商务技术演进提供了量化分析框架, 并对未来沉浸式交互、全局智能化及规范化治理等发展方向进行了展望。

关键词

专利分析, 技术挖掘, BERTopic, 时间演化

Research on E-Commerce Technology Topic Mining and Temporal Evolution Analysis Based on BERTopic

Qinyou Zhu

State Key Laboratory of Public Big Data, Guizhou University, Guiyang Guizhou

Received: December 18, 2025; accepted: December 31, 2025; published: January 28, 2026

Abstract

This study analyzes 5750 e-commerce patents from the China National Intellectual Property

文章引用: 朱勤尤. 基于 BERTopic 的电子商务技术主题挖掘与时间演化分析[J]. 电子商务评论, 2026, 15(1): 591-601.

DOI: 10.12677/ec.2026.151073

Administration from 2015 to 2024, employing the BERTopic deep topic modeling approach to mine and analyze technological themes and their evolution. Ten distinct technological topics were identified and categorized into four characteristic clusters based on their temporal evolution patterns: “life-cycle-type” hardware automation, “peak-and-decline-type” core platforms, “steady-state-type” product informatization, and “low-frequency-maturity-type” comprehensive support technologies. The temporal analysis reveals that e-commerce technology development exhibits phased characteristics, progressing from foundational construction to intelligent advancement. Based on these findings, optimization pathways are proposed from the perspectives of differentiated corporate technology investment and industrial ecosystem cultivation. The study provides a quantitative analytical framework for understanding e-commerce technology evolution and offers insights into future development directions such as immersive interaction, global intelligentization, and standardized governance.

Keywords

Patent Analysis, Technology Mining, BERTopic, Temporal Evolution

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

电子商务已成为驱动全球数字经济发展的核心引擎，其技术架构与创新模式经历了从简单线上交易平台向复杂智能商业生态系统的深刻演进。在这一进程中，海量的专利文本、学术论文及行业报告持续涌现，构成了记录和反映电商技术发展轨迹的“多模态知识库”。如何从这些非结构化的文本大数据中，自动、高效且动态地识别出关键技术主题结构，并刻画其随时间演化的规律，对于学术界把脉技术前沿、产业界规划研发路径、以及政府部门制定产业政策，均具有重要的理论价值与现实意义。

传统上，对技术主题的梳理多依赖于专家归纳或基于固定分类体系的统计分析，存在主观性强、难以发现新兴交叉领域、以及无法捕捉动态趋势等局限[1]。近年来，以 BERTopic 为代表的深度主题模型，凭借其结合预训练语言模型强大语义理解能力与概率主题模型可解释性的优势，为从大规模文本中挖掘细粒度、语义连贯的技术主题提供了全新的方法论工具。然而，现有研究多将 BERTopic 应用于新闻[2]、社交媒体[3]等通用领域，针对电子商务这一垂直且技术密集领域的系统性主题挖掘与演化分析尚显不足。特别是，如何将模型输出的原始主题进行合理的业务化归并与诠释，并置于时间维度下揭示其兴衰更替的内在逻辑，是一个有待深入探索的研究课题。

2. 相关研究

在技术主题挖掘方法方面，研究历程从早期的潜在狄利克雷分布(LDA)及其变体[4]，发展到如今融合深度语义表示的模型。LDA 类模型虽被广泛应用于专利和技术文献分析[5]，但其基于词袋的假设难以捕捉现代技术文本中复杂的语义关系。以 BERTopic 为代表的第三代主题模型，通过利用 Sentence-BERT 生成文档嵌入，并在降维后的语义空间中进行聚类，极大地提升了主题的语义一致性和区分度[6]。已有学者将其初步应用于科技前沿探测[7]，证实了其在处理专业领域文本上的潜力。然而，在电子商务技术领域，如何针对其“软硬结合”(软件平台与物流硬件)、“前后端融合”(用户交互与后端系统)的特点，对 BERTopic 生成的主题进行有效的后处理，仍缺乏成熟的范式。

在电子商务技术演化研究方面,现有成果多从宏观视角定性论述发展趋势[8],或基于单一技术指标(如专利申请量)进行定量分析[9]。这些研究提供了有价值的洞察,但往往未能从细粒度技术主题的层面,揭示不同技术集群之间相互关联、协同演化的动态图景。少数研究尝试利用文本挖掘方法[10],但所采用的主题模型在语义深度上存在局限,导致对复杂技术概念(如“多模态信息融合”、“分布式系统模组”)的识别不够精准,难以支撑深层次的演化机理分析。

综上所述,当前研究存在一个明显的交叉空白:即缺乏运用像 BERTopic 这样先进的深度主题模型,对电子商务领域进行细粒度技术主题的动态挖掘、层次化梳理与时间序列演化分析的综合性研究。具体而言,如何从海量文本中自动识别出兼具技术内涵与应用场景的主题,如何理解这些主题之间的结构关系,以及它们如何随时间相互作用和演变,是亟待回答的科学问题。

