基于大语言模型的废弃技术价值再创造机会识别方法设计与前提性验证

李天柱，肖 静

（贵州财经大学工商管理学院，贵阳 550025）

摘要：废弃技术在体量上远超正在使用的技术，其中蕴藏着重要的价值再创造潜力。然而，由于废弃技术已脱离原有的知识网络，如何系统识别其在新场景中的应用可能性，仍缺乏可操作的方法论支持。提出一种基于大语言模型（LLM）的废弃技术价值再创造机会识别方法，构建“LLM 驱动的技术特征提取与语义扩展→知识搜索空间构建→基于语义嵌入模型的广度搜索→ LLM 驱动的证据挖掘与语义推理”4 阶段方法框架。工作重点在于方法框架的设计逻辑及其前提性环节的可行性验证。以废弃技术 Enzastaurin 为例，对前两个前提性环节开展概念验证。结果表明，LLM 能够从专利文本中提取废弃技术的多维特征，完成概念抽象、语义扩展与知识搜索空间界定，从而验证了该方法前两个前提性环节的可行性，为该方法框架的成立提供了前提性支撑。

关键词：废弃技术；价值再创造；机会识别；大语言模型（LLM）；专利文本分析

中图分类号：F062.3 **文献标志码**：A **文章编号**：1671－1807(2026)17－0019－10

技术创新是经济发展的内生动力，但并非所有技术都能顺利实现从研发到市场的价值转化。事实上，大量技术在其演化过程中失去了来自社会的资源和能力的支持，被弃之不用，沦为“废弃技术”。废弃技术形成的内在原因是技术向价值转化的进程出现了中断或终止，因此不具有或不再具有满足创新主体需要的能力。人类技术活动的规律决定了，历史上积累起来的废弃技术在数量上远超正在获得广泛使用的技术，尤其是当前世界主要国家普遍加大科技投入、新旧技术更替的节奏越来越快，废弃技术的体量呈现加速增长的长期趋势。

废弃技术是否完全丧失了价值？产业创新实践早已给出了否定的回答。例如，在生物制药领域，Thalomid 曾因严重的致畸副作用在全球范围内被禁用，后来却被重新开发为治疗多发性骨髓瘤的重磅药物。光伏发电、电动汽车等当今的重要技术，追本溯源都是在被长期搁置或抛弃的技术基础上重新发展起来的[1]。这些案例表明，为数不少的废弃技术蕴藏着不容忽视的应用潜力，考虑到废弃技术的巨大体量，即便其中只有很小一部分能够再次进入价值创造的过程，其经济意义和社会意义也应该是极为显著的。然而，废弃技术的价值再创造面临前置难题，即如何识别基于废弃技术的创新机会？

从知识重组的视角来看，技术创新的本质是对已有要素的重新组合[2]。废弃技术是由于各种原因脱离了已有知识组合的技术，不能在原有的知识组合中发挥作用，或者从未被纳入某种有效的知识组合之中。因此，废弃技术价值再创造机会识别的实质，是为废弃技术寻找一个新的知识重组机会。这要求不仅能够准确把握废弃技术的机理特征和功能特征，还要能在广泛的知识空间中发现废弃技术与新应用场景之间的潜在关联。然而，目前缺乏以废弃技术价值再创造为主题的方法论，虽然药物重定位领域已出现计算驱动的特征提取技术[3]，专利计量分析领域也发展出专利地图[4]、SAO（subject-action-object，主体-动作-客体）网络分析[5]、融合大语言模型与语义嵌入的专利语义扩展智能检索策略[6]等技术创新机会识别手段，但尚未被整合为面向废弃技术价值再创造机会识别的系统方法。

大语言模型（large language model，LLM）的快速发展为弥合上述方法缺口提供了可能。LLM 在

收稿日期：2026-04-27

基金项目：辽宁省社会科学规划基金重点项目（L19AGL008）

作者简介：李天柱（1975—），男，辽宁沈阳人，博士，教授，研究方向为技术创新管理、科技管理；通信作者肖静（1985—），女，贵州贵阳人，博士研究生，研究方向为技术创新管理、科技管理。

预训练过程中编码了海量的跨领域知识，具备强大的语义理解、概念抽象和跨领域推理等能力，有望为废弃技术价值再创造机会识别提供有效的工具支持。基于上述认识，本文的工作分为两层：第1层是方法框架的设计，提出“特征提取与语义扩展→搜索空间构建→广度搜索→证据挖掘与语义推理”的4阶段递进式框架；第2层是前提性概念验证，以废弃技术 Enzastaurin 为例对前两个阶段的可行性进行检验。之所以将概念验证聚焦于前两个阶段，是因为四阶段框架在逻辑上呈递进依赖关系，后两个阶段的有效性以前两个阶段的产出为输入条件，因此对前两个前提性环节的可行性验证，构成对整个框架成立性的必要支撑。基于这一定位，试图回应一个具有方法论意义的问题：能否为原本高度依赖偶然性的废弃技术价值再创造机会发现过程提供系统化的方法框架，并为其成立性提供前提性的经验支撑。

1 方法提出的理论基础

1.1 废弃技术价值再创造机会识别的性质

废弃技术价值再创造的理论线索来自关于技术创新失败项目挽救的研究。创新失败并不意味着技术价值的终结，对失败项目的分析和合理处置能够产生新的创新[7]，传统意义上“失败的创新”却可能蕴藏着很大的潜力[8]。在产业实践上，生物制药发展历史中的很多明星药物都是从废弃药物中再次发现价值的结果[9]，创新失败药物的挽救和废弃药物的再利用，其实质就是对废弃技术的价值再创造[10]。这一现象并不局限于单一产业，在更广泛的技术领域同样存在，具有普适性的理论研究意义。

从理论基础看，废弃技术价值再创造的工作重点不是使技术本身发生显著变化，而是发现技术的新应用领域或新功能，这与扩展适应（Exaptation）理论高度契合[11-12]。扩展适应有“弗兰克林型”和“弥尔顿型”两种基本类型[13]，前者是技术在原有功能之上发展出新功能，后者是技术转移到新领域后才展现出价值。这意味着废弃技术的价值可能通过截然不同的路径涌现，机会识别的搜索空间极为广阔且缺乏先验方向。事实上，扩展适应驱动的创新通常是通过 Serendipity 实现的[14]，这既揭示了废弃技术价值再创造的实现路径，也暗示了其机会识别在很大程度上要靠“碰运气”。如何降低这种偶然性，以系统的方法支撑废弃技术价值再创造的机会识别，是亟须回答的方法论问题。

