

数据分析与知识发现
Data Analysis and Knowledge Discovery
ISSN 2096-3467, CN 10-1478/G2

《数据分析与知识发现》网络首发论文

题目：基于集成学习的颠覆性技术早期识别研究——以量子计算领域为例
作者：赵一鸣，刘顺生，吕璐成
网络首发日期：2025-02-27
引用格式：赵一鸣，刘顺生，吕璐成. 基于集成学习的颠覆性技术早期识别研究——以量子计算领域为例[J/OL]. 数据分析与知识发现.
<https://link.cnki.net/urlid/10.1478.G2.20250227.1446.006>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于集成学习的颠覆性技术早期识别研究 ——以量子计算领域为例

赵一鸣^{1, 2, 3}, 刘顺生^{2, 4}, 吕璐成^{5, 6}

- (1. 武汉大学信息资源研究中心 武汉 430072
2. 武汉大学信息管理学院 武汉 430072
3. 武汉大学大数据研究院 武汉 430072
4. 武汉大学图书情报国家级实验教学示范中心 武汉 430072
5. 中国科学院文献情报中心 北京 100190
6. 中国科学院大学经济与管理学院信息资源管理系 北京 100190)

摘要：[目的]通过分析关键技术领域下的专利数据，对具备高颠覆性潜力的技术专利进行早期识别。[方法]基于技术生命周期理论搭建颠覆性技术早期识别指标体系，以智慧芽专利数据库中量子计算领域的专利数据作为研究对象，搭建集成学习模型对该领域具有高颠覆性潜力的技术专利进行早期识别。[结果]通过 BERTopic 主题建模框架发现五个相关颠覆性研究方向：量子加密技术、量子处理器、超导量子比特、半导体技术与量子神经网络，验证了本文方法的有效性与可行性。[局限]实证分析仅围绕量子计算领域展开，未能全面涉及不同的关键技术领域；框架构建与指标提取仅依赖专利数据，可拓展支撑数据的来源种类。[结论]本研究有助于早期识别出高颠覆性潜力的技术专利，进而分析主要的颠覆性研究方向，为国家重大科技战略的制定实行提供研判依据。

关键词：颠覆性技术；集成学习；技术生命周期；量子计算

分类号：G306

1 引言

颠覆性技术能够对主流技术产生替代，对市场结构产生变革^[1]。对于企业乃至国家层面而言，精准识别颠覆性技术承载着举足轻重的战略价值。然而，颠覆性技术的发展具有发展周期长、不确定性高等特点，传统的技术评估手段难以在技术发展早期精确衡量其潜在价值。因此，如何在新技术步入市场初期、完成格局重塑前即有效识别其颠覆性潜力，已成为亟待解决的重要现实问题。

近年来学者们在颠覆性技术识别方面进行了诸多尝试，采用专家意见^[2]、网络分析^[3]、科学计量^[4]、文本挖掘^[5]、机器学习^[6]等方法展开系列研究，利用主观或客观的方式搭建颠覆性技术识别框架，提出各种场景下的指标及其计算方法，综合熵权法、层次分析法、相似性度量法等方法赋予指标权重。颠覆性技术识别研究现阶段仍存在以下不足之处：①颠覆性技术识别指标体系的科学性完整性有待商榷，无法良好地涵盖技术展现颠覆性历程的所有特征；②颠覆性技术识别指

标值的提取往往因计算的简便性仅选用某一类特征，浅层外部特征、引文特征与深层语义特征等多维度特征的结合程度有待提升；③颠覆性技术识别模型的构建过分依赖单一模型在数据集上的优越性，没有充分发挥各种通用算法模型对不同数据特征的代表能力。

鉴于此，本研究基于生命周期理论，聚焦专利文献中各阶段特征对技术颠覆性的早期预判能力，综合使用文本向量表示、引文分析、科学计量等方法提取技术识别特征，利用颠覆性指标筛选高颠覆性潜力技术，通过集成学习算法建立技术识别特征与颠覆性之间的联系，以量子计算领域为例开展颠覆性技术早期识别的实证研究。

2 相关研究

颠覆性技术概念的提出最早可以追溯到 1995 年《哈佛商业评论》的破坏性创新理论，Bower^[7]言简意赅地概括了该类技术的市场化历程：“往往从低端或边缘市场切入，以简单、方便、便宜为初始阶段特征，并随着性能与功能的不断改进与完善，最终取代已有技术，开辟新市场，形成新价值体系”。随着科技不断革新，数据获取难度逐渐降低，颠覆性技术识别研究也逐渐从主观向客观过渡，以下将从研究方法的角度进行文献回顾。

2.1 基于专家意见的识别方法

在研究初期，学者倾向采用领域专家的已有知识和过往经验来评判技术的颠覆性潜力。Matti S 等^[8]采用视觉模拟量表收集机械制造行业专家的意见，针对数字化对该行业可能引发的未来颠覆性变革进行了实证性分析。该方法很大程度上依赖于专家个体的主观臆断，专家间达成共识颇为困难，极有可能导致某些具有颠覆性潜力的新技术遭到忽视。此外，该方法需要多名专家协同参与，耗时耗力，难以在复杂技术领域进行大规模应用。

2.2 基于网络分析的识别方法

网络分析方法依赖网络这一形态提取可计算的特征值，一般将文献或关键词视为节点，通过引文、共现等关系确定连边来搭建网络，分析网络中各节点的中心性特征，进而识别颠覆性技术。Jia 等^[9]就使用科学文献关键词共现网络测度

颠覆性技术的前沿性指标,采用知识图谱呈现中国电子信息领域中的颠覆性技术发展路线图。邢晓昭^[10]等就运用社会网络分析方法定义技术演化过程中出现的新主题出现、原主题趋热和多主题融合三类突变情况,识别出类脑领域中的颠覆性技术。该类方法所采纳的某些理论框架未能充分考虑颠覆性技术的独特属性,所运用的评价指标相对单一,其基于的简单规律仅能有效识别出已明显展现出颠覆性的技术,而对于那些潜藏着颠覆性可能性的技术则缺乏前瞻性的预测能力。

2.3 基于科学计量的识别方法

科学计量方法开始探索使用定量指标来测度技术颠覆性程度,这些指标一方面可以是直接取自数据源的,苏敬勤等^[11]就采用专利前三年引用率计算得来的专利影响因子对技术颠覆性进行研判。另一方面,也可以是根据问题场景提出科学化计量的维度指标,这些维度指标的归纳主要考虑颠覆性技术在科学、技术、市场等^{[12][13]}方面与传统技术的差异,李乾瑞等^[14]就从技术融合性、新颖性、扩张性和影响力四个维度构建系统性的颠覆性技术识别模型。该类方法通过定量手段评估技术的颠覆性程度,能够系统性且高效地揭示颠覆性技术的内在特征,但指标选取多是围绕颠覆性技术的趋势、影响和应用展开的,无法预见那些处于试验探索阶段、颠覆性现象尚未外化的技术。

2.4 基于文本挖掘的识别方法

文本挖掘方法尤为关注非结构化文本数据中的潜在语义特征,并将数据源扩展到公众媒体数据。Xiang 等^[15]利用改进的 LDA2Vec 模型识别颠覆性技术中的主要话题,并基于时间序列判断技术的发展趋势。Pan 等^[16]使用词频向量构建用于颠覆性技术认定的专利演化矩阵。Li 等^[17]运用文本情感分析工具提取网络新闻数据中公众对技术的期望变化模式,判别新兴技术专利是否具有颠覆性发展趋势。Qiao 等^[18]从技术替代的视角出发,使用 SAO 语义分析和社区检测算法筛选候选技术,综合进步性、适用性等特征识别技术的颠覆性。该类方法聚焦于文本数据中词频、情感及语义特征的关注度,内核在于利用时间趋势识别颠覆性技术,但因其往往需要基于技术的长时间演化来总结规律,故在预测能力上存在较大的局限性。

