

电子商务领域临时专利申请技术主题识别研究 ——基于BERTopic的分析

胡崇文

江苏大学科技信息研究所, 江苏 镇江

收稿日期: 2026年8月3日; 录用日期: 2026年8月17日; 发布日期: 2026年10月16日

摘 要

随着电子商务持续向数字支付、智能推荐、平台治理和沉浸式交互等技术密集场景延伸, 相关创新活动呈现出迭代快、变化快和早期技术形态不稳定等特征, 研发早期技术信息的识别与分析因此具有现实意义。文章以电子商务领域临时专利申请专利为研究对象, 基于Incopat全球专利数据库, 采用IPC分类号检索与优先权信息识别相结合的方式提取样本, 并运用BERTopic模型对专利标题与摘要进行主题识别与结构分析。结果显示, 相关专利主要集中于交易支付、平台支撑、内容分发、智能推荐、场景交互和安全认证六类主题。研究表明, 临时专利申请专利能够较好反映电子商务领域研发早期的技术信息, BERTopic方法能够有效呈现其主题分布与技术特征, 可为专利情报分析和企业早期专利布局提供参考。

关键词

电子商务专利, 临时专利申请, BERTopic, 技术主题识别

Research on Technical Topic Identification of Provisional Patent Applications in the E-Commerce Field —An Analysis Based on BERTopic

Chongwen Hu

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

Received: August 3, 2026; accepted: August 17, 2026; published: October 16, 2026

Abstract

As e-commerce continues to expand into technology-intensive scenarios such as digital payment,

intelligent recommendation, platform governance, and immersive interaction, innovation in this field is characterized by rapid iteration, fast change, and unstable technological forms in the early stages. Identification and analysis of technological information from the early stages of research and development therefore carry practical significance. This study takes provisional patent applications in the field of e-commerce as the research object. Based on the Incopat global patent database, samples are collected through IPC-based retrieval and priority information identification, and BERTopic is used to identify technological themes and analyze their structure from patent titles and abstracts. The results show that these patents are mainly concentrated in six themes: payment and transaction, platform support, content distribution, intelligent recommendation, scenario interaction, and security authentication. The study indicates that provisional patent applications can reflect fairly well early-stage technological information from research and development in e-commerce and that BERTopic can effectively reveal thematic distribution and technological characteristics, providing support for patent intelligence analysis and early corporate patent layout.

Keywords

E-Commerce Patent, Provisional Patent Applications, BERTopic, Technology Topic Identification

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

电子商务已由早期的平台交易逐步扩展至智能推荐、跨境支付、供应链协同、数字营销、用户行为分析和平台治理等多个技术场景[1]。相关创新高度依赖算法、数据处理、交互流程和系统架构的持续迭代,技术更新快、应用转换快,许多创新在形成初期即具有较强的不稳定性与探索性。围绕这类创新,如何在研发早期完成技术占位并为后续完善预留空间,已成为企业竞争中的重要问题。

临时专利申请在这一阶段具有较高适配性。该制度允许申请人在技术方案尚未完全定型时先行锁定优先权日,为后续补充和正式申请保留时间。对于电子商务领域而言,这一点尤为重要,因为相关创新更多表现为业务流程重构、平台机制设计、支付与履约协同以及智能化服务嵌入等复合型技术方案,专利文本分散、主题交叉明显、边界变化较快,仅依赖传统分类号或人工筛选,难以准确把握其技术结构。

现有研究多关注临时专利申请的制度功能、创新激励及企业绩效,对专利文本本身的技术主题缺少深入分析。聚焦电子商务领域后,这一不足更为明显:一是相关专利边界较模糊,仅靠关键词检索容易遗漏样本[2];二是现有研究较少识别这些专利主要集中于哪些技术主题,以及主题之间存在何种结构关系[3];三是传统 IPC 统计或共词分析方法在处理大规模、跨主题专利文本时,较难兼顾语义识别与主题解释[4]。

基于此,本文以电子商务领域临时专利申请专利为研究对象,首先构建识别与筛选规则,提取相关专利样本;随后结合专利标题与摘要文本,运用 BERTopic 模型开展技术主题识别与结构分析;进一步从主题分布、关键词表征和主题关系等维度,揭示电子商务领域临时专利申请专利的主要技术方向及其特征。本文旨在回答三个问题:电子商务领域临时专利申请专利如何更有效识别与提取?这些专利主要集中于哪些技术主题?这些主题呈现出怎样的结构特征?