1.2 知识重组与废弃技术价值再创造机会识别

如前所述，废弃技术价值再创造的机会来自扩展适应，知识重组则为此提供了有益的视角。知识重组理论将技术创新视为对已有知识要素的重新组合过程[2,15]，核心逻辑是要素之间的交互作用将产生新的意义和功能[16]。研究表明，发明的新颖性和不确定性根植于对新要素和新组合的搜索过程[17]，而知识要素的多样性和异质性是产生突破性创新的关键条件[18]。然而，现有关于知识重组的研究主要关注企业如何在自身知识基础上搜索和整合新要素，较少涉及已脱离原有应用情境的技术如何重新进入重组过程，这正是废弃技术价值再创造机会识别所面临的独特挑战。

由于废弃技术已经脱离了原有的知识组合情境，其“技术-价值”转化链条已经断裂，因此废弃技术价值再创造机会识别的重点不是在既有的知识组合框架内进行改进，而是要在新的知识空间中发现知识重组的可能性。这一任务对机会识别提出了3方面要求：①对废弃技术的作用机理和功能特征形成准确理解，而非停留在表面描述；②能够从具体的技术特征中抽象出更高层级的功能概念，以突破原有应用领域的认知边界；③能够发现废弃技术与其他领域知识之间的潜在关联，从而识别出新的知识重组机会。这3方面要求共同指向一种兼具深度语义理解、概念抽象和跨领域知识关联能力的方法工具。

1.3 LLM 在知识创造中的应用

在 MOOSE-Chem 框架中的研究表明，LLM 能够依据研究背景知识从大量文献中检索出与背景看似无关但实际上能够激发新假说的“灵感文献”，并将其与背景知识关联，重新发现已有的化学假说的核心创新[19]。这一发现的关键启发在于，LLM 能够识别人类尚未认识到的潜在知识关联。这与废弃技术价值再创造机会识别的任务高度同构，二者都需要在广阔的知识空间中发现隐藏的跨领域关联。此外，Boiko 等[20]开发的 Coscientist 系统展示了 LLM 具备文本理解能力的同时，能进行多步推理并自主设计实验方案，这表明 LLM 具备从技术描述中提取深层语义、进行抽象推理的能力，而非仅限于关键词层面的浅层匹配。Zhang 等[21]在对 LLM 科学应用的系统评述中进一步指出，LLM 正在从辅助性的文本工具发展为能够跨越领域术语壁垒、整合多学科知识的研究工具。

鉴于此，LLM 及其语义嵌入技术为废弃技术价

值再创造的机会识别提供了方法论层面的突破性工具选择,有望将原本依赖 Serendipity 的偶然性发现过程转变为系统性、可操作的识别方法。具体来说:①针对深度语义理解的要求,LLM 在大规模科技文献上的预训练使其能够理解技术的作用机理和功能特征,而不仅是匹配关键词[22];②针对概念抽象的要求,LLM 的推理能力使其能够从具体技术描述中提炼出更高层级的功能概念,为突破原有应用领域的认知边界提供了条件;③针对跨领域知识关联的要求,基于 Transformer 架构的语义嵌入模型能够将不同领域的技术描述映射到统一的语义空间中,通过向量相似度计算发现跨领域的潜在关联[6,23],恰好对应废弃技术知识重组中最为关键的跨领域连接需求。

2 基于 LLM 的废弃技术价值再创造机会识别方法设计

2.1 方法框架的设计逻辑

如前所述,从知识重组的视角来看,废弃技术是脱离了原有知识组合的技术要素,其价值再创造的关键是寻找新的知识重组可能性。这一认识构成了本文方法设计的理论起点。具体而言,废弃技术价值再创造机会识别的任务可以分解为 4 个递进的子问题:①废弃技术的哪些特征构成了搜索知识重组可能性的依据?②在什么知识范围内搜索知识重组的可能性?③如何在广泛的知识空间中发现与废弃技术存在潜在关联的知识要素?④被发现的关联是否真正指向有价值的知识重组机会?

针对上述 4 个子问题,可以根据“LLM 驱动的技术特征提取与语义扩展→知识搜索空间构建→基于嵌入模型的广度搜索→LLM 驱动的证据挖掘与语义推理”的 4 阶段递进式方法展开(图 1)。该方法框架的核心理念在于,将废弃技术的价值再创造机会识别视为基于 LLM 的语义理解与推理,而非仅通过词汇表面匹配发现已有概念体系内的显性关联,LLM 能够理解技术文本的深层含义,进行概念层次的抽象与泛化,并在跨领域的知识空间中发现非显而易见的潜在联系,从而拓展机会识别的搜索边界。

(1)在 LLM 驱动的技术特征提取与语义扩展阶段,主要解决“搜索什么”的问题。技术手段是借助 LLM 从废弃技术的核心专利文本中提取多维度的技术特征,并进行概念抽象和语义扩展,形成多层次的搜索特征集。

(2)在知识搜索空间构建阶段,主要解决“在哪里搜索”的问题。技术手段是依据第 1 阶段对废弃技术特征的理解,构建后续搜索所需的知识搜索空间。

(3)在基于嵌入模型的广度搜索阶段,主要解决“如何搜索”的问题。技术手段是利用语义嵌入模型将搜索特征集与知识搜索空间中的专利文本进行语义匹配,识别与废弃技术存在深层语义关联的候选知识要素。

(4)在 LLM 驱动的证据挖掘与语义推理阶段,重点解决“是否构成知识重组机会”的问题。技术手段是通过 LLM 对候选知识要素进行语义推理,研判其是否实质性地支持废弃技术的价值再创造,从而完成从“潜在关联”到“重组机会”的最终判断。

2.2 方法的操作过程

2.2.1 LLM 驱动的技术特征提取与语义扩展

针对“搜索什么”的任务功能,从废弃技术的核心专利文本中提取构成搜索依据的技术特征,并通过语义扩展拓宽搜索的概念边界。具体操作分为 3 个步骤:第 1 步是特征提取,将废弃技术的核心专利全文输入 LLM,提取物理特征(如化学结构、分子靶点)和功能特征(如作用机制、靶向通路、涉及的生物学过程、治疗策略);第 2 步是语义扩展,LLM 对提取的特征进行多层次的概念抽象与泛化,生成