2.5 基于机器学习的识别方法

文本挖掘方法在实际施行过程中往往需要个人智慧的介入，技术识别精度与技术领域熟悉程度直接挂钩，研究人员尝试采用机器学习算法来实现技术识别工作的一体化、自动化^[19]，Chen X^[20]开展特征工程初步提取颠覆性技术的大部分特征，并使用 Gartner Hype Cycle 作为训练数据，尝试进行可控的技术预测。He GX^[21]设计新型神经网络软提示感知共享 BERT，比较目标论文与其参考文献在摘要上的语义差异性，预测即时发布论文的潜在颠覆性指数。该系列方法侧重于模式发现的相关性探索，却往往忽视了规律背后的因果机制，这导致模型在可解释性方面存在明显局限，影响了其在实际应用中的全面评估与优化。

2.6 研究述评

通过对已有研究的系统性梳理，不难发现用于颠覆性技术识别的评估指标选取向可测度方向演变，从概念化的维度特征到简单具象的外部特征再到深层抽象的语义特征；数据的获取途径从单一到多源，逐渐囊括科学论文、技术专利、网络舆情等数据类型。然而不可忽视地，依赖专家智慧对候选技术进行颠覆性认定消耗过多人力资源不能在现实中大规模应用，网络分析中单一化指标难以涵盖颠覆性发展中的关键特征，科学计量与文本挖掘方法采用趋势性规律面临的耗时性问题使得技术识别的前瞻性无法得到很好保障。相较而言，机器学习方法善于依据特征从大量数据中发现潜在模式，能够有效缩短技术识别工作的时间跨度，但关键特征选用一般缺乏因果性的理论支撑。本研究基于技术生命周期理论，期望从颠覆性技术生命周期入手构建早期识别框架，采用集成学习策略搭建颠覆性技术识别模型，对具有高颠覆性潜力的技术专利进行精准识别。

3 研究设计

3.1 颠覆性技术早期识别的框架构建

技术生命周期理论^[22]认为技术的发展与演化是一个渐变过程，反映了技术从创意到商业化应用，再到市场饱和与最终衰退的全过程，可以总结为六个阶段：

技术开发阶段、技术验证阶段、技术应用启动阶段、技术扩张阶段、技术成熟阶段、技术退化阶段。基于技术自身的内在扩张力,技术专利的发明与扩散历经“理念引入-开发验证-市场应用-成长扩张”的发展路径。本文认为,技术发展进程中同类技术的演化可视为相似技术专利的发明与继承链条,这些专利围绕共同需求目标相互衔接,构成技术颠覆性演进的基本单元,推动技术体系不断发展。

本文类比技术生命周期理论,结合技术发展进程中的内在扩张力,考虑到颠覆性技术专利从理念设计到成长扩张的内在循环发展路径,搭建颠覆性技术早期识别框架,如表 1 所示,期望在颠覆性现象出现前识别出技术的核心研究方向及其代表性专利。技术引入阶段体现的是颠覆性技术的诞生初衷是源于社会重大需求,其设计理念有别于主流技术,期待在成果上取得突破性进展。技术验证阶段体现的是研究人员如何将先验知识与专家智慧融入到技术专利的开发中,其技术潜力是区别于噪声数据的重要因素。技术应用阶段体现的是技术颠覆性成果向市场实践的转化,是技术实现其最终价值的关键环节,其成效将通过未来市场上的结果反馈得以反映。技术扩张阶段体现的是单一技术专利通过向其他技术专利扩张,本质上是技术专利中的颠覆性潜质的继承与改进,关系到技术颠覆性潜力能否在专利间平稳过渡,如果没能得到很好的传承,那么这项技术专利只能沉淀在领域专利的数据汪洋中,失去发挥颠覆性潜能的机会。技术搭乘专利的载体不断演变并逐步累积颠覆性潜质,达到某种程度便步入技术成熟阶段,触发颠覆性现象,对市场格局产生巨大冲击,获取非比寻常的价值收益。

表 1 颠覆性技术早期识别框架

Table1 Early Identification Framework for Disruptive Technologies				
阶段	维度	特征层次	计算指标	参考文献
颠覆性技术引入	异轨性 TT	引文特征	引证文献差异性	苏鹏 ^[23] ,刘云 ^[24] ,等
	新颖性 TN	语义特征	往期文献主题差异性	王莉晓 ^[25] ,等
	突破性 TB	外部特征	IPC 分类号大组	陈育新 ^[12] ,黄鲁成 ^[26] ,等
颠覆性技术验证	技术知识量 TK	外部特征	①页数 ②IPC 分类号数量	冉从敬 ^[27] ,等
	理论支撑性 TS	引文特征	①专利引用量 ②非专利引用量	Kong J ^[28] ,等
	团队贡献度 TC	外部特征	①专利权人实力 ②发明人实力	Xu S ^[29] ,等
颠覆性技术应用	市场吸引力 MA	外部特征	专利家族数量	邢晓昭 ^[30] ,等
	市场竞争力 MC	外部特征	专利权利要求数	陈育新 ^[12] ,等
颠覆性技术扩张	替代性 TA	语义特征	未来文献主题差异性	王超 ^[31] ,等
	影响力 TI	引文特征	①短期被引频次 ②被引领域数量	Jia WF ^[32] ,等

在颠覆性技术引入阶段，主要考察技术在理念设想上是否潜藏颠覆性，具体维度有异轨性、新颖性、突破性。技术最初凭借其独有的创新性思维展现出颠覆性特质，或是通过技术轨道的分化，或是研究焦点的转移，研究者希望从这两种途径出发构思技术原型以取得突破性成就。

在颠覆性技术验证阶段，主要考察技术颠覆性在落地应用中是否可行，具体维度有技术知识量、理论支撑性、团队贡献度。此阶段，研究团队致力于将原先构思的技术原型转化为可实施的技术专利，这一过程实质上涉及知识的创新与价值的转化。团队凭借自身实力，并融合其他既有知识，通过知识的解构、重组、吸收、凝聚，为技术专利赋予了知识层面的内在价值，而其外在特征则表现为技术的颠覆性。

在颠覆性技术应用阶段，主要考察技术进入市场中带来的价值提升，具体维度有市场吸引力、市场竞争力。凭借其独特的优越性，颠覆性技术在步入市场初期即吸引特定用户群体注意力。在经历与市场既有技术的激烈竞争后，该类技术能够成功占据一个稳固的市场生态位，并逐步对主流市场实现渗透与拓展。

在颠覆性技术扩张阶段，主要考察技术对同领域技术发展产生的影响，具体维度有替代性、影响力。一方面，颠覆性技术凭借自身高效益挤占生存空间，对原有同定位技术产生取代作用；另一方面，颠覆性技术作为领域内上游技术打通下游技术先前存在的堵点，促进全领域技术循环发展。

3.2 颠覆性技术早期识别的指标提取

（1）颠覆性技术引入阶段

① 异轨性

针对某一技术领域内具备奠基性质与开创意义的研究成果而言，其所援引的参考文献往往属于该领域内鲜少被其他研究者引用的范畴。某一专利的被引用频次，对其引证文献的异轨性具有直接影响；具体而言，若某项专利的被引次数相对较低，则表明该专利后续的引证文献所蕴含的技术路径，与其他技术路径存在着较大的差异性。基于上述思路，作者尝试引入专利引用网络计算技术的异轨性，提出以下公式（1）和（2）。

$$TT_p = \frac{1 + \sum v_{\bar{r}}}{1 + \text{num}(v_{\bar{r}})} \quad (1)$$

$$v_{前} = \frac{TT_p}{\text{num}(v_{前})} \quad (2)$$

专利往往拥有一定的创新性，对引文网络中所有专利节点初始值设为 1。对于某一具体专利 p 而言，其异轨性指数被后向引用关系修正，并借由前向引用关系传递到下一节点。公式中 TT_p 表示专利节点 p 的异轨性指数， $v_{前}$ 表示专利节点 p 前向引用关系值的集合， $v_{后}$ 表示专利节点 p 后向引用关系值的集合。

