项目结题验收单

专家验收表（主持人所在单位组织 3-5 名专家对项目进行验收、自评。）

项目名称	中国丘陵山区农业机械技术创新机会识别研究——基于多源科研数据融合视角				
主持人	潘颖	职务/职称	主任/副研究馆员		
所在单位	（加盖单位公章）				
专家意见	多源数据融合研究有助于提升技术创新机会识别的全面性、准确性、前瞻性，有利于优化资源配置，提升研发效率。该项目融合期刊、专利两类科研数据，基于 LDA-GTM 模型和对比学习技术，构建了技术创新机会识别模型与评估体系，并聚焦丘陵山区农业机械化领域开展实证研究。 模型构建思路清晰，研究方法具有一定的创新性。首先提出了两类语义对齐框架，通过对比学习解决论文与专利的术语异构性问题；再次构建 LDA-GTM 联合模型，实现技术主题挖掘与空白点识别，克服了人工主管判断误差；最后建立 STR（科学-技术关联度）、TGV（技术空白度）、TRL（成熟度）三维指标评估体系，量化技术机会优先级，提供策略依据。 实证分析严谨。数据来源客观权威，清洗与预处理流程规范，构建了领域专用词库和停用词库，基于识别模型提出新能源动力系统、智能视觉导航等 7 类技术空白点，明确对应前沿型、潜在型、应用型开发策略，对企业和科研机构的技术研发具有一定的参考意义。 未来可以增加市场数据、政策文本等数据维度，增强技术机会评估的产业化适配性。增加时间序列维度，研究技术演进动态，识别技术生命周期中的突变点。 本项目研究目标明确，方法创新，逻辑严谨，为融合多源数据开展技术创新机会识别研究提供了研究框架和实证案例。研究成果具有一定的参考性、推广性。 经评审，专家一致同意验收通过，准予结题。 （如需要可增加页数）				
专家签字					
职务/职称					

项目编号：2024031

CALIS 全国农学文献信息中心研究项目
结题报告

项目名称：中国丘陵山区农业机械技术创新机会识别研究——基于多源科研数据融合视角（2024031）

项目关键词：丘陵山区　农业机械　多源数据　机会识别

项目单位(盖章)：江苏大学

通信地址：江苏省镇江市学府路 301 号江苏大学图书馆学科服务中心，212013

项目主持人：潘颖

联系电话：18796086763

电子邮件：5283443@qq.com

提交日期：2025.5.15

题目：中国丘陵山区农业机械技术创新机会识别研究——基于多源科研数据融合视角（2024031）

关键词：丘陵山区　农业机械　多源数据　　机会识别

1 研究背景、目的及意义

1.1 研究背景

科技创新是加快推进丘陵山区农业机械化的动力。丘陵山区是我国重要的果蔬茶和粮油生产基地，但由于地形复杂、地块碎小、坡陡路窄，导致传统的农业生产方式效率低下、用工成本增高[1]，丘陵山区农业机械化水平远低于平原地区，并且在耕作、种植、田间管理、收获等环节可用机具较少。从 2018 年中央一号文件提出“加快研发经济作物、丘陵山区农林机械装备研发制造”[2],到 2023 年中央 1 号文件指出，加快先进农机研发推广，加紧研发大型智能农机装备、丘陵山区适用小型机械和园艺机械,同时国家高度重视丘陵山区机械化发展问题，投入大量的科研人员和专门的科研经费进行关键零部件技术攻关[3]。**技术创新机会是技术创新的起点，技术创新机会识别有助于优化技术创新战略布局。**技术机会识别是通过运用市场调研、技术趋势分析、数据获取、文本挖掘和技术图谱等方法，来分析和预测不同领域中的技术机会，揭示技术发展的趋势和未来可能的突破点，帮助决策者和研究者确定最具潜力的技术领域和方向[4-6]。通过技术机会识别能够把握科技发展机遇，提前布局，取得竞争优势。**多源数据融合研究能够更全面和准确地挖掘技术机会。**由表 1 可知,当前研究在技术识别方面存在数据源单一的局限性,单一数据源研究虽然有其价值,但忽视了其它数据来源的潜在信息，所以单一数据源目前已经无法满足技术机会识别分析的需求。由于技术创新受到政府政策、科技研发水平、经济条件、社会因素等多方面的影响，形成的数据也是多源的，因此随着数据量的增加，对于多源数据的挖掘来识别技术机会显的尤为重要。

表 1 技术创新识别方法及数据源（部分列举）

作者	论文名称	年份	方法	数据源
康宇航;苏敬勤	技术创新机会的可视化识别——基于专利计量的实证分析	2008	共现分析、聚类、关联分析、网络分析	专利
吕一博;康宇航;王淑娟	基于共现分析的技术机会发现与可视化识别	2012	共现分析	专利
Naoki Shibata, Yuya Kajikawa, Ichiro Sakata	Extracting the commercialization gap between science and technology — Case study of a solar cell	2010	共现网络	论文、专利
Lee, Y ; Kim, SY ; Song, I ; Park, Y; Shin, J	Technology opportunity identification customized to the technological capability of SMEs through two-stage patent analysis	2014	迭代分析	专利
Lee, C ; Kang, B ; Shin, J	Novelty-focused patent mapping for technology opportunity analysis	2015	专利地图	专利
王金凤;吴敏;岳俊举;吴汉争;冯立杰	创新过程的技术机会识别路径研究——基于专利挖掘和形态分析	2017	专利挖掘及形态分析	专利
PARK Y,YOON J	Application technology opportunity discovery from technology portfolios:use of patent classification and collaborative filtering	2017	专利引文网络与技术知识流动网络	专利
王 坤, 王京安, 汤月, 校姜文	基于专利和科技论文的技术机会识别研究——以金属 3D 打印技术为例	2018	共现网络、多维度聚类	论文、专利
慎金花;闫倩倩;孙乔宣;万召侗	基于专利数据挖掘的技术融合识别与技术机会预测研究——以电动汽车产业为例	2019	共现网络	专利
张维冲 , 王 芳, 赵 洪	多源信息融合用于新兴技术发展趋势识别——以区块链为例	2019	数据融合、主题关联分析	论文、专利
王翠; 波熊坤; 刘文俊	基于CiteSpace的技术预测研究的可视化分析	2020	知识图谱	论文
许学国; 桂美增	基于 GTM 逆向映射的技术创新机会识别——以新能源汽车为例	2021	深度学习、自然语言处理和技术地图	SCI 论文、专利
张振刚;罗泰晔	基于RFM模型和随机行动者导向模型的技术机会识别	2021	随机行动者导向模型、知识网络	专利
冯立杰;曾小红;王金凤;张珂	一种三级技术机会识别方法及其应用——基于 SAO 语义分析和多维技术创新地图	2021	SAO(Subject-Action-Object,主语-谓语-宾语)语义分析	专利
宋凯;冉从敬	基于主题挖掘与专利评估的技术机会识别研究——以智慧农业为例	2023	主题挖掘	专利

1.2 研究目标

随着国内外技术的快速更迭，技术竞争更加激烈，如何占领技术高地成为必争之处。如果想要在竞争中获得优势，就必须从技术这个角度入手才能获得新的机遇。因此本研究通过**技术知识的挖掘和整合**等方法，获得潜在的技术机会，提供未来技术的发展方向，从众多的技术知识数据中准确、全面的识别技术机会进行创新活动是非常关键的。