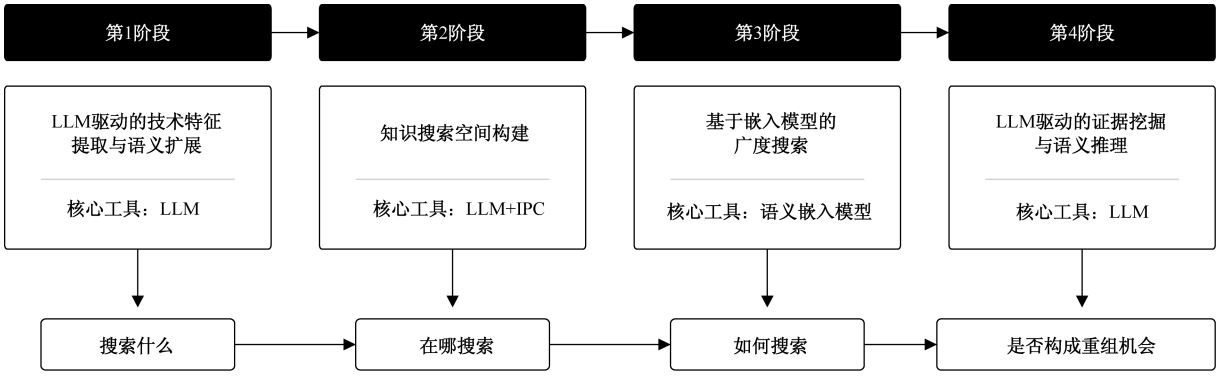


图 1 基于 LLM 的废弃技术价值再创造机会识别方法框架

同义概念、近义概念和上位概念，从而将搜索空间从具体技术描述拓展到更一般性的功能和机理层面；第 3 步是术语适配，面向目标专利库的语言环境，对第 2 步扩展后的特征集进行表述调整，生成可直接用于后续语义搜索的多层次特征集。

LLM 在本阶段承担核心执行角色，其在预训练中编码的跨领域知识使其能够系统性地沿概念层级进行抽象，将搜索依据从具体的技术术语拓展到更一般性的功能和机理描述，为后续跨领域的知识重组搜索提供更广的语义覆盖。

2.2.2 知识搜索空间构建

针对“在什么知识范围内搜索知识重组可能性”的任务功能，依据第 1 阶段对废弃技术特征的理解，构建后续搜索所需的知识搜索空间，LLM 的跨领域知识储备有助于识别出研究者可能忽略的关联领域，拓展搜索空间的覆盖范围。选择专利文献作为知识搜索空间的载体，这是因为专利文本对技术方案的作用机理、功能特征和应用场景有系统化的规范描述，是技术知识密集的文本形式；同时，专利体系自带的 IPC（International Patent Classification）分类号为搜索空间的领域界定提供了标准化的筛选工具。知识搜索空间的构建包含两个步骤：第 1 步是类型过滤，仅保留发明专利，以确保搜索空间中的技术方案具有足够的新颖性和技术含量；第 2 步是领域筛选，由 LLM 依据第 1 阶段提取的技术特征，结合 IPC 国际专利分类号，推荐与废弃技术的作用机理和功能特征可能相关的分类号范围，据此从专利库中提取候选专利集。

2.2.3 基于语义嵌入模型的广度搜索

针对“怎么搜索知识重组可能性”的任务功能，在知识搜索空间中识别与废弃技术存在深层语义关联的候选知识要素。知识重组的可能性往往隐藏在表述不同但机理相关的技术之间，因此搜索需要在语义层面而非词汇层面进行。基于嵌入模型的广度搜索，采用 BGE（BAAI general embedding）等语义嵌入模型，将第 1 阶段产出的多层次特征集与搜索空间中的专利文本分别映射为语义向量，通过余弦相似度计算识别语义关联度高的候选专利，按相似度降序筛选形成候选集。语义嵌入模型能够捕捉文本间深层的语义关系，使表述不同但含义相近的技术描述在向量空间中彼此接近，从而发现仅凭词汇匹配无法触及的潜在关联。

本阶段的语义嵌入模型与第 1 阶段的 LLM 构成协作关系，LLM 负责理解废弃技术并扩展其语义表征，语义嵌入模型负责在广泛的知识空间中执行高效的语义匹配与筛选。

2.2.4 LLM 驱动的证据挖掘与语义推理

针对“是否构成知识重组机会”的任务功能，对第 3 阶段发现的潜在语义关联进行深度研判，判断其是否实质性地指向有价值的知识重组机会。这是因为，语义相似度高意味着候选专利与废弃技术在特征层面存在关联，但关联本身不等于重组机会，需要从技术机理层面做进一步的判断。LLM 在理解候选专利技术内容的基础上进行语义推理，识别那些在文本表述中并不直观，但在机理层面与废弃技术存在实质性关联的证据，汇总形成机会识别的研判结果。证据挖掘与语义推理的具体操作是，将候选专利的关键段落输入 LLM，通过结构化提示词引导其进行研判：该专利中的技术知识是否与废弃技术构成有意义的知识重组可能？具体的关联机理是什么？关联的强度如何？

LLM 在本阶段承担核心执行角色。但与第 1 阶段 LLM 面向废弃技术自身进行“自我理解”不同，本阶段 LLM 面向废弃技术与候选知识要素之间的关系进行“关系研判”，完成从“潜在关联”到“重组机会”的最终判断。

3 概念验证：以 Enzastaurin 为例

本节并非对 4 阶段方法框架进行全链条实证检验，而是围绕其前提性环节开展概念验证。原因在于，研究提出的 4 阶段方法具有递进依赖关系，第 3、第 4 阶段的广度搜索与机会研判均以前两个阶段所生成的技术特征表征和知识搜索空间为输入。若前两个阶段无法成立，则后续阶段将失去运作基础。因此，首先检验 LLM 能否准确理解废弃技术的多维技术特征，并据此完成概念抽象、语义扩展与搜索空间界定。换言之，概念验证的目标，不是完成从候选技术筛选到重组机会判定的全流程实证，而是为方法框架提供前提性支撑，检验其得以成立的关键前提是否成立。鉴于语义嵌入模型在专利语义检索中的应用已具有较明确的方法基础[6,23]，同时 LLM 在多步推理、潜在关联识别和跨领域知识整合方面的能力也已得到相关研究支持[19-21]，因此将第 3、第 4 阶段作为后续研究继续展开。

