

情报理论与实践
Information Studies:Theory & Application
ISSN 1000-7490,CN 11-1762/G3

《情报理论与实践》网络首发论文

题目：专利颠覆性潜力识别的双维融合框架：以新型储能技术为例
作者：赵志洋，杨杰，吴娜，王昊，邓三鸿
网络首发日期：2026-07-08
引用格式：赵志洋，杨杰，吴娜，王昊，邓三鸿．专利颠覆性潜力识别的双维融合框架：以新型储能技术为例[J/OL]．情报理论与实践.
https://link.cnki.net/urlid/11.1762.G3.20260708.1127.004



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。

网络首发时间：2026-07-08 14:52:23　网络首发地址：https://link.cnki.net/urlid/11.1762.G3.20260708.1127.004

●赵志洋[1,2]，杨　杰[1,2]，吴　娜[1,2]，王　昊[1,2]，邓三鸿[1,2]
（1．数据工程与知识服务重点实验室（南京大学），江苏　南京　210023；2．南京大学信息管理学院，江苏　南京　210023）

专利颠覆性潜力识别的双维融合框架：以新型储能技术为例*

摘　要：[目的/意义]新型储能是我国双碳战略落地与能源安全保障的核心赛道，如何从该领域海量专利中早期识别具有颠覆性潜力的技术，是技术情报领域的难题。[方法/过程]本文构建了融合文本突破性（KI 指数）与引文颠覆性（CD 指数）的双维识别框架，系统比较熵权法、CRITIC 法与主成分分析法，基于数据结构检验分析确立融合方案，构建专利综合颠覆性指数（PCDI），并以新型储能领域的 21671 件专利为样本进行前瞻筛选与历史回溯验证。[结果/结论]两指标近乎正交且方差接近时，三种赋权方法排序相似，主成分分析自然收敛于等权解，据此形成“检验数据结构—内生选择融合方案”的方法论原则。PCDI 的 AUC 值为 0.673，Precision@5%达 0.356，较 CD 指数提升 23.6%，较 KI 指数提升 53.4%，历史重要专利平均排序分位数低至 0.0286，优势在跨领域（集成电路、复合高分子材料）验证中保持稳健。[创新/价值]制定基于数据结构检验的双维指标融合规程，并据此构建了专利综合颠覆性指标（PCDI），为优化专利颠覆性潜力的定量测度方法提供了经验证据与决策参考。

关键词：颠覆性技术识别；专利分析；双维融合框架；新型储能技术；跨领域

Two-dimensional Fusion Framework for Identifying Patent Disruptive Potential: A Case Study of New Energy Storage Technology

Zhao Zhiyang[1,2]，Yang jie[1,2]，Wu Na[1,2]，Wang Hao[1,2]，Deng Sanhong[1,2]
（1. Key Laboratory of Data Engineering and Knowledge Services in Provincial Universities，Jiangsu Nanjing 210023; 2. School of Information Management，Nanjing University，Jiangsu Nanjing 210023）

Abstract: [Purpose/significance] New energy storage is a core area for implementing China’s dual-carbon strategy and ensuring energy security. Identifying potentially disruptive technologies at an early stage from the vast number of patents in this field is a key challenge in technology intelligence. [Method/process] This study constructs a two-dimensional identification framework that integrates textual breakthrough (KI Index) and citation disruptiveness (CD Index). It examines the applicable boundaries of the entropy weight method, the CRITIC method, and principal component analysis (PCA). An integrated fusion scheme is endogenously determined based on data structure testing, and a Patent Comprehensive Disruption Index (PCDI) is established. Using 21,671 patents in the emerging energy storage field as the research sample, prospective screening and historical retrospective verification are further conducted. [Result/conclusion] When the two indicators are nearly orthogonal and have similar variances, the three weighting methods yield consistent ranking results, and principal component analysis naturally converges to an equal-weight solution. On this basis, a methodological principle of “testing data structure: endogenously selecting the fusion scheme” is formulated. The PCDI achieves an AUC value of 0.673 and a Precision@5% of 0.356, representing a 23.6% improvement over the CD Index and a 53.4% improvement over the KI Index. The average ranking quantile of historically important patents is as low as 0.0286, and these performance advantages remain robust in cross-domain validation covering integrated circuits and composite polymer materials. [Innovation/Value] This paper proposes a two-dimensional index fusion specification based on data structure testing and constructs the Patent Comprehensive Disruption Index (PCDI). It provides empirical evidence and decision-making references for optimizing the quantitative measurement of patent disruptive potential.

Keywords: disruptive technology identification; patent analysis; dual-dimensional integration framework; new energy storage; cross-domain verification

0 引言

颠覆性技术通过开辟新的技术路径，突破现有范式的边界，重塑产业发展格局，对经济结构变革、国家科技竞争力乃至社会发展模式均产生深远影响[1]。新型储能是双碳战略的核

* 本文为国家社会科学基金重点项目“基于领域集体智慧挖掘的颠覆性范式变革预测研究”的成果之一，项目编号：25ATQ008。

心赛道，从海量专利中早期识别颠覆性技术是技术情报领域的关键难题。专利颠覆性的概念本身具有多维性，既涉及技术内容对已有范式的背离，也涉及该技术对后续创新轨迹的实际重构[2]。如何在方法论层面对其进行有效测度，仍有待进一步解决[3]。

围绕专利颠覆性识别，现有研究主要沿着引文网络与文本语义两条路径展开[4]。两类方法虽各有实证效度支持，但存在三方面局限。其一，单一维度测度存在难以覆盖的盲区。引文路径依赖后验的“产业跟随”信号，易遗漏内容已突破但引用滞后的早期发明[5]；文本路径则面临“语义跃迁未必等同于技术替代”的困境[6]。本文中基于新型储能领域的实证表明，KI 与 CD 指数近乎正交（Pearson 系数仅 0.017），至少在新型储能领域证实两者可以看作独立的指标，单一指标极易导致对颠覆性潜力的误判。其二，双维融合的研究虽有推进，但尚未形成具有统计基础的方法论框架[7]。现有研究多采用单一客观赋权或预设先验权重，鲜有针对熵权法、CRITIC 法、主成分分析等不同方案在颠覆性识别场景下的适用边界进行系统比较，融合过程的客观性与可复现性有待提升[8]。其三，现有融合框架多植根于技术范式稳定的成熟领域，其在引用网络稀疏、语义空间快速漂移的新兴技术领域中的适用性与有效性，尚缺乏充分的实证检验。

针对上述缺口，本文提出“检验数据结构—内生选择融合方案”的双维识别框架，并行考虑熵权法、CRITIC 法与主成分分析法，基于指标的相关性与方差特征确定最优融合权重，构建专利综合颠覆性指数（Patent Comprehensive Disruption Index，PCDI）。以新型储能领域 21671 件专利为样本，通过前瞻预测与历史回溯交叉检验 PCDI 的有效性，并在集成电路、复合高分子材料领域开展跨领域验证。本研究的核心贡献在于提供了一套可迁移的内生融合方法论，而非固定的等权公式。

1 相关研究

1.1 专利颠覆性识别方法

专利作为技术创新最核心的载体，其颠覆性水平是衡量技术变革潜力的一个核心标尺。现有针对专利颠覆性的识别研究，可大致分为定性识别与定量识别两类[9-10]。

定性识别是该领域早期的主流研究范式[11]，主要依托同行评议与专家研判来甄别颠覆性技术[12]。此类基于同行评议的识别方法权威性突出，判别结果的行业认可度较高，但其局限性也较为明显：一是主观性强，不同背景评审者的判断标准不同；二是难以适配海量专利批量筛选，效率低且较难早期识别，不能支撑前瞻布局需求。

