

基于“论文-专利”的主题关联与互动演化研究——以氢燃料电池汽车领域为例*

张军荣 乔建智

(中南民族大学法学院 湖北武汉 430074)

摘要: 科学论文与专利技术之间存在密切且复杂的交互关系,但具体的关联与演化规律仍未完全明晰。深入解析论文与专利之间的相互作用机制,有助于揭示知识流动的内在规律,并准确预测技术转化路径。研究从 WOS 与 Incopat 中提取论文与专利数据单元,采用 BERTopic 主题模型,挖掘各时间切片的论文与专利主题,并基于语义相似度,分析论文与专利之间关联和互动趋势。结果表明:氢燃料电池汽车的部分研究主题表现出较强的协同创新趋势;论文与专利的关联强度随着时间推移减弱,反映出技术从基础研究向应用开发的转变过程;论文与专利之间存在吸收式演化、存续式演化和创设式演化三种演化机制,通过创新理论剖析演化机制,为未来技术变革和创新战略提供理论指导。

关键词: 主题关联; 互动演化; BERTopic 主题模型; 主题相似度; 氢燃料电池汽车

中图分类号: G250.2

0 引言

科技创新是生产力发展的核心要素。正如约瑟夫·熊彼特在其著作《经济发展理论》中所阐述的,通过引入新的生产方法,重新对生产要素进行组合,从而提升生产效率,促进经济发展。这些新的生产方法通常源于科学知识的探索与应用技术的突破,在科技创新价值链中,科学研究为新技术的研发奠定了理论基础,以专利为代表的技术则是将科学成果转化为实际应用的关键形式。科学论文代表前沿知识的创造与积累,还反映了基础研究的最新成果;专利技术则体现了知识向实际应用的转化能力。尽管科学论文与专利技术在推动科学研究与技术进步方面共同发挥着重要作用,但它们之间的互动与转化并非简单的线性关系。揭示科学论文与专利技术之间的主题关联与互动演化规律,有助于深入理解知识从理论到实践的转化过程,对预测技术趋势和加速技术创新具有重要意义。尤其是在科技快速发展的背景下,厘清这一规律显得尤为重要。

尽管学术界在揭示科学论文与专利技术的关联性方面取得了一些进展,但现有研究大多停留在基于共引关系或引文网络分析的方法上,侧重于静态视角下的固定关联模式,未能充

* 基金项目:国家自然科学基金项目“产权结构对高校专利转化绩效的影响研究”(项目编号:72004238)。

作者简介:张军荣(1986—),男,山西平陆人,博士,中南民族大学法学院副教授、硕士研究生导师,研究方向为知识产权、科技管理;乔建智(2000—),男,湖北宜昌人,中南民族大学法学院硕士研究生,研究方向为知识产权。本文通讯作者:乔建智。

收稿日期:

分揭示科学与技术之间的动态演化机制。随着人工智能和自然语言处理技术的发展,基于文本内容的动态分析方法正逐渐成为探讨这一关联的前沿领域。通过结合 BERTopic 主题建模、语义相似度和演化机制分析,可以更深入地研究科学论文与专利技术的关联机制及其演化过程,进一步揭示科学知识向技术应用转化的内在逻辑,并为技术创新管理提供理论支持。

1 相关研究进展

1.1 论文—专利主题关联研究

知识创造和知识流动的有效性在高级经济体中已然成为重要的竞争因素^[1]。当前,学界在论文与专利的主题关联研究方面已取得一定进展,研究方法大致可分为基于引用关系、基于文本内容以及基于映射关系的三类研究方法。

基于引用关系的研究方法通过分析论文与专利之间的引用或者耦合关系,揭示科学与技术之间的联系。例如, M. Ahmadpoor 等^[2]提出了一种基于引用距离的度量方法,揭示科学研究与技术发明之间的关联性。X. Chen 等^[3]基于共被引关系进一步衡量了科学研究与技术知识之间的互动。此外,宁子晨等^[4]探讨了“数据挖掘”主题下专利主体与技术主题以及文献关键词的耦合关系及其演化规律。

基于文本内容的研究方法则通过语义分析或主题建模,挖掘论文与专利文本中的相似性或关联性。例如, L. X. Chen 等^[5]通过向量空间模型(VSM)和 WF-IDF 加权法,测量了专利引文间的文本相似性,检验了其是否能够指示知识关联。张凯等^[6]结合 Termolator 算法与 GPT 提示学习,创新性地对技术术语抽取和聚类分析,以识别新兴技术。

基于映射关系的研究方法通过学科与技术领域的映射,研究论文与专利之间的知识流动与相互影响。例如,唐露源等^[7]基于维基百科分类树构建了知识本体,并将论文与专利映射到该本体中。通过配对样本 T 检验,验证了论文与专利在知识演化方面的时间滞后性与知识流动的特征。R.B. Ferreira 等^[8]则构建了科学研究与技术应用成果之间的映射关系,揭示了基础科学对技术转化的影响。

上述三类研究方法各具优劣。基于引用关系的研究方法能够从宏观层面揭示论文与专利的结构性联系,但难以克服数据时滞性问题,且未能深入挖掘文本内容,导致结果的细粒度不足。基于文本内容的方法能够从微观层面分析论文与专利的内容,提供更为精细的结果,但其效果易受样本类型、模型质量和参数设置的影响,从而对实验结果的稳定性与可靠性提出了较高要求。基于映射关系的研究方法在跨学科或跨技术领域研究中,由于学科分类标准的差异,难以建立有效的对应关系,限制了其应用效果。为克服这些局限性,混合研究方法为主题关联分析提供了新的解决思路。冉从敬等^[9]结合社会网络分析、科学知识图谱与

“BERT+Kemans+LDA”模型，提出了一种新的分析框架，为科学与技术的互动研究提供了理论与实践上的创新。

1.2 互动演化研究

互动演化研究旨在系统分析和追踪特定领域内主题、概念或关键词随时间变化的过程，以揭示其发展趋势、相互关系及演变机制。传统方法包括领域专家主导的定性分析、基于计算机技术的定量分析，以及二者结合的混合分析方法^[10]。B. Glaser 与 A. Strauss^[11]提出的扎根理论通过系统的定性研究，从数据中生成理论，被广泛应用于社会科学研究。此外，K.Krippendorff^[12]的内容分析法通过对文本或媒体内容进行编码与分类，识别潜在的模式和主题。然而，这些定性分析方法由于过度依赖领域专家，逐渐暴露出主观性强、效率低、人力成本高等缺点，限制了其应用的广泛性。

随着计算机技术的进步，G. Salton 等^[13]发展了向量空间模型（VSM），为主题自动化识别与相似性计算提供了技术支持。此后，D. M. Blei 等^[14]提出的潜在狄利克雷分配模型（LDA）极大地促进了计算机辅助的主题识别。在此背景下，学者们逐渐应用计算机技术开展主题演化研究，例如，H. Zhang 等^[15]融合 LDA 主题模型、生命周期理论与文本相似度，评估特定技术的演化路径。寇园园等^[16]从 IPC 专利分类的视角，应用马尔可夫链和显性技术比较优势指数，识别技术的演进特征与演变结果。此外，胡泽文等^[17]提出了基于 LDA2Vec—BERT 的主题识别与演化模型，用以识别区块链技术的演化趋势与特征。