3.1 案例背景

Enzastaurin 是礼来公司开发的 PKCβ 选择性抑制剂，曾在弥漫性大 B 细胞淋巴瘤、胶质母细胞瘤等多个适应证上开展临床试验。2012 年 III 期临

床试验失败后，礼来停止了对 Enzastaurin 的开发，Enzastaurin 及其相关专利技术由此成为典型的废弃技术。此后，中国索元生物收购了 Enzastaurin 的全球权益，通过发现新的生物标志物 DGM1 实现了基于患者分层的精准用药策略，使其重新进入价值创造过程。实际上，Enzastaurin 的核心专利（WO 2010/074936 A3）中已明确提出“生物标志物驱动的患者分层”这一策略方向，并以 HDAC2 表达水平作为具体标志物设计了分层治疗方案。然而，HDAC2 并非最终实现 Enzastaurin 价值再创造的标志物。这表明，策略方向可能已嵌入 Enzastaurin 自身的知识中，但实现该策略所需的具体知识要素，如更有效的替代标志物，则可能分布在更广泛的外部知识存量中。

选择 Enzastaurin 案例对废弃技术价值再创造机会识别方法开展概念验证，目标在于检验该方法能否准确理解废弃技术的多维技术特征，并据此形成合理的知识搜索空间界定，从而验证其前两个前提性环节的可行性。概念验证将知识搜索空间的时间范围限定为 2012 年底之前公开的中国生物医药发明专利，以还原 Enzastaurin 被废弃时的知识情境，排除事后信息的干扰。

3.2 对“LLM 驱动的技术特征提取与语义扩展”的验证过程

本阶段以 Enzastaurin 核心专利的英文全文为输入，通过多轮递进式 LLM 对话完成技术特征提取、语义扩展和术语适配，每轮对话均采用结构化提示词引导 LLM 执行特定任务。实际验证结果表明，LLM 在本阶段展现了 3 方面能力：从专利文本中准确提取多维度技术特征的能力；通过同义、近义、上位和相关概念 4 个维度的语义扩展，从术语覆盖、语义边界、概念抽象和跨域关联 4 个层面系统性地拓展搜索空间的能力；面向目标语料库进行术语适配、生成可直接用于后续搜索的多层次特征集的能力。具体验证结果展示如下。

3.2.1 特征提取

第 1 轮对话要求 LLM 从专利全文中提取物理特征和功能特征，并附上原文依据以确保可溯源性。LLM 提取的结果涵盖了物理特征和功能特征两个维度（表 1），每项特征均可追溯至专利原文的具体段落，提取准确率高。物理特征包括化学结构类型（双吲哚-马来酰亚胺类杂环化合物）、分子靶点（PKCβ）、IPC 分类号 4 项；功能特征包括核心作用机制（PKCβ 选择性抑制）、靶向信号通路（PKC 通路、HDAC 通路）、涉及的生物学过程（细胞周期进程、肿瘤转移、细胞凋亡、血管生成）、核心治疗策略（基于 HDAC2 的患者分层）、生物标志物（HDAC2）等。

3.2.2 语义扩展

第 2 轮对话要求 LLM 对表 1 中的每个特征项进行同义概念、近义概念、上位概念和相关概念 4 个维度的扩展。其中：同义概念是指表述不同但含义相同；近义概念是指含义相近但不完全相同；上位概念是指更抽象、更广泛的概念类别；相关概念是指在生物学机制上存在关联。鉴于篇幅所限，以上位概念扩展和相关概念扩展为例进行解释。

上位概念扩展体现了 LLM 沿概念层级进行系统性抽象的能力。以分子靶点“PKCβ”为例，LLM 将其逐层抽象为“蛋白激酶 C 家族→丝氨酸/苏氨酸激酶→激酶→信号转导调控”，使搜索依据从具体靶点拓展到一般性的信号调控功能层面。

表 1 Enzastaurin 核心技术特征提取结果

特征维度	特征类别	提取内容	专利原文依据
物理特征	化学结构类型	双吲哚-马来酰亚胺类杂环化合物，含吡啶-哌啶侧链	3-(1-methyl-1H-indol-3-yl)-4-[1-[1-(pyridin-2-ylmethyl) piperidin-4-yl]-1H-indol-3-yl]-1H-pyrrole-2,5-dione
	分子靶点	蛋白激酶 Cβ（PKCβ）	Enzastaurin is a PKC Beta selective inhibitor
	IPC 分类号	A61K 31/4545；A61K 31/4406；A61K 31/00；A61P 35/00	专利首页 International Patent Classification
功能特征	核心作用机制	PKCβ 选择性抑制	Enzastaurin is a PKC Beta selective inhibitor
	靶向信号通路	蛋白激酶 C（PKC）信号通路；组蛋白去乙酰化酶（HDAC）通路	HDACs belong to the histone deacetylase superfamily...
	涉及的生物学过程	细胞周期进程；肿瘤转移；细胞凋亡；血管生成	Class I HDACs are associated with cell cycle progression, metastasis, and apoptosis
	核心治疗策略	基于 HDAC2 表达水平的患者分层治疗	When the level of HDAC2 is low or undetectable, Enzastaurin alone is expected to be particularly effective
	生物标志物	HDAC2（组蛋白去乙酰化酶 2）	The present invention employs HDAC2 as the preferred biological marker

相关概念扩展则体现了 LLM 发现跨知识域关联的能力。以核心治疗策略“基于生物标志物的患者分层”为例，LLM 关联出伴随诊断、临床试验富集设计、药物敏感性预测等在机制上相关但分属不同知识领域的概念，为后续搜索提供了更广的语义入口。这种从特定技术要素向通用功能机理的抽象化处理，为后续在广泛的知识空间中搜索能够实质性支撑技术复活的新型重组要素提供逻辑接口。