为填补上述研究缺口,本文开展了一项基于 BERTopic 的电子商务技术主题挖掘与时间演化分析研究,研究内容如下:

主题识别与图谱构建:利用 BERTopic 模型,从大规模电子商务相关文本语料中,自动挖掘潜在的关键技术主题,构建该领域细粒度技术主题图谱。

动态演化规律分析:基于时间切片,分析主题在近十年(2015~2024)的关注度演化轨迹,揭示电子商务技术从数字化到智能化、从线上化到线上线下融合的阶段演进规律与内在驱动逻辑。

3. 数据与实验

3.1. 数据来源

本研究的数据来源于中国国家知识产权局的专利检索系统。为系统追踪电子商务领域的技术创新轨迹,我们在该数据库中以“电子商务”为关键词,于中国范围内对专利名称进行检索。在此基础上,进一步将申请期限定在 2015 年 1 月 1 日至 2024 年 12 月 31 日之间,以覆盖近十年该领域的技术发展全周期。经过严格筛选与去重,最终构建了一个包含 5750 项有效专利的时序数据集,作为本主题挖掘与演化分析的实证基础。其专利数量分布如图 1 所示。

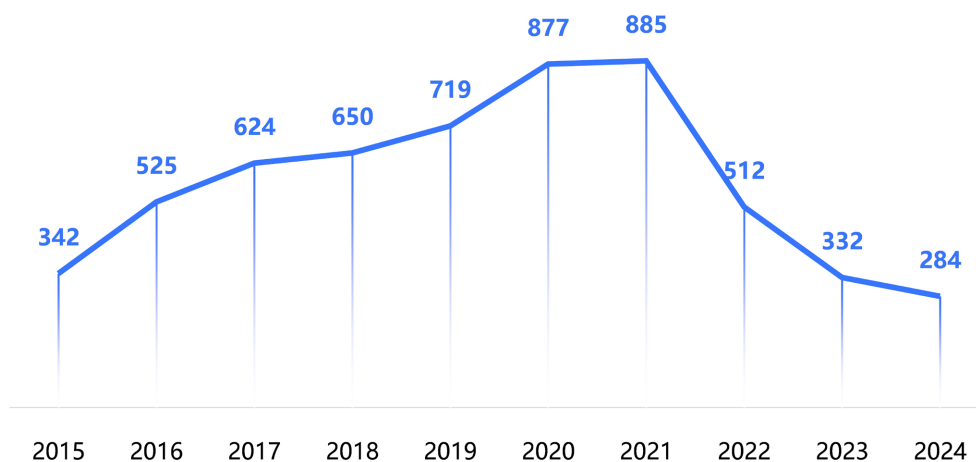


Figure 1. Distribution of e-commerce patent data among 2015~2024

图 1. 2015~2024 年电子商务专利数据分布情况

3.2. 主题建模实验设计

为从专利文本中提取有意义的、语义连贯的技术主题,本研究采用基于深度语义的 BERTopic 主题

建模框架[6]。其实验流程主要包括文档嵌入生成、降维、聚类及主题表征四个核心步骤，具体参数设置如下：

(1) 文档嵌入生成

首先，针对中文专利文本的语义特点，本研究选用在中文自然语言处理任务中表现优异的预训练模型 BERT-Chinese 作为基础编码器。该模型基于 Transformer 架构，在大规模中文语料上进行了预训练，能够深刻理解中文技术术语的语境与语义关联。我们将每条专利的文本内容(标题与摘要)输入该模型，生成高维度的语义向量表示，作为后续主题分析的基础。

(2) 降维处理

为缓解高维向量空间中的“维度诅咒”(指当数据特征维度极高时，数据点变得异常稀疏、距离计算失效，导致机器学习算法性能急剧下降的现象)，并提升后续聚类过程的效率与稳健性，采用统一流形逼近与投影(UMAP)算法对文档嵌入进行降维。具体参数设置为：'n_neighbors = 15'，以平衡局部与全局结构的保留；'n_components = 4'，在保留足够语义信息的同时控制计算复杂度；'min_dist = 0.0'与'metric = 'cosine'，确保基于余弦相似度的语义关系在低维空间中得以最大程度保持。

(3) 聚类分析

在降维后的低维语义空间上，使用层次密度聚类算法(HDBSCAN)进行无监督聚类。该算法能够自动识别簇的数量，且对噪声点(即不属于任何明确主题的文档)具有天然的识别能力。关键参数设定为：'min_cluster_size = 10'与'min_samples = 5'，旨在确保每个形成的技术主题都具有最低的文档规模支撑和足够的内部一致性；距离度量采用'metric = 'euclidean'。