总体而言，主题演化研究经历了从定性分析到定量分析，再到混合分析的发展演变。随着计算机技术和数据分析方法的进步，混合分析方法在提升研究效率和深度上展现出巨大潜力。BERTopic 等新兴模型为主题演化提供了灵活工具，能更好地捕捉主题的细微变化和动态演变，逐渐成为当前研究的主流趋势。

2 研究思路与研究方法

2.1 研究思路

本研究以智能化数据处理为核心方法。研究主线包括“确定研究范围→BERTopic 主题建模→主题关联与互动演化分析”，针对特定领域进行深入分析。具体而言，研究流程分为以下几个步骤：首先，选定研究主题，确定时间范围，制定检索策略，分从 Incopat 数据库提取专利数据单元及从 WOS 数据库提取论文数据单元，并对提取的文本进行预处理；其次，应用 BERTopic 主题建模，对论文与专利的摘要数据集进行文档嵌入、UMAP 降维、HDBSCAN 聚类及 C-TF-IDF 主题表示，提取主题聚类结果；最后，从两个维度展开分析：一是通过余弦相似度评估论文与专利之间的主题关联关系，二是基于主题对的关联度分析互

动演化过程与趋势。研究构建的“论文-专利”主题关联与互动演化分析框架如图 1 所示。

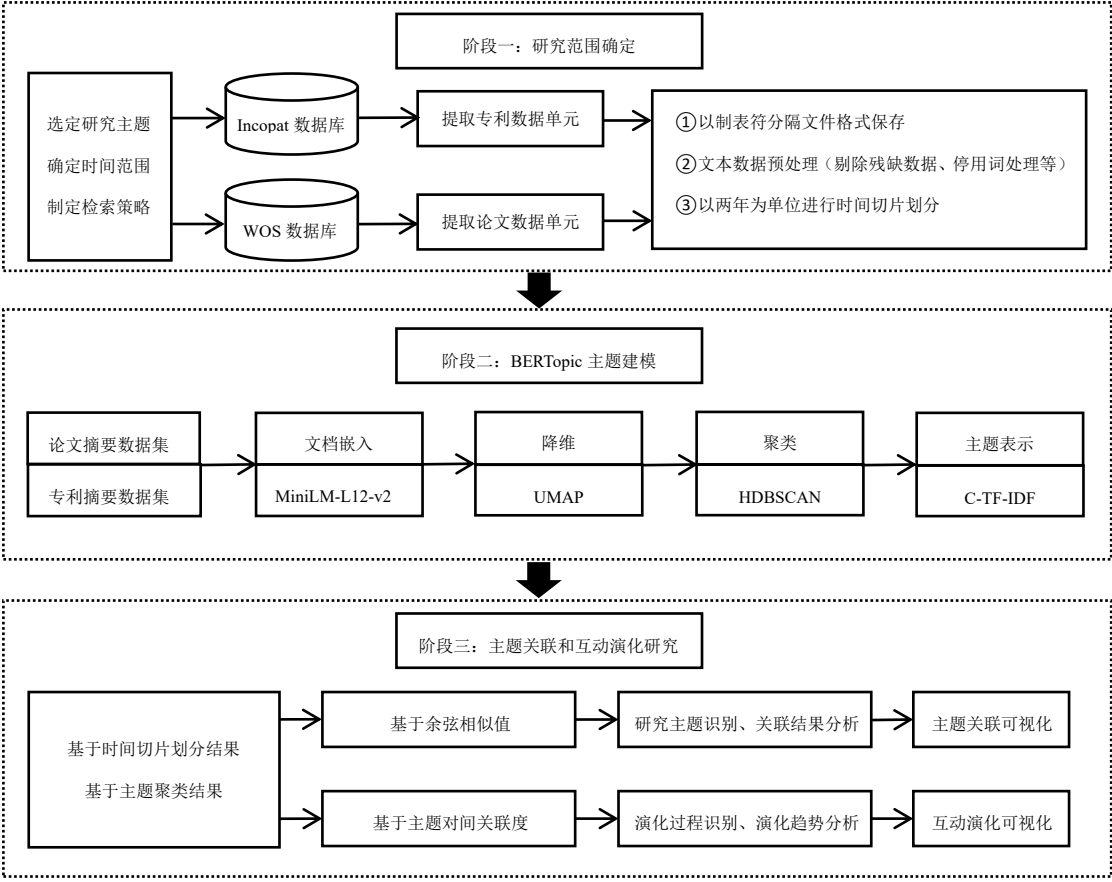


图 1 基于“论文-专利”的主题关联与互动演化分析框架

Fig.1 Analysis Framework for the Evolution of Theme Associations and Interactions Based on 'Paper-Patent'

2.2 研究方法

2.2.1 BERTopic 主题建模的研究方法

本研究采用 BERTopic 模型对论文与专利摘要数据集进行主题建模。相较传统 LDA 模型，该方法克服了忽略文本上下文语义关联的局限^[18]，通过预训练语言模型生成文档嵌入，结合 UMAP 降维与 HDBSCAN 聚类实现主题划分，并利用 C-TF-IDF 进行主题表示，在复杂文本的主题挖掘与语义分析中表现出较强的适用性。

文档嵌入是将文本转换为向量表示的一种技术，通常通过预训练的语言模型生成。本研究使用了微软开发的轻量级预训练语言模型“all-MiniLM-L6-v2”。该模型基于 BERT 架构，包含 6 层 Transformer，每层有 384 个隐藏单元，具备高效处理大规模文本数据的能力。文档嵌入的过程可概述为以下几步：首先，对输入文本进行分词，将文本分解为词汇或子词单元。对于每个词汇，模型会查找其对应的预训练嵌入表示，假设词汇 w_i 对应的嵌入为 e_i 。接着，模型通过加入位置编码，保留词语在文本中的位置信息。假设位置嵌入为 p_i ，则最终的

输入嵌入表示为词嵌入 e_i 与位置嵌入 p_i 之和。在 Transformer 编码器部分，模型采用多层结构，每层包含自注意力机制和前馈神经网络。自注意力机制计算序列中每个词语与其他词语的依赖关系，其计算过程如下所示：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \text{公式 (1)}$$

式中， Q 表示查询向量， K 表示键向量， V 表示值向量， d_k 为键向量的维度。通过softmax函数将点积结果转换为注意力权重，用以衡量词语间的相关性。softmax 的输出与值向量 V 相乘，得到加权求和后的注意力输出 $\text{Attention}(Q, K, V)$ 。该输出被送入前馈神经网络，每个词嵌入经过一层线性变换、ReLU 激活函数和第二层线性变换的处理，计算公式如下：

$$h = \max(0, xW_1 + b_1) \quad \text{公式 (2)}$$

$$y = hW_2 + b_2 \quad \text{公式 (3)}$$

式中， W_1 和 W_2 分别表示前馈神经网络中第一层和第二层的权重矩阵， b_1 和 b_2 为偏置向量，ReLU 激活函数定义为 $\max(0, x)$ 。最后，通过提取对应[CLS]标记的输出向量，作为整个文档的嵌入表示。

UMAP 是一种非线性降维技术，旨在保留高维数据的局部和全局结构，将高维数据嵌入低维空间。其降维过程分为两步：首先，使用 k-NN 算法找到每个数据点的 k 个最近邻居；然后，将这些高维距离转换为概率，表示数据点之间的连接强度。相似度计算公式如下：

$$s(i, j) = \exp\left(-\frac{d(i, j) - \rho_i}{\sigma_i}\right) \quad \text{公式 (4)}$$