3.2.3　术语适配

第 3 轮对话要求 LLM 将扩展后的英文特征集转换为中文生物医药领域的标准术语，并生成面向中文专利库的分层查询集（表 2）。查询集按抽象层级递进设计，从最具体的分子靶点层（如“PKCβ 抑制”）逐步上升到信号通路层、生物学过程层、治疗策略层、联合用药层，最终形成整合关键机理特征的综合查询层，共 6 个层次。这样的层级化设计，有助于将废弃技术的核心机理特征转化为可操作的检索表达，并为后续语义嵌入搜索提供多层次的语义入口。

3.2.4　与已有方法的对照分析

从特征提取的维度覆盖来看，Cousins 等[24]对计算药物重定位方法的系统梳理表明，现有方法可分为基于蛋白质结构的对接方法、基于转录组签名的匹配方法、基于生物网络的拓扑分析方法和基于真实世界临床数据的分析方法 4 大类。这些方法各自从单一维度提取药物特征：对接方法依赖蛋白质三维结构信息，签名匹配方法依赖基因表达谱数据，网络方法依赖药物-靶点-疾病的交互关系数据，临床数据方法依赖电子健康记录等数据[3,24]。每一类方法不仅受限于单一维度，还需要预先构建特定的数据库或数据管线作为分析前提（如转录组签名方法需要 CMap 等基因表达谱数据库，网络方法需要蛋白质交互网络数据库等）。相比之下，本文方法在技术特征提取环节以单篇专利全文为输入，借助 LLM 的跨领域知识储备，直接从非结构化的技术文本中同时提取物理特征和功能特征两大维度下的多个子类（表 1），在这一环节无须预先构建或依赖特定领域的专用数据库（如蛋白质结构库、基因表达谱数据库或药物-靶点交互数据库等），降低了特征提取的数据准备门槛。需要指出的是，本文方法在后续的搜索空间构建环节仍需借助 IPC 分类体系等结构化知识系统，但这种依赖发生在特征提取与语义扩展完成之后，而非作为特征提取的前置条件。

从语义扩展的深度来看，在专利分析领域，Lee 等[4]提出的关键词专利地图方法通过文本挖掘提取关键词向量，利用关键词共现关系识别技术空白区域，但其分析停留在词频统计层面，无法触及词汇背后的概念层级关系。Choi 等[5]提出的 SAO 网络分析方法进一步将专利文本解析为“主语-动作-宾语”三元组，通过网络分析揭示技术趋势，相比纯关键词方法能够表达技术的功能关系（即“技术如何被使用”和“技术用于什么目的”）。然而，SAO 方法的概念发现仍受限于专利文本中已出现的显性语义关系，无法实现向更高概念层级的抽象，也无法发现文本中未直接出现的跨领域关联。本文的 LLM 方法则在保留显性语义关系提取能力的基础上，通过上位概念扩展实现了沿概念层级的系统性抽象（如将“PKCβ”逐层抽象为“蛋白激酶 C 家族→丝氨酸/苏氨酸激酶→激酶→信号转导调控”），通过相关概念扩展发现了跨知识域的潜在关联（如从“患者分层”关联到“伴随诊断”“药物基因组学”等在机制上相关但分属不同领域的概念）。这两个维度的扩展是关键词方法和 SAO 方法均无法实现的。

表 2　面向中文专利库的分层查询集

查询层次	查询内容	设计意图
层 1：分子靶点	PKCβ 抑制剂；蛋白激酶 Cβ 选择性抑制；PKCβ 激酶活性调控；HDAC2 表达调控	锚定废弃技术的核心分子靶点，匹配靶点层面直接相关的专利
层 2：信号通路	PKC 信号通路调控；组蛋白去乙酰化酶通路调控；PKC/HDAC 协同调控；信号转导通路调控	从具体靶点上升到通路层面，覆盖机制相关但靶点不同的技术
层 3：生物学过程	肿瘤细胞周期调控；肿瘤细胞凋亡诱导；肿瘤血管生成抑制；肿瘤转移抑制	从通路上升到生物学过程层面，覆盖功能相关但通路不同的技术
层 4：治疗策略	生物标志物驱动的患者分层；基于基因表达水平的精准用药；伴随诊断指导的个体化治疗	从生物学机制上升到治疗策略层面，覆盖策略相关的跨领域技术
层 5：联合用药	激酶抑制剂联合 HDAC 抑制剂；靶向药物联合表观遗传调控药物；多靶点协同抗肿瘤策略	覆盖联合用药方案中涉及的协同机制和互补靶点技术
层 6：综合查询	PKCβ AND（患者分层 OR 生物标志物）；激酶抑制 AND 组蛋白去乙酰化 AND 肿瘤	整合关键机理特征进行组合查询，提高匹配精度

此外,王琛等[6]提出的融合大语言模型与语义嵌入的专利检索策略同样使用 LLM 进行语义扩展,但其扩展对象是用户输入的检索查询词,扩展维度为同义术语、下位概念、技术实现路径和应用场景 4 个方面,目的是提升专利检索的查全率和查准率。本文的 LLM 应用与之存在根本区别:本文的扩展对象是废弃技术本身的多维技术特征,扩展维度包含了对废弃技术进行深度语义理解后的上位概念抽象和跨领域关联发现,目的不是优化检索效率,而是为废弃技术在未知知识空间中寻找知识重组可能性提供语义基础。简言之,王琛等[6]的方法回答的是"如何更准确地找到已知类型的专利",本文方法回答的是"如何理解一项废弃技术并为其发现未知的重组方向"。

综上,与已有方法相比,LLM 方法在特征提取上实现了从单一维度到多维度的跨越,在语义扩展上实现了从词汇层面到概念层级的跨越,在应用目标上实现了从检索优化到知识重组机会发现的跨越。