2. 研究设计

本文以电子商务领域临时专利申请为研究对象,具体步骤包括:首先,根据电子商务领域的技术特

征和业务场景设定专利检索规则，从专利数据库中筛选相关专利申请数据；再结合优先权信息识别其中经历美国临时专利申请的专利，形成研究样本。在样本构建完成后，提取专利标题、摘要、申请时间和优先权号等关键信息，并对原始数据进行去重、清洗和标准化处理，以保证后续分析的数据质量。

在文本分析阶段，本文以专利标题和摘要作为主题识别的主要文本来源，对样本进行预处理，包括文本清洗、分词、停用词剔除和语义规范化等操作。随后，将每件专利的标题与摘要合并，形成用于主题识别的完整文本。这样处理的目的在于，尽可能保留技术语义信息，将发明名称与技术描述统一纳入分析范围，从而更完整地反映专利文本所承载的技术内容。

在此基础上，本文采用 BERTopic 模型对专利标题与摘要进行主题识别[5]。首先，利用预训练语言模型对文本进行语义嵌入，将专利文本转化为高维语义空间中的向量表示，为后续识别技术主题提供输入。考虑到高维语义向量不便于直接聚类，本文进一步借助 UMAP 进行降维处理，在尽量保持局部语义结构的基础上，使语义相近的专利在低维空间中更加集中，从而提高潜在技术群组识别的可操作性。

完成降维后，本文采用 HDBSCAN 算法对专利向量进行密度聚类，以识别电子商务领域临时专利申请中的潜在技术主题。相较于传统将每篇文本严格划入单一类别的硬划分方式，BERTopic 能够保留文本在不同主题之间的概率分布信息，因此更适合刻画电子商务领域中常见的技术交叉、功能融合、路径并行和边界尚未完全稳定的创新特征。若某件专利主要集中于某一个主题，说明其技术指向较为清晰；若同时在多个主题上具有较高权重，则说明其技术内容具有较强的跨主题属性。

为进一步刻画不同主题之间的关系结构，本文提取各主题的关键词集合，并基于其语义表示分析主题之间的接近程度。在此基础上，通过比较不同主题之间的语义差异，揭示电子商务领域临时专利申请中各技术主题之间的关联关系和结构分布。之后，再结合各主题的高频关键词和代表性文本内容，对识别出的潜在技术主题进行归纳与命名。

在结果分析阶段，本文进一步从主题分布和时间变化两个方面展开讨论。一方面，统计不同技术主题对应的专利数量和关键词特征，分析电子商务领域临时专利申请主要集中在哪些技术方向；另一方面，结合专利申请年份考察各主题在不同时间上的变化情况，揭示该领域临时专利申请技术主题的动态演进特征。通过上述研究设计，本文较为系统地呈现了电子商务领域临时专利申请的技术主题结构，也为相关专利情报研究和技术趋势识别提供了方法支持。

3. 实证研究

3.1. 数据来源

本文专利数据来源于 Incopat 全球专利数据库。该数据库收录范围较广，覆盖多个国家和地区的专利申请信息，能够提供专利标题、摘要、申请时间、申请人、优先权号、IPC 分类号等较为完整的专利字段，适合用于电子商务领域临时专利申请专利的检索、筛选与后续文本分析。

在样本获取过程中，本文主要采用 IPC 分类号检索方式对电子商务领域相关专利进行识别。IPC 分类号能够较为稳定地反映专利所属的技术领域，相较单纯依赖关键词检索，能够在一定程度上减少因表述差异带来的遗漏与误判。在此基础上，进一步结合优先权信息识别其中经历美国临时专利申请的专利样本，最终形成本文的研究数据集共有专利 2446 条。

3.2. 结果可视化

3.2.1. 主题词权重分布可视化

为了进一步展示各主题的核心内容，本文调用 BERTopic 模型中的 visualize_barchart() 函数，对识别得到的技术主题进行可视化展示，结果见图 1。图中展示了各主题下的高频关键词及其权重分布情况，纵

坐标为各主题中具有代表性的关键词，横坐标表示对应关键词在该主题中的权重大小。关键词权重越高，说明该词在相应主题中的表征能力越强，也越能够体现该主题的核心内容。

从图 1 可以看出，不同主题之间的关键词构成存在较明显差异，权重较高的关键词多与交易处理、平台服务、数据分析、用户识别、推荐机制、物流协同等内容相关，说明该领域临时专利申请主要集中在平台运行、交易支撑与智能化服务等方面。一些主题中出现的场景化关键词，也表明电子商务技术正不断向更具体的业务环节延伸，形成较细的应用分支。