为弥补定性方法的不足，学界逐步发展出定量识别路径，主要围绕引文结构与文本内容两个维度展开[4]。引文路径通过引用网络结构变化量化技术轨迹影响，以 CD 指数为代表并衍生出多种优化变体[13]；文本路径从专利文本提取创新特征，弥补引文指标的时间滞后性[6]。在此基础上，部分学者尝试融合引文结构与文本内容等多类特征，构建多维度的专利颠覆性识别模型，但现有的融合研究仍存在改进空间[14]。

1.2 专利引文颠覆性测度研究

专利引用关系是技术知识流动与创新轨迹演化的直接表征，涵盖直接引用、共被引与引文耦合三类核心关联形式[3]。

基于引用的颠覆性测度核心是通过引文网络结构变化，量化目标专利对后续创新轨迹的重构效应[15]。CD 指数是该领域最具代表性的指标，通过三类施引专利的数量占比区分技术的颠覆与巩固作用[16]，突破了传统被引次数仅能衡量影响力的局限，成为核心测度工具[6]。后续学者从多维度提出优化变体，完善了测度体系[17]。但单一引文指标无法捕捉技术内容突破，易误判“高影响力低突破性”专利，存在固有测度盲区[18]。

1.3 专利文本突破性测度研究

专利文本突破性测度核心是量化专利与既有知识的语义差异，捕捉技术对先前范式的背离与未来创新的引领能力[19]，专利中的文本为其提供了天然数据基础[8]。

现有专利文本突破性测度研究呈现出从表层形式到深层语义的演进脉络[6]。早期研究多基于词袋模型统计新颖术语占比，难以捕捉技术术语间的语义关联。SAO 三元组方法虽实现了从“词汇层”到“功能层”的升级，但对文本长度要求高，在专利摘要短文本场景下区分度不足[7]。

词嵌入技术推动基于语义相似度的测度成为主流，其中 KI 指数已得到广泛验证，在长时段技术创新测度研究中表现出较好的效度[19]。然而现有研究对文本指标在颠覆性测度体系中的权重定位及其与引文指标的互补关系讨论不足，由此形成本研究的切入点[4]。

1.4 多指标融合方法研究

为弥补单一维度指标的缺陷，学界逐步探索多指标融合的专利颠覆性综合测度方法，客观赋权法是实现多指标融合的核心路径[20]。现有研究中应用较广泛的客观赋权方法主要包括熵权法、CRITIC 法、主成分分析法（PCA）三类[21]。然而，现有专利颠覆性测度领域的融合研究，多直接采用单一赋权方法，缺乏对不同赋权方法在双维融合场景中适用性的系统比较与论证[7]。权重的确定多为人为预设，缺乏内生于数据的统计依据，且针对新型储能这类技术路线多元、范式尚未收敛的新兴领域，融合框架的适配性与有效性仍缺乏深入的实证检验。基于此，本研究系统比较了熵权法、CRITIC 法与主成分分析法三种客观赋权方法的适用性，构建融合文本突破性与引文颠覆性的专利综合衡量框架，并以新型储能技术为实证场景，开展全面的有效性与稳健性检验。

2 研究设计

2.1 研究框架

本文旨在构建一套内生于数据特征的专利颠覆性潜力双维识别方法流程，以实现对目标技术领域颠覆性专利的早期、精准筛选。整体研究框架如图 1 所示，核心研究逻辑分为 4 个递进阶段：

第一，双维信号提取与预处理阶段。分别从文本内容与引文网络两个维度，采用 KI 突破性指数与 CD 颠覆性指数，量化专利的文本突破性水平（TextScore）与引文颠覆性水平（CiteScore），并通过 Z-score 标准化使两指标具备对等的初始尺度。

第二，数据结构检验与融合方案内生选择阶段。并行比较熵权法、CRITIC 法与主成分分析法（PCA）三种客观赋权方法，依据“条件检验优先于公式套用”的核心原则，基于两指标的相关性结构与方差特征内生确立融合方案，构建专利综合颠覆性指数（PCDI）。

第三，多维度效度验证阶段。①前瞻预测效度验证：以专利授权后 5 年被引次数为外部效标，从 Spearman 相关性分析、分组梯度检验、AUC 与 Precision@5%评估三个层次交叉检验 PCDI 的识别有效性。②历史重要专利回溯验证：依据获奖[8]、参与行业标准指定、学科奠基与产业里程碑[22]4 类外部标准，遴选公认的颠覆性专利，反向检验各指标能否将其系统性地置于分布顶端。

第四，跨领域稳健性检验阶段。将整套分析框架迁移至集成电路与复合高分子材料两个技术范式差异显著的领域，验证文本突破性与引文颠覆性的正交结构、PCDI 的效度增益模式，以及互补信号增强效应的跨领域一致性，初步确保核心结论对领域选择不敏感。

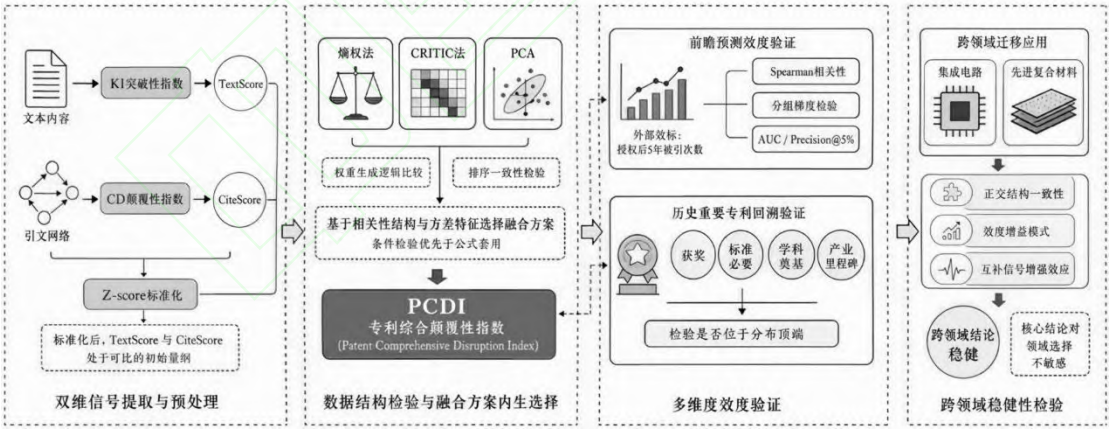


图 1 研究框架示意

Fig.1 Research framework

2.2 核心变量测度

2.2.1 文本突破性测度：KI 突破性指数 本研究采用 Kelly 等提出的 KI 突破性指数[19]（Breakthrough Index）作为文本突破性维度的测度指标。研究的主要目标是构建一套可迁移、可复现的大规模专利颠覆性潜力初筛框架。基于标题与摘要的 KI 指数在计算效率、数据可获得性及跨领域可比性方面具有显著优势，能够满足产业界与科研机构对海量专利进行快速初筛的实际需求。KI 指数核心逻辑为：突破性专利需同时满足“背离既往技术范式”与 “引领后续技术方向”，通过“瞻前顾后”的语义比较实现量化。其采用词嵌入技术将专利标题与摘要转换为高维语义向量，定义目标专利的过去 5 年参照集与未来 5 年参照集，计算目标

专利与两集合的平均余弦相似度，KI 指数为二者之比的自然对数[19]：

$$\mathrm{KI}=\ln\left(\frac{\mathrm{sim_forward}}{\mathrm{sim_backward}}\right) \tag{1}$$

