doi:10.3772/j.issn.2095-915x.2026.04.005

基于 LDA-DistilBERT-LOF 模型的技术创新机会识别研究
—— 以区块链领域为例

开放科学（资源服务）标识码（OSID）

郝世博　刘怡麟

南京理工大学知识产权学院 南京 210094

摘要：[目的 / 意义] 技术创新机会的挖掘与实际技术需求紧密关联，但二者在多数研究中被割裂开来，构建融合技术需求的技术机会识别框架可以实现高效的技术挖掘，系统识别潜在技术创新机会。[方法 / 过程] 构建基于技术创新系统（TIS）理论的框架，在技术需求层，对专利背景技术字段进行 LDA 主题聚类以识别技术需求主题；在技术探索层，结合轻量化 BERT 算法和 LOF 算法识别潜在技术弱信号；在技术评估层，构建创新机会指标评估技术弱信号的潜在价值，最终总结出区块链领域的技术创新机会。[结果 / 结论] 区块链领域技术需求集中在共识算法等八个关键领域，未来技术创新将围绕去中心化、跨链互操作性等展开，研究结果可为企业研发、政府政策制定提供有针对性的决策支持。

关键词：技术机会识别；技术需求；技术创新系统理论；LDA 主题识别；BERT-LOF 模型

中图分类号：G350.7

Research on the Identification of Technological Innovation Opportunities Based on the LDA-DistilBERT-LOF Model: A Case Study of the Blockchain Technology

HAO Shibo　LIU Yilin

School of Intellectual Property, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract: [Purpose/Significance] The mining of technological innovation opportunities is closely related to actual technological needs. However, these two aspects are often treated separately in existing research. Constructing a technology opportunity identification framework that integrates technological needs can enable efficient technology mining and systematically identify potential technological innovation opportunities. [Method/Process] A framework based on the Technological Innovation System (TIS) theory is constructed to identify technological innovation opportunities in the blockchain field. The framework begins by extracting blockchain-related patents from the PatSnap database. At the technological needs layer, the background

作者简介　郝世博（1987-），通信作者，博士，副教授，主要研究方向为知识产权管理、专利计量与数据治理，E-mail：635613519@qq.com；刘怡麟（2001-），硕士研究生，主要研究方向为知识产权管理、专利计量。

引用格式　郝世博 , 刘怡麟 . 基于 LDA-DistilBERT-LOF 模型的技术创新机会识别研究 —— 以区块链领域为例 [J]. 情报工程 , 2026, 12(4): 53-64.

doi:10.3772/j.issn.2095-915x.2026.04.005

technology sections of patents are analyzed using LDA topic clustering to identify key technology demand themes. At the technology exploration layer, a lightweight BERT algorithm combined with the LOF algorithm is employed to detect potential weak technological signals. At the technology assessment layer, multiple innovation opportunity indicators are constructed to evaluate the potential value of these weak signals. Ultimately, technological innovation opportunities in the blockchain field are summarized. [Result/Conclusion] Technological needs in the blockchain field are concentrated in eight key areas, including consensus algorithms. Future technological innovations are expected to focus on decentralization, cross-chain interoperability, and related directions. The research findings can provide targeted decision-making support for enterprise R&D and government policy formulation.

Keywords: Technological Opportunity Identification; Technological Demand; Technological Innovation System (TIS) Theory; LDA; BERT-LOF

引言

在加快实现科技自立自强的国家战略背景下，发展新质生产力对于我国实现高质量发展和现代化建设目标至关重要。2023 年 9 月，习近平总书记在黑龙江调研期间指出要“整合科技创新资源，引领发展战略性新兴产业和未来产业，加快形成新质生产力”。2025 年 3 月，习近平总书记在参加江苏代表团审议时进一步强调“让创新链和产业链无缝对接”。实现创新链与产业链深度融合，亟需建立技术需求识别与技术创新机会发掘的有效联结机制。精准识别特定领域的技术创新机会，不仅能够揭示技术需求缺口，更能为政策制定者提供数据驱动的决策依据，加速关键核心技术自主化进程。

专利数据是分析技术发展趋势与创新机会的重要数据源 [1]。现有研究在技术创新机会识别中仍存在一定局限：一是未能系统整合宏观技术需求与微观创新线索，导致技术机会识别缺乏靶向性；二是过度依赖跨源数据关联 [2]，因数据异构性导致结果可解释性降低，多源数据融合的固有障碍加剧了分析过程对专家经验的依赖 [3]。三是对技术创新机会的评估多停留于定性层面，缺乏多维量化指标支撑。

针对以上问题，本文构建分层递进的技术创新机会识别框架，选用专利文本多字段协同分析策略，运用背景技术字段解析需求，标题与摘要双驱动识别技术弱信号，规避跨源数据对齐难题。创新设计综合技术创新指标体系，为潜在技术创新机会的筛选提供客观依据。以区块链领域为例，验证所提框架的可实践性，为技术创新机会识别提供方法论参考。

1 相关研究

1.1 技术需求挖掘相关研究

随着科技发展，技术需求逐渐成为全球技术创新的核心驱动力，不少研究者也展开了对技术需求的挖掘 [4]。王雪等 [5] 基于 KANO 模型对相关商品的用户在线评论进行需求分析，为企业提供产品改进策略。谢瑶瑶等 [6] 从网站收集用户评论，采用主题依赖注意力模型挖掘技术需求。然而，社交媒体数据往往不能很好地代表技术需求。为避免上述问题，本文选择专利的背景技术字段来进行技术需求识别，专利