② 新颖性

对主题新颖性的测度，学术上普遍的做法是由技术主题所含文档的平均出现时间^[25]来进行测度，本文仿照传统思路，限定主题聚类时间为 1 年，利用固定时间窗口内主题文档的数量来表征单篇技术专利的主题新颖性，某一主题下的文档数目越多则主题新颖性欠佳。

预训练语言模型 BERT 能够捕捉词汇的上下文语义信息，且由多任务目标训练而来，具备不俗的泛化能力，操作直接有效。针对某一专利 p ，本文在对专利标题和摘要进行停用词剔除后输入 BERT 自带的词嵌入工具转换为词向量，采用基于密度的层次聚类算法 HDBSCAN 对过往 1 年内的专利文本进行自适应主题聚类，用专利所属主题具有的专利数除以专利总数来表征技术的新颖性。

③ 突破性

技术的突破性创新^[33]不仅注重从来源上与现有主导技术的差异性，同时也强调不同技术间的融合发展^[34]。作者迁移陈育新^[12]源于逆文档率的技术突破性计算公式，并增加技术融合修正项以补足技术相互融合带来的突破性。

$$TB_p = \frac{\sum_{i \in IPC_p} \log \frac{N}{N_i}}{\text{num}(IPC_p)} + \frac{\text{num}(NC_p)}{\text{num}(C_p)} \quad (3)$$

p 表示某篇具体的专利文献， TB_p 表示专利 p 的突破性指数。 IPC_p 表示专利 p 拥有的 IPC 大组集合， N 表示在专利 p 申请日之前某一固定时间窗口内的专利总数， N_i 表示其中拥有 i 这一 IPC 大组的专利数量。 C_p 表示专利 p 含有的 IPC 大组成对组合集合， NC_p 表示相较于此前专利新出现的成对组合集合。特别地，如果 C_p 为空集，那么此公式的后项值取 0。前项固定时间窗口是考虑到颠覆性技术的不连续性，与渐进性创新有所差异，本文设置为 1 年。

(2) 颠覆性技术验证阶段

① 技术知识量

从知识载体的角度，专利文献全面介绍发明结构、要点及用法，其篇幅长度

直观反映专利技术复杂度与深度。从专利分类角度，IPC 分类号从主题应用的角度出发，准确表示了发明的技术创新点及其应用范畴，能够从侧面反映专利中的技术知识。本文借鉴冉从敬^[27]的专利质量度量指标，将专利文献篇幅和 IPC 分类号作为技术知识量维度下的直接测度指标。

② 理论支撑性

根据知识流动理论^[35]，文献间的直接引用关系说明知识发生了流动，而知识流动程度可以用引用频次来量化。本文考虑技术颠覆性创新过程中起主导作用的技术知识和科学知识^[28]来测度技术的理论支撑性，专利引用量反映当前技术对以往技术知识的吸收程度，非专利引用量反映科学知识向技术领域流动程度。

③ 团队贡献度

从作者署名位次入手，根据公式（4）^[29]来分配贡献度，并作归一化处理，某个专利的发明人贡献度得分则是其具有的所有发明人贡献度之和；而专利权人的贡献度则取位次第一的专利权人所发表的专利数量。

$$c_{p,i} = \begin{cases} \frac{1}{i}, & i \leq 10 \\ 0.05, & i > 10 \end{cases} \quad (4)$$

其中，p 表示某篇专利文献，i 表示署名位次， $C_{p,i}$ 表示专利 p 排名 i 的发明人贡献度。对于名称消歧，智慧芽数据库对专利权人字段进行了专门标注，保证了名称的唯一性，本文认为同一技术人员短时间不会频繁切换工作单位，并且工作单位的变动往往是存在合作关系的，将具有相同作者名称缩写且专利权人存在合作关系的作者视为同一人。

（3）颠覆性技术应用阶段

本阶段采用的维度有市场吸引力和市场竞争力。专利家族凭借自身在多国的优先权抢先占领市场份额^[30]，故其规模能够衡量技术专利的市场吸引力。高价值专利为确保创新主体能够充分享有其创新成果所带来的权益，往往申请的保护范围广，权利要求条数多^[12]，故权利要求数能够衡量技术专利的市场竞争力。

（4）颠覆性技术扩张阶段

本阶段采用的维度有替代性和影响力。替代性用未来 1 年内专利主题的差异性进行表征，专利聚类与计算方式同主题新颖性，而影响力则用短期被引频次和被引领域数量^[32]进行表征。

3.3 集成学习模型搭建

为避免单一分类算法因对数据集不同特征适应程度有所差异带来的评估结果偏差问题，本研究尝试采用集成学习策略来进一步提高模型的分类精度。图 1 展示了本文采用的集成学习框架，各个基学习器分别学习颠覆性技术特征，通过模型融合层综合各个类别结果，最后得到颠覆性技术识别结果。集成学习策略综合不同基学习器的训练效果来给出最终结果，在实际应用中应尽可能采用差异化的算法模型作为基学习器，而模型融合层常用方法有投票法 voting 和堆叠法 stacking。投票法既允许少数服从多数的硬投票机制，也可以是按分类概率加权平均的软投票决策；堆叠法则将基学习器输出的分类结果作为特征，利用某种机器学习算法重新训练一个分类器用于得出最终识别结果。集成学习策略在应对分类结果的分散性方面展现出有力优势，通过有效聚合不同模型的分类结果，不仅强化了多个分类模型的鲁棒性，还显著提升了其泛化效能^[36]。

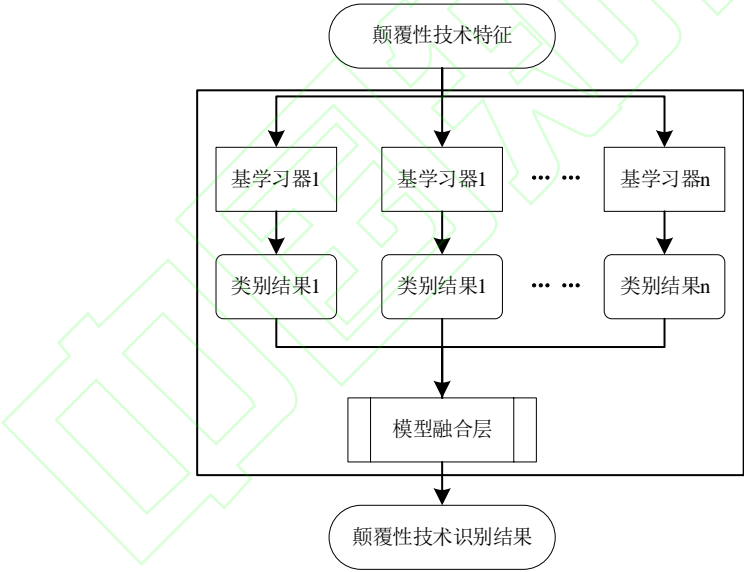


图 1 颠覆性技术识别的集成学习流程

Fig.1 Integrated Learning Process for Disruptive Technology Identification

4 实证分析

量子计算作为一种革命性的计算技术，凭借其独特的海量信息存储与高效处理能力，预示着对现有经典计算机算力界限的突破。本文选取量子计算领域开展实证分析，识别高颠覆性潜力技术专利，发掘重要的技术研究方向，具体实验流程如图 2 所示。