式中， $d(i, j)$ 表示数据点 i 与 j 之间的距离， ρ_i 表示点 i 到其最近邻居的最小距离， σ_i 是用于控制局部邻域大小的参数。UMAP 通过最小化高维空间与低维空间相似度的差异来优化降维效果，其损失函数（交叉熵）如下所示：

$$CE(i, j) = \sum_{(i, j) \in \text{edges}} s(i, j) \log\left(\frac{s(i, j)}{s_{\text{low}}(i, j)}\right) + (1 - s(i, j)) \log\left(\frac{1 - s(i, j)}{1 - s_{\text{low}}(i, j)}\right) \quad \text{公式 (5)}$$

式中， $s(i, j)$ 表示点 i 与点 j 在高维空间中的相似度， $s_{\text{low}}(i, j)$ 表示它们在低维空间中的相似度。集合 edges 表示基于 k-NN 算法生成的邻近点对。UMAP 通过最小化高维与低维空间相似度的差异来优化降维效果。降维后，使用 HDBSCAN 进行聚类分析，该算法通过识别数据点的密度差异和噪声点来提高聚类的准确性和细粒度。主题表示采用 C-TF-IDF 方法，其通过增强类别内词频的差异来突出主题，公式如下：

$$W_{t,c} = \text{tf}_{t,c} \times \log\left(1 + \frac{N}{f_t}\right) \quad \text{公式 (6)}$$

式中， $\text{tf}_{t,c}$ 表示词语 t 在类别 c 中的词频， f_t 表示词语 t 在所有类别中的总出现次数， N 表示每个类别的平均词数， $W_{t,c}$ 为词语 t 在类别 c 中的权重。根据权重排序，选取词语作为类别的主题表示。

2.2.2 主题关联与互动演化的研究方法

为揭示科学论文与专利技术之间的主题关联,本研究采用了基于语义相似度的关联分析方法。通过计算论文和专利的主题嵌入向量之间的余弦相似度,衡量它们在语义空间中的接近程度。余弦相似度作为一种常见的相似性度量方法,通过计算两个向量夹角的余弦值来评估它们的方向相似性,具体计算公式如下:

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{\|A\| \times \|B\|} \quad \text{公式 (7)}$$

式中, $A \cdot B$ 表示向量 A 与向量 B 的点积,计算方法见公式 (8)。 $\|A\| \times \|B\|$ 表示向量 A 与向量 B 的模的乘积,计算方法见公式 (9)。

$$A \cdot B = \sum_{i=1}^n (A_i \times B_i) \quad \text{公式 (8)}$$

$$\|A\| \times \|B\| = \sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2} \quad \text{公式 (9)}$$

为确保研究结果的准确性,按照两年为单位划分时间切片,并计算不同时间段内论文和专利主题的语义相似度。参考刘春丽等^[19]在相关研究中的方法,将相似度阈值设定为 0.8。当相邻时间切片的主题相似度大于 0.8 时,表明主题关联度较高,反映出科学或技术主题在演化过程中存在向下一个主题转移的现象。另外,得益于桑基图在呈现复杂信息流动过程中的直观性和准确性,其已被广泛应用学术研究中^[20]。为系统分析论文与专利主题在时间维度上的互动演化路径,本研究基于主题聚类结果与主题相似度构建了主题演化网络,节点代表特定时间段的主题,边则表示主题间的相似度关系,通过桑基图对该网络进行可视化。在此基础上,进一步结合创新理论,深入探索其内在的演化机制。

3 “论文-专利”主题关联与互动演化实证分析

3.1 阶段一: 研究范围确定

节能降碳是推动碳达峰碳中和战略实施的核心举措,也是实现美丽中国建设和推动经济社会绿色转型的关键路径。在此背景下,氢能源凭借清洁、低碳和可再生的优势,成为能源转型的重点方向之一。我国在《氢能产业发展中长期规划(2021-2035 年)》中明确指出氢能将在未来国家能源体系中的重要地位,强调其在终端能源使用中的绿色低碳转型作用,并将氢能产业视为未来发展的战略性新兴产业。作为氢能应用的重要领域之一,氢燃料电池汽车不仅是减少碳排放的有效途径,还将在全球交通系统中带来深远变革。氢燃料电池汽车的广泛推广,不仅能够有效应对全球气候变化、促进能源清洁化,还将提升国家能源安全,减少对化石燃料的依赖,助力实现人与自然和谐共生。因此,深入开展氢燃料电池汽车领域的科学与技术研究,对于加速该领域的产业化进程具有重要的战略意义。

本研究选定氢燃料电池汽车领域为例进行研究。专利技术研究数据来源于 Incopat 专利数据库, 科学论文研究数据来源于 Web Of Science 学术数据库, 检索全球氢燃料电池汽车领域的相关专利和论文。综合国家知识产权局发布的《氢能产业技术分类与国际专利分类 IPC 对照及检索应用》与学者已有研究^[23], 将检索策略表达式确定为(((IPC-LOW=H01M4/86 OR IPC=H01M8*) OR TIABC=("hydrogen fuel cell vehicle" OR HFCV OR "fuel cell vehicle" OR "proton exchange membrane fuel cell vehicle" OR PEMFCV OR "solid oxide fuel cell vehicle" OR SOFCV OR "solid polymer fuel cell vehicle" OR SPFCV OR "alkaline fuel cell vehicle" OR AFCV OR "phosphoric acid fuel cell vehicle" OR PAFCV OR "molten carbonate fuel cell vehicle" OR MCFCV OR "polymer electrolyte fuel cell vehicle" OR PEFCV)) AND (TIABC=(氢燃料 OR 氢能 OR 氢气 OR H2 OR 液氢 OR 固体氢 OR hydrogen)) OR TIABC=(氢 OR 质子交换膜 OR 碱性 OR 磷酸 OR 熔融碳酸盐 OR 固体高分子型 OR 固体氧化物 OR 氢燃料电池 OR 燃料电池电动汽车 OR 氢动力汽车))(2N)(燃料电池 OR 汽车))AD=[20120101 TO 20211231], 检索日期为 2024 年 9 月 2 日, 检索范围为 2012 年 1 月 1 日—2021 年 12 月 31 日, 经扩展同族合并后得到专利数据单元(包含 IPC 分类号、摘要、优先权申请等信息)共计 32 874 条。论文数据单元(包含共引频次、发表时间、摘要等信息)共计 4 108 条, 作为研究数据集。最后, 编写 Python 代码对文本数据进行筛除空值、去停用词等预处理操作, 为主题建模分析奠定基础。

3.2 阶段二: BERTopic 主题建模

对预处理后的论文和专利摘要进行主题建模。实验设置阶段首先利用 all-MiniLM-L6-v2 模型生成文档嵌入, 并基于公式(4)与(5)进行 UMAP 降维, 随后利用 HDBSCAN 算法进行聚类分析。最终, 通过公式(6)计算出 C-TF-IDF 值, 以完成主题表示。由于篇幅限制, 仅抽取部分主题建模结果进行展示。表 1 列出了用于计算 2012—2013 年期间专利摘要的 BERTopic 主题模型的关键参数设置。参数数值的确定基于已有研究成果^[25], 并经过多次实验调试和优化得出。