1.3 研究意义

本研究**在理论上**丰富了技术创新机会识别相关研究方法理论体系，当前研究大多以一种数据源来展开研究，对于技术机会识别来说，存在数据损失风险，无法全面、客观进行分析、判断、预测。因此本研究建立包含专利数据和科学文献数据多源科研数据融合模型，基于 LDA-GTM 的技术开展技术创新机会挖掘和识别研究，丰富技术创新机会识别方法。**在实践上**，以丘陵山区农业机械领域为例，寻找潜在的技术创新点，为丘陵山区农业机械领域科研人员提供技术研发方向，为研究创新工作提供参考。

2 研究思路与方法

2.1 研究方法

本研究基于科学（期刊文献）与技术（专利数据）双维度，构建多源科研数据融合框架，结合 LDA 主题建模与 GTM 生成式拓扑映射，识别丘陵山区农业机械领域的技术创新机会。具体研究内容如下：

1. 多源科研数据预处理及融合模型

针对期刊文献与专利数据在术语体系（前者侧重理论原理，后者强调技术特征）、语义结构和内容粒度上的异质性，提出基于对比学习（Contrastive Learning）的跨模态语义对齐模型：

（1）数据预处理：采用 NLP 技术对文本进行分词（Jieba 工具）、去停用词（自定义农业机械停用词库）、同义词合并（构建领域本体词典）及词向量化（Word2Vec）。

（2）语义对齐：通过双塔神经网络架构，分别编码论文与专利文本的语义特征，利用对比损失函数（Contrastive Loss）优化模型，使描述相同技术概念的跨模态文本

在向量空间中距离最小化。

2. 技术的主题分析研究

从微观角度深入分析丘陵山区农业机械领域的关键技术的技术主题，研究丘陵山区农业机械关键技术有哪些问题。通过采用潜在狄利克雷分布 LDA（Latent Dirichlet Aiiocation）模型，从微观层面解析丘陵山区农业机械技术主题分布：

（1）主题识别：对预处理后的文本数据（期刊、专利摘要）进行主题建模，提取每篇文档的隐含主题分布。

（2）主题-技术关联矩阵：利用关键词共现网络构建二进制技术词矩阵，量化主题内技术特征强度

3. 基于 GTM 技术创新机会识别

通过 GTM 生成式拓扑映射构建了关键技术主题地图，将高维技术词矩阵投影至二维拓扑空间，形成技术主题聚类区域，对地图的真空部分进行识别和解读所包含的技术信息，发现关键技术上的技术空白点，在通过对技术空白点的识别研究中，对技术机会做出更深入的讨论。

4. 构建技术机会评估体系

结合“科学-技术关联度”、“技术空白度”与“技术成熟度”构建技术创新优先级决策模型，用于筛选上述技术空白点的技术价值，从而明确技术创新方向。

2.2 研究路线

在全球化和信息化加剧经济竞争的背景下，新技术、新业态与新模式不断涌现，无论企业还是科研机构都面临机遇与挑战并存的环境。技术创新成为获取竞争优势、实现可持续发展的核心路径，而技术机会识别作为关键环节，通过分析市场、技术与竞争信息，帮助企业或者科研单位发现新商业机会、创新技术，从而能够应对市场的激烈竞争。然而，目前社会处于信息化高度集中的年度，信息集中且复杂多样，面对海量的数据仅依靠人工判断很难得到准确全面的结果，因此要借助数据挖掘的方法对数据进行处理。

基于上述分析，本研究从科学和技术两个角度出发，以丘陵山区农业机械领域为例，构建了“异构数据融合-主题挖掘-机会识别”框架，通过对期刊文献和专利数据相关信息进行挖掘和分析，实现对该领域潜在的技术创新机会的识别（图

1 技术路线图）。考虑到期刊文献以及专利数据在结构、形式上各有不同，本文在进行研究时，首先采用自然语言处理相关技术对多源异构数据进行处理，建立构建跨模态语义对齐模型，实现异构数据在语义上同构。通过分词、去停用词、构建自定义词库、同义词合并、语义对齐等步骤，实现对文献和专利的标题和摘要文信息的挖掘，为后续目标领域技术主题提取和技术机会识别奠定基础。其次，采用潜在狄利克雷分布 LDA（Latent Dirichlet Aiiocation）主题模型对期刊文献和专利数据进行主题识别，利用识别出的关键词术语，构建二进制表示的关键词表示向量，最终形成论文和专利数据的技术词矩阵。随后，基于 GTM 生成式拓扑映射法绘制技术地图，对地图出中低密度区域的点进行详细分析，从而实现技术创新机会的识别。最后，依据构建的技术机会评估体系，量化上述技术空白点，从而明确高价值的技术创新方向。