背景技术对于创新过程中产生需求情况的描述更加详细和专业，为后续技术创新机会识别奠定基础。

1.2 离群专利相关研究

离群专利作为技术弱信号的载体，其识别方法经历了从传统统计模型到深度学习模型的转变。早期研究依赖 LOF 算法，有研究利用 LOF 算法结合 Doc2vec 向量化，成功识别出农业物联网中的边缘计算专利，其技术路线与主流方案显著不同 [8]。近年来，BERT 模型显著提升了语义表示能力，能够捕捉专利文本中的深层语义关联 [9]。冯立杰等 [10] 验证了利用 BERT-LOF 算法获取离群专利的有效性。丁敬达等 [11] 利用 BERT 模型向量化专利摘要，通过相关离群因子检测算法筛选出新能源汽车领域的离群专利和主流专利。邓娜等 [12] 基于 Sentence-BERT 与孤立森林算法获得离群专利集对专利新颖性进行评估。然而，现有方法大多未充分结合市场需求和技术实际应用场景进行研究。

1.3 技术机会识别相关研究

技术机会分析主要是指通过技术手段和方法从科学情报中挖掘技术机会的过程 [13]，是技术预测领域的新兴分支 [14]。目前主要集中在借助专利地图、网络分析、文本挖掘等手段开展研究 [15]。王金凤等 [16] 利用形态分析和专利文本挖掘获得形态矩阵，并采用技术效益、科研投入等指标进行技术评价；此外链路预测方法也被引入技术机会分析中 [17]。翟东升等 [18] 从宏观层面对技术融合进行网络指标分析，构建运用技术融合演化来分析技术创新机会的框架。侯艳辉等 [19] 考虑技术生命周期的阶段性，对不同生命周期的专利进行文本挖掘和测度，识别出量子信息领域的颠覆性技术。虽然技术机会发现的相关研究快速发展，但从技术需求角度展开技术机会发现的实践较少 [20]。对处于发展初期、隐蔽且具有潜力的技术创新机会识别能力的缺乏，导致技术机会发现的视角不够全面。已有研究将 LOF 方法应用于技术机会识别 [8-10]，表明 LOF 在捕捉技术创新机会方面具有较强适用性。然而，现有方法多侧重于技术本身的离群特征，尚未系统将技术需求与创新价值纳入分析框架。

2 研究设计

技术创新系统（TIS）理论及其系统功能分析法最初是针对特定新兴技术提出的。Carlsson 率先将 TIS 定义为：在特定制度背景下，围绕新兴技术的产生、扩散和应用而相互作用的主体所构成的网络 [21]。随后，Jacobsson[22]、Bergek 等 [23] 进一步丰富和发展了这一定义，认为 TIS 核心在于通过关键系统功能，如知识开发与扩散、创业实验、引导搜索、市场形成、资源动员和合法性建立等促进技术从开发到广泛应用。基于该理论，本文构建了三层分析框架：技术需求层对应 TIS 的“引导搜索”功能，识别技术发展方向与外部需求；技术探索层涵盖“知识开发”和“知识扩散”，负责新知识的生成与传播；技术评估层对应“市场形成”与“资源动员”，关注产业化过程中的市场资源支持。该框架嵌入 TIS 系统功能体系，增强对技术机

doi:10.3772/j.issn.2095-915x.2026.04.005

会识别机制的解释力。

2.1 技术需求层

技术需求层聚焦于从专利文本中精准挖掘技术需求，主要过程涵盖数据采集与预处理、特征工程与模型训练、技术需求识别三个核心环节。首先，将专利说明书中的背景技术字段作为分析对象。在专利撰写工作中，该部分内容需重点阐述技术瓶颈与技术需求。技术需求洞察的具体步骤包括：首先对原始文本进行分词处理及停用词过滤。接着，构建词频 - 逆文档频率矩阵表征文本特征，并采用困惑度与主题一致性指标优化 LDA 主题模型参数，保证主题结果的稳定性与可解释性，据此训练得到高区分度的主题模型。最后，结合主题语义解析，从背景技术中识别出技术需求。

2.2 技术探索层

技术探索层旨在以需求主题为语义背景，识别与其高度相关却尚未形成规模的技术突破点，具体流程涵盖数据预处理、专利语义向量构建及离群检测三个核心环节。首先，对原始专利数据预处理。接着，基于 HuggingFace Transformers 库的 DistilBERT 模型 [25] 构建高维专利语义向量空间。DistilBERT 模型的核心优势源于知识蒸馏技术，相较于原 BERT 模型，DistilBERT 在显著降低模型复杂度与计算资源需求的同时，仍能保持优异的语义性能。最后，使用 scikit-learn 库中局部离群因子算法 LOF 在语义向量空间中实施离群检测，具体为，设第 p 个对象的局部可达密度（lrd）为其 k 近邻的平均可达距离的倒数，记为：

$$lrd_k(p)=\frac{1}{\frac{1}{k}\sum_{o\in N_k(p)}reach_dist_k(p,o)} \quad (1)$$

其中，$reach_dist_k(p,o)$ 为 p 的局部可达距离，$\sum_{o\in N_k(p)}$ 为 p 的 k 近邻集合。

p 的局部离群因子（LOF）定义为：

$$LOF_k(p)=\frac{\frac{1}{k}\sum_{o\in N_k(p)}lrd_k(o)}{lrd_k(p)} \quad (2)$$

计算完毕后返回分类标签，其中被标记为 -1 的专利被界定为潜在技术弱信号的关键载体。

2.3 技术评估层

技术评估层旨在构建多维指标体系识别潜在技术创新机会。传统技术创新指标体系选取的指标围绕被引次数、IPC 数量、非专利引用文献数量展开 [28]。在此基础上，本文以国家知识产权局发布的《专利评估指引》为依据，整合技术、法律与经济三维度，构建综合技术创新指标体系，包括技术影响力指数（TII）、技术多样性指数（TDI）、经济潜力指数（EPI）和技术新颖性指数（TNV）。四项指标具体计算方法如式（3）至式（6）所示。