表 1 BERTopic 主题模型关键参数

Tab.1 Key Parameters of the BERTopic Model

参数名称	参数内涵	参数值
n_neighbors	控制每个数据点局部邻域的大小	25
n_components	降维后的目标维度	70

metric	设定计算数据点间距离的度量标准	cosine
min_dist	设定数据点之间的最小距离	0.04
n_epochs	设定 UMAP 优化过程中的迭代次数	300
learning_rate	学习率	0.1
spread	数据点的扩散程度	0.8
min_cluster_size	定义聚类的最小簇大小	25
min_samples	定义每个点在簇中的最小密度阈值	3
allow_single_cluster	定义 HDBSCAN 是否允许生成单个聚类	False
prediction_data	是否允许新的数据点进行聚类预测	True
calculate_probabilities	计算每个文档属于不同主题的概率分布	True

基于上述预设参数的主题模型，挖掘出 2012—2013 年期间氢燃料电池汽车领域的 41 个专利研究主题（Topic0—Topic40）。专利文档的主题分布情况如图 2 所示，各主题所属文档呈现出明显的聚类趋势。此外，特征词权重变化趋势的分析显示，当特征词数量超过 6 个后，大多数主题的特征词权重趋于稳定，表明额外的特征词对主题区分的贡献度较小。权重变化趋势的具体结果如图 3 所示。设定主题特征词的提取数量为 6，调用 `topic_model.visualize_barchat()` 函数生成的部专利主题的特征词结果如图 4 所示。

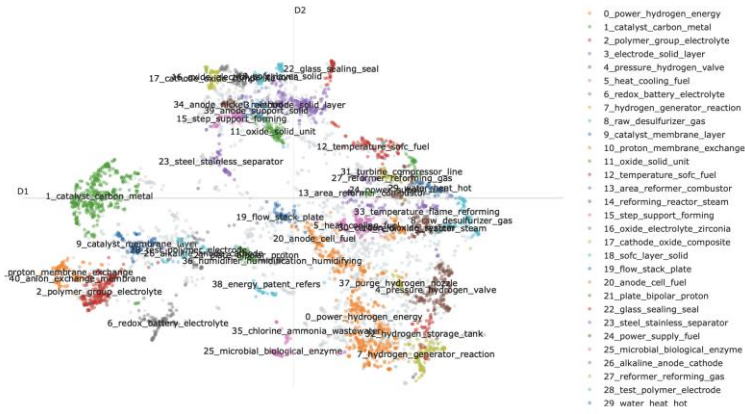


图 2 专利主题文档可视化分布

Fig.2 Visualization of Patent Theme Document Distribution

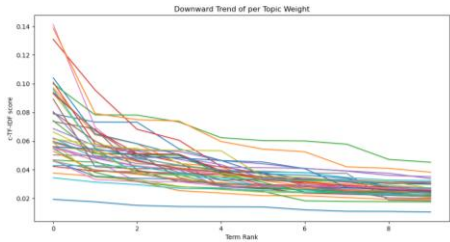


图 3 特征词权重下降趋势

Fig.3 Declining Trend of Feature Word Weights

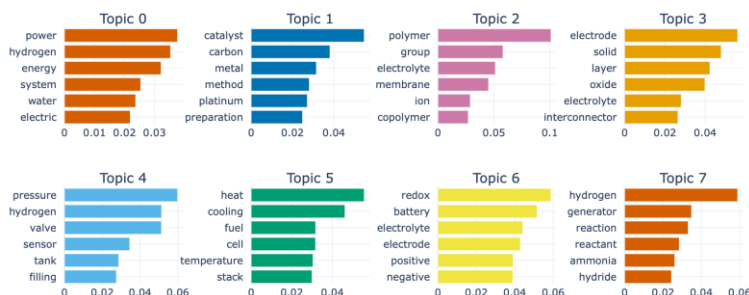


图 4 部分专利主题特征词提取结果

Fig.4 Extracted Results of Feature Words from Selected Patent Themes

调用 `topic_model.visualize_hierarchy()` 函数生成的 2012—2013 年期间专利主题的层次聚类结果如图 5 所示。结合层次聚类分析与专利主题特征词，可以将该时期的 41 个专利研究主题归纳为五大研究主题：膜技术与电解质材料应用、固体氧化物燃料电池材料与结构设计、燃料供应与能源转化管理技术、氢能生产与存储优化、电极材料创新与高效应用技术。

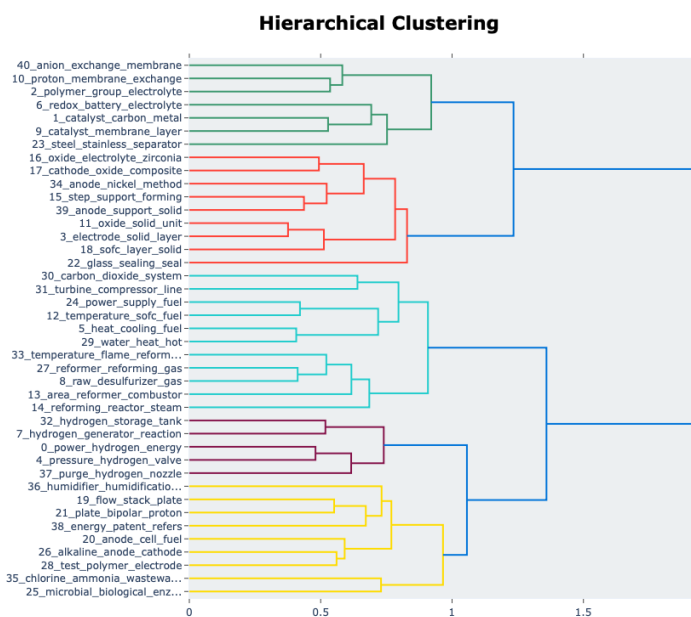


图 5 2012—2013 年期间专利主题层次聚类

Fig.5 Hierarchical Clustering of Patent Themes from 2012 to 2013

3.3 阶段三：主题关联和互动演化分析

3.3.1 主题关联分析

基于 BERTopic 主题建模结果，对各时间切片的主题进行关联分析。首先，根据层次聚类结果依次确定其他时期的论文与专利研究主题，并汇总（详见表 2）。其次，基于公式(7)、

（8）和（9），分别计算表 2 中归纳的论文与专利主题在相同及不同时间切片中的语义相似度，并根据计算结果确定关联强度。最后，将余弦相似度的数值以区间形式呈现（详见表 3），从而清晰地揭示论文与专利研究主题之间的关联程度。