式中，KI 指数取值含义明确：KI＜0 为衰退/巩固性技术；KI≈0 为技术正常延续；KI＞0 为突破性技术且值越大突破性越强。本研究采用预计算的 5 年窗口 KI 指标（字段名 lqsim05），记为 TextScore。

2.2.2　引文颠覆性测度：CD 颠覆性指数　本研究采用 Funk 和 Owen-Smith 提出的 CD 指数（Consolidation/Disruption Index）[3]作为引文颠覆性维度的测度指标。该指标的核心逻辑是，颠覆性专利会切断后续创新与先前知识基础的联系，使后续专利无须依赖其参考文献。统计目标专利授权后 5 年的施引专利，按引用行为分为三类：*n_i*（仅引目标专利，代表完全依赖其新知识）、*n_j*（同时引目标专利及参考文献，代表补充既有知识）、*n_k*（仅引参考文献，代表绕过目标专利）。

基于上述三类施引专利计数，CD 指数的计算公式为[3]：

$$\mathrm{CD}=\frac{n_i-n_j}{n_i+n_j+n_k} \tag{2}$$

式中，CD 指数取值范围[－1，1]：CD＞0 为颠覆性技术（值越大颠覆性越强）；CD≈0 为技术正常延续；CD＜0 为被削弱或绕开。本研究从 PatentsView 专利数据库中提取专利 5 年窗口的 *n_i*5、*n_j*5、*n_k*5 变量，分母加极小常数避免除零，结果记为 CiteScore。

2.3 验证策略与评估指标

PCDI 的指标效度须接受外部效标的交叉验证。为系统检验 PCDI 的效度，本文构建了“前瞻预测+历史回溯”的双向验证框架，并在新型储能、集成电路、复合高分子材料三个技术范式差异显著的领域同步执行，以构成相互印证的效度证据链。

在前瞻预测部分，外部效标选取专利授权后 5 年被引次数（cit5）。需要说明的是，本研究认为高被引不等同于颠覆，但颠覆性技术通常具备高影响力，因此被引次数是极具实践价值的检验标准。本部分的研究不直接判定专利颠覆性，而是检验 PCDI 筛选未来高影响力专利的效率。以 cit5 前 10%为高被引正例，采用 Spearman 秩相关系数、AUC、Precision@5%、分组首尾平均被引倍差为评价指标，并据此设计验证策略。

在历史回溯部分，本研究从相反的时间方向切入：依据获奖、标准必要、学科奠基与产业里程碑 4 类外部公信标准，在三个领域各遴选 10 件左右历史重要专利。将其 KI 指数、CD 指数与 PCDI 得分分别映射为全样本排序分位数，检验融合指标能否将已获公认的颠覆性标杆系统性置于分布顶端，并特别关注它能否矫正因单指标盲区而被低估的“内容先行、引用待发”型早期突破性发明。

3 实验过程与结果分析

3.1 实验准备

3.1.1 数据获取与样本筛选

1）数据来源。本研究全部专利数据来源于 PatentsView 专利数据库 2024 年 3 月快照。根据研究设计，从数据库中提取的核心数据表包括：CPC 分类信息表（cpc.parquet）、专利授权年份表（patent_year.parquet）、专利前向/后向引用数据表（patent_ncit.parquet、patent_nref.parquet、patent_references.parquet）、CD 指数中间变量表（Disruption.parquet）、KI 突破性指数表（PatentSimilarityImportanceBreakthrough_forPost2022.parquet）、专利标题与摘要文本表（patents_title.parquet、patent_abstract.parquet）。

2）技术领域界定。本研究以新型储能技术为核心实证场景，同时选取集成电路、复合高分子材料两大领域作为补充研究样本，基于技术内涵与应用边界，在 CPC 分类体系的 cpc_group 级别下完成技术领域的精准界定与分类号筛选，具体如下：①新型储能：覆盖电化学储能、电力储能系统等六大方向及全技术链条，筛选 H01M（电化学/氢能储能器件）、H02J（电力储能系统）、Y02E（综合能源与节能技术）等 24 个 CPC 分类号。②集成电路领域：覆盖半导体器件、集成电路设计、制造工艺、封装测试全技术流程，筛选得到 H01L（半导体器件）等 59 个 CPC 分类号。③复合高分子材料领域：涵盖碳纤维复合高分子材料、高分子薄膜、功能涂层、纳米复合高分子材料四大方向，筛选得到 C08J（高分子组合物与加工）等 20 个 CPC 分类号。完整分类号清单见见本文中国知网增强出版相关数据。

3）样本筛选规则。样本筛选采用精确 CPC 匹配规则，仅当专利的 cpc_group 字段与上述分类号完全一致时才纳入初始样本池。相较于前缀模糊匹配，精确匹配可避免因分类号包含关系导致的样本边界模糊与技术领域混杂问题。

4）时间窗口控制。KI 指数与 CD 指数的计算均需完整的时间观测窗口：KI 指数需目标专利授权后 5 年的专利文本作为未来参照集；CD 指数需授权后 5 年的引用数据以保证计数完整，避免数据右截断导致的指标失真。以数据库快照年份 2024 年为截止时点，仅保留授权年份不晚于 2019 年的专利，确保所有样本均具备完整的 5 年观测窗口。经相同的精确 CPC 匹配与时间窗口筛选（授权年份不晚于 2019 年），最终得到三个领域的有效研究样本，分别为新型储能 21671 件、集成电路 19141 件和复合高分子材料 6936 件。

3.1.2　核心变量描述性统计　表 1 展示了 TextScore 与 CiteScore 在新型储能领域 21671 件全样本中的分布特征。两项指标均呈现典型的右偏长尾分布，符合技术创新领域“少数成果实现突破、多数为常规渐进改进”的一般规律。超过 50%的样本 TextScore 取值为零或负，表明在本研究样本中，仅有少数专利在文本内容上显著背离既往范式；CiteScore 的分布相对连续，但零值附近同样存在较高集中度。

表 1 核心变量描述性统计结果

Tab.1 Descriptive statistical results of core variables

指标	均值	中位数	标准差	最小值	最大值
TextScore（KI 指数）	0.098	0.000	0.190	- 0.902	2.036
CiteScore（CD 指数）	0.110	0.028	0.187	- 0.919	0.994

进一步分析显示，TextScore 与 CiteScore 的 Pearson 相关系数仅为 0.017，近乎正交且无统计显著性。该结果具有重要方法论意义：在新型储能领域，文本突破性与引文颠覆性是相互独立且信息不重叠的维度。这意味着单一维度存在盲区，文本指标无法反映技术轨迹实际冲击，引用指标易遗漏早期突破性专利，双维融合具有必要性。该数据特征为后续 PCA 分析自动生成等权结果提供了统计基础。

3.2 双维指标融合方法的系统比较与选择

3.2.1　各方法权重结果与统计特性分析　表 2 汇总了三种赋权方法在新型储能领域的权重分配结果及排序一致性指标。

表 2 三种赋权方法的权重分配与排序一致性

Tab.2 Weight distribution and ranking consistency of the three weighting methods

赋权方法	TextScore 权重	CiteScore 权重	PC1 方差解释率	与 PCA 排序的 Spearman ρ
熵权法	0.563	0.437	-	0.93
CRITIC 法	0.389	0.611	-	0.98
PCA 法	0.500	0.500	0.508	-