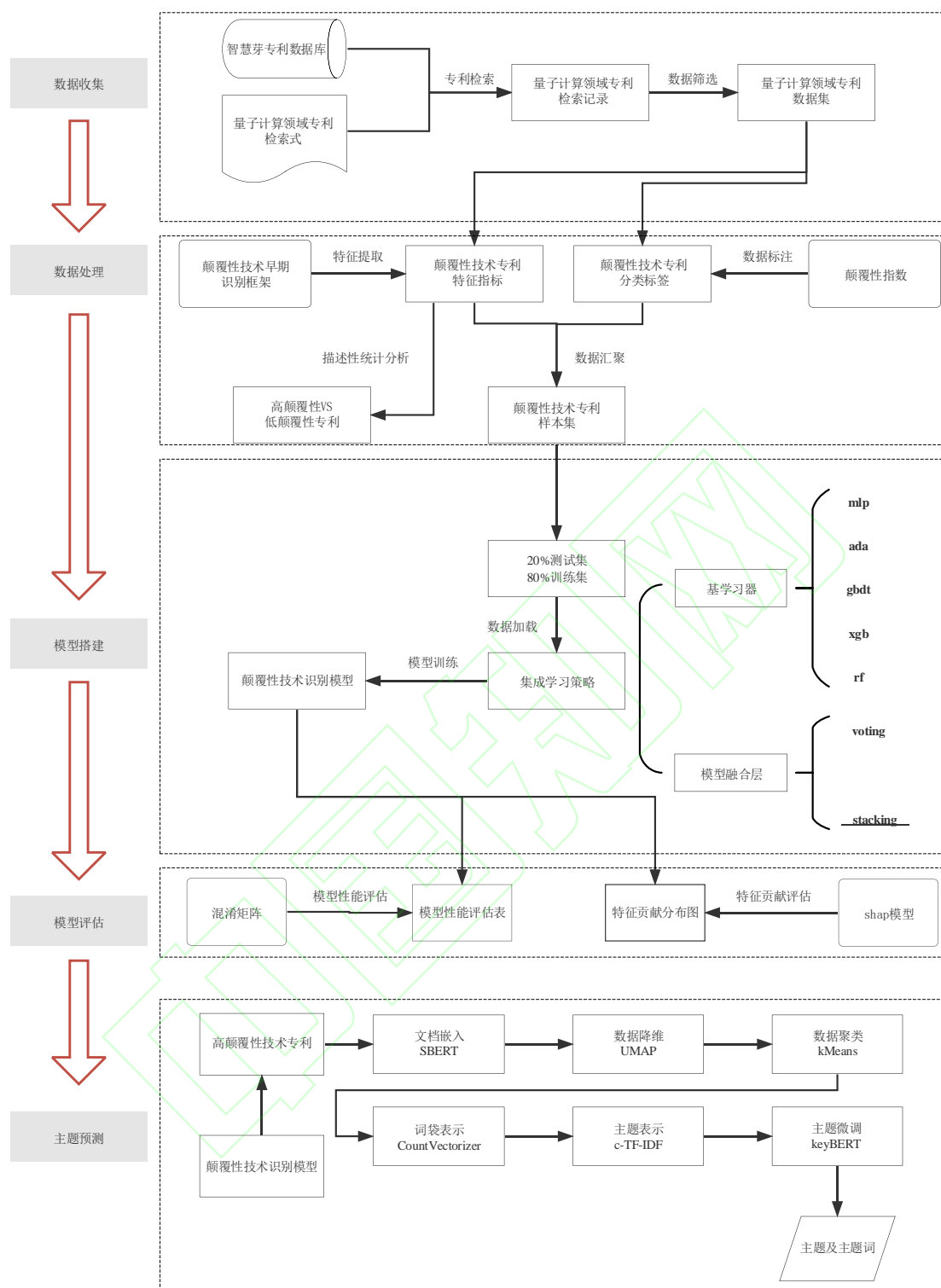


图 2 实验流程
Fig.2 Experimental Process

4.1 数据收集与预处理

本文选取量子计算为技术领域，在智慧芽专利数据库中收集数据，检索截止申请日设置为 2023 年 12 月 31 日，结合国家知识产权局办公室发布的《关键数

字技术专利分类体系（2023）》确定检索式，检索式为：TA:(comput* OR chip OR calculat* OR superconduct* OR ion trap OR integrated circuit OR Silicon semiconductor OR silicon isotope OR processing unit OR silicon-based spin bit OR optical bit OR topology OR annealing OR nuclear magnetic resonance OR cold atom OR diamond color center OR neutral atom OR spin wave OR algorithm OR software OR coding OR programming OR compilation OR measurement OR Hamiltonian OR machine learning OR neural network OR fault-tolerant OR fidelity OR CRSS OR flipping correction OR analogy OR Simulation OR Circuit design) AND TA:(quantum) AND APD:[* TO 20231231]。

共检索到 45215 条专利文本，25202 组简单同族，采用每个专利家族一个代表专利进行专利索引记录的导出，并进行二次检索以收集被引专利数据和引用专利数据。对专利数据进行去重编号处理后，最终收集到量子计算领域的目标专利 25198 条，被引专利 62399 条，引用专利 83558 条，抽取专利的最早申请日、最早公开日、引用文献等相关字段，并进行相应技术识别特征提取。数据的最早申请日字段有所缺失，共计 39 条专利索引数据，占总体数据比例比较小，因专利申请日与专利公开日在时间上临近，拟采用最早公开日进行数据插补。

4.2 数据标注与特征提取

在数据标注环节，本文采用 Funk^[37]基于引文网络提出的颠覆性指数，该指数通过衡量目标专利家族公开后受到专利引用中仅受自身而非其参考专利的影响程度，其有效性在多学科、大数据场景^[38]应用中已得到学术界的一致认可。参考颠覆性指数^[37]给予专利数据固定的分类标签，公式见（5）。将专利技术的颠覆性潜力分为高、低两个等级，通过设定阈值来进行数据样本的标注。

$$Disrupt = \frac{s_i - s_j}{s_i + s_j + s_k} \quad (5)$$

式中，*Disrupt* 代表目标专利的颠覆性指数， s_i 表示只引用目标专利但不引用目标专利的引用专利的专利数量， s_j 表示同时引用目标专利及其引用专利的专利数量， s_k 表示不引用目标专利但引用其引用专利的专利数量。由于专利的颠覆性指数通常在专利公开 5 年后趋于稳定^[39]，因此本文仅对最早公开日距离检索日期超过 5 年的专利计算颠覆性指数。