表 2 各时间切片论文与专利研究主题

Tab.2 Research Themes of Papers and Patents in Different Time Slices		
时间区间	论文研究主题	专利研究主题
2012—2013 年	Paper#1（氢燃料电池汽车能效提升与排放控制）Paper#2（氢燃料电池能源管理与控制策略优化）Paper#3（氢燃料电池催化剂开发与电化性能优化）Paper#4（燃料电池膜技术改进与水管理优化）Paper#5（高压储氢与温度调控）	Patent#1（膜技术与电解质材料应用）Patent#2（固体氧化物燃料电池材料与结构设计）Patent#3（燃料供应与能源转化管理技术）Patent#4（氢能生产与存储优化）Patent#5（电极材料创新与高效应用技术）
	Paper#1（氢燃料电池系统与能源管理优化）Paper#2（低排放新能源汽车研究）Paper#3（燃料电池堆与膜材料技术改进）Paper#4（氢气储存与加注系统优化）Paper#5（燃料电池电极与催化性能优化）	Patent#1（电化学材料与燃料电池技术创新）Patent#2（固体氧化物燃料电池设计与材料创新）Patent#3（氢能生产、存储与系统集成优化）Patent#4（高压储氢与安全管理技术）Patent#5（新能源与混合动力汽车技术集成）
	Paper#1（氢燃料电池汽车能效提升与排放管理）Paper#2（燃料电池系统能量控制与优化策略）Paper#3（高效质子交换膜燃料电池研究）Paper#4（高压储氢与燃料管理技术优化）Paper#5（新能源与混合动力系统管理策略）	Patent#1（新能源与燃料电池在交通和无人系统中的应用技术）Patent#2（燃料电池能源转换与管理技术）Patent#3（固体氧化物燃料电池材料优化与结构设计）Patent#4（电解质膜创新与应用技术）Patent#5（新型催化剂材料的开发与应用）
2014—2015 年	Paper#1（混合动力系统能源优化与控制技术）Paper#2（氢燃料电池汽车排放控制与成本管理）Paper#3（质子交换膜燃料电池性能提升与耐久性优化）Paper#4（高压氢气储存与温度管理研究）	Patent#1（燃料电池材料与储能技术）Patent#2（聚合物电解质膜技术改进）Patent#3（氢能与可再生能源系统集成优化）Patent#4（氢气储存与加注系统技术优化）Patent#5（燃料电池控制与检测技术）Patent#6（燃料电池机械组件与结构优化）
	Paper#1（质子交换膜燃料电池性能优化）Paper#2（能源系统智能化控制与管理）Paper#3（氢燃料电池汽车能效提升与排放控制）Paper#4（氢气储存安全管理技术）Paper#5（混合动力系统能源管理与优化）Paper#6（可再生能源与氢能系统集成优化）	Patent#1（燃料电池核心材料与结构优化）Patent#2（新能源与化学转化技术集成）Patent#3（燃料电池控制与安全管理技术）Patent#4（车辆能源管理与无人系统集成）Patent#5（燃料电池测试与热管理技术）Patent#6（燃料电池堆叠结构优化设计）Patent#7（电动部件与机械结构设计改进）

表 3 “论文—专利”研究主题关联度

Tab.3 Association Degree of 'Paper-Patent' Research Themes

Patent \ Paper	2012—2013 年	2014—2015 年	2016—2017 年	2018—2019 年	2020—2021 年
2012—2013 年	0.42—0.82	0.38—0.79	0.35—0.80	0.41—0.79	0.37—0.80
2014—2015 年	0.50—0.85	0.45—0.86	0.43—0.86	0.46—0.86	0.44—0.85
2016—2017 年	0.46—0.77	0.40—0.78	0.34—0.78	0.46—0.76	0.38—0.77
2018—2019 年	0.37—0.78	0.37—0.82	0.32—0.79	0.40—0.77	0.36—0.78
2020—2021 年	0.38—0.80	0.35—0.76	0.31—0.79	0.39—0.74	0.39—0.75

通过分析表 3 中的余弦相似值，发现共有 17 对论文与专利主题的关联度较高（相似度大于 0.8），其中，2014 至 2015 年期间的专利技术研究主题 Patent#5 与多个时期的科学论文研究主题表现出显著的关联性。具体而言，Patent#5 与 2012—2013 年的 Paper#1 和 Paper#5，2014—2015 年的 Paper#2 和 Paper#4，2016—2017 年的 Paper#1 和 Paper#4，2018—2019 年的 Paper#2 和 Paper#4，以及 2020—2021 年的 Paper#3、Paper#4 和 Paper#6 之间均存在紧密联系。这种关联性凸显了氢燃料电池汽车技术的跨领域整合趋势，即 Patent#5（新能源与混合动力汽车技术集成）并非孤立发展，而是与氢燃料电池汽车能效提升与排放控制、高压储氢与温度调控、可再生能源与氢能系统集成优化等多个领域紧密协同。通过整合多领域的创新成果，氢燃料电池汽车技术不断优化系统效率和安全性，实现整体技术的持续提升。此外，Patent#5 与不同时期论文研究主题的关联性揭示了氢燃料电池汽车技术的累积性演化特征，即每个时期的研究基于前一阶段的成果，通过不断解决实际问题并优化系统，逐步推进技术进步。同时，这种高度关联性还反映了氢燃料电池汽车领域研究热点的长期持续性。无论是早期的排放控制、能效提升、还是后期的氢气存储与安全管理等核心问题，各阶段的研究均围绕这些关键挑战展开，显示出这些问题具有持久的研究价值和现实挑战性。

进一步分析表 3 发现，论文研究主题与专利研究主题之间的关联度呈现逐步下降的趋势。数据显示，在 2012—2015 年期间，论文与专利的关联度较高，随后随着时间的推移逐渐降低。这一趋势揭示了氢燃料电池汽车技术从基础研究向应用开发和产业化演进的自然过程。在技术发展的早期阶段，科学论文对专利创新具有较强的推动作用，专利与论文的关联度较高；然而，随着技术逐渐成熟，产业化需求开始主导技术创新，专利更依赖于现有技术的优化与整合，科学论文的影响力相应减弱。这一现象反映了氢燃料电池汽车技术演化过程中，应用创新逐渐占据主导地位的实际情况，并进一步揭示了基础研究与应用开发之间复杂而动态的互动关系。

3.3.2 主题互动演化分析

基于 2.2.2 节中的互动演化研究方法，通过绘制论文与专利主题的互动演化趋势图（如图 6 所示），揭示主题随时间变化过程中的分裂、吸收与消亡等演化规律。在图中，“Paper”和“Patent”分别代表论文与专利，括号中的第一个数字表示时间切片，第二个数字表示主题序号。例如，Patent（1,1）表示 2012—2013 年期间的第一个专利主题 Patent#1。通过分析各时间切片中论文与专利的主题数量及其在演化过程中的吸收、分裂、衰灭和新兴现象，可以揭示“论文—专利”互动演化的主要类型。