(4) 主题表征与提取

对于每个由 HDBSCAN 识别出的文档簇，采用基于类的 TF-IDF (c-TF-IDF)方法提取最具区分度的关键词。该方法通过比较类内词频与整体语料词频，筛选出最能代表该技术主题的术语列表，从而形成语义明确、可解释性强的技术主题标签。

采用上述参数配置，模型聚类结果中离群值(噪声点)数量仅为 6 个，表明绝大多数专利(约 99.9%)均被有效分配到相应的技术主题中。这一结果证明了本研究所采用的嵌入模型、降维策略与聚类参数对电子商务专利文本具有良好的适配性，能够实现高覆盖、低噪声的主题划分，为主题识别与演化分析提供了可靠的数据基础。

3.3. BERTopic 模型关键参数选择与评估

为获得最优的主题建模效果，本研究采用主题一致性得分作为评估指标，对 BERTopic 模型的核心参数进行了系统性的网格搜索与择优。参数调优主要围绕影响聚类粒度与主题稳定性的三个关键参数展开：UMAP 降维维度(n_components)、HDBSCAN 最小聚类大小(min_cluster_size)以及 HDBSCAN 最小样本数(min_samples)。我们在以下参数空间进行组合实验：

n_components: [3, 4, 5, 6]

min_cluster_size: [8, 10, 12, 15]

min_samples: [3, 5, 7]

对于每一种参数组合，运行 BERTopic 模型并计算其生成所有主题的平均归一化点互信息(Normalized Pointwise Mutual Information, NPMI)得分和主题多样性(Topic Diversity)得分。NPMI 衡量主题内部关键词的语义一致性，值域为[-1, 1]，值越高表示一致性越好[11]。主题多样性通过测量所有主题中独特关键词的比例来评估，值域为[0, 1]，值越高表示主题间的冗余度越低[12]。

实验结果表明，当参数设置为 n_components = 4, min_cluster_size = 10, min_samples = 5 时，模型在

主题数量合理性(生成 10 个主题)、平均 NPMI 得分(0.072)与主题多样性得分(0.87)上达到最优平衡。此参数组合下,主题间的区分度良好,主题内部语义一致性强,且主题间关键词重复率低。

采用上述最优参数训练最终模型后,我们计算了识别出的 10 个技术主题各自的 NPMI 得分(如表 1 所示)。所有主题的 NPMI 得分均为正值,表明各主题内部的关键词之间存在显著的语义关联。其中,Topic 1 (商品推荐与订单管理)、Topic 6 (自动化仓储与物流调度)和 Topic 9 (分布式系统与多节点架构)得分较高(均大于 0.08),说明这些主题的定义更为清晰、内部一致性更强。整个模型的主题多样性得分为 0.87,表明生成的 10 个主题中,有 87%的关键词是独特的,主题间的冗余度较低,覆盖了电子商务领域内多样化的技术方向。

Table 1. NPMI scores of 10 topics

表 1. 10 个主题 NPMI 得分

主题编号	代表性关键词(前 5)	主题释义	NPMI 得分
0	装置、分拣、电机、顶部、底部	智能物流与硬件自动化	0.058
1	商品、订单、推荐、平台、用户	商品推荐与订单管理	0.085
2	照片、名称、支付、立体图、终端机	商品视觉展示与支付	0.061
3	色彩、形状、结合、信息、模型	商品多模态信息融合	0.069
4	资料、模组、系统、集成、数据	电商数据与系统集成	0.073
5	视频、监控、报警、无线、安全	电商安防与监控系统	0.066
6	储存、物流、基座、液压、调度	自动化仓储与物流调度	0.082
7	资讯、页面、会员、SKU、内容	电商内容与会员管理	0.071
8	扫码、扫描、检索码、识别、图像	扫码与图像识别技术	0.076
9	模组、资料库、复数节点、分店、架构	分布式系统与多节点架构	0.081

3.4. 与 LDA 模型对比分析

为客观评估 BERTopic 模型的性能优势,本研究在相同数据集上,应用传统 LDA 模型(设定主题数 $K = 10$, 参数经网格搜索优化)进行对比。我们采用 NPMI 和主题多样性这两个通用的量化指标对两种模型生成的主题质量进行评估。

如表 2 所示,在相同的专利文本语料上,BERTopic 模型在两项指标上均显著优于 LDA 模型。BERTopic 的平均 NPMI 得分(0.072)高于 LDA (0.041),表明其生成的主题内部关键词具有更强的语义一致性。同时,BERTopic 的主题多样性得分(0.87)也高于 LDA (0.72),说明 BERTopic 能挖掘出更丰富、更少冗余的技术主题。