按照公式（5）计算颠覆性指数，考虑到阈值对模型预测准确率的影响，测

试不同阈值下数据集对逻辑回归模型的训练效果，从而确立最佳阈值。如图 3 所示，模型准确率达到稳定时的阈值为 0.8。

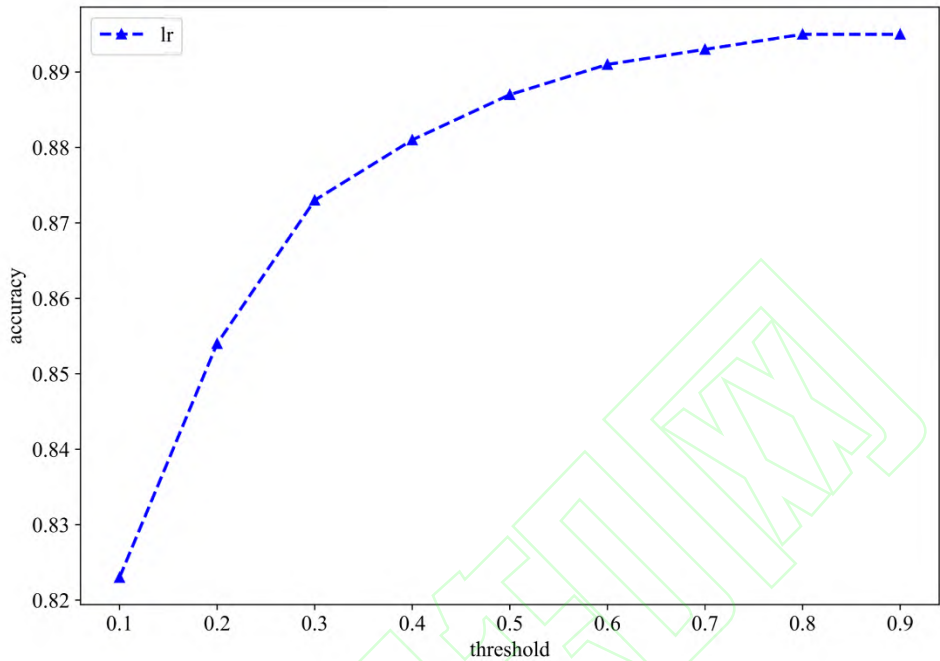


图 3 不同颠覆性阈值大小对模型预测的影响

Fig.3 The Impact of Different Disruptive Threshold on Model Prediction

通过对比高颠覆性技术专利与低颠覆性技术专利，本文探讨技术生命周期各阶段特征值对颠覆性潜力的表征能力，并进一步验证颠覆性技术识别体系框架的适用性和指标计算方式的合理性，得到的专利样本各特征指标描述性统计分析结果如表 2 所示。在颠覆性技术引入阶段，高颠覆性技术专利的异轨性（TN）与突破性（TB）得分较高，而与往期文献的主题相似性（TN）得分较低，表明颠覆性技术在初期设想上就展现出差异化特征。在颠覆性验证阶段，高颠覆性技术专利的篇幅（TK_1）较长、IPC 涵盖范围（TK_2）较窄，表明颠覆性技术研究人员聚焦于某些领域展开深入研究；高颠覆性技术对往期文献的摘引（TS_1、TS_2）次数较少，表明颠覆性技术往往是开拓性的，思路与前人研究鲜有相似；高颠覆性技术的研究团队隶属机构在研究领域的实力雄厚，而核心成员的数量较少，表明高实力的小团队是颠覆性技术研究的主力军，这也与 Wu 等人^[38]的小团队高创新等研究结论相呼应。在颠覆性技术扩张阶段，高颠覆性技术在早期的市场吸引力（MA）较弱，可能是因技术研发周期较长对市场应用场景的还有待深

入挖掘，但凭借顽强的市场竞争力（MC）也能后来居上。在颠覆性技术扩张阶段，高颠覆性技术与未来文献的主题相似度较高（TA），对研究主题产生趋同作用，并吸引更多相关领域人员展开进一步研究工作（TI_1，TI_2）。

表 2 专利样本各特征指标描述性统计分析

Table2 Descriptive Statistical Analysis of Various Characteristic Indicators					
阶段	指标	高颠覆性专利		低颠覆性专利	
		均值	标准差	均值	标准差
颠覆性 技术引入	TT	0.98	0.08	0.85	0.21
	TN	0.21	0.21	0.27	0.21
	TB	0.90	0.49	0.77	0.46
颠覆性 技术验证	TK_1	3.41	2.68	2.73	2.28
	TK_2	12.9	12.26	16.0	17.79
	TS_1	2.24	5.98	4.71	12.61
	TS_2	1.15	3.64	2.18	10.33
	TC_1	61.42	97.47	54.25	95.25
	TC_2	3.43	7.76	5.23	14.17
颠覆性 技术应用	MA	1.57	1.82	1.84	2.92
	MC	9.66	11.51	8.98	10.15
颠覆性 技术扩张	TA	0.32	0.23	0.27	0.22
	TI_1	1.78	3.13	1.02	3.59
	TI_2	3.98	7.59	1.79	6.13

4.3 模型训练及评估

由于数据标注要求专利公开时间不少于 5 年，因此将数据集中最早申请日 2019 年 1 月 1 日的专利数据作为样本集，随机划分 20%的样本作为测试集用于模型评估，剩余样本作为训练集调整模型参数。本文选取在样本集训练中表现较好的 5 个经典机器学习模型作为基学习器，分别是多层感知机 mlp、自适应增强 ada、梯度提升决策树 gbdt、极端梯度提升 xgb 和随机森林 rf。voting 集成学习模型的分​​类投票器采用少数服从多数的硬投票机制，stacking 集成学习模型则选用逻辑回归作为分类投票器，经过训练评估后得到的模型最佳性能对比如表 3 所示。

表 3 颠覆性技术识别模型训练评估结果

Table3 Evaluation Results of Disruptive Technology Recognition Model Training							
模型	准确率	精确率	召回率	F1 值	FPR	TPR	AUC
mlp	0.91	0.66	0.30	0.41	0.02	0.30	0.64
ada	0.92	0.82	0.40	0.54	0.01	0.40	0.69
gbdt	0.93	0.80	0.43	0.56	0.01	0.43	0.71
xgb	0.92	0.76	0.41	0.53	0.02	0.41	0.70
rf	0.93	0.83	0.41	0.55	0.01	0.41	0.70
voting	0.92	0.86	0.40	0.54	0.01	0.40	0.69
stacking	0.94	0.83	0.46	0.59	0.01	0.46	0.72

由于样本集中高颠覆性专利与低颠覆性专利的比例为 1:10，存在严重的类别

不均衡现象，导致模型的召回率普遍偏低，本研究认为准确甄别出颠覆性技术对我国在关键领域中的技术突破方向确定尤为重要，模型应在保证精确率不低的前提下适当平衡召回率，同时 **stacking** 集成学习模型的 F1 值和 AUC 得分排名最高，故采用 **stacking** 集成学习模型作为分类器进行最终训练来识别 2019-2022 这 4 年颠覆性技术专利。

本研究还尝试使用可解释机器学习模型 **SHAP** 来了解集成学习器中各特征的贡献比，结果如图 4 所示。**SHAP** 提供了统一的框架来解释不同模型，利好本文集成策略下由多个模型训练得到的颠覆性技术识别模型，图中的每行代表一个特征，特征贡献由所有样本的平均 **SHAP** 值进行测度，贡献度排名高的特征位于上方，贡献度排名越高其在颠覆性技术识别任务中的重要性越强。分析结果显示，在基于生命周期理论搭建的颠覆性技术早期识别框架中，利用语义特征计算的新颖性 **TN** 和替代性 **TA** 与利用引文特征计算的异轨性 **TT** 对颠覆性技术的识别作用表现突出。