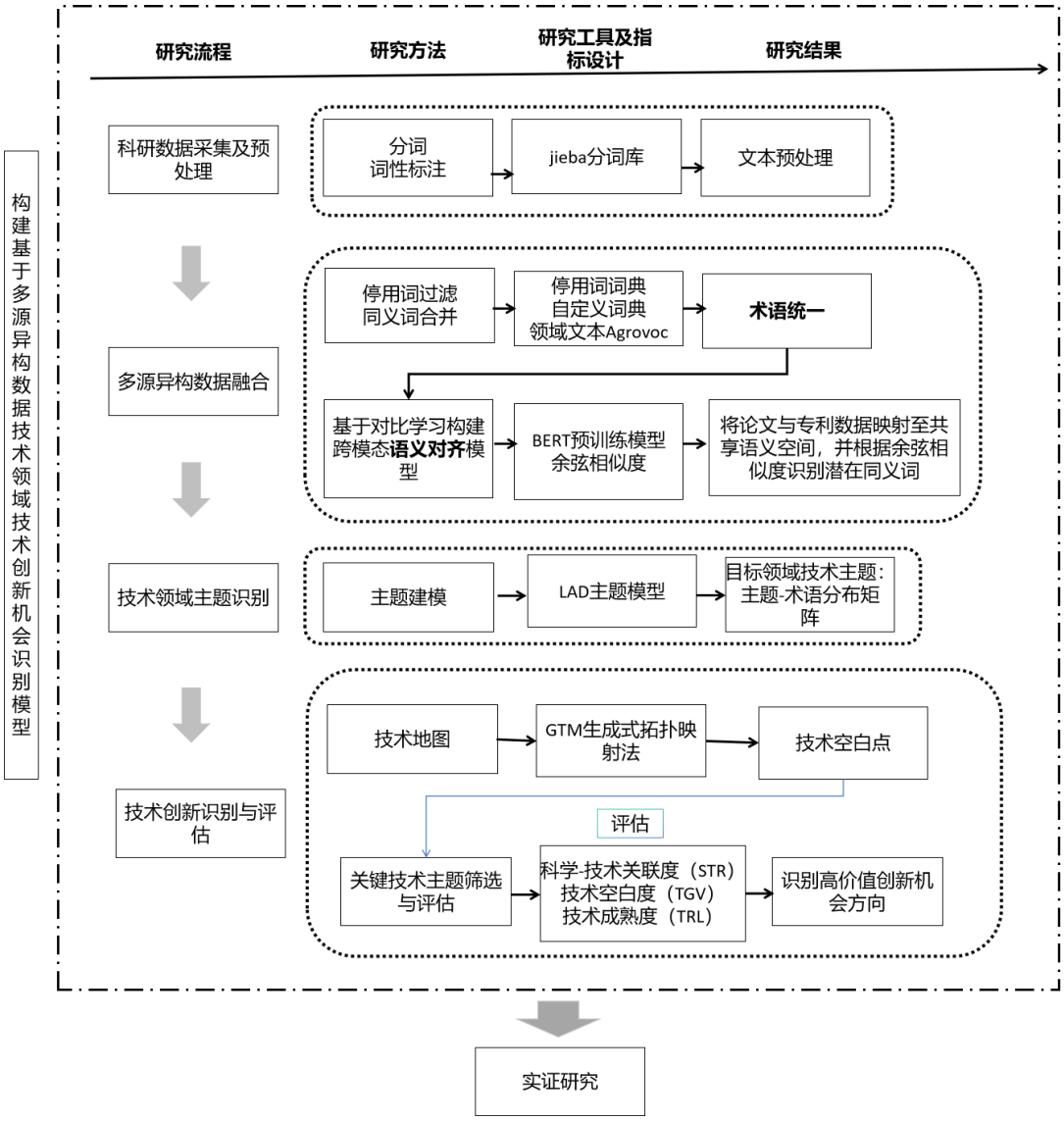

图 1 技术路线图

2.3 主要创新点

本研究创新型主要包括以下几方面：

1.提出跨模态语义对齐框架：首次将对比学习引入农业机械领域多源数据分析，解决论文与专利的术语异构性问题。

2.构建 LDA-GTM 联合模型：通过主题建模与拓扑映射的动态耦合，实现技术演化路径的可视化解析，精准识别技术空白点。

3.构建技术机会评估体系：结合“科学-技术关联度”“技术空白度”与技术成熟度指标，建立丘陵山区农业机械创新优先级决策模型，为研发资源分配提供

量化依据。

2.4 相关理论和方法

1. Transformer 编码器

2017 年，Google 提出了一种名为 Transformer 的模型，能够有效的提取文本的特征，被应用在自然语言处理领域中，通过采用注意力机制来联系输入和输出。它利用了全连接的多头自注意力机制来捕捉全局信息，以便更有效地提取特征并提高可解释性。每个 Transformer 编码器由两个组件组成：多头自注意力和前馈神经网络，对于每个子结构，都会进行残差链接和层归一化，这些结构的组合构成了 Transformer 编码器的整体结构，如图 2 所示。

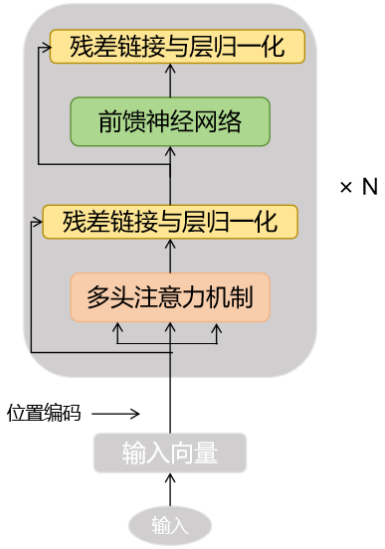

图 2 Transformer 编码器结构

2. BERT 预训练语言模型

预训练语言模型是指基于大规模数据训练的语言模型，它的核心思想是首先在大规模的语料库上进行无监督学习，通过训练来学习到语言的内在规律和表示，然后在此基础上进行特定任务的微调，使模型能够适应各种不同的自然语言处理任务。BERT（Bidirectional Encoder Representations from Transformers）由 Google 在 2018 年提出。BERT 通过深度双向 Transformer 架构进行预训练，旨在通过考虑文本的左右上下文来提升语言理解能力。BERT 模型[7]是最常用的预训练语言模型之一，主要利用了 Transformer[8]的编码器，如图 3 所示，E_1，...，E_N 表示文本输入，经过双向 Transformer 编码器可以得到含有丰富上下文语义

信息的文本表示向量 T_1，…，T_N。BERT 模型自发布起，不仅在多个自然语言处理任务上取得了较好的性能，同时也推动了自然语言处理技术的迅猛发展。

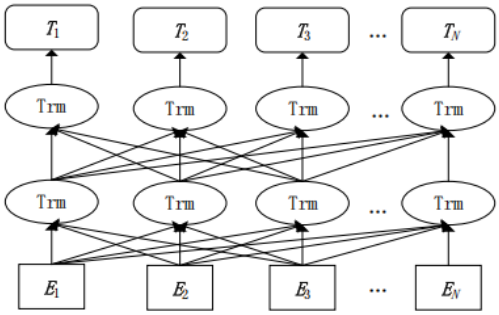

图 3 BERT 预训练模型

3. 对比学习

近年来，自监督对比学习在自然语言处理领域和计算机视觉领域等多个领域都取得显著成效。它利用规则生成正负例样本来学习样本之间的相似度度量。对比学习的重点是学习样本的特征表示，使得相似的样本在特征空间中距离更近，而不相似的样本距离更远。这种方法主要应用于文本表示学习，以便更好地度量文本之间的相似性。其核心目标是通过学习有效的特征表示，使得相似的文本在特征空间中距离更近，从而实现文本分类、聚类等任务。总之，对比学习是一种有效的自监督学习方法，它可以在没有标注数据的情况下学习到有效的特征表示，并通过特征距离度量实现文本相似性的度量和文本处理任务的优化。

4. LDA 主题模型

LDA（Latent Dirichlet Allocation）主题模型也称为三层贝叶斯概率模型，是 Blei D M、Ng A Y 和 Jordan 在 2003 年提出的一种文档主题生成模型[9]，其整体包含词、主题和文档三层结构，在科学文献知识挖掘、科学研究热点发现、新兴主题探测、科学研究主题演化和学术评价等研究方向有广泛的应用[10]。LDA 是一种典型的词袋模型，它将每篇文档视为一个词频向量，其中包含许多随机分布的主题，词与词之间没有顺序以及先后的关系，通过使用 LDA 主题模型，可以从每篇文档中提取出一个主题，并使用该主题来提取关键词。经过多次重复，直到遍历整个文档中的每个单词，从而构建主题模型。LDA 是一种无监督学习方法，其训练过程无需依赖手工标注的训练集，仅需 提供文档集并指定主

题数量，LDA 便能自动分析文档中的词汇统计信息，从而挖掘出文档集合中潜在的主题结构。

3 研究内容

3.1 技术创新机会识别模型构建

1. 多源异构数据融合模型

缩小多源异构数据语义的方法就是将异构数据映射至共同语义空间，捕获共享语义。本研究的模型首先选择双塔神经网络架构，由两个独立的 Transformer 编码器组成，分别处理论文（科学文献）和专利（技术文档）两种异构文本，可以得到具有全局特征和上下文语义信息的动态文本表示向量。其次，构建共享投影层，将不同模态的文本表示向量映射到同一语义空间。投影层投影层是由全连接层、ReLU 激活函数层和全连接层串联组成，能够起到进一步挖掘重要特征、提高模型性能和降维的作用，将文本表示向量输入到投影层，经过多重线性组合与激活后，最后输出的文本表示向量维度将由 768 变为 128。在该空间中，基于对比学习机制的文本特征距离来度量文本相似性，拉进相似特征的文本，拉远不同特征的文本。并通过比损失函数 InfoNCE Loss 来对模型进行优化，使得模型能够学习到深层的不同类别的文本情感特征表示。最终，模型进行异构数据融合，并输出最终将专利和论文向量合并，实现两者文本的语义对齐。多源异构数据融合框架图，如图 4 所示。