Table 2. Comparison of topic quality metrics between BERTopic and LDA models

表 2. BERTopic 与 LDA 模型主题质量量化指标对比

评估指标	BERTopic 模型	LDA 模型
平均 NPMI 得分	0.072	0.041
主题多样性得分	0.87	0.72

主题质量差异的主要原因在于模型机制。LDA 基于“词袋”假设,忽略了词语顺序和深层语义,导致其生成的主题常包含语义松散或宽泛的关键词组合(如“系统”、“方法”、“装置”等高频通用词频

繁共现), 降低了主题的一致性和独特性。而 BERTopic 利用 BERT 生成深度语义嵌入, 能够捕捉技术术语间精细的上下文关联, 从而形成语义凝练、边界清晰且更具区分度的主题(如表 1 所示), 更适用于电子商务这类专业性强、术语丰富的垂直领域技术文本分析。

4. 实验结果

4.1. 主题聚类

本研究利用 BERTopic 主题建模技术, 对收集的电子商务领域文本语料进行无监督挖掘, 最终生成 10 个具有清晰技术内涵的主题(Topic 0 至 Topic 9), 基于 BERTopic 模型内建的 UMAP 降维算法, 将高维度的主题 - 词项分布投影至二维空间, 生成主题间的距离可视化图谱(见图 2)。在该图中, 每个点代表一个主题, 点之间的距离近似反映主题语义上的相似性。初步观察可见, 部分主题在空间中呈聚集态势, 为后续的聚类分析提供了直观依据。

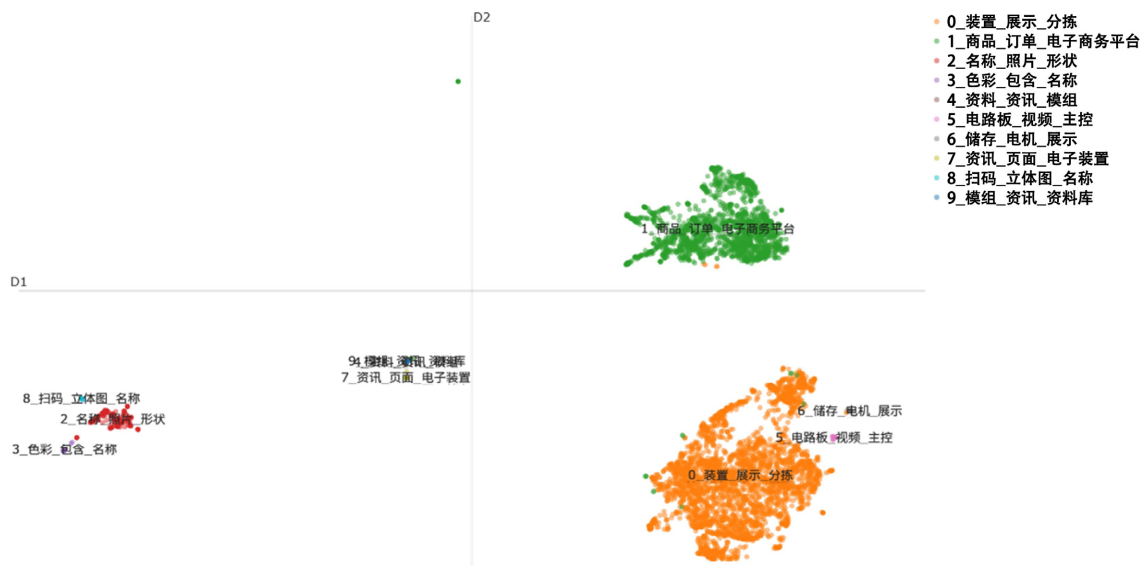


Figure 2. Two-dimensional spatial distribution of topic clusters in e-commerce technology
图 2. 电子商务技术主题聚类二维空间主题点分布图

图 3 展示了每个主题的前 8 个主题词及每个主题词的出现频次, 这些主题系统性地覆盖了电子商务从前端交互、核心业务到后端支撑与基础设施的全链条技术体系, 反映了该领域技术与应用深度融合的现状。

Topic 0 的关键词“装置”、“分拣”、“电机”、“顶部”、“底部”等, 揭示该主题聚焦于智能物流中的机电一体化硬件技术, 涉及分拣装置的机械结构设计、电机驱动控制与自动化分拣算法的集成, 其典型应用场景为电商仓储中心与快递中转场的高速自动化分拣流水线[13]。Topic 1 围绕“商品”、“订单”、“推荐”、“平台”等关键词展开, 属于“商品推荐与订单管理”主题。其领域核心技术包括协同过滤、内容推荐等个性化推荐算法, 以及高并发、高可用的订单处理系统架构。该主题直接支撑电子商务最核心的交易流程与用户体验优化。Topic 2 与 Topic 3 均与商品视觉呈现密切相关。Topic 2 作用于“商品视觉展示与支付”侧重于通过“照片”、“立体图”、“支付终端机”等技术实现商品的直观展示与线上线下支付融合。Topic 3 则倾向于“商品多模态信息融合”, 其关键词“色彩”、“形状”、“结合”等揭示了多模态检索与特征融合技术的应用, 旨在通过图像、三维模型等多维度信息提升商品搜索与匹