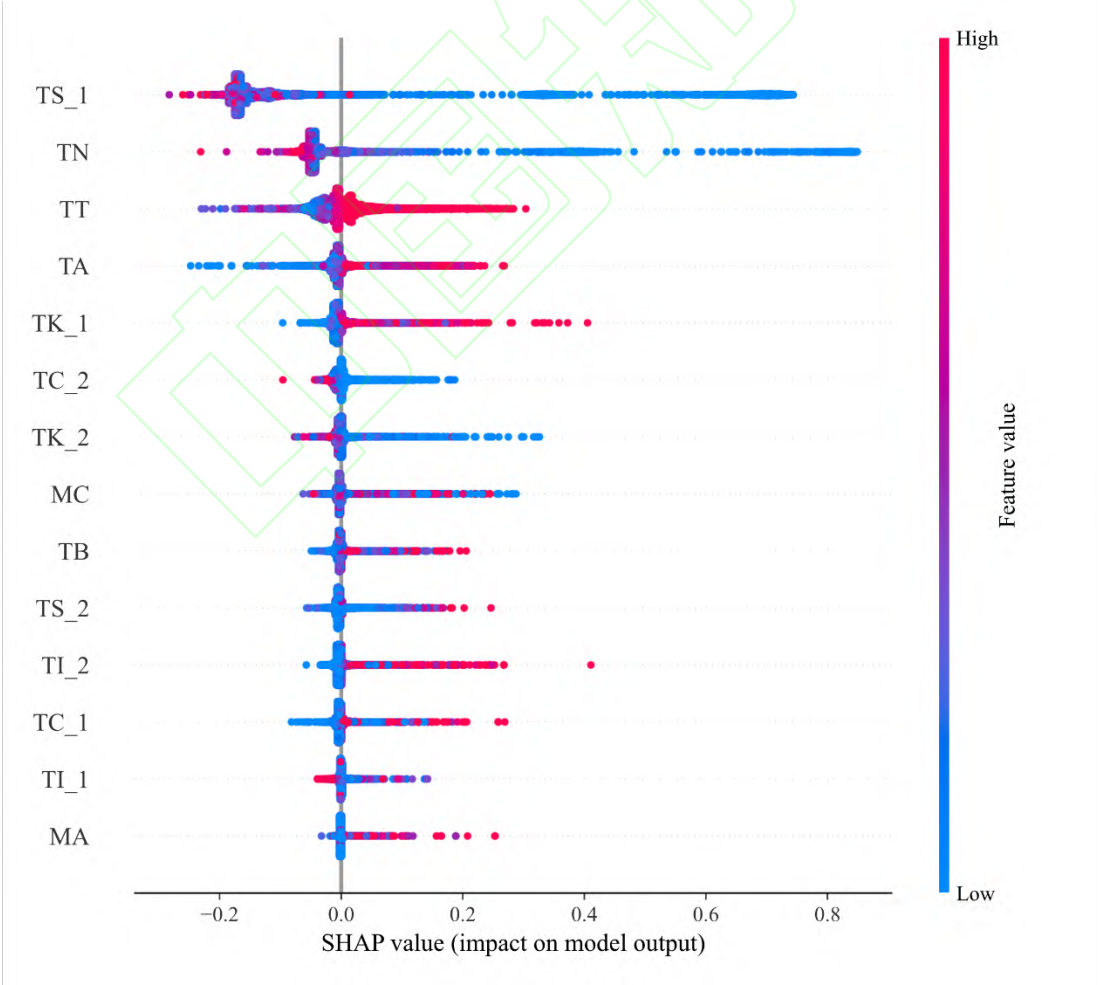


图 4 颠覆性技术早期识别特征的 SHAP 值

Fig.4 SHAP Values for Early Identification Features of Disruptive Technologies

4.4 识别结果分析

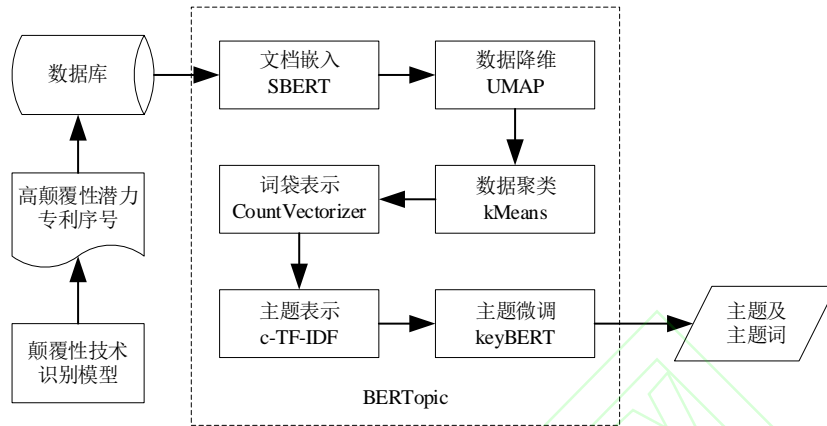


图 5 BERTopic 的工作流程

Fig.5 The Workflow of BERTopic

本文使用 BERTopic 框架进行文本主题建模，流程见图 5。利用颠覆性技术识别模型得到的高颠覆性潜力专利序号，从数据库中读取对应年份的潜在高颠覆性专利文本，借助 SBERT 将自然语言转换成词嵌入向量，进而通过 UMAP 在保留数据关键特征的同时将高维词向量映射到低维空间，然后使用 kMeans 进行聚类，对聚类结果的词汇词频矩阵使用 c-TF-IDF 进行主题表示，并利用 keyBERT 筛选出更具主题表现力的词汇，最后输出主题及主题词。

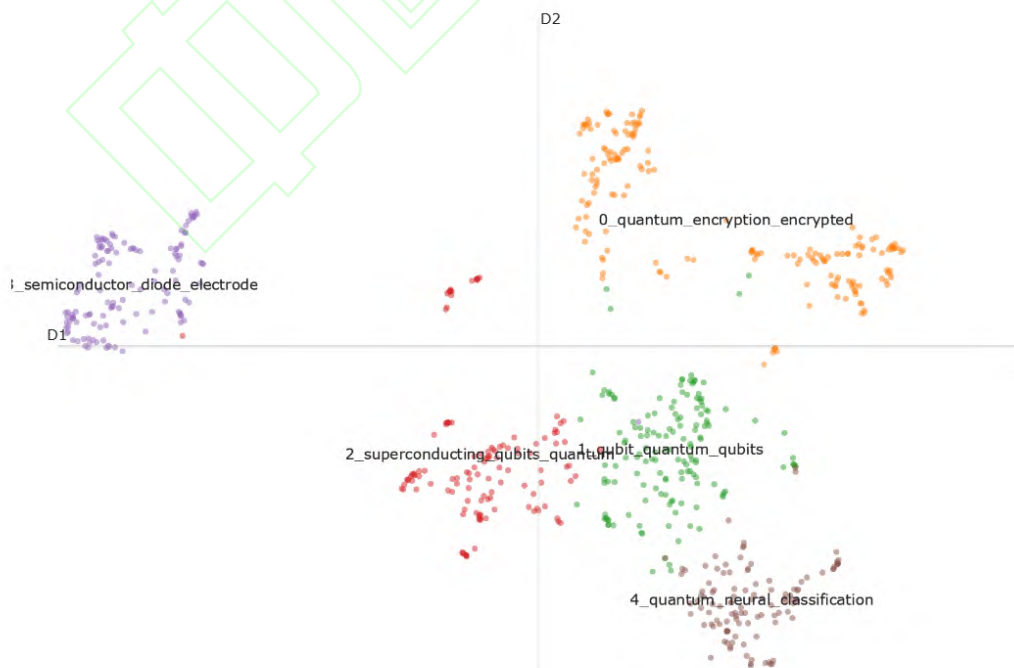


图 6 高颠覆性潜力专利文档聚类结果

Fig.6 Clustering Results of High Disruptive Potential Patent Documents

通过进行多次主题建模，发现当主题数为 5 时，文档聚类效果突出，提取出的主题说服力较强，图 6 展示了文档主题建模结果，表 5 则展示了具体的主题及主题词。为了精细化识别结果，本文进一步分析每个技术主题的关键词，来凝练出技术主题潜在的代表性颠覆性技术小类。

技术主题 0 量子加密技术研究方向下，具备颠覆性潜力的技术小类主要有量子签名、量子随机数生成、量子态分发。量子签名基于量子态的唯一性，确保数据的永久准确性和不可伪造性；量子随机数生成器尝试利用量子力学原理产生真正的随机数，有望实现密钥生成的模块化设计；量子态分发期望通过光纤等介质，将量子态从源端发送到接收端，实现跨区域量子密钥分发的高效性。

技术主题 1 量子处理器研究方向下，具备颠覆性潜力的技术小类主要有量子任务调度、量子程序执行、量子计算优化。在量子任务调度方面，关注如何利用优化算法确定目标节点的任务执行顺序，以提高计算效率和资源利用率；在量子程序执行方面，怎样利用拓扑图进行节点和量子门的映射进行操作系统层面的服务设计有待取得突破性成果；在量子计算优化方面，关注利用哈密顿量进行量子态的计算和旋转，结合公式推导和数学归纳，以求解复杂场景下经典方法难以处理的问题。

技术主题 2 超导量子比特研究方向下，领域内相关研究主要围绕超导约瑟夫森结量子比特设计与优化、微波频率下超导量子比特测试与表征、超导量子比特电路干扰与噪声抑制等方面展开，以实现量子比特的稳定操作和高效耦合。

技术主题 3 半导体技术研究方向下，研究者期望将成熟的半导体技术作为量子计算发展的助推器，量子点技术对于可扩展量子计算具有不可忽视的里程碑意义，氮化镓量子光源芯片可作为量子计算设备的新型能源材料。

技术主题 4 量子神经网络研究方向下，具备颠覆性潜力的技术小类主要有：
①量子神经网络训练与优化，利用量子算法找寻神经网络中的最优参数，加速学习过程，提高复杂场景下模型的预测精度；②量子卷积神经网络与分类，利用量子计算的特性处理高维输入数据，提取更具关键性的特征向量，有望实现高效的图像或信号处理；③量子神经网络故障诊断与预测，通过分析实际场景中的样本数据，实现故障的早期预警和定位功能，在交通、风电、电机等现实系统中有着广泛的应用前景。

表 4 量子计算领域下的颠覆性专利主题
Table4 Disruptive Patent Themes in the Field of Quantum Computing