3.3 对"知识搜索空间构建"的验证过程

3.3.1 知识搜索空间的构建过程及结果

本阶段基于第 1 阶段提取的技术特征,由 LLM 结合 IPC 分类号体系推荐知识搜索空间应覆盖的领域范围。

LLM 依据第 1 阶段提取的技术特征推荐了 7 个IPC 分类号(表 3)。基于物理特征中的化学结构类型(双吲哚-马来酰亚胺类杂环化合物),推荐 A61K 31(有机活性成分的医药制剂)和 C07D(杂环化合物),覆盖分子结构层面的关联技术;基于分子靶点和生物标志物的蛋白质属性(PKCβ 和 HDAC2 均为蛋白质)以及涉及的生物学过程(细胞周期调控、凋亡、转移等),推荐 C07K(肽类与蛋白质)和 A61P 35(抗肿瘤药物),覆盖靶点蛋白与治疗活性层面的关联技术;基于功能特征中的治疗策略和生物标志物(基于 HDAC2 的患者分层),推荐 C12N(微生物或酶与遗传工程)、C12Q(涉及酶或微生物的测定方法)和 G01N(材料测试与分析),覆盖基因调控、分子检测和伴随诊断层面的关联技术。

3.3.2 知识搜索空间构建方法与已有方法的对照分析

搜索空间构建的核心在于判定哪些技术领域应被纳入搜索范围,这一判定本质上是对废弃技术与候选领域之间技术相关性的评估。在已有文献中,Breschi 等[25]提出的基于 IPC 共分类的技术相关性测度方法是这一问题最具代表性的分析工具,将其作为对照基准。该方法的核心逻辑是:如果两个 IPC 分类号在大量专利中被频繁地同时赋予同一篇专利文件,则可以认为这两个分类号所代表的技术领域存在知识相关性。这一方法已被广泛应用于技术距离测量和技术溢出研究,是技术相关性测度领域具有影响力的方法之一。

若以该方法界定 Enzastaurin 的知识搜索空间(以下对照在各分类号的原始粒度上进行,以保留技术特征的区分度),则以原始专利所标注的 A61K 31 和 A61P 35(表 1)为种子分类号,依据其与其他分类号的历史共分类频率划定搜索边界。本文未实际运行共分类计算,而是基于该方法的统计逻辑与生物医药专利的一般分类结构进行推断:该方法所覆盖的领域将主要集中于 C07D(杂环化合物)、A61P 35(抗肿瘤药物)和 C07K(肽类与蛋白质)等与种子分类号高频共分类的领域,这些关联是显性的、在既有专利分类实践中已建立统计联系的。C12N

表 3 LLM 推荐的 IPC 分类号及技术特征关联

IPC 分类号	官方描述	技术特征关联
A61K 31	医用、牙科用或梳妆用配制品,含碳化合物的活性成分(有机有效成分的医药制剂)	物理特征(化学结构类型),覆盖有机活性成分的分子结构层面
C07D	杂环化合物	物理特征(化学结构类型),Enzastaurin 为双吲哚-马来酰亚胺类杂环化合物,覆盖杂环类药物分子骨架
A61P 35	化学化合物或药用制剂的特定治疗活性,抗肿瘤(抗癌药)	功能特征(涉及的生物学过程:细胞周期调控、凋亡、转移、血管生成),覆盖抗肿瘤治疗活性层面
C07K	肽类;蛋白质(包括其制备方法、衍生物、载体结合或固定化形式)	物理特征与功能特征(分子靶点 PKCβ 和生物标志物 HDAC2 均为蛋白质),覆盖靶点蛋白与相关分子层面
C12N	微生物或酶;其组合物;繁殖、保藏或维持微生物;变异或遗传工程;培养基	功能特征(治疗策略、HDAC2 患者分层),覆盖基因调控、酶工程与遗传工程层面
C12Q	涉及酶、核酸或微生物的测量或测试方法;其组合物或试纸;制备该组合物的方法	功能特征(生物标志物、HDAC2 表达水平检测),覆盖分子检测与酶/核酸层面的测定方法
G01N	通过测定化学或物理性质来研究或分析材料	功能特征(伴随诊断、患者分层所需的诊断检测),覆盖生物标志物检测与材料分析层面

(遗传工程)由于涉及基因调控与酶工程,与药物制剂分类号之间存在一定程度的共分类关系,属于该方法可能部分覆盖的边缘领域。

然而,LLM 推荐的 7 个 IPC 分类号中,C12Q(涉及酶或核酸的测定方法)和 G01N(材料测试与分析)两个则难以被共分类方法识别。原因在于,药物制剂专利(A61K 31)与分子检测方法专利(C12Q)或材料分析专利(G01N)在功能上分属不同的技术领域,在分类实践中很少被同时赋予同一篇专利。LLM 之所以推荐这两个分类号,是因为它在第 1 阶段中识别了 Enzastaurin“基于 HDAC2 表达水平的患者分层治疗”这一功能特征,并通过语义推理将“患者分层”关联到“生物标志物检测”“伴随诊断”等跨领域概念,进而映射到 C12Q 和 G01N。

Enzastaurin 的实际价值再创造路径为上述对照提供了回溯验证的依据。索元生物(Denovo Biopharma)在获取 Enzastaurin 后,利用其全基因组扫描平台从礼来公司此前开展的 PRELUDE 三期临床试验样本中发现了一组全新的基因组生物标志物 DGM1(Denovo Genomic Marker 1),该标志物能够有效预测 Enzastaurin 在弥漫性大 B 细胞淋巴瘤患者中的疗效[26]。DGM1 的发现与应用涉及两类关键技术:其发现依赖全基因组变异筛查与基因组关联分析,在 IPC 分类体系中对应 C12Q(涉及核酸的测定方法);其临床转化则需要开发面向患者分层的伴随诊断检测方案,涉及对生物样本的系统化分析与测定,在 IPC 分类体系中对应 G01N(材料测试与分析)。其分类方法能否覆盖这两个领域,取决于 A61K 31 与 C12Q、G01N 在历史专利中的统计共现程度,LLM 并非通过统计共现发现这两个领域,而是基于对 Enzastaurin“患者分层”功能特征的语义理解,经由“生物标志物检测”和“伴随诊断”等概念推导而来,这一推导路径与 DGM1 实际的发现和应用路径高度吻合,为 LLM 搜索空间推荐的合理性提供了独立的回溯验证证据。