配的精度[14]。

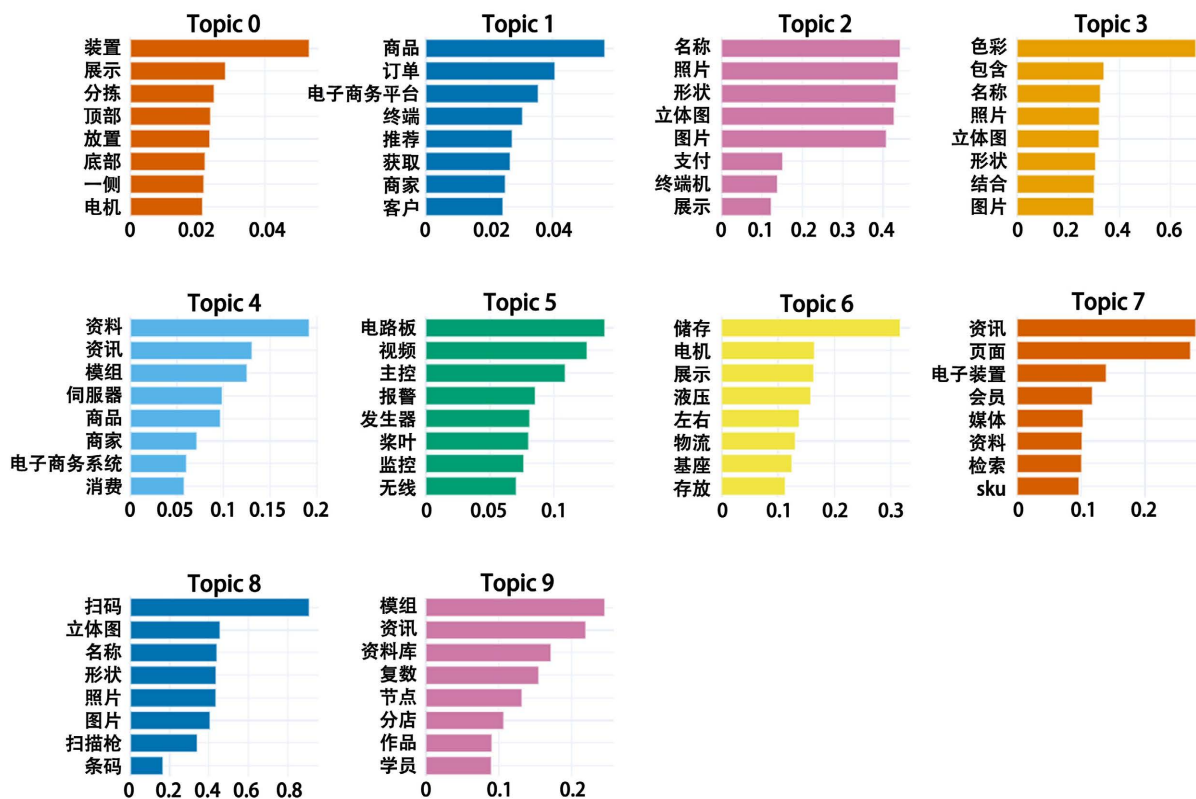


Figure 3. Top 8 topic terms and their frequencies in e-commerce technology clustering

图 3. 电子商务技术主题聚类 top8 主题词及其频数

Topic 4 与 Topic 7 共同指向电商数据与内容管理。Topic 4 的关键词“资料”、“模组”、“系统”等指向了“电商数据与系统集成”领域，体现了模块化系统架构与数据中台思想，支撑多商家、多业务的集成管理。Topic 7 涉及“电商内容与会员管理”则更具体地关注“资讯”、“页面”、“会员”、“SKU”等要素，涉及内容管理系统与用户画像技术，服务于精准营销与媒体化电商运营。Topic 5 的关键词“视频”、“监控”、“报警”、“无线”明确了其“电商安防与监控系统”的属性。该主题融合了物联网传感技术、视频智能分析与无线通信技术，主要应用于仓储物流环境的物理安全防控与直播电商等内容安全监管。Topic 6 的关键词“储存”、“物流”、“基座”、“液压”等，与 Topic 0 形成呼应，共同构成后端物流自动化板块。Topic 6 关注于“自动化仓储与物流调度”，其技术核心在于自动化存取系统、液压传动设备及物流路径优化算法，应用于智能立体仓库与配送中心。Topic 8 则聚焦于“扫码”、“扫描”、“检索码”，代表了“扫码与图像识别”技术主题。该主题依托二维码解码与计算机视觉技术，实现了从线下到线上的关键入口连接，广泛应用于扫码购、商品溯源与防伪等场景。Topic 9 的关键词“模组”、“资料库”、“复数节点”、“分店”揭示了其“分布式系统与多节点架构”的本质。该主题关注支撑大型电商平台的底层技术基础设施，包括分布式数据库、微服务架构等，是系统具备高扩展性与高可用性的基础[15]。