技术影响力指数（TII）

$$TII=\frac{C_{total}\cdot\log(N_{cite}+1)\cdot N_{5y}}{N_{3y}} \quad (3)$$

其中，C_{total} 表示总被引次数，N_{cite} 表示引用专利数量，N_{3y} 和 N_{5y} 分别表示 3 年和 5 年内被引用数量。

技术多样性指数（TDI）

$$TDI=\sqrt{C_{IPC}}\times\log\left(1+\frac{N_{non-patent}}{N_{total}}\right) \quad (4)$$

其中，C_{IPC} 表示 IPC 分类数量，代表专利覆盖的技术领域广度。$N_{non\text{-}patent}$ 表示非专利引用文献数量，反映基础研究与跨学科融合度。

N_{total} 为总引用文献数 (专利 + 非专利)。

经济潜力指数（EPI）

$$EPI = \left(\frac{N_{countries}}{10}\right)^{0.6} \times \left(\frac{N_{INPADOC}}{100}\right)^{0.4} \quad (5)$$

其中，$N_{countries}$ 表示简单同族国家 / 地区数量，反映技术市场布局广度。$N_{INPADOC}$ 表示 INPADOC 同族专利申请数量，体现全球保护强度。

技术新颖性指数（TNV）

$$TNV = \frac{N_{non-patent}}{N_{cite}+1} \times \left(1-\frac{C_{IPC}}{C_{max}}\right) \quad (6)$$

其中，$N_{non\text{-}patent}$ 表示非专利引用文献数量，反映专利与科学文献的关联度。N_{cite} 表示引用专利数量。C_{IPC} 表示 IPC 分类数量，反映专利的技术集中度。C_{max} 为最大 IPC 分类数量，用于标准化不同技术领域的专利分类数量。

结合上述指标计算综合技术创新机会指数（CTOI）的指标权重分配如下：技术影响力指数（TII）占比 35%，因其通过引用时效性与总量直接量化技术扩散强度，为创新价值实现的核心路径；技术多样性指数（TDI）占比 25%，其 IPC 广度与非专利引用反映技术融合潜力；经济潜力指数（EPI）与技术新颖性指数（TNV）各占 20%，前者通过同族专利布局评估短期市场价值，后者依托科学关联与 IPC 集中度捕捉长期突破潜力，二者平衡体现技术机会识别中现实收益与前瞻布局的协同需求。具体研究流程如图 1 所示。