以下依次阐释各方法的权重生成逻辑，并结合本文的数据特征分析其统计学含义。

1）熵权法。熵权法依据指标的信息熵分配权重：指标离散程度越高，信息熵越小，对应的权重越大。在本研究样本中，TextScore 的变异系数（CV=1.94）略高于 CiteScore（CV = 1.70），因此 TextScore 被赋予 0.563 的权重，CiteScore 获得 0.437。该结果由数据自身的离散度差异驱动，在信息论框架内自洽。

需指出的是，本研究同时观测到 TextScore 与 CiteScore 近乎正交（r=0.017）。在这一数据结构下，两个指标携带的是不重叠的独立信息，此时若仅以离散度作为赋权依据，权重必向离散度更大的指标倾斜，但这部分额外倾斜的权重并非来源于该指标与颠覆性构念更强的关联，而仅来源于其内部数值差异更大。该特性是否影响其对综合颠覆性的刻画效度，须结合排序结果加以检验。

2）CRITIC 法。CRITIC 法在信息熵的基础上进一步引入指标间的“冲突性”：指标与其他指标的相关性越低，其提供的独立信息越多，应获得越高权重。在本研究中，TextScore 与 CiteScore 近乎正交，二者的冲突性值均接近 1。同时，CiteScore 的对比强度（归一化后标准差 0.107）高于 TextScore（0.068），因此 CRITIC 判定 CiteScore 的独立信息量更大，赋予 CiteScore 的权重为 0.611，TextScore 为 0.389。

该结果反映了 CRITIC 法对“独立变异信息”的敏感性。在近乎正交的双指标场景中，

CRITIC 权重分配主要取决于二者的对比强度之比。该机制在数学上是明确的，但其推论“CiteScore 的重要性约是 TextScore 的 1.6 倍”在现有颠覆性研究文献中尚无明确共识。

3）PCA 法（第 1 主成分）。PCA 法对经 Z-score 标准化后的双指标矩阵进行特征值分解。标准化后，TextScore 与 CiteScore 均服从标准正态分布 N（0，1），初始方差完全对等。在本研究的数据结构下，两变量近乎正交（r≈0.017），其方差—协方差矩阵趋近于单位矩阵 $\boldsymbol{I}$。在此条件下，两个特征值近乎相等，第一主成分方向恰好位于 45°对角线方向，载荷向量为（1/√2，1/√2），约（0.707，0.707），归一化后两指标权重均为 0.500。其统计学含义是：两个方差相等且近乎正交的标准化变量，对第一主成分的贡献天然对称。图 2 展示了在新型储能领域内专利 CD 指数与 KI 指数之间的正交程度，其中 PC1 载荷方向近乎 45°。

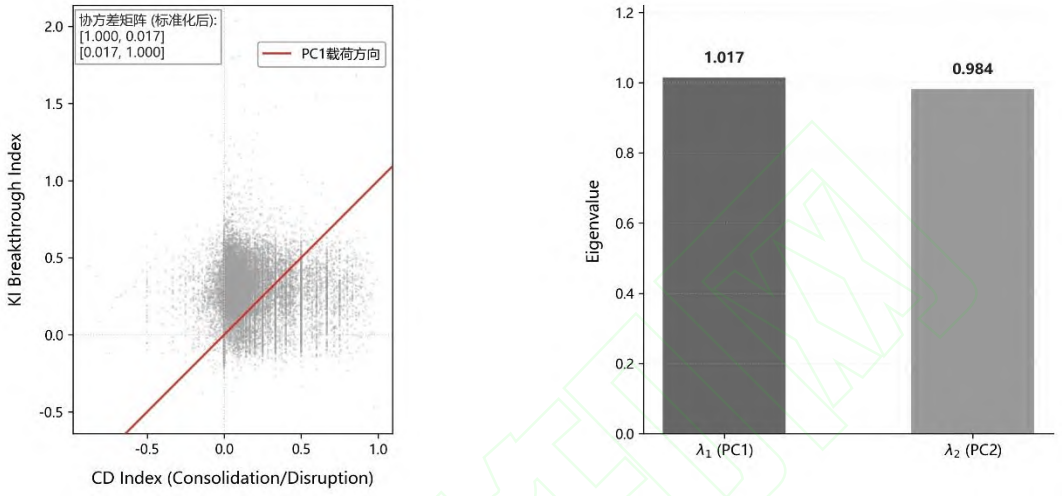


图 2 新型储能领域 KI 文本突破性与 CD 引文颠覆性的正交分布特征

Fig. 2 Orthogonal distribution characteristics of KI text breakthrough and CD citation disruptiveness in the new energy storage field

3.2.2 融合方案的选择：一份基于数据结构检验的推荐规程 上述分析揭示了一个关键的方法论洞见：在本数据集中，PCA 之所以收敛于等权，并非方法的固有属性，而是“TextScore 与 CiteScore 近似正交且标准化后方差相等”这一特定数据结构下的结果。该发现表明，不应简单地将等权公式（$PCDI = 0.5 \times Z(\text{TextScore}) + 0.5 \times Z(\text{CiteScore})$）作为最终结论，而应将其提炼为一份更通用的、有前提条件的分析规程。

因此，本文更希望提供的并非某个具体的融合公式，而是如下所示的双维融合分析规程：

1）条件检验。在进行双维融合前，研究者须首先检验其所在技术领域内 TextScore 与 CiteScore 两指标的相关性结构与方差特征。

2）方案选择。情境 A（如本研究）：若两指标近乎正交（如|r|＜0.1）且离散度接近，则 PCA 第一主成分将自然收敛于等权方向。此时，等权融合方案不仅可视为 PCA 等常见客观赋权方法在该数据结构下的共同近似解，且具备较好的统计解释性，可作为首选基准方案。情境 B：若两指标存在中度及以上相关（如|r|＞0.3），或方差悬殊，则 PCA 将产生非对称载荷。此时，等权公式不再具有最优方差解释性质，研究者应采用 PCA 第一主成分载荷，或经系统比较后选择的 CRITIC 法等更优赋权方案作为融合权重。

需要特别说明的是，本研究后续在十余个不同技术领域超二十万件专利数据中广泛实证表明，情境 A 是专利创新测度中的主导性情境，而情境 B 则属于理论上存在但实证中较为罕见的特殊情况（详见本文中国知网增强出版相关数据）。本文据此提出一个假设性推论：在专利创新衡量中，KI 指数与 CD 指数普遍存在近似正交的数学结构。若此假设在未来更多领域得到验证，则本文提出的基于等权融合的 PCDI 将可能成为一种具有领域鲁棒性的初筛工具。本文认为这一发现进一步凸显了本文提出的等权融合方案在绝大多数实际应用场景中的适用性与便捷性。然而，本文的规程之所以包含“情境 B”，正是基于对可能存在的“特殊领域”的审慎考量。在某些领域，KI 与 CD 或许不再正交，而是呈现出高度正相关。届时，基于本规程的非对称赋权方案将凸显其价值。本研究中观察到的 KI 指数与 CD 指数近乎正交的特征，仅指二者在统计数学层面的变异相互独立，即一个指标的数值变化无法线性

预测另一个指标的数值变化，该实证发现并不否定二者在理论逻辑上的内在关联。

3.3 前瞻预测效度验证：基于未来被引的筛选效能

须明确一个共识：专利被引次数并非颠覆性的等价物，而是受专利权人声誉、领域引用密度等多种非颠覆性因素干扰的混合信号。将其作为外部效标，并非基于二者等价，而是基于“颠覆性专利通常具有持续高影响力”这一必要非充分条件。故本节验证目标并非直接判定颠覆性，而是检验 PCDI 作为前瞻性筛选工具检索未来高影响力专利的效率。据此，选用 AUC（排序区分能力）与 Precision@5%（候选集命中率）作为核心评价指标。