4.2. 时间演化分析

静态的主题列表仅能揭示技术要素的存在，无法反映其动态重要性。为进一步探究各主题在时间维

度上的关注度变化,我们基于模型输出的主题-时间分布,绘制了2015年至2024年间10个主题**的强度演化曲线**(见 图 4)。为客观刻画各技术主题的演化模式,本研究基于其年度强度时间序列,计算了四项核心量化指标(见表 3): 1. 峰值年份: 强度达到局部或全局最大值的年份。2. 平均增长速度: 从起始年(2015)至峰值年,强度值的年度复合增长率(CAGR)。对于在整个观察期内无显著峰值或峰值强度极低(<5%)的主题(如 Topic 6 与 Topic 9),其增长率接近或等于零,表明其缺乏增长趋势。3. 峰值强度: 时间序列中的强度最大值,反映了该技术主题在观测期内受到的相对最高关注度。4. 峰值后衰减系数: 强度达到峰值后,其逐年相对变化率的算术平均值,用于量化关注度的衰退速度。

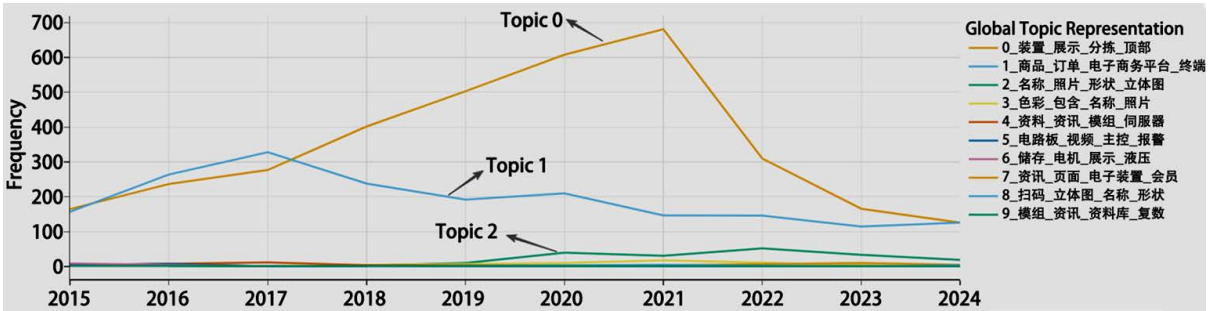


Figure 4. Evolution curves of the strength of 10 topics
图 4. 10 个主题强度演化曲线

Table 3. Temporal evolution analysis of 10 topics
表 3. 10 个主题时间演化分析表

主题	峰值年份	平均增长速度	峰值强度	峰值后衰减系数
0	2021	12.82%	76.95%	-16.66%
1	2017	26.28%	52.56%	-9.34%
2	2023	1.47%	10.27%	-34.85%
3	2022	0.29%	2.03%	-37.65%
4	2017	0.96%	1.92%	-32.54%
5	2016	1.53%	1.53%	-27.18%
6	2015	0.00%	2.65%	-48.82%
7	2023	0.43%	3.02%	-41.73%
8	2021	0.09%	0.56%	-26.58%
9	2015、2017、2018、2022	0%	0.59%	-5.61%

通过精确分析各曲线的峰值、趋势与转折点,我们可以识别出四类具有鲜明生命周期特征的宏观技术集群:

“生命周期型”硬件自动化集群

以 Topic 0 为代表,该类主题(Topic 0: 装置、展示、分拣、顶部)的强度演化呈现出一个完整的“兴起-成熟-衰退”技术生命周期。其关注度自约 2015 年起迅速增长,在 2021 年左右达到顶峰,随后显著下降。这一轨迹精准对应了电商物流自动化设备从技术引入、大规模投资建设到逐步进入标准化和部署放缓阶段的全过程,表明该领域可能已从爆发性创新投入期,进入以优化和迭代为主的成熟应用期。

“达峰缓降型”核心平台集群

以 Topic 1 为代表, 该类主题(Topic 1: 商品、订单、电子商务平台、终端)的曲线特征为先升后降, 在 2017 年达到关注度顶峰后, 呈波动式缓慢下降趋势。这反映了电子商务核心交易平台的基础架构与模式, 在移动互联网红利期达到建设高潮后, 其作为独立技术热点的关注度逐渐让位于更前沿、更细分的创新领域。但其总体强度仍保持较高水平, 说明其作为技术基底的稳定性。