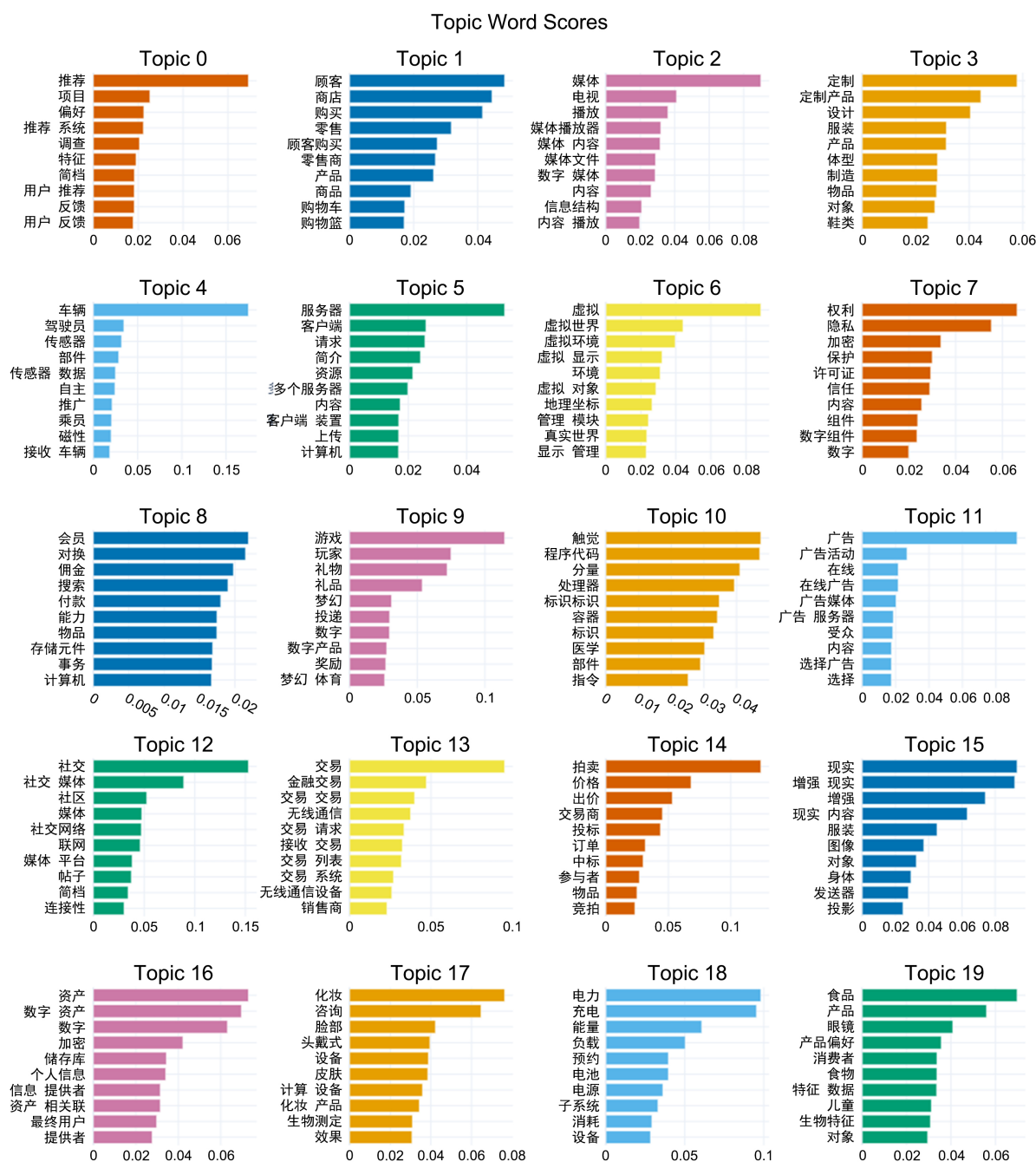


Figure 1. Topic word weight distribution chart

图 1. 主题词权重分布图

3.2.2. 文档 - 主题分布可视化

为了更直观地呈现各主题下的文档分布特征，本文调用 BERTopic 模型中的 visualize_documents()函数进行可视化展示，结果见图 2。图中每一个散点代表一篇专利文本，不同颜色对应不同技术主题，横轴和纵轴分别表示文档在降维空间中的相对位置。散点之间距离越近，说明文本语义相似性越高；散点聚集越明显，说明相应主题内部一致性越强。

从图 2 可以看出，各技术主题在二维语义空间中形成了若干相对集中的聚类区域，说明 BERTopic 模型能够较好地地区分电子商务领域临时专利申请中的不同技术主题。部分主题簇分布较为紧密，散点集中度较高，表明这些主题内部文本语义相近，已经形成较清晰的技术边界。部分主题簇呈现出一定程度的交叉和邻近分布，说明这些主题之间存在较强的关联性，反映出电子商务领域技术内容本身具有明显的交叉融合特征。