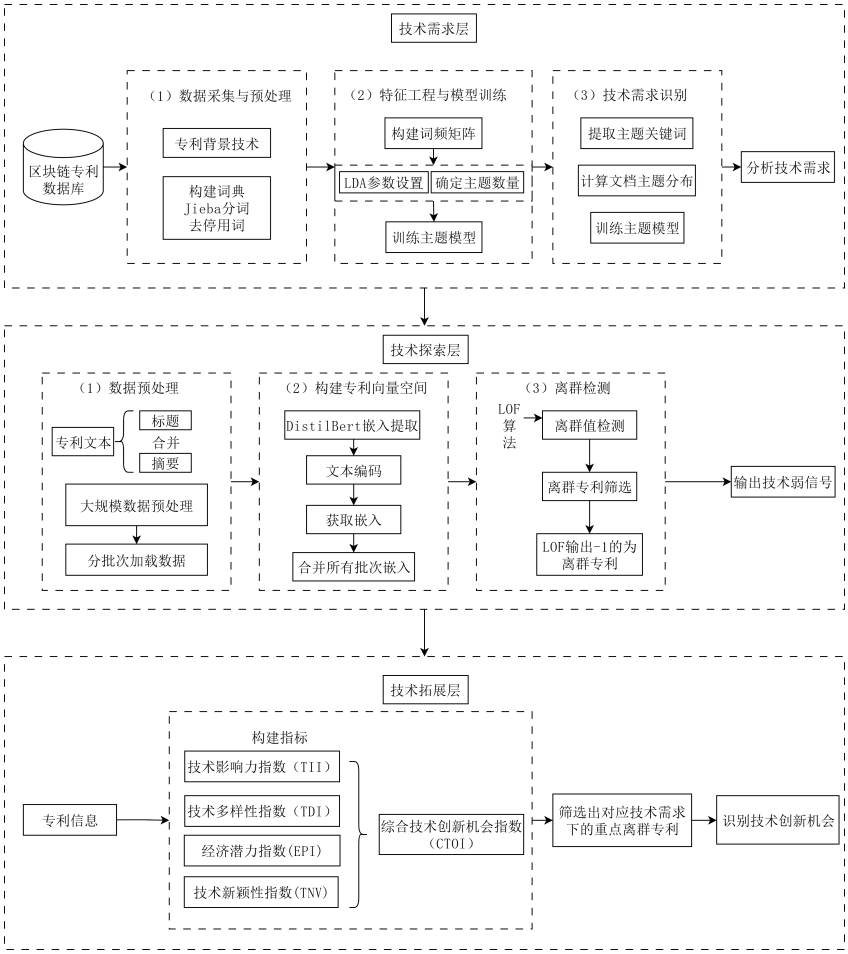


图 1　研究框架

doi:10.3772/j.issn.2095-915x.2026.04.005

3 研究过程

3.1 数据准备及预处理

在数据收集环节，通过 PatSnap 智慧芽全球专利数据库获取数据，采用分类号与关键词结合的检索策略。初步检索得到全球区块链相关专利 55496 项，简单去重清洗处理后获得 55420 项专利，其中受理局为中国的专利数量为 41457 项。具体信息如表 1 所示。

表 1 区块链技术的专利检索信息

项目	内容
平台	智慧芽（PatSnap）专利数据库
检索时间	2025 年 1 月 19 日
检索范围	2005 年 1 月 19 日 —2025 年 1 月 19 日
合并类型	简单同族合并
检索式	(IPC: (G06F OR H04L OR G06Q OR H04N OR H04W OR G06K OR H01L OR G02F OR H04M OR H04B OR H01S OR G01D OR G02B OR G10L)) AND (TA_ALL: (区块链 OR “block-chain” OR “block chain” OR blockchain OR 比特币 OR Bitcoin OR 分布式账本 OR “distributed ledger” OR 非对称加密 OR “asymmetric encryption” OR 智能合约 OR “smart contract” OR 共识机制 OR “consensus mechanism”))
简单法律状态	审中、有效
专利类型	发明申请、授权发明

为了确保专利数据的有效性，本文选择了发明申请与授权发明作为研究对象，并对法律状态进行筛选，仅保留“审中”与“有效”状态的专利。同时，通过合并同族专利，剔除了重复数据，确保数据唯一性。区块链领域专利近 20 年的年度申请数量如图 2 所示。

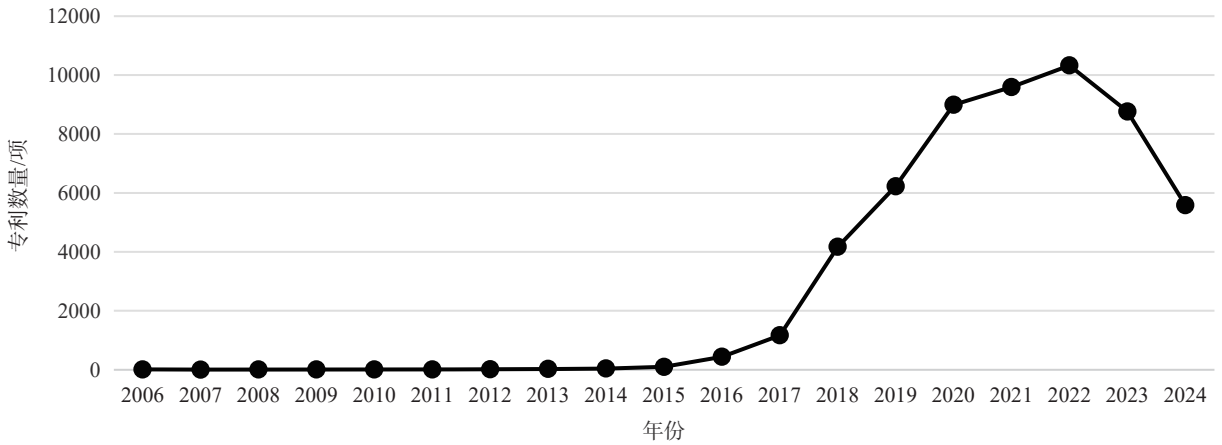


图 2 区块链领域近 20 年年度专利数量

3.2 技术需求主题识别

采用受理局为中国的专利背景技术文本作为 LDA 主题识别的语料。模型训练过程中，为确保选择最优的主题数量和模型参数，结合主题困惑度进行评估，困惑度曲线如图 3 所示。在得到主题词可视化结果后，通过主题之间的重叠度进行调整，最终确定主题数量 K=8。如图 4 所示，各主题分布间重叠较小，各自呈现出独特的关键词。