3.3.1 研究设计与多领域结果概览 本文在新型储能、集成电路、复合高分子材料三个技术范式迥异的领域同步检验 PCDI 预测专利授权后 5 年被引次数（cit5）的能力。三个领域在技术成熟度、引用网络密度和语义空间特征上存在结构性差异：新型储能属新兴交叉领域；集成电路为高度成熟、引用网络密集的代表；复合高分子材料则介于二者之间。以 cit5 排名前 10%的专利为“高被引专利”标签，表 3 整合报告了三个领域中 KI 指数、CD 指数与 PCDI 的全部效度指标对比。

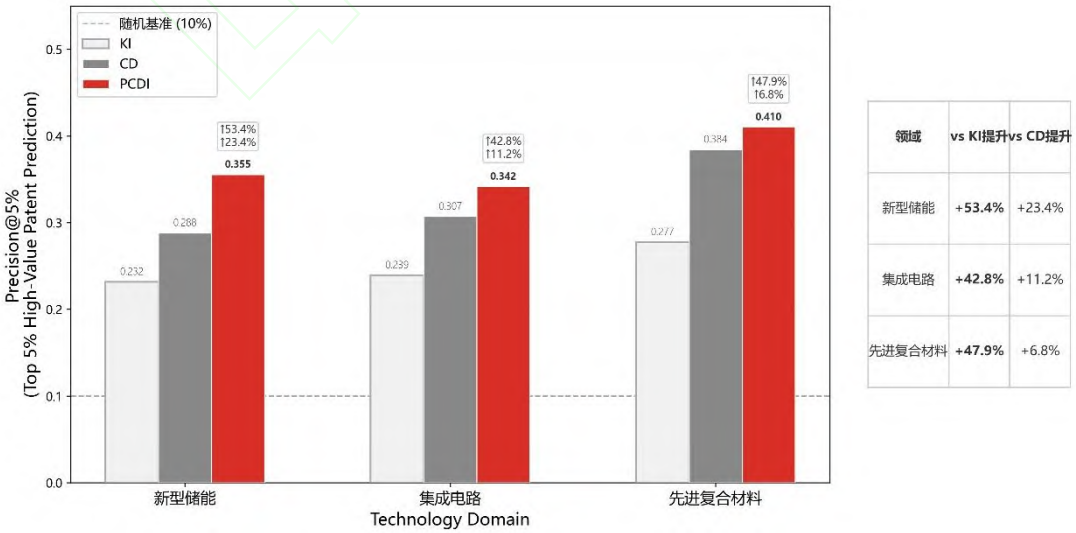
表 3 三领域单指标与 PCDI 的预测效度对比

Tab.3 Comparison of predictive validity between single indicators and PCDI in three fields

领域	指标	Spearman ρ	AUC	Precision@5%
新型储能（*N*=21671）	KI	0.218	0.623	0.232
	CD	0.477	0.631	0.288
	PCDI	0.435	0.673	0.356
集成电路（*N*=19141）	KI	0.207	0.603	0.239
	CD	0.422	0.637	0.307
	PCDI	0.391	0.668	0.342
复合高分子材料（*N*=6936）	KI	0.187	0.605	0.277
	CD	0.526	0.691	0.384
	PCDI	0.456	0.708	0.410

3.3.2 互补信号增强效应的跨领域一致性分析 表 3 揭示了一个贯穿三个领域的一致模式：PCDI 的 Spearman ρ 均介于 KI 与 CD 之间，但在 AUC 与 Precision@5%上均超越任一单指标。此模式的深层机制值得仔细辨析。

基于引用结构的 CD 指数与被引总量天然相关，这解释了其在 Spearman ρ 上的稳定优势。但 Spearman ρ 衡量全局排序质量，而实践中研究者仅关注前 5%～10%的头部命中率，因此 AUC 与 Precision@5%更具参考价值。KI 指数单独预测被引能力有限（三领域 Spearman ρ 在 0.187～0.218 之间），印证了文本新颖性不等于产业变革的论断。但 KI 与被引行为近乎正交的特性，使其携带了引用维度无法捕捉的技术内容跃迁独立信号。等权融合后虽未提升全局排序精度，却显著增强了顶尖高影响力专利的识别能力，这种“互补信号增强”效应在三领域均得到一致验证。图 3 展示了 PCDI 在 Precision@5%指标上相对于 KI 与 CD 的提升。



领域	vs KI提升	vs CD提升
新型储能	+53.4%	+23.4%
集成电路	+42.8%	+11.2%
先进复合材料	+47.9%	+6.8%

图 3 多技术领域下 KI、CD 与 PCDI 的 Precision@5%指标性能对比

Fig.3 Comparison of precision@5% Performance of KI，CD and PCDI across multiple technological fields

进一步通过分组梯度检验考察各指标对专利未来影响力的区分度。将全样本分别按 KI、CD 和 PCDI 得分从小到大五等分，计算每组平均 cit5，结果汇总见表 4。

表 4 三领域三种指标五分组的平均 5 年被引次数对比

Tab.4 Comparison of average 5-year citation counts for five groups of three indicators across three fields

领域	指标	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	首尾倍差
新型储能	KI	3.00	4.74	5.89	7.85	7.65	2.55
	CD	3.44	5.08	5.46	6.54	9.23	2.68
	PCDI	3.45	4.56	4.45	6.67	10.00	2.90
集成电路	KI	4.45	5.50	6.13	6.86	7.64	1.72
	CD	3.74	4.44	5.18	6.63	10.72	2.87
	PCDI	4.36	4.09	4.92	5.97	11.23	2.58
复合高分子材料	KI	2.92	3.18	3.54	3.65	4.72	1.62
	CD	1.76	3.01	3.37	4.06	6.41	3.64
	PCDI	2.28	2.42	2.82	3.99	6.49	2.85

在区分最高分组与最低分组的能力上，PCDI 在新型储能领域的首尾倍差达到 2.90，超越 KI 的 2.55 和 CD 的 2.68。集成电路领域 PCDI 的首尾倍差（2.58）虽略低于 CD（2.87），但 PCDI 最高分组的绝对被引均值（11.23 次）已超越 CD（10.72 次）。复合高分子材料领域 PCDI 的首尾倍差（2.85）同样略低于 CD（3.64），但最高分组绝对被引均值（6.49 次）达到了 CD 的水平（6.41 次）。三个领域中，PCDI 最高分组的绝对被引均值均达到或超越了表现最优的单指标，意味着当研究者使用 PCDI 筛选排名前 20%的专利时，这组专利未来的平均被引表现优于或等价于用任一单指标筛选的同比例专利。

值得注意的是，中间分组的非线性波动在三个领域中普遍存在。新型储能领域第 3 组（4.45）略低于第 2 组（4.56），集成电路领域第 2 组（4.09）甚至低于第 1 组（4.36），复合高分子材料领域第 2 组（2.42）与第 1 组（2.28）差距甚微。此跨领域一致的现象表明，融合指标在中等得分区间的区分力有所衰减，但其在两端的强区分能力（尤其是对高潜力候选集的有效圈定）恰恰是前瞻性筛选工具最受关注的性能特征。本研究认为，能够在两端做出清晰判断，比在中间段实现精细排序具有更高的实践价值。

3.3.3 实践筛选效能的量化解读 基于表 3 与表 4 的数据，对 PCDI 的预测效度进行跨领域量化解读。在 AUC 指标上，PCDI 在三个领域的取值分别为 0.673、0.668 和 0.708，均系统性地高于 CD 和 KI 的对应值。以新型储能领域为例，0.673 的 AUC 值意味着随机抽取一件高被引专利与一件普通专利，PCDI 将高被引专利排在前列的概率为 67.3%，较 CD（63.1%）高出 4.2 个百分点，较 KI（62.3%）高出 5.0 个百分点。