上述对照揭示了两种搜索空间界定逻辑的本质差异:基于 IPC 共分类的方法以历史专利数据中已建立的统计关联为依据,更擅长识别“已经被关联过的”技术领域;LLM 方法则以对技术机理特征的语义理解为依据,能够发现历史数据中尚未建立统计关联但在机理上存在潜在相关性的跨域领域。综合来看,第 2 阶段的概念验证表明,LLM 能够依据 Enzastaurin 的多维技术特征,将对废弃技术的语义理解进一步转化为对知识搜索空间的合理界定。其所推荐的 IPC 分类号较好地覆盖了与 Enzastaurin 分子结构、作用靶点、生物学过程及患者分层策略相关的主要知识领域,说明该方法能够有效回答“在哪里搜索”这一问题。为还原 Enzastaurin 被废弃时的知识情境并避免事后信息干扰,将知识搜索空间限定为 2012 年底之前公开的中国生物医药发明专利。至此,已完成对 4 阶段方法框架中两个前提性环节的概念验证,证明 LLM 能够形成支撑后续广度搜索与机会研判所需的技术表征和知识搜索空间。

4 结论与展望

4.1 主要结论

针对废弃技术价值再创造机会识别缺乏可操作方法的问题,从知识重组视角出发,将机会识别视为“为废弃技术寻找新的知识重组可能性”,并将这一任务分解为“搜索什么、在哪里搜索、如何搜索、是否构成知识重组机会”4 个递进的子问题,据此设计了 4 阶段方法框架:LLM 驱动的技术特征提取与语义扩展解决“搜索什么”的问题;知识搜索空间构建解决“在哪里搜索”的问题;基于语义嵌入模型的广度搜索解决“如何搜索”的问题;LLM 驱动的证据挖掘与语义推理解决“是否构成知识重组机会”的问题。在此基础上,以 Enzastaurin 为例进行初步概念验证。该方法框架的核心思想是将机会识别建立在 LLM 的语义理解与推理能力之上,而非依赖词汇层面的表面匹配。以 Enzastaurin 为例的概念验证表明,LLM 能够支撑技术特征提取、语义扩展与知识搜索空间界定两个前提性环节,从而说明该方法框架在“搜索什么”与“在哪里搜索”两个关键问题上具有可行性。

4.2 理论贡献

(1)拓展了知识重组理论的应用边界。现有知识重组研究主要关注企业如何在自身知识基础上搜索和整合新要素以实现创新[16-18],较少涉及那些已经脱离原有知识组合的废弃技术。废弃技术价值再创造的机会识别,本质上是在知识空间中搜索重组可能性,而非在既有组合框架内进行改进,这拓展了知识重组理论的应用边界。

(2)重新定位了知识重组研究中“搜索空间”的理论角色,将其从隐含于搜索主体认知的前置条件显化为可被独立建构的分析对象。在已有的知识重组研究中,搜索过程通常以企业为主体,搜索空间由企业既有知识基础及其邻近领域自然界定:Fleming[17]关于本地搜索与远程搜索的讨论、Savi-

no 等[18]的系统综述以及 Xiao 等[16]的研究回顾关注的核心问题是“如何在搜索空间内更好地搜索”,搜索空间本身在企业情境下不构成需要独立回答的问题。然而,废弃技术已经脱离了原有的知识组合情境,缺少一个由原持有者认知所界定的搜索边界,“搜索空间从何而来”由此浮现为机会识别必须首先回答的独立问题。基于这一观察,将“搜索空间的建构逻辑”视为知识重组研究中一个值得显化的分析维度,它既不同于搜索内容(“搜索什么”),也不同于搜索结果(“找到了什么”),而是规定搜索得以发生的前置条件。为操作化这一理论命题,以专利文献为知识载体,通过类型过滤、时间还原与 LLM 驱动的领域推荐共同界定搜索空间。这一具体方案在生物制药案例中得到了有效性的初步证据(详见 3.3 节),但其更重要的意义在于示范了“搜索空间作为建构对象”在方法论上的可操作性,使得这一理论维度不仅可被讨论,也可被实证。上述命题为知识重组研究开辟了一条尚未被充分挖掘的探索路径,除了继续追问“如何在搜索空间内搜索”,还可以追问“不同情境下搜索空间应如何被建构”“不同建构逻辑会导向何种重组结果”等问题,这一拓展尤其适合那些原持有者认知框架已不再约束知识要素流动的情境(如废弃技术、过期专利、组织间技术转移等)。

(3)拓展了扩展适应理论在技术创新管理中的方法应用。Garud 等[14]指出扩展适应驱动的创新通常依赖偶然性(Serendipity),从扩展适应理论关于偶然性发现的讨论出发,尝试提出一种将废弃技术价值再创造机会识别由偶然发现转化为系统搜索的方法框架,从而为扩展适应理论在技术创新管理中的方法论操作化提供了新的思路。

(4)丰富了 LLM 在技术创新管理中的应用情境。LLM 在科学发现领域已展现出识别潜在知识关联的能力[19,20],在专利分析领域则主要用于技术趋势识别和专利检索[4-6,23]。将 LLM 定位为废弃技术机会识别的核心工具,展示其语义理解、概念抽象和跨域关联能力如何作用于面向废弃技术的知识重组搜索。这一应用场景不同于既有的技术预测或文献分析任务,而是指向“人机混合智能”如何支持突破性创新的问题[27]。

4.3 实践启示

本文的方法框架为废弃技术的持有者和潜在获取者提供了可操作的分析工具。首先,对于持有大量废弃技术资产的企业(如大型制药公司),可借助 LLM 系统性地评估废弃技术的潜在价值,为技术资产的处置决策(如对外许可、出售或内部再开发)提供依据。其次,对于专注于废弃技术再开发的企业(如案例验证中的索元生物),本方法能够辅助其在收购决策前快速识别废弃技术的潜在重组方向,降低搜索成本和决策风险。最后,LLM 驱动的语义扩展和搜索空间构建思路,也可推广至更广阔的跨领域技术机会发现场景,不限于废弃技术的价值再创造。

4.4 局限与展望

本文存在以下局限:①以 4 阶段方法框架的构建为主,概念验证为辅,仅完成了前两个阶段的概念验证,第 3 阶段(语义嵌入广度搜索)和第 4 阶段(LLM 驱动的证据挖掘与语义推理)的实证尚未开展,方法框架的完整有效性有待后续研究验证。②概念验证基于单一案例和单一领域(生物制药),方法的跨领域适用性尚需检验。③LLM 的产出质量受提示词设计和模型能力的影响,当前提示词设计尚有优化空间。

后续研究将沿 3 个方向展开。①完成第 3、第 4 阶段的完整实证,检验方法框架从“特征理解”到“机会研判”的全链条有效性;②在多案例、多领域层面检验方法的适用性和稳健性;③探索提示词设计的系统性优化策略,以提升各阶段产出的质量。

参考文献

[1] ADNER R. The wide lens: a new strategy for innovation [M]. London: Penguin Uk, 2012.