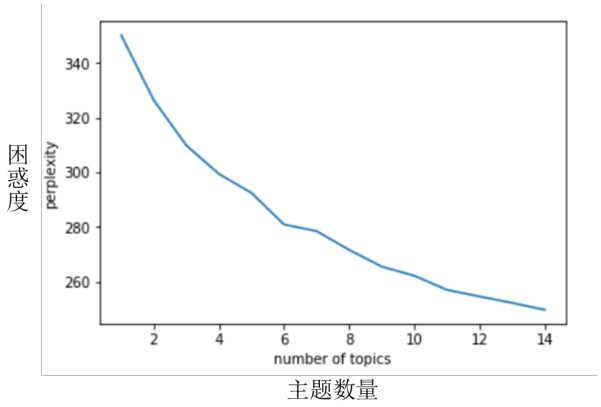
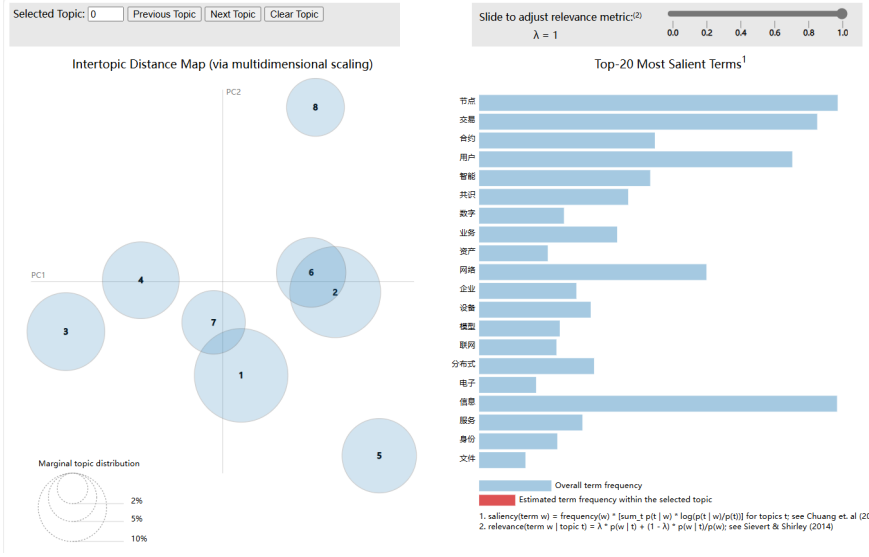


图 3　困惑度曲线



(a) 主题间距离分布图　　(b) 前 20 个高频特征词统计

图 4　LDA 主题模型可视化

通过对区块链技术专利文本的 LDA 主题建模分析，识别出 8 个技术需求主题，如表 2 所示。这 8 个技术需求主题可以看成三个层次：（1）基础层技术需求。Topic1 涉及共识算法，主要关注效率优化与可扩展性需求。Topic2 则集中于网络安全，强调可信环境的构建需求。（2）核心层技术需求。Topic6 关注数字资产的金融基础设施重构需求。Topic7 涉及智能合约的可信代码工程需求。（3）应用层场景需求。Topic3 探讨了供应链管理中的产业协同升级需求。Topic4 关注医疗数据治理需求，体现了现有医疗区块链在非结构化数据处理方面的技术短板以及临床研究场景需要可验证的数据共享机制。Topic5 则聚焦于用户服务，反映了用户对区块链应用的便捷性、安全性等服务具有较高需求。Topic8 涉及能源账本的需求，特别是现有能源系统与区块链的融合需要新型市场机制的设计。

doi:10.3772/j.issn.2095-915x.2026.04.005

表 2　主题识别结果统计

序号	主题名称	主题特征词
Topic1	共识算法	节点 共识 网络 算法 系统 机制 证明 任务 协议 性能 投票 过程 时间 分布式 消息 状态 算力 方法 效率 数量 广播 记账 资源 一致性
Topic2	网络安全	设备 联网 信息 网络 系统 用户 身份 服务器 密钥 方案 方法 管理 车辆 传输 传统 过程 证书 服务 中心 私钥 环境 公钥 信任 移动
Topic3	供应链管理	企业 信息 管理 系统 产品 过程 发展 平台 生产 监管 方式 供应链 风险 传统 机构 方法 效率 物流 成本 影响 情况 评估 流程 商品
Topic4	医疗数据	电子 文项 信息 医疗 方法 检测 方式 视频 图像 模型 特征 过程 分析 内容 领域 系统 文本 患者 效率 大量 传统 发展 时间 证据
Topic5	用户服务	用户 业务 信息 服务 平台 账户 系统 方式 地址 客户 对应 资源 模块 互联网 银行 终端 内容 身份 网络 目标 服务器 对象 体验 过程
Topic6	数字资产	数字 资产 模型 金融 训练 货币 联邦 价值 交易 版权 参与方 凭证 资金 银行 发展 机器 融资 过程 方法 领域 风险 作品 贷款 金融机构
Topic7	智能合约	合约 智能 代码 跨链 漏洞 测试 方法 检测 部署 调用 系统 逻辑 功能 软项 程序 方式 过程 自动 平台 分析 业务 效率 函数
Topic8	能源账本	交易 分布式 数据库 系统 信息 账本 方式 电力 记录 能源 机制 利用 分片 时间 电网 信任 密码学 传输 方法 共识 模式 传统 网络 特点 顺序

3.3　技术弱信号感知

使用技术探索层所构建的方法识别区块链领域的技术弱信号，文本嵌入阶段采用 DistilBERT 模型，设置最大序列长度为 512；使用 [CLS] token 对应的输出向量作为句嵌入表示，维度为 768 维。在 LOF 离群识别阶段设置 n_neighbors 参数为 20，使每个数据点基于其 20 个近邻完成密度特征计算，这一设置在交叉验证与网格搜索中被证明具有稳健性。同时，contamination 参数设定为 0.05，专利数据的长尾效应显示，技术创新在主流方向与边缘领域的非均衡分布使得约 5% 的专利可能属于边缘创新范畴。

提取 LOF 值为 –1 的专利作为离群专利，共得到 2771 项离群专利，具体如表 3 所示。为了分析离群专利在不同技术需求主题下的分布，需要将技术探索层识别出的离群专利映射到技术需求层所定义的八个主题。本文依据 LDA 模型生成的文档 – 主题概率分布进行映射。将每一项离群专利归属于其概率值最高的技术需求主题。数量对应关系如图 5 所示。