“稳态关注型”商品信息化集群

以 Topic 2 为代表: 该类主题(Topic 2: 名称、照片、形状、立体图)的强度在整个时间范围内维持在一个相对稳定的中等水平, 并在小范围内波动。这表明, 围绕商品基础信息数字化(名称)与可视化(照片、形状)的技术, 已成为电商运营中一项常态化、必备的支撑能力。其关注度未出现爆发式增长, 可能意味着该领域的技术范式相对成熟, 创新主要集中在既有框架内的体验优化(如从照片到立体图)。

“低频成熟型”综合支撑集群

以 Topic 3~9 为代表: 其余主题(Topic 3 至 Topic 9)的强度曲线整体上长期处于低频区间, 且波动幅度平缓。该集群涵盖了数据管理、内容服务、安防监控及特定交互技术等多元化的支撑性技术。这种长期低频且平稳的特征表明, 这些技术大多属于电商生态中早已完成基础普及、进入稳定应用阶段的成熟模块。它们构成了行业稳健运行的“技术基座”, 其迭代通常是渐进式和场景化的, 因此难以引发持续的高关注度, 但却是生态系统不可或缺的组成部分。

以上所呈现的四类演化模式, 从动态视角揭示了电子商务不同技术模块截然不同的创新扩散节奏与成熟阶段: 硬件自动化完成了从热潮到平稳的完整周期, 核心平台技术已过关注度顶峰, 商品信息化处于稳态应用期, 而众多支撑技术则早已沉淀为成熟的“基座”能力。

5. 面向电子商务技术发展的优化路径建议

基于上述技术演化规律及主题集群特征, 本研究从企业创新战略与产业生态培育两个层面, 提出以下具象化的优化路径建议, 旨在为相关主体把握技术节奏、配置研发资源、构建可持续竞争力提供决策参考。

5.1. 面向技术生命周期的企业投资策略

企业技术投资策略需与特定技术主题的演化阶段紧密耦合。本研究识别的四大技术集群(如表 3 所示), 为差异化投资提供了明确的量化依据。

对于已进入成熟阶段的“达峰缓降型”与“低频成熟型”技术(如 Topic 1 及 Topics 3~9), 其关注度峰值已过或长期处于低位(峰值强度均低于 5%), 表明其技术范式趋于稳定。企业研发重点应从突破性创新转向效率优化与生态加固。具体而言, 可对核心交易平台(Topic 1)进行微服务化与云原生改造以持续降本增效; 同时, 推动数据管理、内容服务等综合支撑模块(Topics 3~9)的标准化与 API 化, 积极接入行业 SaaS 生态, 从而将内部资源从重复性基础建设释放至更具增长潜力的创新领域。

对于处于稳定应用期的“稳态关注型”技术(如 Topic 2), 其强度序列长期平稳(年均增长率仅 1.47%), 已成为支撑电商运营的常规能力。该类技术直接关联用户体验, 投资应聚焦于体验增值与数据闭环。这意味着企业需推动技术从基础商品展示向沉浸式交互(如 AR/VR)演进, 并构建“交互-数据-洞察”的闭环系统, 使商品可视化数据反哺精准营销与供应链优化, 实现稳态技术的动态价值挖掘。

对于已完成完整生命周期、进入衰退阶段的“生命周期型”技术(如 Topic 0), 其呈现显著的“兴起-达峰-衰减”轨迹(衰减系数达-16.66%), 预示着大规模投资热潮已过。企业应采取敏捷跟随与场景创新策略。一方面, 避免在技术衰退期进行重型资产投资, 转而采用与专业公司合作、租赁或“机器人即服务”(RaaS)等灵活模式获取自动化能力, 保持组织弹性。另一方面, 应深耕冷链物流、高端品仓储等高

价值细分场景，开发定制化解决方案，在局部形成差异化竞争优势。

5.2. 促进产业协同演进的生态培育方向

为引导电子商务技术体系长期健康发展，产业与政策层面需构建分层、动态的支持生态系统。

首先，产业界需共同筑牢“基座层”。针对已广泛普及的“低频成熟型”综合支撑技术(Topics 3~9)，行业协会与标准组织应牵头制定统一的数据接口、安全协议及合规标准，并鼓励头部企业将经过验证的技术模块开源。此举可降低全行业的互联互通与合规成本，避免重复投入，使产业整体受益于规模效应。

其次，应着力激活“创新层”的前沿探索。建议设立专项孵化基金与真实业务场景下的测试场，鼓励在 XR 购物、低碳物流机器人、隐私计算等前沿交叉领域进行探索。通过举办“产业-学术”联合挑战赛，围绕“双十一”极致履约、动态定价等真实业务痛点公开征集解决方案，可以实际问题驱动创新。