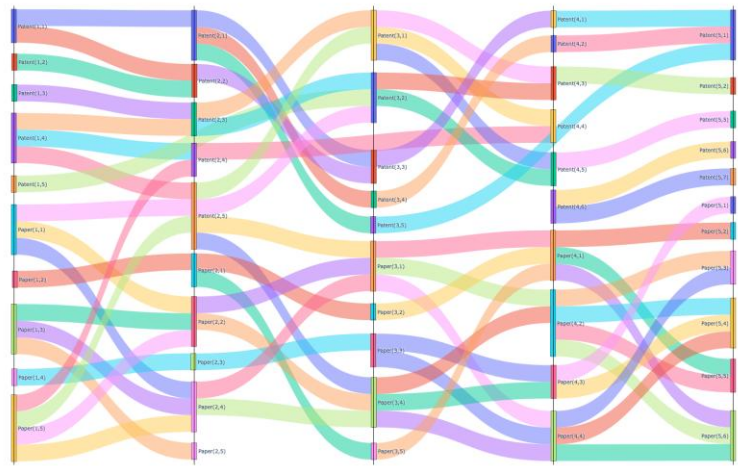


图 6 “论文—专利主题演化趋势

Fig.6 Evolution Trend of 'Paper-Patent' Themes

一是吸收式演化。此种演化模式下，论文与专利的研究主题从多个研究方向逐渐聚集吸收演变为新的研究主题，通过整合与优化前期科学与技术成果，从而实现主题的完善与提升。例如，氢气储存与加注技术的互动演化研究涉及 9 个研究主题，展现出吸收式演化的特征，通过整合前期研究逐步聚焦于更具综合性的创新领域。吸收式演化模式如图 7 所示。在 2012—2013 年，氢燃料电池汽车的科学研究主要集中于能效提升、排放控制、高压储氢和温度调控等关键技术环节，同时对氢燃料的供应、能源转化管理以及氢能源生产和存储优化等问题进行了广泛技术探索。特别是 350 bar 与 700 bar 的高压储氢技术及燃料电池堆的热管理成为提升氢燃料电池性能的核心研究方向。到了 2014—2015 年，科学研究与技术开发的关联度显著增强，氢能生产、存储与系统集成优化、高压储氢与安全管理技术，以及新能源与混合动力汽车技术集成广泛吸收并应用了前期的科研成果。在 2016—2017 年，新能源与燃料电池在交通和无人系统中的应用技术继续从前一阶段的研究成果中获益。2018—2019 年，氢气存储与加注技术经过前期研究积累与技术演进逐渐成熟，标志着氢燃料电池汽车产业化进程进入新阶段。尤其是 700 bar 高压储氢瓶和快速加注站技术的突破，大幅度提升了氢燃料电池汽车的实用性和经济性，丰田 Mirai 和现代 Nexo 等车型的推出，标志着氢燃料

电池汽车的商业化进程迈出了重要一步。

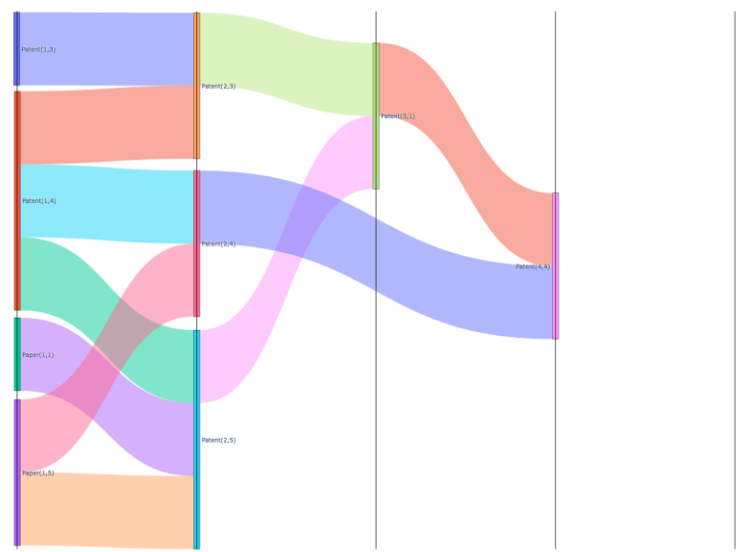


图 7 氢气存储与加注技术的主题演化

Fig.7 Theme Evolution of Hydrogen Storage and Refueling Technology

二是存续式演化。在此模式下，某一科学或技术研究主题在发展过程中保持其核心框架的稳定性，并基于此框架进行逐步的改进与优化。科学论文或专利技术的主题通常表现出较强的独立性与稳定性，反映了主题演化路径的延续性和递进性。例如，燃料电池核心材料与结构优化研究涉及 9 个主题均为专利主题，显示了该领域的存续式演化特点，即技术创新在核心框架内稳步推进。存续式演化模式如图 8 所示。在技术研究早期阶段（2012—2013 年），研发人员主要聚焦于电化学储能系统中的膜技术与电解质材料应用，尤其是在膜交换、电解质和催化剂层等方面的创新性材料和技术开发。该阶段的主要目标是提升这些核心组件的性能和效率。到 2014—2015 年，研究重心逐渐转向电化学材料与燃料电池技术的创新，特别是在质子交换膜、催化剂和电极材料的改进上，旨在提高燃料电池的整体性能和耐久性。同时，固体氧化物燃料电池的材料创新和结构设计也逐渐受到重视，标志着该领域技术的持续发展。至 2016—2017 年，燃料电池双极板设计与电解质膜技术取得了显著进展，推动了氢燃料电池汽车的发展。与此同时，新型催化剂材料的研发成为研究热点，尤其是在减少对铂等贵金属催化剂依赖的背景下，碳基催化剂与石墨烯复合材料等新材料的应用探索成为关键方向。进入 2018—2019 年，研发进一步聚焦于提升燃料电池关键材料的性能与储能。相比于此前阶段，这一时期的研发特别关注提高聚合物电解质膜技术的改进，尤其是在提升膜的导电性、耐化学性及其高温环境下的稳定性方面，以确保燃料电池的可靠性和长期运行性能。到 2020—2021 年，技术进步不仅体现在材料层面的持续优化，还反映出行业对可持续性和

经济性需求的深刻认知。通过减少贵金属（尤其是铂）的使用，燃料电池的制造成本得以显著降低，为氢燃料电池汽车未来的广泛应用奠定了基础。同时，双极板设计的创新和电解质膜材料的改进，为燃料电池在高效性、可靠性和使用寿命方面提供了坚实的技术。这些技术进展不仅推动了氢燃料电池汽车产业的可持续发展，更为实现能源转型和构建低碳未来提供了关键的技术基础。

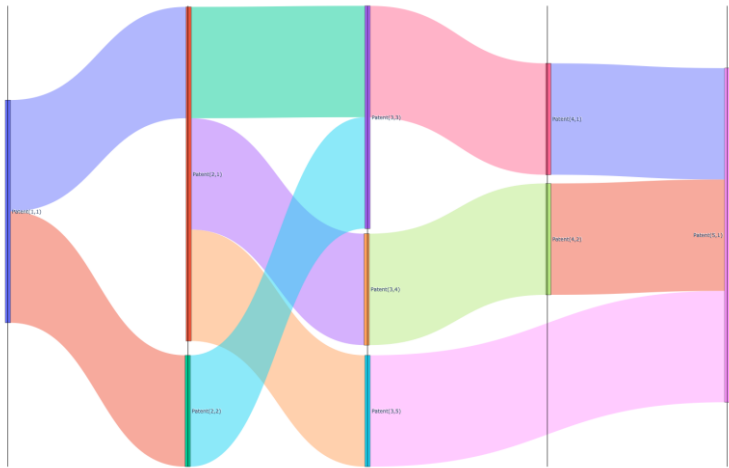


图 8 燃料电池材料与结构设计主题演化

Fig.8 Theme Evolution of Fuel Cell Materials and Structural Design

三是创设式演化。这种演化模式以创新和分化为核心，推动新的研究主题不断涌现，进而促进科学与技术的多元化发展。新主题不仅继承了现有的研究基础，还通过创新拓展了科学和技术的前沿。例如，氢燃料电池汽车的能源效率与排放控制方面的研究涵盖 23 个主题，突显了通过持续创新和分化拓展研究领域的创设式演化特征。创设式演化模式如图 9 所示。在 2012—2013 年，科学研究的重点集中在提升氢燃料电池汽车的能源效率与排放控制，尤其是通过改进燃料电池系统的能效来减少碳排放。到 2014—2015 年，低排放新能源车辆研究、氢气存储与加注技术成为学术界的研究热点。尤其在高压储氢技术和加注系统优化方面，取得了显著进展。此外，新能源与混合动力汽车技术逐渐融合，研究人员开始探讨太阳能、甲醇重整和电动模块与燃料电池混合动力汽车的整合，以提升整体系统的可持续性。在 2016—2017 年，科学研究的重心转向减少车辆碳排放及提高氢燃料电池的利用效率。高压存氢技术及燃料管理的优化，特别是在不同温度和压力条件下的调控，成为核心课题。与此同时，燃料电池在新型交通工具及无人系统中的应用日益增多，研究领域进一步扩展至氢气生成、微生物电解反应及燃料气体发电等技术，尤其是潜艇等特殊领域的应用也开始受到更多关注。到 2018—2019 年，科学研究的重点转向了混合动力系统的能源优化与控制，进一步降低氢燃料电池汽车的排放并优化成本。质子交换膜燃料电池的性能提升和耐久性优化成