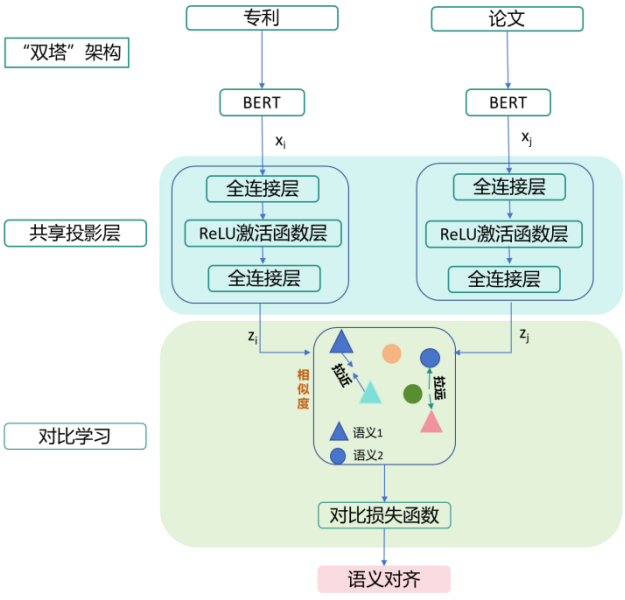

图 4 多源异构数据融合框架图

对于专利和论文文本而言，可以将同一技术主题的专利、论文文本当做正样本，将所有不同技术主题的专利、论文文本作为负样本，通过计算与正负样本间的相似度函数，来缩小与正样本对的特征距离，拉大与粗样本对的特征距离，相似度公示如下：

$$\text{sim}（z_i, z_j） = \frac{z_i \cdot z_j}{\|z_i\| \cdot \|z_j\|}$$

式中：z_i：专利文本的向量表示，z_j：论文文本的向量表示，$\|z\|$：向量模长，计算向量到原点的欧氏距离，sim（z_i, z_j）：范围[-1,1]，值越大表示语义越相似。

本文的多源异构数据融合模型使用对比损失函数 InfoNCE Loss 来对模型进行优化，InfoNCE Loss 损失函数的公式如下：

$$\mathcal{L} = -\log \frac{\exp（\frac{\sin （z_j, z_+）}{\tau}）}{\sum_{i=1，\ i\neq j}^{2N} \exp(\frac{\sin(z_j, z_i)}{\tau})}$$

式中：$\mathcal{L}$：对比损失值，衡量政府样本对的相似度分布差异；z_i：专利文本的向量表示；z_j：论文文本的向量表示；z_+：正样本专利的向量表示，与z_j描述同一技术概念；τ：温度参数，控制相似度分布的陡峭程度；N：批次中的总样本数，包括 1 个正样本和 N-1 个负样本。

2. 技术创新识别 LAD-GTM 模型

LAD-GTM（Latent Dirichlet Allocation - Generative Topographic Mapping）是一种融合主题建模与拓扑映射的技术创新识别框架，旨在从语义对齐的多源科研数据（如期刊论文与专利）中挖掘技术主题，并通过可视化技术地图识别潜在创新机会。基于 LDA 主题模型挖掘技术数据中的技术主题及其特征词，形成技术特征词矩阵，通过 GTM 模型将矩阵映射到二维空间，绘制技术地图并识别技术空白，再将技术空白点通过 GTM 逆向映射回原始空间。

3. 基于 LDA 主题模型的技术特征词识别

主题一致性得分是一个衡量主题质量与估计主题数量的方法，相较于困惑度，主题一致性更能衡量 LDA 的主题提取性能[11]。因此，本文首先采用主题

一致性确定最优主题数。其次，读取技术数据进行 LDA 主题挖掘，得到技术主题及其对应技术特征词。根据技术特征词对每条技术数据进行标记，用 1 表示对应数据中包含该词，0 表示不包含。若存在 n 个技术特征词，则得到 n 维技术特征词矩阵。

4. 基于 GTM 的专利空白挖掘方法

生成式拓扑映射（GenerativeTopographicMapping，GTM）作为一种广泛用于分类、聚类及可视化的算法或潜在变量模型[12-13]，将多维的专利数据转化成低维的数据空间，呈现在每一个网格上，空白的网格自动识别为专利空白区域，同时可以基于贝叶斯理论的反向绘图方法，将空白区域还原成原始数据要素。具体原理是，将 n 维技术特征词矩阵导入 GTM 模型中，实现 n 维矢量至二维空间的降维，使得每项技术数据被映射至一个二维平面中。 其中，生成地图中的特定点表示技术领域内已存在的技术，技术领域内的空白区域即被视为潜在技术机会[14]。在此基础上，通过设置阈值将特征词向量转化为二进制形式，实现 GTM 反向映射到原始技术空间，进而获得空白技术的技术组合进而实现对潜在技术机会的解读。由于其克服了主观识别与解释技术的不足，因此常被用于技术机会识别。

3.2 潜在技术创新机会评估体系

1. 科学-技术关联度（STR）指标

用于衡量某一技术主题在科学研究（期刊论文）与技术应用（专利）之间的协同程度。其核心目标是揭示“基础研究”与“产业应用”的匹配关系，为技术机会识别提供依据。具体公式如下：

$$STR(T_i) = \frac{\sum_{K=1}^{M} p_{论文}（T_i，k）\cdot p_{专利}（T_i，k）}{\sqrt{\sum_{K=1}^{M} P_{论文}（T_i，k）^2}\cdot\sqrt{\sum_{K=1}^{M} P_{专利}（T_i，k）^2}}$$

式中：T_i：第 i 个技术主题；$p_{论文}（T_i，k）$：论文中第 k 篇文档属于主题T_i的概率，范围[0,1]；$p_{专利}（T_i，k）$：专利中第 k 篇文档属于主题T_i的概率，范围[0,1]；M：文档总数（论文和专利合并后的总数）。

2. 技术空白度（TGV）指标

用于衡量某一技术主题在科学文献（论文）与技术应用（专利）中的覆盖差异。其核心目的是识别科学研究领先于技术应用（或反之）的领域，从而揭示潜在的技术转化机会或市场空白。具体公式如下：

$$TGV = \left| \frac{N_{论文}（\omega） - N_{专利}（\omega）}{N_{论文}（\omega） + N_{专利}（\omega）} \right| \times 100\%$$

式中：$N_{论文}（\omega）$：术语ω在论文中标准化频次；$N_{专利}（\omega）$：术语ω在专利中标准化频次。

3. 技术成熟度（TRL）指标

技术生命周期将某项技术的发展过程按照不同特点划分为不同阶段，通常包括导入期、成长期、成熟期和衰退期 4 个阶段。技术生命周期能够帮助人们客观了解某项技术领域的发展阶段和当前发展情况，并判断该技术领域的未来发展趋势，为企业战略布局和产业转型发展提供重要的理论支撑。专利指标法是一种定量和定性相结合的方法，通过引入 4 个专项指标来分析技术生命周期，且相关数据检索较为容易。专利指标法引入技术生长率(ν) 、技术成熟系数(α) 、技术衰老系数(β) 和新技术特征系数(N) 等参数来分析某一技术领域生命周期。技术生长率(ν) 是指某技术领域当年发明专利申请量或授权量占过去 5 年发明专利申请量或授权量总和的比例，技术成熟系数(α) 是某技术领域当年发明专利申请量或授权量占该技术领域当年发明专利和实用新型专利申请量或授权量总和的比例，技术衰老系数(β) 是指某技术领域当年发明专利和实用新型专利申请量或授权量总和占该技术领域当年发明专利、实用新型专利和外观设计专利申请量或授权量总和的比例，新技术特征系数(N)是由技术生长率(ν) 和技术成熟系数(α) 推算而来。依据技术生长率、技术成熟系数、技术衰老系数等参数变化趋势，结合技术生命周期 图示法可以对某一技术领域生命周期进行准确划分。评估指标参数计算公式和含义如表 2 所示:

表 2 评估指标公式及含义

专利指标参数	计算公式	参数说明	含义
技术生长率	$\nu = \frac{a}{A}$	a 是某技术领域当年发明专利申请量或授权量，A 是该技术领域过	若 ν 续几年增大，说明该技术处于生长阶段

		去 5 年发明专利申请量或授权量总和	
技术成熟系数	$\alpha = \frac{a}{a+b}$	b 是某技术领域当年实用新型专利申请量或授权量	若 α 逐年减小，说明该技术处于成熟期
技术衰老系数	$\beta = \frac{a+b}{a+b+c}$	c 是某技术领域当年外观设计专利申请量或授权量	若 β 逐年减小，说明该技术处于衰老期

通过多角度量化分析，精准定位具有创新潜力的技术方向，具体分类原则如图 5 所示：

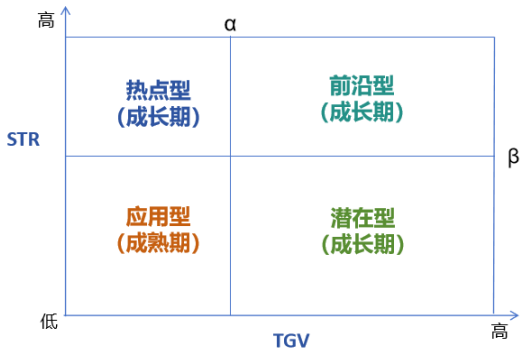

图 5 技术创新机会识别方式

热点型技术主题是 STR>β(阈值)，TGV<α(阈值)，TRL=成长期，此技术主题属于学术和产业都处于活跃阶段；前沿型是 STR>β，TGV>α，TRL=成长期，此技术主题理论研究领先，应用较为滞后；应用型是 STR<β，TGV<α，TRL=成熟期，此技术主题技术成熟，在该技术领域以基本产业化；潜在型是 STR<β，TGV>α，TRL=成长期，早期理论研究已突破，但产业化应用尚未实现。

3.3 实验与分析

3.3.1 数据来源及预处理

1. 数据来源

考虑到现代科学与技术之间的结合越来越紧密，科技发展越来越迅速，本文在进行目标领域技术创新分析时，以期刊文献和专利数据为基础，综合考虑科学和技术两个维度，通过对专利数据和期刊文献进行数据分析和挖掘，为目标领域技术创新机会识别提供相应依据。本研究期刊文献来源于中国知网（CNKI）数据库，检索方式为高级检索，通过查阅相关文献以及定义，设置检

索条件为“(丘陵+山区+山地+低洼+高地) * （农业+农田+田地） * (机械+装备+设备+植保+灌溉+喷雾+耕+收获+收割+施肥+喷施+开沟+播种+插秧+移栽+底盘+行走 +驱动)”，并提剔除“水资源+土壤养分+水利”,同时选择“同义词扩展”，检索时间截至 2024 月 31 日，检索出 1537 件论文。

专利数据来源于 incoPat 专利检索平台。检索策略采用地形、技术关键词、IPC 分类号等相结合方法[15]，合享价值度 9-10 分的专利，同族专利合并，检索时间截至 2024 月 31 日，检索出 4391 件专利。

2. 数据清洗

对得到的期刊文献进行数据清洗，同时为了保证研究的准确性和代表性，剔除新闻报道、会议文献、通知启事和与主题不相关论文等文献，共得到 1338 条数据。对专利文献进行数据清洗，删除与技术主题不相关专利，最终得到 3918 件专利。后续研究将以上述数据为基础展开。

3. 数据预处理

采集完成数据之后，还需要进一步提取这些数据源的标题和摘要部分并对其进行相应的预处理，具体如下：

（1）生成丘陵山区农业机械期刊文献词库和专利词库

利用 Python 的 jieba 函数对丘陵山区农业机械领域期刊和专利进行分词处理，得到该领域的期刊文献词库和专利词库。

（2）分别构建论文和专利的自定义词库和停用词库

首先，通过查阅丘陵山区农业机械相关文献、专著以及专利等资料，对丘陵山区农业机械基本定义和结构进行了解，同时结合中国《农业机械分类》标准对照表，统计丘陵山区农业机专业术语构建该领域自定义词库，对专业术语进行保护；其次，针对专利和文献中常见的虚词（如冠词、介词、连词等）以及无实际意义的、不会为文档提供价值的词汇和标点等设置相应的停用词库，在后续进行数据处理时删除。随后，通过反复的试验和调整，不断改进自定义词库和停用词库，不断提高分词结果的代表性和准确性。

（3）数据清洗及标准化处理

利用停用词表移除数据中的非技术内容，并根据自定义词库、联合国国际粮农组织发布的 AGROVOC 控制词汇库和《农业机械分类》标准对照表统一技

术术语，将期刊文献和专利的词库进行技术术语对齐，进行标准化操作。

3.3.2 丘陵山区农业机械异构数据融合

1. 技术词提取

通过上面对数据预处理，对论文和专利摘要文本进行挖掘，进而识别出丘陵山区农业机械领域的技术特征词。本研究通过 TextRank 算法提取关键词，参数设置窗口大小为 5，阻尼因子为 0.9，迭代次数 100 次，提取关键词数量 100。提取结果如表 3：

表 3 关键词提取结果举例

文献形式	名称	摘要关键词
论文	基于视觉的丘陵整地作业导航方法与避障规划研究	导航,障碍物,旋耕机,整地,检测,避障,田埂,边界,路径,拟合,测量,多项式,视觉感知,自主,路径规划,标记,偏差,图像,分割,颜色,差分法,卫星,航向,精度,绿色植物,分量,立体匹配,横向,不规则,测距,像素,灰度,测量误差,三维建模,窗口,边界点,识别率,距离,运算,相对误差,仿真,切线,逻辑,相机,传感器,双目,误检率,高度,土壤结构,肥力
专利	一种平地机导航方法和系统	平地机,导航,路径规划,控制参数,距离,地势,平整

2. 基于对比学习的 BERT 预训练模型的语义对齐

（1）模型参数设置

本研究使用的 BERT-Base 模型的深层语义表征能力，基于对比学习优化模型，合并论文和专利关键词，构建统一的领域技术术语。模型参数如下表 4：

表 4　BERT-Base 模型参数设置

参数项	配置值
预训练模型名称	BERT-Base
隐藏层数	12
隐藏层维度	768
注意力头数	12
最大序列长	32
激活函数	ReLU 激活函数层
温度系数 τ	0.1
相似度阈值	0.82
批量大小	32
学习率	2e-5

（2）模型架构

① 双塔 BERT 模型：

论文塔：基于 BERT-Base-Chinese，输入论文关键词序列，输出 768 维动态向量。

专利塔：共享 BERT-Base-Chinese 权重，输入专利关键词序列，输出 768 维动态向量。

共享投影层：全连接层（768→256）+ ReLU +全连接层（256→128），输出 128 维语义向量。

② 对比学习机制

损失函数：基于温度缩放（Temperature Scaling）的交叉熵损失，温度系数设为 0.1。

正样本：同一技术主题的论文-专利对（如“履带底盘优化”论文与对应专利）。

负样本：随机采样不同技术主题的论文-专利组合。

3. 运行结果

（1）基于 LDA 主题模型技术特征词识别

通过 Python 搭建 LDA 主题模型，将数据导入后测试发现，聚类结果如图 7 所示，对主题结果进行优化，例如自动（316）、控制系统（288）、控制器（261）、传感器（379）、定位（256）、检测（256）、算法（93）、仿真（212）、软件（211）、导航（53）、智能化（66）、远程控制（21）、路径规划（24）、自动调平（26）、图像处理（23）、控制算法（34）、仿真软件（32）、自动驾驶（8）、人工智能（6）、模糊控制（15）、人机交互（15）、激光雷达（14）等技术词，归纳总结为智能控制技术主题。根据聚类结果并人工调整优化，得到 6 组关于丘陵山区农业机械技术主题，共计 168 个特征词，见表 6。然后对论文及专利摘要文本数据进行处理，构建二进制表示的关键技术词表示向量。当某项论文或专利的摘要文本中包含了所选定的特征词，则该技术词向量中对应的元素值为 1，否则为 0，最终形成论文与专利数据的技术词矩阵。