[2] SCHUMPETER J A. The theory of economic development[M]. Cambridge: Harvard University Press, 1934.

[3] PUSHPAKOM S, IORIO F, EYERS P A, et al. Drug repurposing: progress, challenges and recommendations [J]. Nature Reviews Drug discovery, 2019, 18(1): 41-58.

[4] LEE S, YOON B, PARK Y. An approach to discovering new technology opportunities: keyword-based patent map approach[J]. Technovation, 2009, 29(6/7): 481-497.

[5] CHOI S, YOON J, KIM K, et al. SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells[J]. Scientometrics, 2011, 88(3): 863-883.

[6] 王琛, 赵燕清, 李登宇, 等. 融合大语言模型与语义嵌入的专利检索策略研究[J]. 情报理论与实践, 2026, 49(6): 190-196, 180.

[7] LEONCINI R. Learning-by-failing. an empirical exercise

on CIS data[J]. Research Policy, 2016, 45(2): 376-386.
[8] 李天柱, 侯锡林, 马佳. 技术创新成功与失败的纯理论思考[J]. 技术经济与管理研究, 2016(3): 24-28.
[9] ANTOSZCZAK M, MARKOWSKA A, MARKOWSKA J, et al. Old wine in new bottles: drug repurposing in oncology[J]. European Journal of Pharmacology, 2020, 866: 172784.
[10] 李天柱, 董晓东. 技术创新失败挽救的理论界定[J]. 自然辩证法通讯, 2019, 41(2): 69-73.
[11] MOKYR J. Evolutionary phenomena in technological change[M]. ZIMAN J. Technological innovation as an evolutionary process. Cambridge: Cambridge University Press, 2000.
[12] GOULD S J, VRBA E S. Exaptation: a missing term in the science of form[J]. Paleobiology, 1982, 8(1): 4-15.
[13] GOULD S J. The structure of evolutionary theory[M]. Cambridge: Harvard University Press, 2002.
[14] GARUD R, GEHMAN J, GIULIANI A P. Serendipity arrangements for exapting science-based innovations[J]. Academy of Management Perspectives, 2018, 32(1): 125-140.
[15] NELSON R R, WINTER S G. An evolutionary theory of economic change[M]. Cambridge: Harvard University Press, 1985.
[16] XIAO T, MAKHIJA M, KARIM S. A knowledge recombination perspective of innovation: review and new research directions[J]. Journal of Management, 2022, 48(6): 1724-1777.
[17] FLEMING L. Recombinant uncertainty in technological search[J]. Management Science, 2001, 47(1): 117-132.
[18] SAVINO T, MESSENI PETRUZZELLI A, ALBINO V. Search and recombination process to innovate: a review of the empirical evidence and a research agenda[J]. International Journal of Management Reviews, 2017, 19(1): 54-75.
[19] YANG Z, LIU W, GAO B, et al. Moose-chem: large language models for rediscovering unseen chemistry scientific hypotheses[J]. arXiv preprint arXiv: 2410.07076, 2024.
[20] BOIKO D A, MACKNIGHT R, KLINE B, et al. Autonomous chemical research with large language models[J]. Nature, 2023, 624: 570-578.
[21] ZHANG Y, KHAN S A, MAHMUD A, et al. Exploring the role of large language models in the scientific method: from hypothesis to discovery[J]. npj Artificial Intelligence, 2025, 1(1): 14.
[22] ZHAO W, YANG X, LYU Z, et al. Road of large language model: source, challenge, and future perspectives[J]. Research, 2025, 8: 0655.
[23] 周潇, 高雅倩, 樊嘉逸. 基于 BERT 词嵌入的专利检索策略研究[J]. 情报学报, 2023, 42(11): 1347-1357.
[24] COUSINS H C, NAYAR G, ALTMAN R B. Computational approaches to drug repurposing: methods, challenges, and opportunities[J]. Annual Review of Biomedical Data Science, 2024, 7(1): 15-29.
[25] BRESCHI S, LISSONI F, MALERBA F. Knowledge-relatedness in firm technological diversification[J]. Research Policy, 2003, 32(1): 69-87.
[26] NOWAKOWSKI G S, ZHU J, ZHANG Q, et al. ENGINE: a Phase III randomized placebo controlled study of enzastaurin/R-CHOP as frontline therapy in high-risk diffuse large B-cell lymphoma patients with the genomic biomarker DGM1[J]. Future Oncology, 2020, 16(15): 991-999.
[27] SHI J. Does hybrid intelligence trigger breakthrough innovation?[J]. Technovation, 2026, 152: 103488.

Design and Prerequisite Validation of a Large Language Model-based Method for Identifying Value Re-creation Opportunities in Abandoned Technologies

LI Tianzhu, XIAO Jing

(School of Business Administration, Guizhou University of Finance and Economics, Guiyang 550025, China)

Abstract: Abandoned technologies far outnumber those currently in use and retain significant potential for value re-creation. However, because they have become detached from their original knowledge networks, there is still a lack of operational methodological support for systematically identifying their possible applications in new contexts. A large language model (LLM)-based method for identifying value re-creation opportunities in abandoned technologies was proposed, and a four-stage methodological framework was constructed: "LLM-driven technical feature extraction and semantic expansion → knowledge search space construction → breadth search based on semantic embedding models → LLM-driven evidence mining and semantic reasoning". The focus was placed on the design logic of the methodological framework and the feasibility validation of its prerequisite stages. A proof-of-concept validation of the first two prerequisite stages was conducted using the abandoned technology Enzastaurin as an example. It is indicated by the validation results that the LLM can extract multidimensional technical features of abandoned technologies from patent texts, complete conceptual abstraction, semantic expansion, and knowledge search space definition. The feasibility of the first two prerequisite stages is thereby validated, and prerequisite support is provided for the viability of the proposed framework.

Keywords: abandoned technology; value re-creation; opportunity identification; large language model(LLM); patent text analysis