为关键研究方向，尤其是在高压氢气储存和温度管理技术方面取得了显著进展。同时，技术研究还聚焦于氢能与可再生能源系统的集成优化、氢气储存与加注技术的改进，以及燃料电池控制与检测系统的开发。至 2020—2021 年，质子交换膜燃料电池的性能优化、智能控制与能源管理，以及氢燃料电池汽车的能效提升与排放控制成为研究重点。高压氢气储存的安全性及温控问题依然备受关注，混合动力系统的能源管理也成为焦点。技术研究进一步深化了新能源与化学转化技术的集成，特别是在燃料电池的热管理及与可再生能源系统的协同应用方面取得显著进展。通过这些阶段性的发展，氢燃料电池技术在能效、储能安全性以及与可再生能源整合方面取得了显著成果，推动了清洁交通技术的快速发展。

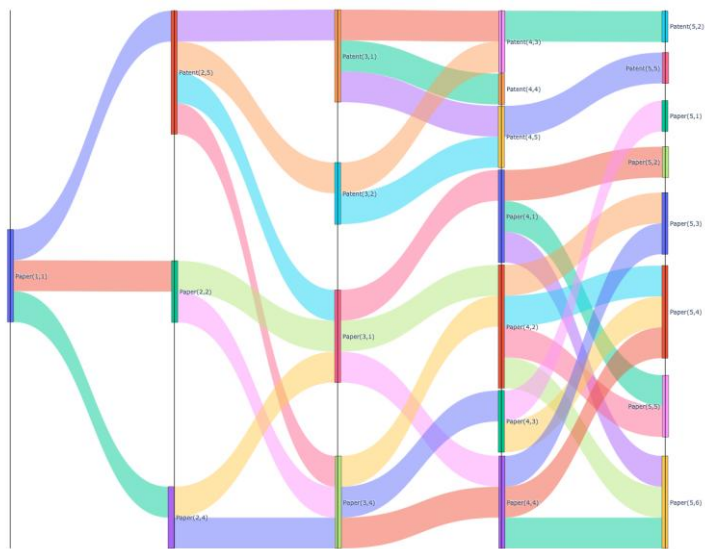


图 9 燃料电池能源效率与排放控制的主题演化

Fig.9 Theme Evolution of Fuel Cell Energy Efficiency and Emission Control

3.3.3 主题关联与互动演化的机制解析

论文与专利作为科学研究与技术开发的重要载体，其主题关联与互动演化揭示了知识流动与技术创新的复杂机制。科技创新通常呈现两种演化模式：渐进式创新与颠覆式创新^[26]。渐进式创新依托知识积累与经验整合，推动技术的持续优化与完善。Z. Griliches^[27]指出，科技进步依赖于早期研究的积累，通过整合已有的知识与经验，持续促进创新。颠覆式创新则通过引入全新科学技术打破现有格局，实现跨越式发展，并重新定义科技演进路径。C.M. Christensen^[28]指出，新兴科技起通常源于边缘市场，随着成熟度提升，最终有可能颠覆现有市场格局。基于上述创新理论的视角，进一步剖析“论文—专利”在科技创新中的驱动作用。

吸收式演化通过整合前期研究成果与知识外溢效应，推动技术的持续改进，是渐进式创新的典型特征。其核心在于论文与专利的双向互动：论文为专利开发提供理论支撑，专利的

工程实践反哺基础研究。例如，氢气储存与加注技术从 350 bar 提升至 700 bar 的突破，正是论文揭示材料特性与专利工程实践结合的成果。这一机制不仅推动了单领域技术改进，还通过跨领域的知识整合提升了技术创新效率。吸收式演化在知识积累与流动中形成渐进性改进，为多学科协同与技术整合提供了理论支持，体现了渐进式创新从细化知识到优化路径的内在逻辑。

存续式演化聚焦于现有科学或技术框架的局部优化，是渐进式创新的延续形式。通过论文揭示细节问题或专利提出解决方案，该机制提升了技术的市场竞争力与可行性。例如，燃料电池催化剂与电极材料的持续优化，显著提升了燃料电池的效率与耐久性，同时增强了技术的稳定性与经济性。存续式演化在核心框架稳定的基础上不断优化细节，为实验室技术向产业化转化提供了坚实支撑，展现了渐进式创新从细节优化到规模化应用的路径逻辑。

创设式演化通过探索全新科学技术路径，为颠覆式创新提供动力。与吸收式和存续式演化不同，创设式演化注重从根本上突破现有技术框架。例如，在燃料电池能源效率与排放控制领域，相关论文提出了新型燃料电池架构的理论模型，专利则通过多元化技术实践，分化并催生出新的研究主题。这一演化模式重塑了技术路径，不仅推动了技术跨越式发展，还拓展了行业边界，为未来技术格局的创新注入新动能。

4 结论和建议

4.1 结论

本文围绕氢燃料电池汽车领域的论文—专利主题关联与互动演化规律展开研究，主要有以下发现和结论：一是分析不同时间切片间论文与专利主题的语义相似度，发现 17 对强关联性主题，表明科学与技术的协同推动了技术领域的累积演化；二是论文与专利的关联度随时间减弱，反映氢燃料电池汽车从基础研究向技术应用的转型过程，研究重心逐步由理论探索转向实践开发，应用创新成为主要驱动力；三是论文与专利主题在横向时间切片上的关联度变化揭示了吸收式、存续式和创设式演化三种互动演化机制；四是从科学技术创新视角剖析主题关联与互动演化机制，加深了对技术演化多样性及创新驱动因素的理解。

4.2 理论贡献

本研究选定特定技术领域，依托 BERTopic 主题建模和语义相似度等方法，构建了以“研究领域选定→主题模型设计→实证研究分析”为主线的系统分析框架。该框架能够快速识别不同知识来源间的关联性，揭示科学研究与技术应用的互动逻辑，并为新兴技术的监测、评估和发展规划提供理论与实践支持。此外，通过模块化设计，提高了框架的通用性与迁移价值，为其他行业及多学科交叉领域的类似研究提供了重要参考路径。