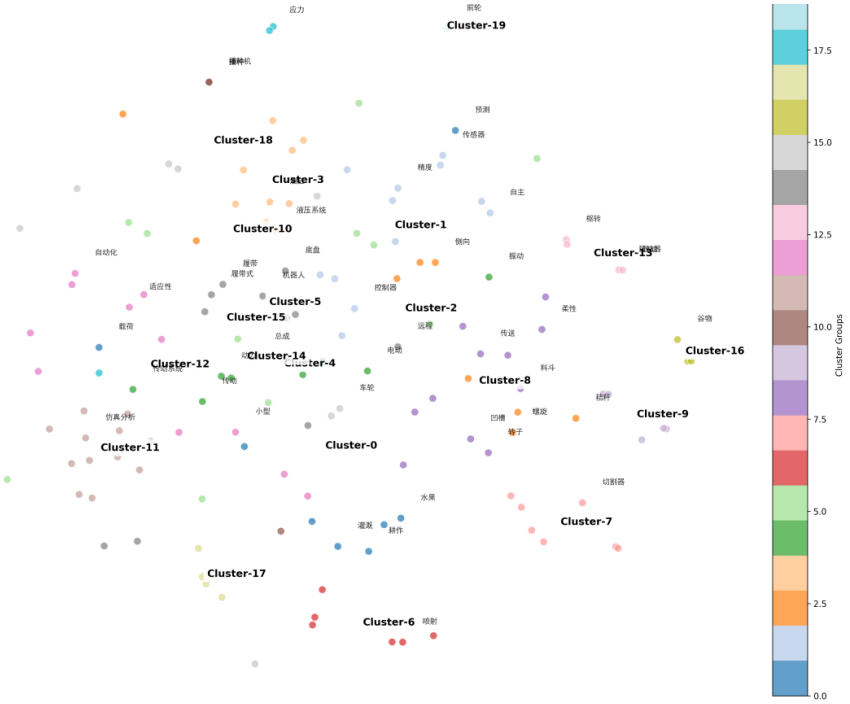

图 6　LDA 主题模型丘陵山区农业机械技术主题聚类图

表 6 丘陵山区农业机械技术主题

技术主题	技术特征词
机械类型方向	小型化农业机械、轻量化机械设计、履带式行走机构、四轮驱动技术、电动农业机械、油电混合动力机械、遥控无人操作机械、山地拖拉机、微型耕整机、坡地播种机、山地收割机、果园管理机械、茶园专用机械、林果采摘机械、药材采收机械、梯田作业机械、履带式旋耕机、自走式喷雾机、山地运输机械、折叠式机械臂、无人机植保、割草机、穴盘育苗移栽、高精度播种器、陡坡果园运输车、山地植保无人机、窄行距播种机、小型谷物干燥机、便携式采茶机、微型挖坑栽树机、多地形割捆机、自适应开沟机、履带式除草机、山地秸秆打捆机、多功能、喷灌机、自走式喷灌机、多功能机、轮式机械、排种器、播种器
动力传动与能源技术	新能源动力、锂电池动力、氢燃料电池、太阳能、能量回收、低排放柴油机、永磁同步电机、变频调速技术、扭矩自适应分配、动力换挡技术、静液压传动（HST）、双离合变速器、能量管理系统、无线充电、混合动力、动力学、硅碳负极锂电池、油电切换的控制算法、光伏-柴油混合架构、扭矩耦合器、电池增效、微电网拓扑结构、拓扑结构、化学气相沉积法、磁吸定位系统、流体力学、车辆动力学、液压系统
地形适应与行走机构	坡度自适应底盘、重心动态调整、多自由度底盘、铰接式转向、差速锁技术、坡道驻车制动、防侧翻预警、地形扫描建模、越障能力优化、软地面通过性、履带式底盘、四轮独立驱动、坡地自适应、仿生机械、防倾覆稳

	定机构、抗振减震、姿态调整、自动调平、地形跟随算法、姿态传感器、仿生柔性结构、动态仿生、粒子阻尼、仿真分析（软件）、动态补偿算法、运动平稳、自动调平
智能控制	导航、激光扫描、多光谱成像、即时定位与地图构建、红外热成像、超声波、定位系统、无人机、激光雷达、机器人、机器视觉、三维语义地图、摄像机、退化补偿算法、地形建模、边缘计算、作物生长监测、无人驾驶、多传感器、智能避障、路径规划、物联网、远程控制、机器学习、无线网络、神经网络、多光谱、故障知识图谱、有限元、离散元法、控制算法、模糊控制、非线性优化模型、人机交互、自动驾驶、图像处理、脉冲神经网络处理器
节能增效与绿色技术	水肥一体化、精准控制、绿色能源驱动、太阳能、精准变量施肥、智能滴灌、农药喷洒、地膜回收机器人、节水灌溉、激光对靶喷雾、精准变量施药、变量施肥、分布式优化方法、多目标优化函数、时序协同
区域特色机械	水稻插秧机、玉米收获机、坡地马铃薯收获机、柑橘采摘平台、油茶果采收机、花椒采摘机械、魔芋挖掘机械、食用菌栽培机械、中药材烘干机、小型茶叶杀青机、蓝莓采摘机、猕猴桃授粉机、甘蔗收割机、咖啡鲜果脱皮、松子采收机械、百合挖掘机械、三七分选装置、核桃清洗设备、板栗去壳机械、竹笋采收机械。

（2）利用 GTM 算法绘制技术地图

本研究采用 20*20 的二维网格作为潜在空间基底，计算 RBF 基函数，通过带正则化的最小二乘优化和高斯径向基函数映射，将获得的技术词矩阵输入到使用 Python 语言编译 GTM 算法中，使得高维技术特征投影到二维空间，从而绘制技术地图。运用论文和专利数据形成的技术地图中体现试验发展阶段的专利技术点（●）、基础研究的论文研究点（★）和交叉点（×）三种情况。图 7 是以丘陵山区农业机械技术为例绘制的技术地图，可以发展部分技术处于基础研究和试验发展阶段并行时期，也有不少技术处于基础研究或者试验发展阶段。

图 7 丘陵山区农业机械技术地图

从图中点的聚集程度可以看出有些比较聚集属于高密度区域，而有些则比较分散，分散区域为低密度区域，而这些很大可能成为未来重点的发展方向，因此可以确定这技术空白点。再将这几个技术点通过 GTM 算法反向映射到原始数据空间（见图 5）。根据技术地图可以获得 7 个低密度区域的技术特征词，从而确定该技术所涉及的领域。

4. 识别结果分析

通过 GTM 对丘陵山区农业机械低密度技术区域的技术点逆向映射识别出的技术创新机会。每一个技术创新机会中可能涉及多个关键技术词，这些关键技术词可能是未来技术创新的关键点，技术研发人员可以通过它们来判断技术未来发展方向，进行技术创新。表 7 为丘陵山区农业机械技术创新机会，如技术空白点 6 涉及的关键技术词有：机器学习、神经网络、路径规划、动力学、姿态调整、动态补偿等技术关键词，根据涉及的关键词，可以分析出未来丘陵山区农业机械研究方向集中于能源动力系统革新、地形自适应智能控制技术升级、智能化运维体系构建和集群作业协同技术突破四个方面。下面对本文识别出的技术进行分析，空白点 2、3、6 涉及到地形自适应及智能控制技术升级，