主题 序号	专利 数量	技术内容	主题词	代表专利
0	175	量子加密技术	quantum, encryption, encrypted, decryption, computing, security, ciphertext, hash, cloud, authentication	167725340: 基于量子密钥加 密的移动数据存储终端
1	153	量子处理器	qubit, quantum, qubits, computing, processor, processing, scheduling, binary, computer, operating	183570741: 一种量子计算任 务计算方法、装置及量子计算 机操作系统
2	124	超导量子比特	superconducting, qubits, quantum, qubit, superconductive, circuits, interconnection, coupling, junction, resonator	182526795: 超导电路、量子 比特装置及其制备方法
3	120	半导体技术	semiconductor, diode, electrode, backlight, led, silicon, substrate, emitting, epitaxial, infrared	184560456: 一种倒装高压发 光二极管芯片
4	120	量子神经网络	quantum, neural, classification, computing, algorithm, prediction, learning, models, layer, fuzzy	187172295: 一种基于量子模 糊神经网络模型的大数据处 理方法及系统

5 结语

本文从技术生命周期理论入手搭建颠覆性技术早期识别框架,提出了一种基于集成学习模型来识别潜在颠覆性技术的方法,学术贡献包括:

一是丰富和完善了颠覆性技术识别理论,基于技术生命周期理论提出了技术颠覆性展现的主要历程,包括颠覆性技术引入、颠覆性技术验证、颠覆性技术应用和颠覆性技术扩张,并阐释了用于识别技术潜在颠覆性的具体维度及计算指标。

二是深化和拓展了颠覆性技术识别的指标提取方法,综合考虑技术专利数据中含有的外部特征、引文特征与语义特征等多层次特征,使用科学计量、引文分析与文本向量表示等多种方法提取指标特征,并在实证分析中验证其有效性。

三是利用集成学习思路搭建预测模型,聚合了不同经典机器学习算法模型在单一数据集上的差异化学习能力,颠覆性技术识别模型的性能得到改善,能够有效识别出关键技术领域的颠覆性研究方向。

本研究也存在一定的局限性,可以在未来进一步开展研究。一是实证研究仅选用量子计算领域为例进行分析,研究样本有限。二是数据来源仅为专利数据库,未来可以开展开源情报收集工作,融合多源数据来进一步细化识别结果。

参考文献:

- [1] 中国科学院颠覆性技术创新研究组. 颠覆性技术创新研究: 生命科学领域[M]. 北京: 科学出版社, 2020. (Subversive Technology Innovation Research Group, Chinese Academy of Sciences. Research on Disruptive Technological Innovation: Life Sciences [M]. Beijing: Science Press, 2020.)
- [2] 罗素平, 寇翠翠, 金金, 等. 基于离群专利的颠覆性技术预测——以中药专利为例[J]. 情报理论与实践, 2019, 42(07): 165-170. (Luo Suping, Kou Cuicui, Jin Jin, et al. Disruptive Technology Prediction Based on Outlier Patents: Traditional Chinese Medicine Patents as an Example [J]. Information Studies: Theory & Application, 2019, 42(07): 165-170.)
- [3] Bhatt P C, Lai K K, Drave V A, et al. Patent Analysis Based Technology Innovation Assessment with the Lens of Disruptive Innovation Theory: A Case of Blockchain Technological Trajectories[J]. Technological Forecasting and Social Change, 2023(Nov.):196.
- [4] Zhang JW, Wu KB, Done Y. Topic Identification Index of Disruptive Technology Based on Patent Characteristics[C]//18th International Conference on Scientometrics and Informetrics(ISSI), KU Leuven, 2021: 1295-1306.
- [5] 刘俊婉, 庞博, 徐硕. 基于弱信号的颠覆性技术早期识别研究[J]. 情报学报, 2023, 42 (12): 1395-1411. (Liu Junwan, Pang Bo, Xu Shuo. Early Identification of Disruptive Technology Using Weak Signals [J]. Journal of the China Society for Scientific and Technical Information, 2023, 42 (12): 1395-1411.)
- [6] 梁镇涛, 毛进, 李纲. 融合“科学-技术”知识关联的高颠覆性专利预测方法[J]. 情报学报, 2023, 42(06): 649-662.
- [7] Bower J L, Christensen C M. Disruptive Technologies: Catching the Wave[J]. Harvard Business Review, 1995, 73(1): 43-53.
- [8] Sommarberg M, Makinen S J. A Method for Anticipating the Disruptive Nature of Digitalization in the Machine-building Industry[J]. Technological Forecasting and Social Change, 2019, 146(Sep.):808-819.
- [9] Jia W, Xie Y, Zhao Y, et al. Research on Disruptive Technology Recognition of China's Electronic Information and Communication Industry based on Patent Influence[J]. Journal of Global Information Management, 2021, 29(2): 148-165.
- [10] 邢晓昭, 任亮, 雷孝平, 等. 基于专利主题演化的颠覆性技术识别研究——以类脑智能领域为例[J]. 情报科学, 2023, 41(03): 81-88. (Xing Xiaozhao, Ren Liang, Lei Xiaoping, et al. The Identification of Disruptive Technology Based on Patent Theme Evolution: Taking the Field of Brain-Inspired Intelligence as an Example [J]. Information Science, 2023, 41(03): 81-88.)
- [11] 苏敬勤, 刘建华, 王智琦, 等. 颠覆性技术的演化轨迹及早期识别——以智能手机等技术为例[J]. 科研管理, 2016, 37(3) : 13-20. (Su Jingqin, Liu Jianhua, Wang Zhiqi, et al. The Evolution Trajectory and Early Identification of Disruptive Technology by Taking Smartphones and Other Technologies as an Example[J]. Science Research Management 2016, 37(3) : 13-20.)
- [12] 陈育新, 卢俊, 韩毅. 基于专利文献的颠覆性技术识别研究——以人工智能为例[J]. 情报学报, 2022, 41(11): 1124-1133. (Chen Yuxin, Lu Jun, Han Yi. Topic Prediction for Disruptive Technologies Based on Patent Literature——A Case Study of Artificial Intelligence Patents[J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(11): 1124-1133.)
- [13] Liu XW, Wang XZ, Lyu LC, et al. Identifying Disruptive Technologies by Integrating Multi- source Data[J]. Scientometrics, 2022, 127(9): 53-54.
- [14] 李乾瑞, 郭俊芳, 黄颖, 等. 基于专利计量的颠覆性技术识别方法研究[J]. 科学学研究, 2021, 39(07): 1166-1175. (Li Qianrui, Guo Junfang, Huang Ying, et al. Research on the Method of Disruptive Technology Identification Based on Patent Bibliometrics [J]. Studies in Science of Science, 2021, 39(07): 1166-1175.)
- [15] Xiang MH, Fu D, Lv K. Identifying and Predicting Trends of Disruptive Technologies: An Empirical Study Based on Text Mining and Time Series Forecasting[J]. Sustainability, 2023, 15(6): 53-54.
- [16] Pan D, Ren XJ, Zhang L, et al. A Disruptive Technology Identification Method for New Power Systems Based on Patent Evolution Analysis[J]. Electronics, 2023, 12(9): 53-54.
- [17] Li X, Xie Q, Huang L. Identifying the Development Trends of Emerging Technologies Using Patent Analysis and Web News Data Mining: The Case of Perovskite Solar Cell Technology[J]. IEEE Transactions on Engineering Management, 2022, 69(6): 2603-2618.
- [18] Qiao Y, Wang X, Huang Y, et al. Tech Mining Approach for Identifying Potentially Disruptive Technologies:

From the Perspective of Technological Alternatives[J]. IEEE Transactions on Engineering Management, 2024, 71: 5921-5938.

[19] Choi Y, Park S, Lee S. Identifying Emerging Technologies to Envision a Future Innovation Ecosystem: A Machine Learning Approach to Patent Data[J]. Scientometrics, 2021, 126(7): 1-46.

[20] Chen X, Han T. Disruptive Technology Forecasting Based on Gartner Hype Cycle[J]. IEEE, 2019.

[21] He GX, Lin CX, Ren JY, et al. Predicting the Emergence of Disruptive Technologies by Comparing with References via Soft Prompt-aware Shared BERT[J]. Journal of Informetrics, 2024, 18(4).