在富有实践价值的 Precision@5%上，PCDI 的优势更为显著。三个领域中，PCDI 的 Precision@5%分别为 0.356、0.342 和 0.410，较 CD 的增幅分别为 23.6%、11.4%和 6.8%，较 KI 的增幅分别达 53.4%、43.1%和 48.0%。以新型储能领域为例，PCDI 排名前 5%的专利中，35.6%最终跻身高被引行列，该比例约是随机筛选基准（10%）的 3.6 倍，而 CD 和 KI 分别是随机筛选基准的 2.9 倍和 2.3 倍。对于实践中的专利筛选工作而言，Precision@5%的提升具有直接的成本效益含义：若审查资源允许深度研判前 5%的专利，使用 PCDI 较使用 CD 可多捕获约 24%的高被引专利，较使用 KI 可多捕获约 53%的高被引专利。

跨领域相似的结果支撑了双维融合的核心价值：在仅使用两个独立构建的测度指标、未引入任何直接引用频次信息的条件下，PCDI 识别未来顶尖影响力专利的表现显著优于单维度指标，优势在 AUC 与 Precision 两个独立评价维度均有体现。此跨领域复现具有重要方法论意义：文本与引文维度的互补性并非新型储能特有，而是可以跨领域的。

3.4 历史重要专利回溯验证：基于公认标杆的召回能力

前瞻预测效度验证从“未来”方向检验了 PCDI 的筛选效率，本文则从“过往”方向引入历史重要专利回溯验证，以已获公认的技术里程碑为参照，反向审视 PCDI 能否将既有的颠覆性标杆准确定位于全样本的头部位置。下文从不同时间方向构成效度证据的闭环。

3.4.1 遴选标准与验证设计 遴选标准参照学界对颠覆性技术的定性识别惯例，确定以下 4 类外部判别依据（满足其一即可入选）：①获奖专利：获得国际公认技术奖项，如诺贝尔奖

相关工作转化专利、美国国家技术创新奖、R&D 100 Awards 等。②标准必要专利：被写入 IEC、IEEE 等国际标准且认定为核心技术。③学科奠基性专利：在权威综述（如 Nature Reviews Materials、IEEE Spectrum 历史回顾）中反复被列为领域奠基性成果。④产业里程碑专利：由彭博新能源财经、IC Insights、麦肯锡技术曲线等权威产业报告标识为改写产业格局的关键发明。依据上述标准，在新型储能、集成电路、复合高分子材料三个领域各筛选出 10 件美国授权专利，共 30 件（完整清单见本文中国知网增强出版相关数据）。将所选专利的 KI、CD 与 PCDI 值分别映射到各自领域全样本的排序分位数上（分位数定义为该专利得分在所有专利中的排序位置百分比，值越小代表排名越靠前），以检验各指标能否系统性地将历史重要专利置于分布顶端。

30 件候选专利中有 28 件成功匹配至数据集，US5905000 号与 US8288988 号专利因授权年份超出本研究所要求的前后 5 年观测窗口而缺失 KI 值，未纳入后续计算。表 5 汇总了全部 28 件历史重要专利在三个指标上的排序分位数，表 6 给出了三个领域各指标的平均分位数。从表 6 可以看出，在三个领域中，PCDI 的平均排序分位数均明显低于 KI 和 CD 的对应值。尤其在集成电路和复合高分子材料领域，PCDI 平均分位数低至千分之一量级（0.0028 和 0.0010），表明这些遵循严格外部标准遴选的历史重要专利，在 PCDI 的综合排序中几乎无一例外地被置于全样本的最顶端。相比之下，CD 指数的平均分位数在新型储能和复合高分子材料领域分别为 0.2018 和 0.2307，显著逊色于 PCDI。由此可见，融合文本与引用两维信息的 PCDI，能够比单一维度指标更有效地“召回”公认的颠覆性技术标杆。

表 5 历史重要专利在 KI、CD、PCDI 上的排序分位数

Tab.5 Ranking quantiles of historically important patents on KI，CD and PCDI

领域	专利号	KI 平均分位数	CD 平均分位数	PCDI 平均分位数
新型储能	US5512387	0.0135	0.0034	0.001
	US5541589	0.0001	0.952	0.0002
	US6709566	0.0001	0.4348	0.0002
	US5856773	0.0127	0.0014	0.0009
	US9093718	0.0034	0.0111	0.0006
	US6151330	0.0031	0.0071	0.0005
	US6168884	0.3689	0.2002	0.2257
	US9309124	0.0001	0.0041	0
集成电路	US5301318	0.0001	0.1515	0.0001
	US5330879	0.0124	0.0029	0.0003
	US5586319	0.0005	0.0334	0.0002
	US5595607	0.0178	0.0078	0.0006
	US5625217	0.0134	0.0072	0.0005
	US5744283	0.0773	0.0017	0.0009
	US7107567	0.6232	0.0037	0.0238
	US6129901	0.0003	0.0052	0.0001
	US6184545	0.0851	0.0008	0.0008
	US4933872	0.0392	0.0029	0.0006
复合高分子材料	US5273729	0.0004	0.0697	0.0003
	US5514439	0.0013	0.1968	0.0014
	US4626516	0.0181	0.0082	0.0009
	US4636480	0.0133	0.0043	0.0007
	US4765930	0.043	0.0099	0.0018
	US4801564	0.0011	0.2297	0.0016
	US4810442	0.0134	0.0263	0.0022

	US6593166	0.0001	0.9549	0.0001
	US7132679	0.0003	0.7929	0.0005
	US8216541	0.001	0.0146	0.0004

表 6 三领域历史重要专利平均排序分位数

Tab.6 Average ranking quantiles of historically important patents in three fields

领域	KI 平均分位数	CD 平均分位数	PCDI 平均分位数
新型储能	0.0502	0.2018	0.0286
集成电路	0.0869	0.0217	0.0028
复合高分子材料	0.0092	0.2307	0.001

注：分位数数值越小代表排名越靠前，如 PCDI 平均分位数 0.0286 意味着这些专利在 PCDI 排序中平均位居全样本的前 2.86%。

3.4.2 指标互补模式的个案揭示 表 5 的个案数据进一步揭示了指标互补的具体模式，三个领域均呈现出典型的“高 KI、低 CD”型专利被 PCDI 成功保护的案例。

在新型储能与智能电网技术领域，P. J. Delaney 所发明的 US5541589 号专利《电表数据采集与控制系统》（Power Meter Data Acquisition and Control System），奠定了远程自动抄表（AMR）的核心技术架构，其核心技术方案于 2002 年被正式纳入 IEC 62056 系列智能电表国际标准体系，成为全球智能电表数据采集与控制领域的核心技术基准。其 KI 分位数为 0.0001（几乎最优），但 CD 分位数高达 0.9520，这意味着如果仅使用 CD 指数，该专利的排序将位于全样本的后 5%，完全淹没在普通专利之中。然而，经等权融合处理后，PCDI 的分位数校准至 0.0002，仍稳定处于全样本前 0.02%的极高分位区间。