4.3 启示

吸收式、存续式和创设式演化机制推动了氢燃料电池汽车技术从局部优化到全面突破的多层次演化，不仅揭示了论文与专利之间的互动机制，还为跨领域技术融合与创新战略的优化提供了重要借鉴。本研究对科技管理策略的制定具有以下启示：一是吸收式演化通过整合前期成果推动主题优化，表明管理应加强知识流动与机制整合，例如促进科研机构与企业合作，支持技术的深度开发；二是存续式演化强调技术框架的稳定性与逐步优化，揭示管理需保持技术发展的延续性与递进性，例如制定长期规划以确保关键技术持续投资、通过标准化提升技术适用性和竞争力；三是创设式演化以创新与分化为核心，推动新主题涌现，指明管理需加大对颠覆性创新的支持，例如设立专项基金资助前沿研究，促进跨学科协作与知识融合，加强技术预测与潜在技术识别，为新技术路径的形成提供科学依据和资金支持。

4.4 不足和展望

但本研究仍有局限性：分析方法主要针对氢燃料电池汽车领域，结果精度依赖于模型质量，不同行业技术特性对参数设置提出更高要求；BERTopic 的无监督特性限制了传统可靠性检验方法的适用性。在未来研究中，笔者将进一步优化算法细节，提升框架的跨领域迁移效果，并结合定量与定性研究，系统探讨如何通过政策激励、资源配置等手段，协调三种互动演化机制，以期为科技成果转化提供多元化路径选择。

参考文献：

- [1] MOED H F, GLÄNZEL W, Schmoch U. Handbook of Quantitative Science and Technology Research: The Use of Publications and Patent Statistics in Studies of S&T Systems[M]. Dordrecht: Kluwer Academic Publisher, 2004.
- [2] AHMADPOOR M, JONES B F. The dual frontier: patented inventions and prior scientific advance[J]. Science, 2017, 357(6351): 583-587.
- [3] 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.
- [4] 宁子晨, 魏来. 专利主体视角下专利文献与学术论文关联关系发现研究——以“数据挖掘”主题为例[J]. 图书情报工作, 2020, 64(12): 106-117.
- [5] CHEN L X. Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations[J]. Journal of Informetrics, 2017, 11(1): 63-79.
- [6] 张凯, 吕璐成, 韩涛, 等. “论文—专利”关联视角下的新兴技术识别研究[J]. 情报理论与实践, 2024, 47(9): 183-191.

- [7] 唐露源, 谢士尧, 徐源. 知识演化视角下论文与专利的热点技术方法对比分析——以人工智能自然语言处理领域为例[J]. 科技管理研究, 2024, 44(10): 153-160.
- [8] FERREIRA R B, PARREIRA M R, NABOUT J C. Is there concordance between Science and Technology in Natural Science? Mapping the relationship among number of papers and patents from research on Cerrado plants[J]. World Patent Information, 2022, 69: 102108.
- [9] 冉从敬, 田文芳, 贾志轩. 基于混合方法的“科学论文-专利技术”关联关系模型构建——以生物医药领域为例 [J/OL]. 情报科学, 1-18 [2024-09-20].
<http://kns.cnki.net/kcms/detail/22.1264.g2.20240506.1927.030.html>.
- [10] PORA U, GERDSRI N, THAWESAENGKULTHAI N, et al. Data-Driven roadmapping (DDRM): Approach and Case Demonstration[J]. IEEE Transactions on Engineering Management, 2020, 69(1): 209-227.
- [11] GLASER B, STRAUSS A. Discovery of Grounded Theory: Strategies for Qualitative Research[M]. New York: Routledge, 2017.
- [12] KRIPPENDORFF K. Content Analysis: An Introduction to Its Methodology[M]. California: SAGE Publications, 2018.
- [13] SALTON G, WONG A, YANG C S. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.
- [14] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [15] Zhang H, Daim T, Zhang Y P. Integrating patent analysis into technology roadmapping: A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain[J]. Technological Forecasting and Social Change, 2021, 167: 120729.
- [16] 寇园园, 陈会英, 徐华杰, 等. 海外跨国公司在华人工智能专利布局及竞争态势研究[J]. 情报杂志, 2022, 41(9): 48-54.
- [17] 胡泽文, 王梦雅, 韩雅蓉. 基于 LDA2Vec-BERT 的新兴技术主题多维指标识别与演化分析研究——以颠覆性技术领域: 区块链为例[J]. 现代情报, 2024, 44(9): 42-58.
- [18] 席笑文, 郭颖, 宋欣娜, 等. 基于 word2vec 与 LDA 主题模型的技术相似性可视化研究[J]. 情报学报, 2021, 40(9): 974-983.
- [19] 刘春丽, 臧东宇, 陈爽. 科学—技术—产业关联测度与主题演化规律研究——以生物医药领域为例[J]. 图书情报工作, 2024, 68(14): 95-116.
- [20] 张玲, 恽诚涛, 尹思力, 等. 我国科研诚信政策与文献主题演化对比分析[J]. 现代情报, 2023, 43(6):

108-120.

- [21] 苏婉, 于森, 嵇传阳. 我国数据要素政策主题演化与识别分析[J]. 图书情报工作, 2024, 68(16): 90-103.
- [22] 马海群, 崔文波, 张涛. 我国数据安全政策文本主题挖掘及其演化分析[J]. 现代情报, 2024, 44(8): 28-38.
- [23] 慎金花, 王薇, 张更平, 等. 基于动态主题网络的新兴技术主题识别——以氢燃料电池领域为例[J]. 情报杂志, 2024, 43(9): 92-100.
- [24] 李昌, 吴红, 伊惠芳, 等. 基于改进 LDA 主题模型的中日美氢能产业链技术布局研究[J]. 情报杂志, 2019, 38(7): 78-84+110.
- [25] 杨思洛, 于永浩. 基于 BERTopic 模型的国内信息资源管理研究主题挖掘与演化分析[J/OL]. 情报科学, 1-18 [2024-09-21]. <http://kns.cnki.net/kcms/detail/22.1264.G2.20240506.1701.012.html>.
- [26] 李玉花, 李丹丹. 企业数字化程度促进双元创新了吗?——基于知识多样性视角分析[J]. 商业研究, 2024, (3): 125-133.
- [27] GRILICHES Z. The Search for R&D Spillovers[J]. The Scandinavian Journal of Economics, 1992, 94: 29-47.
- [28] CHRISTENSEN C M. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail[M]. Boston: Harvard Business Review Press, 2015.

The Correlation and Interactive Evolution of 'Papers-Patents' Themes: A Case Study of Hydrogen Fuel Cell Vehicles

Zhang Junrong Qiao Jianzhi

(School of Law, South-Central Minzu University, Wuhan 430074)

Abstract: The interaction between scientific papers and patents is intricate and close, yet the specific relationship and evolutionary patterns remain unclear. A thorough analysis of the interaction mechanisms between papers and patents helps to reveal the underlying patterns of knowledge flow and accurately predict technology transformation pathways. This study extracts paper and patent data units from WOS and Incopat, utilizing the BERTopic model to identify themes from both papers and patents across different time slices. Based on semantic similarity, it then analyzes the correlation and interaction trends between papers and patents. The results indicate that certain research topics in hydrogen fuel cell vehicles exhibit a strong trend of collaborative innovation. The correlation between papers and patents weakens over time, reflecting the transition of technology from basic research to applied development. Furthermore, three evolutionary mechanisms are identified between papers and

patents: absorptive evolution, persistent evolution, and creative evolution. Through an innovation theory lens, these evolutionary mechanisms are analyzed to provide theoretical guidance for future technological changes and innovation strategies.

Key Words: Thematic Correlation; Interactive Evolution; Bertopic Model; Theme Similarity; Hydrogen Fuel Cell Vehicles