表 3　离群专利（节选）

序号	专利标题	LOF 值
1	账本供应链地理定位	–1
2	一种将预订的生命周期放在区块链上的系统	–1
3	使用不可替代代币的人工智能内容生成控制	–1
4	区块链网络中的功能即服务 (FAAS) 平台	–1
…	…	–1
2769	一种基于 diffie-hellman 和智能合约的可验证单往返静态对称可搜索加密方案	–1
2770	基于区块链和 Solidity 的假货识别系统	–1
2771	使用不可替代代币 (NFTS) 的基于区块链的条形码验证系统	–1

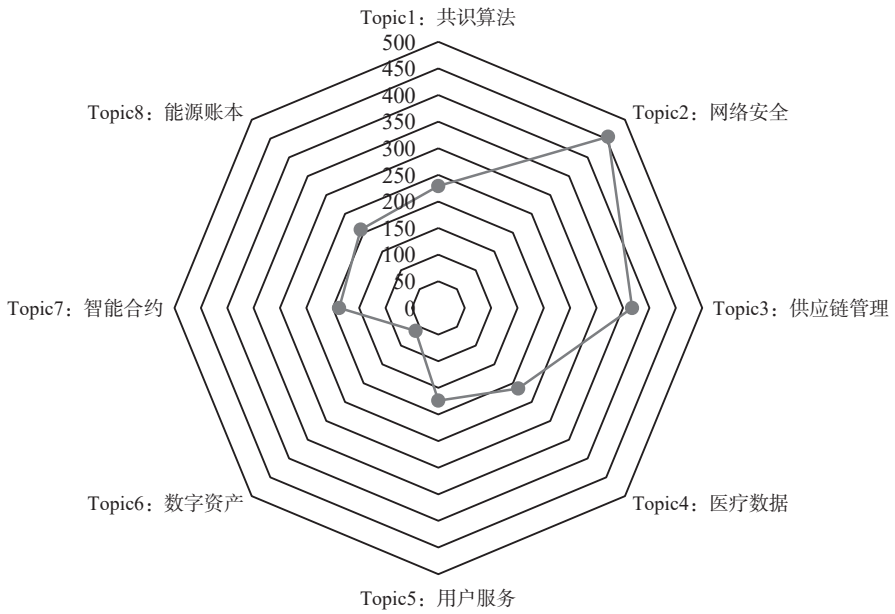


图 5　技术需求主题对应离群专利数量雷达图

总体来看，离群专利数量较多的技术需求领域有更加明显的技术弱信号，拥有快速创新潜力，而离群专利较少的领域则具有更加稳定的技术发展趋势。

3.4　技术机会发现

计算离群专利的综合技术创新机会指数（CTOI），对每个技术需求主题下离群专利的综合技术创新机会指数进行排序，每个技术需求主题选取排名前五的专利进行技术创新机会分析，如表 4 所示。

4　研究结果与分析

根据每种技术需求主题下排名靠前的专利文本内容，识别出以下区块链领域相关技术需求主题对应的技术机会：

①在共识算法技术需求主题下，当前高价值专利主要集中于拜占庭容错机制优化、动态节点管理及信任驱动的 DPoS 方法，并延伸至去中心化身份标识与数据授权融合领域。未来创新应着力构建高效可扩展的轻量级共识机制，提升节点治理自适应能力，推动身份验证与共识协议的深度集成，以增强区块链系统安全性。

②在网络安全技术需求主题下，前沿专利技术聚焦隐私数据保护、身份认证机制及安全架构设计，涵盖属性加密、FIDO 认证等方向。创新重点在于构建精细化数据权限管理体系、多因子去中心化认证机制及支持边缘协同的安全架构，强化数据隐私保护能力。

③在供应链管理主题下，技术创新集中于区块链与艺术品拍卖、防伪溯源、碳排放监管的融合应用，联盟链与隐私保护技术结合展现显著潜力。发展趋势表明，区块链应用正从传统供应链向高价值资产管理、绿色合规审计等复杂场景延伸。未来创新方向包括：多主体协同的链上碳排放追踪平台、具备法规适配能力的供应链治理框架，以及融合可信时间戳与数字水印的艺术品溯源认证体系。