Figure 2. Document-theme distribution chart
图 2. 文献 - 主题分布图

3.3. 主题提取

为了在保留细分技术方向的基础上，进一步提炼电子商务领域临时专利申请的整体技术结构，使主题分布更加清晰，也便于后续开展内容分析与演化讨论，本文进一步对语义相近、功能关联较强的主题进行聚类归纳，最终将 33 个合并主题概括为 6 个大类主题(表 1)。

第一类为交易支付与账户管理主题。该类主题主要包括支付、购买、商家、消费者、价格、电子账户、账单、费用、供应商、买方等内容，涉及交易达成、支付结算、账户管理和市场关系等核心环节。对应的合并主题主要有主题 0、主题 13、主题 21、主题 12 和主题 8。这一类主题反映了电子商务最基础的商业运行逻辑，也是整个技术体系中最核心、最稳定的组成部分。

第二类为数据记录与平台支撑主题。该类主题主要包括事务数据、记录、电子文档、数据库、数据项、质量分析、客户端装置、服务器、内容更新、历史信息等内容，体现了电子商务平台运行中对数据记

录、信息管理和系统支撑的技术需求。对应的合并主题主要有主题 14、主题 30、主题 31 和主题 19。这一类主题承担着平台基础设施和信息流转支撑的功能，是电子商务系统持续运行的重要技术底座[6]。

第三类为内容分发与媒体服务主题。该类主题主要包括内容、媒体、无线网络、音频、视频、许可、客户端内容分发、通信信道等内容，涉及数字内容的传输、展示和管理过程。对应的合并主题主要有主题 3、主题 10 和主题 27。这类主题说明电子商务领域的技术布局已经从单纯的交易功能扩展到内容传播、媒体服务和信息递送层面，反映出平台型电商和数字内容服务的融合趋势[7]。

第四类为智能推荐与营销服务主题。该类主题主要包括推荐系统、偏好、用户反馈、广告活动、目标受众、搜索查询、推介、用户界面、应用程序呈现等内容，涉及用户识别、个性化推荐、广告投放和流量分发等环节。对应的合并主题主要有主题 24、主题 5、主题 4、主题 7 和主题 9。这类主题集中体现了电子商务平台在运营端对用户行为分析、精准触达和智能化服务的持续强化[8]。

第五类为场景交互与沉浸体验主题。该类主题主要包括图像显示、图形用户界面、虚拟环境、增强现实、服装试穿、空间交互、情绪情感识别等内容，反映了电子商务技术向可视化、沉浸式和情境化体验方向的延伸。对应的合并主题主要有主题 6、主题 25、主题 28、主题 16 和主题 32。这类主题说明电子商务领域正在不断吸收虚拟现实、图像识别和情感计算等交叉技术，以提升用户体验和交互效率[9]。

第六类为安全认证与垂直应用主题。该类主题主要包括隐私保护、加密、认证、防伪、区块、地理围栏、车辆传感、电力充电、医疗处方、产品真实性、触觉识别等内容，涵盖了电子商务运行中的安全保障机制以及若干细分场景应用。对应的合并主题主要有主题 17、主题 22、主题 15、主题 20、主题 18、主题 11、主题 2、主题 26、主题 23 和主题 29。这类主题虽然内部差异较大，但都具有较强的场景依附性和功能导向性，体现出电子商务技术不断向多行业、多场景渗透的特点[10]。

Table 1. Topic group information
表 1. 主题类群信息

主题类别	包含的合并主题	核心内容
交易支付与账户管理主题	主题 0、主题 13、主题 21、主题 12、主题 8	支付、购买、商家、消费者、价格、电子账户、账单、费用、供应商、买方等
数据记录与平台支撑主题	主题 14、主题 30、主题 31、主题 19	事务数据、记录、电子文档、数据库、数据项、客户端装置、服务器、历史信息、过程控制等
内容分发与媒体服务主题	主题 3、主题 10、主题 27	内容、媒体、无线网络、音频、视频、许可、通信信道、客户端内容分发等
智能推荐与营销服务主题	主题 24、主题 5、主题 4、主题 7、主题 9	推荐系统、用户偏好、广告活动、目标受众、搜索查询、应用程序、用户界面、预测模型等
场景交互与沉浸体验主题	主题 6、主题 25、主题 28、主题 16、主题 32	图像显示、图形用户界面、虚拟环境、增强现实、服装试穿、情绪识别、捕获图像等
安全认证与垂直应用主题	主题 17、主题 22、主题 15、主题 20、主题 18、主题 11、主题 2、主题 26、主题 23、主题 29	隐私保护、加密认证、防伪溯源、地理围栏、车辆传感、电力充电、医疗处方、产品真实性、协作文档、食品偏好等