[22] Michael G Harvey. Application of Technology Life Cycles to Technology Transfers[J]. Journal of Business Strategy, 1984, 5(2):51-58.

[23] 苏鹏, 苏成, 潘云涛. 基于历史案例的颠覆性技术特征分析[J]. 中国科技论坛, 2019, (08): 1-9. (Su Peng, Su Cheng, Pan Yuntao. Research on Features of Disruptive Technology Based on Historical Cases[J]. Forum on Science and Technology in China, 2019, (08): 1-9.)

[24] 刘云, 桂秉修, 王小黎. 基于引文网络的颠覆性技术发展路径研究[J]. 科研管理, 2022, 43 (11): 32-44. (Liu Yun, Gui Bingxiu, Wang Xiaoli. Research on the Disruptive Technology Development Path Based on Citation Network Analysis [J]. Science Research Management, 2022, 43 (11): 32-44.)

[25] 王莉晓, 陈伟, 邱含琪. 基于机器学习的颠覆性技术弱信号识别模型研究[J]. 数据分析与知识发现, 2024, 8(8-9): 63-75. (Wang Lixiao, Chenwei, Qiu Hanqi. Weak Signal Identification Model for Disruptive Technologies Based on Machine Learning[J]. Data Analysis and Knowledge Discovery, 2024, 8(8-9): 63-75.)

[26] 黄鲁成, 蒋林杉, 吴菲菲. 萌芽期颠覆性技术识别研究[J]. 科技进步与对策, 2019, 36(01): 10-17. (Huang Lucheng, Jiang Linshan, Wu Feifei. The Identification of Disruptive Technology on Emerging Stage[J]. Science & Technology Progress and Policy, 2019, 36(01): 10-17.)

[27] 冉从敬, 田文芳, 贾志轩. 技术生命周期视角下专利技术主题演化分析方法研究——以视频图像处理技术领域为例[J]. 情报理论与实践, 2024, 47(09): 124-133. (Ran Congjing, Tian Wenfang, Jia Zhixuan. Research on the Analysis Method of Patent Technology Topic Evolution in the Perspective of Technology Life Cycle: Taking Video Image Processing Technology as an Example [J]. Information Studies: Theory & Application, 2024, 47(09): 124-133.)

[28] Kong J, Zhang J, Deng S, et al. Knowledge Convergence of Science and Technology in Patent Inventions[J]. Journal of Informetrics, 2024, 17(3).

[29] Xu S, Hao L Y, Yang G C, et al. A topic models based framework for detecting and forecasting emerging technologies[J]. Technological Forecasting and Social Change, 2021, 162: 120366.

[30] 邢晓昭, 陈亮, 梁琴琴. 融合多元方法的颠覆性技术识别——以类脑智能领域为例[J]. 科技管理研究, 2023, 43 (15): 191-199. (Xing Xiaozhao, Chen Liang, Liang Qinqin. Research on Disruptive Technology Identification by Integrating Multiple Methods: Taking the Field of Brain-inspired Intelligence as an Example[J]. Science and Technology Management Research, 2023, 43 (15): 191-199.)

[31] 王超, 马铭, 王海燕, 等. 生命周期视角下颠覆性技术的扩散特征研究[J]. 情报学报, 2022, 41(08): 845-859. (Wang Chao, Ma Ming, Wang Haiyan, et al. Diffusion Features of Disruptive Technology from the Life Cycle Perspective[J]. Journal of the China Society for Scientific and Technical Information, 2022, 41(08): 845-859.)

[32] Jia WF, Wang S, Xie YP, et al. Disruptive Technology Identification of Intelligent Logistics Robots in AIoT Industry: Based on Attributes and Functions Analysis[J]. Systems Research and Behavioral Science, 2022, 39(3): 557-568

[33] Capponi G, Martnell A, Nuvolari A. Breakthrough Innovations and Where to Find Them[J]. Research Policy, 2022, 51(1): 104376.

[34] Fleming L. Breakthroughs and the "Long Tail" of Innovation[J]. MIT Sloan Management Review, 2007, 49 (1): 69-74+93.

[35] 李奉书, 徐莹婕, 杜鹏程, 等. 数字经济时代下联盟管理能力对企业颠覆性技术创新的影响——知识流动的中介作用与知识重构能力的调节作用[J]. 科技进步与对策, 2022, 39 (04): 80-90. (Li Fengshu, Xu Yingjie, Du Pengcheng, et al. The Impact of Alliance Management Capabilities on Enterprise Disruptive Technological Innovation in the Digital Economy Era[J]. Science & Technology Progress and Policy, 2022, 39 (04): 80-90.)

[36] Xu X, Bai Y, Liu Y, et al. MM-UrbFAC: Urban Functional Area Classification Model Based on Multimodal Machine Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 8488-8497.

[37] Funk R J, Owen-Smith J. A Dynamic Network Measure of Technological Change[J]. Management Science, 2017, 63(3):791-817.

[38] Wu L, Wang D, Evans J A. Large Teams Develop and Small Teams Disrupt Science and Technology[J]. Nature, 2019, 566(7744): 378-382.

[39] Jaffe A B, Trajtenberg M. Patents, Citations, and Innovations: A Window on the Knowledge Economy[M]. Cambridge: MIT Press, 2022.

通讯作者（Corresponding author）：吕璐成，ORCID：0000-0002-2318-1073，
E-mail:lvlc@mail.las.ac.cn。

基金项目（不超过三个, 中英文对应）：本文系教育部人文社会科学重点研究基地重大项目“面向国家科技自立自强的颠覆性技术识别方法与创新机制研究”（项目编号：22JJD870004）；国家自然科学基金面上项目“‘搜索即学习’视角下的用户学习状态感知与智能响应研究”（项目编号：72274146）；中央高校基本科研业务费专项资金项目“人工智能时代大国跨境数据战略竞争与中国应对研究”研究成果之一。

The work is supported by Education Ministry's Key Research Base for Humanities and Social Sciences Major Projects(Grant No.22JJD870004), the General Program of National Natural Science Foundation of China (Grant No.72274146), the Fundamental Research Funds for the Central Universities.

作者贡献声明：

赵一鸣：确定研究选题，论文指导与撰写；

刘顺生：数据处理与分析，论文撰写；

吕璐成：论文指导与撰写。

Early Identification of Disruptive Technology Using Ensemble Learning: Taking the Field of Quantum Computing as an Example

Zhao Yiming^{1,2,3}, Liu Shunsheng^{2,4}, Lü Lucheng^{5,6}

(1. Center for Studies of Information Resources, Wuhan University, Wuhan, 430072, China

2. School of Information Management, Wuhan University, Wuhan, 430072, China

3. Big Data Institute, Wuhan University, Wuhan, 430072, China

4. National Demonstration Center for Experimental Library and Information Science Education, Wuhan University, Wuhan, 430072, China

5. National Science Library, Chinese Academy of Sciences, Beijing, 100190, China

6. Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academic of Sciences, Beijing, 100190, China)

Abstract: [Objective] By analyzing patent data in key technological fields, early identification of technology patents with high disruptive potential can be achieved. [Methods]

Building an early identification index system for disruptive technologies based on the technology lifecycle theory, using patent data in the field of quantum computing in the Smart Sprout patent database as the research object, and constructing an ensemble learning model to identify technology patents with high disruptive potential in this field at an early stage. **[Results]** Through the BERTopic topic modeling framework, five disruptive research directions were identified: quantum encryption technology, quantum processors, superconducting quantum bits, semiconductor technology, and quantum neural networks. The effectiveness and feasibility of the proposed method were verified. **[Limitations]** Empirical analysis only focuses on the field of quantum computing and fails to comprehensively cover different key technological areas; The framework construction and indicator extraction rely solely on patent data, which can expand the types of supporting data sources. **[Conclusions]** This study helps to identify high disruptive potential technology patents early on, and then analyze the main disruptive research directions, providing a basis for the formulation and implementation of major national science and technology strategies.

Keywords: Disruptive Technology; Ensemble Learning; Technology Lifecycle; Quantum Computing