最后，建立健全动态监测与预警体系。建议构建国家或行业级的电子商务专利与技术创新动态监测平台，运用本研究类似的时序主题挖掘方法，自动生成技术演化图谱。该体系有助于政策制定者与产业领袖前瞻性地识别处于萌芽期的潜在颠覆性技术(如从“噪声点”中发现的早期信号)，洞察技术供应链中的潜在风险点，从而为宏观产业政策与重点研发计划的制定提供量化、客观的决策支持，确保资源投入与技术发展的实际节奏同步。

6. 结论与展望

本研究基于 BERTopic 模型对中国电子商务领域近十年的专利文本进行了深度挖掘与动态分析，识别出 10 个技术主题，根据其时间演化特征归纳为四大特征集群：“生命周期型”硬件自动化、“达峰缓降型”核心平台、“稳态关注型”商品信息化及“低频成熟型”综合支撑四大技术集群，并系统揭示了其演进轨迹与内在逻辑。这些发现不仅验证了深度主题模型在垂直技术领域分析中的强大能力，更从动态视角完整呈现了电子商务技术生态从基础建设、平台扩张到体验深化与智能协同的三阶段发展规律，为理解数字经济时代技术创新的周期性特征与驱动机制提供了重要实证依据。

值得关注的是，在主题建模过程中，模型识别出少量未被纳入主要聚类的“噪声点”专利(占比约 0.1%)。对这些文档的关键词分析显示，其代表性术语包括“点数”、“代理商”、“资料”、“链结”、“资讯”、“会员”、“经销商”、“商家”、“身分”、“消费”等，且这些专利集中出现在 2015 至 2022 年间。这一关键词组合呈现出与现有 10 大主题不同的技术焦点，暗示着电子商务领域可能存在尚未形成主流但持续探索的技术方向，如基于点数激励与身份认证的复杂渠道管理体系，以及资讯技术在商业交互中的早期融合尝试。尽管规模尚小，但这些信号可能预示着未来技术生态在渠道管理数字化和交互模式多元化方面的潜在增长点。

参考文献

- [1] 韦婷婷, 冯丹钰, 宋世领, 等. 技术机会发现领域专利挖掘方法研究述评[J]. 情报学报, 2023, 42(10): 1238-1250.
- [2] 杨帆, 张鑫. “环境”与“地缘政治”的叙事博弈: 国际报道如何影响中国电动汽车海外销量[J]. 新闻与传播研究, 2025, 32(9): 111-125+128.
- [3] 李思艺, 段丽珍. 基于社交媒体信息分析的公众档案认知特征研究[J/OL]. 浙江档案, 1-9. <https://link.cnki.net/doi/10.16033/j.cnki.33-1055/g2.20251205.002>, 2025-12-15.
- [4] 陈伟, 林超然, 李金秋, 等. 基于 LDA-HMM 的专利技术主题演化趋势分析——以船用柴油机技术为例[J]. 情报学报, 2018, 37(7): 732-741.
- [5] 杨超, 朱东华, 汪雪锋, 等. 专利技术主题分析: 基于 SAO 结构的 LDA 主题模型方法[J]. 图书情报工作, 2017, 61(3): 86-96.
- [6] 周隽如, 刘智勇. 医学人工智能领域专利技术主题发展态势研究[J]. 世界科技研究与发展, 2024, 46(4): 497-510.

-
- [7] 曾文, 肖己晴, 张希, 等. 开源信息环境下的科技情报线索发现方法探析——以科技前沿主题推理为例[J]. 情报理论与实践, 2025, 48(10): 158-162+147.
- [8] 赵宪明. 新技术与数字化转型对农产品电子商务发展的影响[J]. 农业工程技术, 2024, 44(14): 14-15.
- [9] 蔡洁, 李文静, 夏显力. 电商技术、动态能力与小农户增收[J]. 西北农林科技大学学报(社会科学版), 2023, 23(5): 91-101.
- [10] 樊慧, 李傲霜. 文本挖掘技术在我国图书情报领域的发展趋势及热点研究——基于 CiteSpace 的可视化分析[J]. 内蒙古科技与经济, 2024(16): 130-137.
- [11] Röder, M., Both, A. and Hinneburg, A. (2015) Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Shanghai, 2-6 February 2015, 399-408. <https://doi.org/10.1145/2684822.2685324>
- [12] Dieng, A.B., Ruiz, F.J.R. and Blei, D.M. (2020) Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics*, 8, 439-453. https://doi.org/10.1162/tacl_a_00325
- [13] 李雲菲. 乡村全面振兴背景下邯郸市智慧农业发展问题及对策研究[D]: [硕士学位论文]. 邯郸: 河北工程大学, 2025.
- [14] 郑裕泽. 面向图像分类的跨模态对齐算法研究[D]: [硕士学位论文]. 济南: 山东大学, 2024.
- [15] 龚森. 基于微服务架构的电商平台设计与实现[D]: [硕士学位论文]. 北京: 北京交通大学, 2021.