在复合高分子材料领域，Kia Silverbrook 所发明的 US6593166 号、US7132679 号专利《碳纳米管基体材料的构建方法》（Method for Construction of Nanotube Matrix Material）、《单壁碳纳米管电子器件》（Single Wall Carbon Nanotube Electronic Devices），奠定了碳纳米管三维有序结构可控组装的核心技术体系基础，其核心自组装设计思路被全球碳基材料研发领域广泛采纳，成为碳纳米管宏观功能基体材料研发与应用的核心技术基准。US6593166 号的 KI 分位数为 0.0001，CD 分位数高达 0.9549，PCDI 分位数仍为 0.0001；US7132679 号的 KI 分位数为 0.0003，CD 分位数高达 0.7929，PCDI 分位数仍为 0.0005。这两件专利在内容层面做出了极为激进的突破，却因时间尚短或工艺复杂而尚未形成引用重构效应，是单指标系统性盲区的典型受害者。

在集成电路领域，M. A. LeBlanc 所发明的 US7107567 号专利《电子设计保护电路》(Electronic Design Protection Circuit)，奠定了 FPGA 领域 IP 核仿真验证与版权防护的核心技术架构，其核心防护方案被全球主流 FPGA 与 EDA 厂商纳入商用工具体系。其 KI 分位数高达 0.6232，但 CD 分位数仅为 0.0037，PCDI 分位数为 0.0238。该领域虽以“低 KI、高 CD”模式为主，但 PCDI 同样通过双维信息的权衡，避免了单纯依赖任一维度可能造成的偏颇。

综合三个领域的个案可以发现，PCDI 的核心价值在于：当专利在某一维度表现极为突出而另一维度暂时滞后时，等权融合机制能够有效保护这种“内容先行、引用待发”的早期颠覆性发明不被单一指标所湮没。这正是双维融合相较于单指标筛选的主要优势之一。

同时，对于 US9309124 号（石墨烯剥离方法）、US6129901 号（定向碳管可控合成）等“双高”型专利（KI 和 CD 分位数均极低），PCDI 同样给出了优秀的排序位置，并未因等权融合而牺牲对它们的检索能力。整体而言，历史重要专利的回溯验证从“过往”方向进一步加固了前瞻预测效度检验的结论：PCDI 在两个正交维度之间实现了信息的权衡与互补，既能将已获公认的颠覆性先驱准确定位于分布顶端，又能有效捕获那些因单指标滞后或偏颇而被遗漏的早期突破性工作，具备作为大规模专利颠覆性潜力筛选工具的参考价值。图 4 展示了新型储能、集成电路与复合高分子材料领域典型历史核心专利在 KI、CD 单一指标及 PCDI 融合指标下的排序分位数差异，直观反映出单一创新指标存在明显低估缺陷，验证了 PCDI 融合模型对重大技术专利的精准识别与排序修正能力。

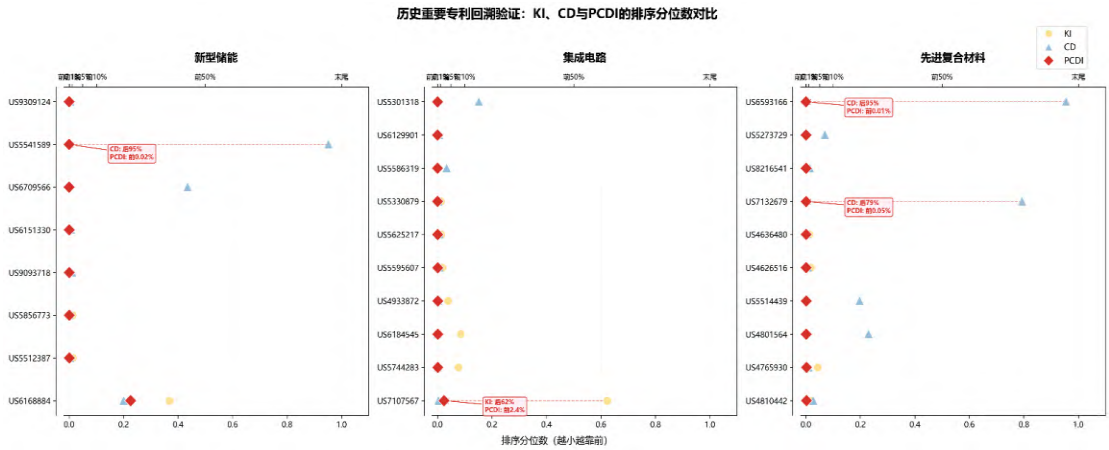


图 4 多领域历史重要专利的 KI、CD 与 PCDI 分位数排序对比

Fig.4 Quantile ranking comparison of KI，CD and PCDI for historically important patents in multiple fields

3.5 讨论

本文以新型储能为主场景、集成电路与复合高分子材料为平行验证场景，完成了“指标测度—融合方案选择—双向效度验证”全流程实证，构建了前向预测与后向回溯相互印证的证据链。

融合方案选择发现：当文本突破性（KI）与引文颠覆性（CD）近乎正交且方差接近时，熵权法、CRITIC 法与 PCA 三种客观赋权法排序大致相同，PCA 自然收敛于等权解。

前瞻预测效度显示，三个技术范式迥异的领域呈现一致稳健模式：PCDI 的全局相关性（Spearman ρ）介于两单指标之间，但衡量实践筛选效率的 AUC 与 Precision@5%均显著领先。三领域 PCDI 的 Precision@5%较 CD 提升 6.8%～23.6%，较 KI 提升 43.1%～53.4%，最高分组绝对被引均值全面超越最优单指标。这种“全局虽不占优但头部命中率领先”的特征很符合对于头部颠覆性专利的筛选。

历史重要专利回溯验证表明，基于获奖等 4 类外部标准遴选的标杆专利中，PCDI 的平均排序分位数显著低于 KI 与 CD。其核心优势在于能将 KI 极高但 CD 极低的“内容先行、引用待发”型早期发明校准至头部位置，同时不牺牲“双高”型颠覆性专利的排序，实现了不同类型信号的合理权衡。

综上，双向效度验证与跨领域复现共同支撑核心结论：在 KI 与 CD 近乎正交的数据结构下，基于 PCA 内生等权融合的 PCDI，在未来高影响力专利检索与已知颠覆性标杆召回两方面均显著优于单维度指标，且优势跨领域稳健，为颠覆性专利早期筛选提供了客观、可复现的实证工具。

4 总结与展望

4.1 核心发现

本文围绕“海量专利中颠覆性技术的早期精准识别”核心问题，以新型储能为主场景、集成电路与复合高分子材料为验证场景，构建并验证了融合文本突破性与引文颠覆性的双维识别规程，得出三项核心发现。

第一，文本突破性与引文颠覆性在实证层面近乎正交，双维融合具有必要性。三领域全样本中，KI 与 CD 指数的 Pearson 相关系数均不显著，表明技术内容语义跃迁与引文网络知识替代是两个独立信号系统。单一维度存在系统性盲区：纯文本路径无法反映技术实际扩散程度，纯引用路径受引文时滞限制，对“内容先行、引用待发”的早期发明缺乏敏感性。

第二，融合方案应内生于数据结构而非先验预设，据此提出“检验数据结构—内生选择融合方案”的方法论路径。当两指标标准化后方差接近且近乎正交时，PCA 自然收敛于等权解，但这是特定数据条件下的结果，而非方法固有属性。熵权法忽略指标相关结构、CRITIC 法在正交条件下权重仅取决于标准差，均存在局限性。研究者应根据领域数据特征选择等权（情境 A）或非对称赋权（情境 B）方案，而非固化等权公式。

第三，PCDI 实现了“互补信号增强”，在前瞻筛选效率与历史回溯召回能力上均显著优于单维度指标。跨领域实证呈现一致稳健模式：PCDI 全局相关性（Spearman ρ）介于两单