4. 结论

本文以电子商务领域临时专利申请专利为研究对象，基于 Incopat 全球专利数据库，采用 IPC 分类号检索与优先权信息识别相结合的方式提取样本，并运用 BERTopic 模型对专利标题与摘要文本进行主题

识别与结构分析。研究结果表明,临时专利申请专利能够较好承载电子商务领域研发早期的技术信息,能够作为观察该领域早期技术布局与主题分布的重要样本窗口。通过语义嵌入、降维聚类 and 关键词提取,本文较为清晰地识别出电子商务领域临时专利申请的主要技术主题,主题划分结果具有较好的区分度与可解释性。

从主题识别结果看,电子商务领域临时专利申请的技术内容围绕交易支付、数据记录、内容分发、推荐营销、场景交互、安全认证等多个方向展开,已经形成较为完整的主题结构。主题词权重分布可视化显示,高频关键词主要集中在支付处理、平台服务、数据管理、用户识别、推荐机制等内容,说明交易支撑与智能化服务构成了该领域临时专利申请的核心技术内容。文档-主题分布可视化结果显示,各主题在语义空间中形成了若干相对集中的聚类区域,部分主题边界较为清晰,部分主题之间保持邻近关系,反映出电子商务技术内部既有明确分工,也存在较强的业务耦合与技术融合。

在主题合并与进一步归类后,本文将 33 个合并主题概括为 6 个大类主题,分别为交易支付与账户管理主题、数据记录与平台支撑主题、内容分发与媒体服务主题、智能推荐与营销服务主题、场景交互与沉浸体验主题、安全认证与垂直应用主题。这一结果表明,电子商务领域临时专利申请已经从基础交易环节延伸到平台治理、内容服务、用户运营、沉浸体验和场景应用等多个层面,展现出较强的技术扩展性和场景渗透性[11]。其中,交易支付与账户管理主题、数据记录与平台支撑主题、智能推荐与营销服务主题构成了电子商务运行的主体框架,场景交互与沉浸体验主题以及安全认证与垂直应用主题则体现出新技术不断进入电子商务场景并推动技术边界持续外扩。

本文的研究说明,基于 BERTopic 的主题识别方法能够较好适用于电子商务领域临时专利申请专利的文本分析,能够从语义层面揭示该领域研发早期专利布局的主要技术方向及其结构特征。研究结果为电子商务领域专利情报挖掘、技术热点识别和企业早期专利布局分析提供了方法支持,也为进一步跟踪电子商务领域技术演进和识别潜在重点布局方向提供了基础。

参考文献

- [1] 范文芳,王千. 个性化智能推荐对消费者在线冲动购买意愿的影响研究[J]. 管理评论,2022,34(12): 146-156,194.
- [2] 张米尔,李海鹏,任腾飞. 数字创新的策略性专利行为及相互作用研究[J]. 科学学研究,2022,40(3): 545-554.
- [3] 柏培文,张云. 数字经济、人口红利下降与中低技能劳动者权益[J]. 经济研究,2021,56(5): 91-108.
- [4] 许海云,董坤,隗玲,等. 科学计量中多源数据融合方法研究述评[J]. 情报学报,2018,37(3): 318-328.
- [5] 刘洋,柳卓心,金昊,等. 基于 BERTopic 模型的用户层次化需求及动机分析——以抖音平台为例[J]. 情报杂志,2023,42(12): 159-167.
- [6] 娄策群,李罗佑,王雪莹. 基于 Kano 模型的 B2C 电商平台信息服务功能研究[J]. 现代情报,2021,41(4): 26-35.
- [7] 蓝庆新,窦凯. 美欧日数字贸易的内涵演变、发展趋势及中国策略[J]. 国际贸易,2019(6): 48-54.
- [8] 高文珺,朱迪. 兴趣电商新业态促进消费的机制研究:以国货消费为例[J]. 新视野,2023(5): 83-90.
- [9] 秦芳,王剑程,胥芹. 数字经济如何促进农户增收?——来自农村电商发展的证据[J]. 经济学(季刊),2022,22(2): 591-612.
- [10] 孙秀蕾. 从亚马逊发展模式看数字经济平台的“自我优待”行为及规制[J]. 南方金融,2021(6): 81-89.
- [11] 戚聿东,蔡呈伟. 数字化对制造业企业绩效的多重影响及其机理研究[J]. 学习与探索,2020(7): 108-119.