空白点 4 是对智能化运维的体系构建，空白 1 是对能源动力系统的研究，空白 5 和 7 涉及到集群作业协同技术突破，表 7 为技术空白点涉及的技术关键词。

表 7 技术空白点及技术关键词

序号	技术空白点	技术关键词
1	新型能源混合动力系统	硅碳负极锂电池、油电切换的控制算法、光伏-柴油混合架构、扭矩耦合器、电池增效技术、微电网拓扑结构、化学气相沉积法（CVD）、SEI 膜负极、高能量密度技术
2	智能视觉导航系统	控制算法、机器视觉、姿态传感器、激光雷达、神经网络处理器、退化补偿算法、机器学习、即时定位与地图构建、三维语义地图、多光谱成像、路径规划、边缘计算
3	仿生装置	仿生柔性结构、控制策略、电活性驱动器、三维语义图像、多光谱成像、即时定位与地图构建、目标检测算法、粒子阻尼、动态仿生、传感器
4	模块化系统	定位系统、自补偿密封环、多协议兼容技术、远程控制
5	数字孪生运维系统	物联网、神经网络算法、机器学习、有限元（FEA）、流体力学（CFD）、仿真软件、传感器、机器视觉、故障知识图谱、三维建模
6	地形适应控制算法	机器学习、神经网络、路径规划、动力学、自动驾驶、离散元法（DEM）、地形建模、动态补偿算法、姿态调整、自适应、运动平稳、模糊控制、控制算法
7	协同作业技术	多功能一体化、非线性优化模型、多目标优化函数、时序协同、分布式优化方法、车辆动力学、自适应交替方向、导航、物联网、GNSS、姿态控制

采用科学-技术关联度（STR）、技术空白度（TGV）和技术成熟度（TRL）三维评估模型（图 8）对上面 7 个技术空白点进行量化分析，STR 阈值是 0.8，TGV 阈值是 0.6，可得到表 8 评估结果：

表 8 技术空白点评估结果

技术空白点	**STR**	**TGV**	**TRL**	**技术类型**	**开发策略**
新能源动力系统	0.62	0.65	4	潜在型	优先开发（基础研究强化）
视觉导航	0.88	0.72	5	前沿型	理论突破与原型验证
仿生装置	0.75	0.71	6	潜在型	跨学科应用开发
模块化	0.81	0.55	7	应用型	产业化推广与标准化
数字孪生	0.68	0.81	3	潜在型	重点攻关（基础理论突破）
地形自适应	0.78	0.69	4	潜在型	重点攻关（技术原理验证）
作业协同	0.86	0.57	7	应用型	工程化场景验证

潜在技术四项（STR<0.8 且 TGV>0.6，TRL≦6）：新能源动力系统，数

字孪生运维控制系统、地形自适应控制技术和仿生装置技术。前沿型技术（STR>0.8 且 TGV>0.6，TRL=5）是视觉导航技术；应用型技术有模块化和作业协同技术。

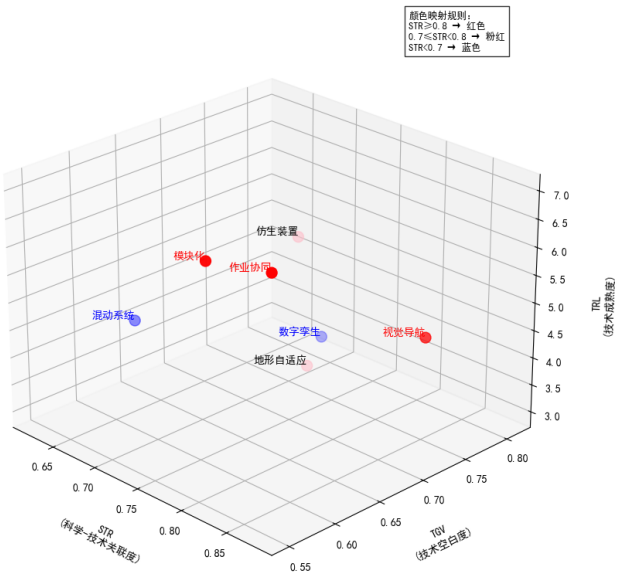

图 8 三维评估模型评估结果

4 结论与建议

在当今技术快速发展的新时期，有效实现技术创新机会识别，有利于国家和企业管理者识别出技术的未来发展方向，从而调整发展战略，为技术竞争占据有利态势。基于此，本研究从论文及专利数据入手，综合运用深度学习、自然语言处理和技术地图等方法实现技术创新机会识别。本研究通过对论文和专利多源异构数据进行语义对齐处理，构建了多源异构数据融合模型，采用基于对比学习的 BERT-Base 文本处理模型，对论文和专利的技术主题进行语义对齐训练，然后利用 LDA 主题模型并结合专家意见识别出要研究的领域的关键技术词，随后运用 GTM 绘制技术地图通过逆向映射实现对技术空白点的识别。并利用科学-技术关联度（STR）、技术空白度（TGV）和技术成熟度（TRL）三维评估模型对技术空白点进行评估，得到最终技术创新机会。最后以丘陵山区农业机械领域为例，验证本研究技术创新机会识别方法。

本研究通过构建基于对比学习机制的 BERT 文本分析模型，在文献主题识别领域取得显著突破。通过构建正负样本对增强语义判别力的创新方法，为人

工智能与大数据分析技术在知识发现领域的应用提供了新的技术路径。在技术机会识别方面，本研究创新性地融合 LDA 主题建模与 GTM 概率可视化算法，构建的多维度技术地图有效整合了论文摘要、专利权利要求等多元数据特征，突破了单一文本特征分析的局限性。构建的"主题强度-技术成熟度-市场关联度"三维评估体系，为我国重点技术领域的创新路径规划提供了科学的决策支持。这套融合深度学习和概率图模型的方法框架，为构建智能化技术创新分析系统提供了重要方法论参考。

本研究在技术创新机会识别维度构建方面仍存在一定局限性，主要体现在数据源的覆盖广度与评估体系的动态适应性两个层面。当前基于学术论文和专利文献的双源分析框架虽能有效捕捉基础研究向应用研究过渡的技术特征，但未能充分整合技术成熟度曲线中的市场需求数据、产业政策文本以及技术转化效益指标，导致技术机会评估体系在产业化适配性方面存在改进空间。后续研究将重点构建"科学-技术-产业"三级联动的动态学习机制：（1）引入产业技术创新图谱中的供应链数据与投融资信息，开发基于多源异构数据的特征融合模块；（2）设计具有时间感知能力的 LDA-GTM 增强算法，通过滑动时间窗机制捕捉技术演进过程中的突变特征；（3）建立包含技术可行性、经济价值性、政策导向性的多维评估矩阵，特别是已启动融合技术标准文件与示范工程报告的多模态分析实验。在实际分析中还需结合更多维度的信息，未来将围绕这个主要问题继续开展研究。

5 项目成果

（1）完成《基于专利视角研究丘陵山区田间管理机械技术演变路径》论文，投稿中。

（2）《丘陵山区农业机械技术创新路径识别——基于多源数据融合的实证研究》撰写中。

6 参考文献

[1] 王晓文,袁寿其,贾卫东.丘陵山区农业机械化现状与发展[J].排灌机械工程学报，2022,40(5):535-540.