指标之间，但实践筛选核心指标显著领先——新型储能领域 AUC 达 0.673、Precision@5% 达 0.356，三领域 Precision@5% 较 CD 提升 6.8%～23.6%，较 KI 提升 43.1%～53.4%，最高分组绝对被引均值全面超越最优单指标。历史回溯中，30 件公认颠覆性专利的 PCDI 平均排序分位数低至 0.0010～0.0286，显著优于单指标。

4.2 研究局限与未来方向

本研究存在以下局限，也为后续工作指明了方向。

第一，文本突破性测度对预计算指标与短文本的依赖。本研究采用 KI 指数，其语义向量基于标题与摘要构建，可能遗漏权利要求书与说明书中承载的关键技术创新细节。未来计划探索引入大语言模型对专利全文进行深度语义编码，并采用可变或领域自适应时间窗口来探索更精细的测度可能。

第二，本研究的实证样本限于美国授权专利，未来计划将框架拓展至 PCT 国际申请或中国发明专利，并探索适用于不完整观测窗口的截断适应方法，以提升框架在真实前瞻场景中的时效性与适用范围。□

参考文献

[1] Christensen C M, McDonald R, Altman E J, et al. Disruptive innovation: an intellectual history and directions for future research[J]. Journal of Management Studies, 2018, 55(7): 1043.

[2] 张琪, 马瑞敏, 杨杰, 等. 融合颠覆性与巩固性引文特征识别领域突破性论文：基于可解释性机器学习方法[J]. 情报理论与实践, 2026, 49(2): 126. （Zhang Qi, Ma Ruimin, Yang Jie, et al. Identifying breakthrough papers by integrating disruptive and consolidating citation features: an interpretable machine learning approach[J]. Information Studies: Theory & Application, 2026, 49(2): 126.）

[3] Funk R J, Owen-Smith J. A dynamic network measure of technological change[J]. Management Science, 2017, 63(3): 791.

[4] Yang A J. Text vs. citations: a comparative analysis of breakthrough and disruption metrics in patent innovation[J]. Research Policy, 2025, 54(8): 105295.

[5] Deng N, Gu X, Fan Y, et al. The critical role of persistent disruption in advancing science[J]. Nature Computational Science, 2025, 5(6): 492.

[6] 吴欣雨, 李涵昱, 张智雄, 等. 科技文献评价中语义新颖性研究综述[J]. 数据分析与知识发现, 2024, 8(3): 29.（Wu Xinyu, Li Hanyu, Zhang Zhixiong, et al. A review of semantic novelty research in scientific literature evaluation[J]. Data Analysis and Knowledge Discovery, 2024, 8(3): 29.）

[7] 吴洁, 刘彦良, 谢小东, 等. 融合专利异质特征的颠覆性技术识别方法研究：以人工智能领域为例[J]. 数字图书馆论坛, 2025, 21(7): 42. （Wu Jie, Liu Yanliang, Xie Xiaodong, et al. Research on disruptive technology identification integrating heterogeneous patent features: a case study of the artificial intelligence field[J]. Digital Library Forum, 2025, 21(7): 42.）

[8] Arts S, Hou J, Gomez J C. Natural language processing to identify the creation and impact of new technologies in patent text: code, data, and new measures[J]. Research Policy, 2021, 50(2): 104144.

[9] Dong J, Yang C, Zhang H, et al. Identifying and understanding disruptive energy technologies: a data-driven integrated approach[J]. Energy Economics, 2026, 155: 109178.

[10] 唐旭丽, 李信. 科研团队多样性对学术颠覆性创新的影响研究：以人工智能领域为例[J]. 情报学报, 2023, 42(1): 43.（Tang Xuli, Li Xin. The impact of research team diversity on academic disruptive innovation: a case study of the artificial intelligence field[J]. Journal of the China Society for Scientific and Technical Information, 2023, 42(1): 43.）

[11] Chen X, Mao J, Li G. A co-citation approach to the analysis on the interaction between scientific and technological knowledge[J]. Journal of Informetrics, 2024, 18(3): 101548.

[12] 赵玉桐, 杨建林. 基于跨领域专利的颠覆性技术识别研究：以人工智能领域为例[J]. 情报理论与实践, 2023, 46(3): 174. （Zhao Yutong, Yang Jianlin. Disruptive technology identification based on cross-domain patents: a case study of the artificial intelligence field[J]. Information Studies: Theory & Application, 2023, 46(3): 174.）

[13] Bornmann L, Devarakonda S, Tekles A, et al. Are disruption index indicators convergently valid The comparison of several indicator variants with assessments by peers[J].Quantitative Science Studies, 2020, 1:1242.

[14] 何郁冰，徐美娟，荆晶. 基于创新采用框架的颠覆性技术识别研究[J]. 科学学研究, 2025, 43(1): 66. （He Yubing, Xu Meijuan, Jing Jing. Disruptive technology identification based on an innovation adoption framework[J]. Studies in Science of Science, 2025, 43(1): 66.）
[15] 潘一如. 基于引文网络的高颠覆性专利知识扩散特征研究[J]. 数据分析与知识发现, 2023, 7(10): 1.（Pan Yiru. Research on knowledge diffusion characteristics of highly disruptive patents based on citation networks[J]. Data Analysis and Knowledge Discovery, 2023, 7(10): 1.）
[16] Park M, Leahey E, Funk R J. Papers and patents are becoming less disruptive over time[J]. Nature, 2023, 613(7942): 138.
[17] Leydesdorff L, Tekles A, Bornmann L. A proposal to revise the disruption index[J]. El Profesional De La Información, 2021: e300121.
[18] Wei C, Li J, Shi D. Quantifying revolutionary discoveries: evidence from Nobel prize-winning papers[J]. Information Processing & Management, 2023, 60(3): 103252.
[19] Kelly B, Papanikolaou D, Seru A, et al. Measuring technological innovation over the long run[J].American Economic Review: Insights,2021,3(3):303.
[20] 李慧，马凯，周潇. 技术传播视角下融合多特征的颠覆性技术识别研究[J]. 现代情报, 2025, 45(11): 18.（Li Hui, Ma Kai, Zhou Xiao. Disruptive technology identification integrating multiple features from a technology communication perspective[J]. Modern Information, 2025, 45(11): 18.）
[21] 俞立平. 客观赋权法本质及在科技评价中的应用研究：以学术期刊为例[J]. 情报理论与实践, 2021, 44(2): 50.（Yu Liping. The essence of objective weighting methods and their application in science and technology evaluation: a case study of academic journals[J]. Information Studies: Theory & Application, 2021, 44(2): 50.）
[22] Mariani M S, Medo M, Lafond F. Early identification of important patents: design and validation of citation network metrics[J]. Technological Forecasting and Social Change, 2019, 146: 644.

作者简介：赵志洋，男，2003 年生，硕士生。研究方向：数据管理。**杨杰**，男，2000 年生，博士。研究方向：科学计量与创新评价。**吴娜，**女，1999 年生，博士生。研究方向：科学计量与创新评价。**王昊，**男，1981 年生，博士，教授，博士生导师。研究方向：自然语言处理。**邓三鸿**（通信作者，Email：sanhong@ nju.edu.cn），男，1975 年生，博士，教授，博士生导师。研究方向：信息处理，科学计量，数字人文。
作者贡献声明：赵志洋，研究设计与论文撰写。**杨杰**，论文写作指导。**吴娜，**数据收集整理。**王昊，**论文框架设计。**邓三鸿**，论文写作指导。
AI 使用声明：研究主体未使用。
录用日期：2026-07-06