doi:10.3772/j.issn.2095-915x.2026.04.005

表 4　综合技术创新机会指数排名前五的离群专利

技术需求主题	专利标题	申请年	CTOI
Topic1：共识算法	实用拜占庭容错共识机制中动态更改共识节点的方法	2017	40.805
	基于信任机制的区块链 DPoS 共识方法	2019	20.850
	基于去中心化标识的数据授权	2020	20.830
	一种用于区块链的拜占庭容错共识优化方法	2018	18.023
	基于私有区块链的资产交易平台系统及方法	2016	15.221
Topic2：网络安全	基于属性加密的区块链隐私数据访问控制方法	2016	32.633
	基于区块链的 FIDO 认证方法、装置及系统	2017	20.499
	内置于 Paas 服务层的区块链云服务框架系统	2016	16.856
	一种基于区块链的身份认证方法及其系统	2019	15.093
	一种分类分级健身健康大数据共享系统及方法	2021	14.197
Topic3：供应链管理	一种基于区块链的艺术品拍卖方法及系统	2020	9.586
	基于区块链的艺术品防伪电子标签系统与方法	2018	9.458
	一种基于联盟链的 GDPR 合规监管方法及系统	2020	8.611
	产品溯源方法、装置、系统和存储介质	2020	8.467
	基于区块链的企业碳排能耗数据管理运营系统及方法	2022	7.497
Topic4：医疗数据	基于区块链技术的档案数据保护系统及方法	2018	17.376
	区块链加密检索方案	2016	9.582
	OCR 图像样本生成、印刷体验证方法、装置、设备及介质	2020	6.989
	基于 ETL 任务的增量数据抽取方法、装置、设备及介质	2020	5.259
	一种地址识别方法、装置、计算机设备及存储介质	2020	4.757
Topic5：用户服务	一种基于手机耳机接口的区块链数字货币钱包	2016	10.014
	一种利用 NFC 查伪溯源的方法及系统	2016	9.279
	基于区块链的组网方法、装置、计算机设备和存储介质	2018	8.392
	基于可链接环签名和多重签名的区块链匿名签名方法	2019	6.574
	基于可信执行环境的操作记录存储方法、装置及设备	2019	5.940
Topic6：数字资产	一种基于 NFR 的权益管理系统	2022	3.568
	基于哈希图和联邦学习的数据安全共享的方法	2020	3.310
	一种基于非同质通证发行同质通证的方法	2019	3.236
	一种基于 bancor 协议的区块链资源兑换方法及装置	2019	3.190
	一种基于区块链的数据湖系统架构	2020	2.434
Topic7：智能合约	一种基于区块链的跨链数据操作方法和装置	2018	14.090
	一种基于 Java 虚拟机的智能合约执行引擎	2017	12.539
	区块链节点部署方法、装置、设备和存储介质	2019	8.798
	一种 Springboot 程序加密和解密的装置及方法	2019	7.910
	一种基于区块链智能合约的数据处理方法及系统	2018	7.329
Topic8：能源账本	一种适合区块链隐私保护的零知识证明方法和介质	2017	50.758
	一种基于区块链的碳交易系统	2018	23.006
	共识任务协调方法及装置、区块链系统、存储介质	2019	19.279
	一种基于区块链的在分布式数据库中进行交易查询的方法	2017	15.172
	一种基于 SQL 数据库的区块链状态存储方法	2017	14.390

④在医疗数据主题中，专利技术集中于医疗档案保护、加密检索、影像识别及增量数据抽取领域，反映对数据安全、结构化处理与智能识别的需求提升。创新重点应聚焦：异构医疗数据的链上集成与跨机构共享方案、隐私保护型数据智能检索机制，以及整合 OCR 识别与链上身份验证的自动化档案管理系统，以提升医疗数据流通安全性与效率。

⑤在用户服务主题下，高价值专利涉及硬件接口钱包、NFC 防伪验证、组网技术及匿名签名方法，凸显终端应用的多元化潜力。未来应开发轻量低功耗硬件钱包系统，融合 NFC 与区块链实现物品可信溯源，并构建基于可信执行环境（TEE）的用户行为数据加密存储与追溯机制。

⑥在数字资产主题中，专利技术集中于 NFT 权益管理、联邦学习数据共享及资源兑换机制，体现新型资产管理与跨域协同的探索方向。需建立统一的多类型数字资产链上表征与交易框架，结合隐私计算实现可信数据价值交换，拓展通证经济范式下的资产发行、激励与流通模式，支撑复杂商业生态创新。

⑦在智能合约技术需求下，重点专利关注跨链数据操作、合约执行引擎优化，尤其注重 JVM 定制化与执行保护。创新方向包括：异构链跨链智能合约交互协议、多语言合约执行的高效虚拟机架构，以及具备隐私保护与行为审计功能的合约安全运行体系。

⑧在能源账本主题中，前沿专利涵盖零知识证明、碳交易系统、SQL 状态存储及任务协调技术，体现区块链在能源数据管理与隐私保护的综合优势。创新应开发基于零知识证明的能耗数据匿名验证机制，构建去中心化碳资产交易平台，设计支持分布式调度与状态管理的链式账本架构，赋能能源系统绿色转型与可信治理。

5 结论与展望

研究结果表明，本文提出的基于 LDA-DistilBERT-LOF 模型的技术创新机会识别框架具有可行性，详细揭示了区块链技术的技术需求与创新机会。在技术需求层，研究通过 LDA 模型从专利文本中提取出八个技术需求主题，涵盖共识算法、网络安全、供应链管理、医疗数据等多个领域。在技术探索层，结合 DistilBERT-LOF 模型，成功识别区块链领域的潜在技术弱信号，捕捉到一些尚未被广泛关注的技术创新点。在技术评估层，通过构建多个创新机会指标，评估技术需求和创新点的潜在价值，最终对应不同技术需求识别出可能推动区块链技术未来发展的创新机会。

本文也存在一些局限性。尽管采用多种先进的数据分析技术，由于技术本身的复杂性，模型精度和结果可解释性仍有待提高。未来研究可进一步优化数据预处理与分析方法，结合多源创新数据打破信息不对称性，提供更加全面的技术创新视野。可进一步结合深度学习等技术提高弱信号识别能力，并结合自然语言处理技术提升技术需求识别精确度。