[2] 中华人民共和国中央人民政府.中共中央国务院关于印发全国农业现代化规划(2016—2020 年)的通知[EB/OL].https://www.gov.cn/zhengce/content/2016-10/20/content_5122217.htm.

[3] 中华人民共和国中央人民政府.中共中央国务院关于做好 2023 年全面推进乡村振兴重点工作的意见[EB/OL].https://www.gov.cn/zhengce/2023-02/13/content_5741370.htm?dzb=true.

[4] MA J,PORTER A L.Analyzing Patent Topical Information to Identify Technology Pathways and Potential Opportunities[J].Scientometrics,2015,102(1):811-827.

[5] 韩晓彤,朱东华,汪雪峰.科学推动下技术机会发现方法研究[J].图书情报工作，2022,66(10):19-32.

[6] 王理,龚妍芸,陈大明,等.技术机会分析方法研究综述[J].情报探索,2024,(03):128-134.

[7] DEVLIN J，CHANG M W，LEE K，et al. BERT：pretraining of deep bidirectional transformers for lan-guage understanding[C]∥Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics，2019:4171-4186.

[8] VASWANI A，SHAZEER N，USZKOREIT K，et al. Attention is all you need［C］∥Proceedings of the Conference of Neural Information Processing Systems，2017,6000-6010.

[9] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3 (4/5): 993-1022.

[10] 关鹏,王曰芬.科技情报分析中 LDA 主题模型最优主题数确定方法研究[J]. 现代图书情报技术, 2016, (09): 42-50.

[11] 马建红,王晨曦,闫林,等.基于产品生命周期的专利技术主题演化分析［J].情 报学报,2022,41(7):684-691.

[12] Bishop C ,Svensén, M, Williams C .GTM: The Generative Topographic Mapping[J].Neural Computation, 2014, 10(1):215-234.

[13] YOON B，MAGEE C L. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction[J].Technological Forecasting and Social Change，2018，132: 105-117.

[14] Teng F , Sun Y , Chen F ,et al.Technology opportunity discovery of proton exchange membrane fuel cells based on generative topographic mapping[J].Technological Forecasting and Social Change, 2021, 169.

[15] 潘颖,孔宝红,袁寿其,等.中国丘陵山区农业机械专利技术发展态势[J].排灌机械工程学报,2024,42(09):965-972.

7 部分代码

```python

# ==================== 模型架构 ====================
class ProjectionHead(nn.Module):
    """共享投影层"""

    def __init__(self, input_dim, output_dim):
        super().__init__()
        self.net = nn.Sequential(
            nn.Linear(input_dim, out_features: 256),  # 第一个全连接层
            nn.ReLU(inplace=True),  # 非线性激活
            nn.Linear( in_features: 256, output_dim)  # 降维层
        )

    def forward(self, x):
        return self.net(x)


class DualTowerModel(nn.Module):
    """双塔模型架构"""

    def __init__(self):
        super().__init__()
        # 从本地路径加载BERT模型
        self.bert_paper = BertModel.from_pretrained(Config.local_path)
        self.bert_patent = BertModel.from_pretrained(Config.local_path)

        # 共享投影层
        self.projection = ProjectionHead(
            Config.projection_in_dim,
            Config.projection_out_dim
        )
        self.to(Config.device)

    def forward(self, paper_inputs, patent_inputs):
        # 论文塔编码
        paper_outputs = self.bert_paper(
            input_ids=paper_inputs['input_ids'],
            attention_mask=paper_inputs['attention_mask']
```

```python
# ==================== 训练流程 ====================
def contrastive_loss(paper_emb, patent_emb):
    """对比损失函数实现"""
    # 计算相似度矩阵
    similarity_matrix = torch.matmul(paper_emb, patent_emb.T) / Config.temperature

    # 创建目标标签（假设配对样本具有相同索引）
    batch_size = paper_emb.size(0)
    labels = torch.arange(batch_size).to(Config.device)

    # 计算交叉熵损失
    loss = nn.CrossEntropyLoss()(similarity_matrix, labels)
    return loss


def train_model(model, train_loaders, val_loaders=None):
    optimizer = torch.optim.AdamW(model.parameters(), lr=Config.learning_rate)
    scaler = torch.cuda.amp.GradScaler()

    for epoch in range(Config.epochs):
        model.train()
        total_loss = 0

        for batch_idx, (paper_batch, patent_batch) in enumerate(zip(*train_loaders)):
            paper_inputs = {
                'input_ids': paper_batch['input_ids'].to(Config.device),
                'attention_mask': paper_batch['attention_mask'].to(Config.device)
            }
            patent_inputs = {
                'input_ids': patent_batch['input_ids'].to(Config.device),
                'attention_mask': patent_batch['attention_mask'].to(Config.device)
            }
```

```python
# ==================== 技术主题聚类模块 ====================
class TechnologyThemeCluster:
    def __init__(self, n_clusters=6):  # 修改为 6 类
        self.n_clusters = n_clusters
        self.lda = LatentDirichletAllocation(n_components=n_clusters, random_state=42)

    def cluster_texts(self, texts):
        """执行聚类并返回主题标签"""
        vectorizer = CountVectorizer()
        X = vectorizer.fit_transform(texts)
        clusters = self.lda.fit_predict(X)
        return clusters

    def get_theme_names(self, texts, clusters, n_keywords=3):
        """为每个簇生成代表性关键词"""
        theme_names = []
        vectorizer = CountVectorizer()
        X = vectorizer.fit_transform(texts)
        feature_names = vectorizer.get_feature_names_out()
        for i in range(self.n_clusters):
            # 获取簇内文本
            cluster_texts = [texts[j] for j in range(len(texts)) if clusters[j] == i]
            if not cluster_texts:
                theme_names.append('')
                continue
            # 提取主题的关键词
            topic = self.lda.components_[i]
            top_keyword_indices = topic.argsort()[-n_keywords:][::-1]
            theme = [feature_names[idx] for idx in top_keyword_indices]
            theme_names.append(';'.join(theme))
        return theme_names

```

```python
# -------------------- GTM算法实现 --------------------
class GTM:
    def __init__(self, n_components=20, sigma=5.0):
        self.n_components = n_components
        self.sigma = sigma

    def fit(self, X):
        self.X = X
        self.n_samples, self.n_features = X.shape

        # 扩展潜在空间范围
        self.grid = np.meshgrid( *xi: np.linspace(-3,  stop: 3, self.n_components),
                                 np.linspace(-3,  stop: 3, self.n_components))
        self.grid = np.vstack([self.grid[0].ravel(), self.grid[1].ravel()]).T

        # 初始化映射矩阵
        self.mapping_matrix = np.random.rand(self.grid.shape[0], self.n_features)

        # 计算RBF基函数
        distances = cdist(X, self.mapping_matrix)
        self.phi = np.exp(-distances ** 2 / (2 * self.sigma ** 2))

        # 计算映射矩阵
        self.W = np.linalg.lstsq(self.phi, X, rcond=None)[0]
        return self

    def transform(self, X):
        distances = cdist(X, self.mapping_matrix)
        phi = np.exp(-distances ** 2 / (2 * self.sigma ** 2))
        return phi @ self.W[:, :2]

# -------------------- 统一模型拟合 --------------------
combined_matrix = np.vstack([papers_tech_matrix, patents_tech_matrix])
gtm = GTM()
gtm.fit(combined_matrix)

papers_gtm = gtm.transform(papers_tech_matrix)
patents_gtm = gtm.transform(patents_tech_matrix)
```