参 考 文 献

[1] ROBINSON D K R, HUANG L, GUO Y, et al. Forecasting innovation pathways (FIP) for new and emerging science and technologies[J]. Technological

doi:10.3772/j.issn.2095-915x.2026.04.005

Forecasting and Social Change, 2013, 80(2): 267-285.
[2] DING Y, CHOWDHURY G, FOO S. Journal as markers of intellectual space: Journal co-citation analysis of information retrieval area, 1987—1997[J]. Scientometrics, 2000, 47(1): 55-73.
[3] 韩晓彤 , 朱东华 , 汪雪锋 . 科学推动下技术机会发现方法研究 [J]. 图书情报工作 , 2022, 66(10): 19-32.
[4] 周泽炯 , 陆苗苗 . 战略性新兴产业自主创新能力的驱动因素研究 [J]. 吉首大学学报 (社会科学版), 2019, 40(1): 30-38.
[5] 王雪 , 董庆兴 , 张斌 . 面向在线评论的用户需求分析框架与实证研究 —— 基于 KANO 模型 [J]. 情报理论与实践 , 2022, 45(2): 160-167.
[6] 谢瑶瑶 , 邓三鸿 , 王昊 , 等 . 基于 SAO 语义挖掘的用户需求 - 专利技术匹配 [J]. 数据分析与知识发现 , 2024, 8(Z1): 213-225.
[7] ÇALDAĞ M T, GÖKALP E. Exploring critical success factors for blockchain-based intelligent transportation systems[J]. Emerging Science Journal, 2020, 1(3): 27-44.
[8] 宋凯 , 冉从敬 . 基于主题挖掘与专利评估的技术机会识别研究 —— 以智慧农业为例 [J]. 图书情报工作 , 2023, 67(3): 61-71.
[9] 孔德婧 , 董放 , 陈子婧 , 等 . 离群专利视角下的新兴技术预测 —— 基于 BERT 模型和深度神经网络 [J]. 图书情报工作 , 2021, 65(17): 131-141.
[10] 冯立杰 , 秦浩 , 张珂 , 等 . 基于离群专利和多维技术创新地图的技术机会识别路径研究 [J]. 情报理论与实践 , 2023, 46(9): 79-86.
[11] 丁敬达 , 汤迎涛 , 邱均平 . 专利离群程度演变视角的高价值专利评估研究 —— 以新能源汽车领域为例 [J]. 情报科学 , 2025, 43(11): 100-108.
[12] 邓娜 , 王雨佳 , 杨洋 , 等 . 基于 Sentence-BERT 与孤立森林算法的专利新颖性评估 [J]. 情报杂志 , 2025, 44(2): 174-182.
[13] PORTER A L, JIN X Y, GILMOUR J E, et al. Technology opportunities analysis: Integrating technology more[J]. SRA Journal, 1994, 26(2): 21-30.
[14] LEE M H, KIM S, KIM H, et al. Technology opportunity discovery using deep learning-based text mining and a knowledge graph[J]. Technological Forecasting and Social Change, 2022(180): 121718.
[15] 伊惠芳 , 刘细文 , 龙艺璇 . 技术创新全视角下技术机会发现研究进展 [J]. 图书情报工作 , 2021, 65(7): 132-142.
[16] 王金凤 , 吴敏 , 岳俊举 , 等 . 创新过程的技术机会识别路径研究 —— 基于专利挖掘和形态分析 [J]. 情报理论与实践 , 2017, 40(8): 82-86.
[17] 王金凤 , 杨慧琳 , 赵伟宇 , 等 . 知识迁移下基于融合知识网络与链路预测的技术机会识别 [J]. 情报杂志 , 2025, 44(5): 165-173.
[18] 翟东升 , 张京先 . 基于专利技术共现网络的无人驾驶汽车技术融合演化研究 [J]. 情报杂志 , 2020, 39(4): 60-66.
[19] 侯艳辉 , 陈荣 , 王家坤 . 技术生命周期视角下颠覆性技术早期识别方法研究 [J]. 情报学报 , 2025, 44(2): 157-170.
[20] 伊惠芳 , 刘宁 , 单晓红 , 等 . 基于需求 — 技术联合分析的技术机会发现研究 [J]. 情报理论与实践 , 2024, 47(5): 18-29.
[21] CARLSSON B, JACOBSSON S, HOLMÉN M, et al. Innovation systems: analytical and methodological issues[J]. Research Policy, 2002, 31(2): 233-245.
[22] JACOBSSON S, JOHNSON A. The diffusion of renewable energy technology: An analytical framework and key issues for research[J]. Energy Policy, 2000, 28(9): 625-640.
[23] HEKKERT M P, SUURS R A A, NEGRO S O, et al. Functions of innovation systems: A new approach for analysing technological change[J]. Technological Forecasting and Social Change, 2007, 74(4): 413-432.
[24] BERGEK A, JACOBSSON S, CARLSSON B, et al. Analyzing the functional dynamics of technological innovation systems: a scheme of analysis[J]. Research Policy, 2008, 37(3): 407-429.
[25] SANH V, DEBUT L, CHAUMOND J, et al. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter[J]. Victor Sanh;Lysandre Debut;Julien Chaumond;Thomas Wolf.CoRR. 2019.
[26] 尚洪涛 , 房丹 . 政府补贴、风险承担与企业技术创新 —— 以民营科技企业为例 [J]. 管理学刊 , 2021, 34(6): 45-62.
[27] 陶锋 , 朱盼 , 邱楚芝 , 等 . 数字技术创新对企业市场价值的影响研究 [J]. 数量经济技术经济研究 , 2023, 40(5): 68-91.
[28] PONTA L, PULIGA G, MANZINI R. A measure of innovation performance: the innovation patent index[J]. Management Decision, 2021, 59(13): 73-98.

（责任编辑：李昶璇）