

基于BERTopic-LSTM模型的 新兴技术主题预测与价值评估*

——以中国人形机器人为例

王珩 杨坤 殷涛

(上海工程技术大学管理学院, 上海 201620)

摘要: 科学研判新兴技术的演化路径对国家战略科技布局具有重要的前瞻启示价值。针对现有研究细粒度主题捕捉能力不足、时序特征融合欠缺以及评估维度单一等局限,本研究构建了“技术主题识别—新兴技术预测—潜力价值评估”分析框架,旨在为技术优先排序与资源配置提供更为精准、可解释的决策依据。首先,提出考虑语义演化与时序融合的BERTopic-LSTM技术预测模型,克服传统模型难以兼顾文本深层语义与演化连续性的难题;其次,设计二维评估矩阵,透视新兴技术的成长动能与生态势能,系统识别其潜在战略价值;最后,以人形机器人领域为例进行实证分析。结果显示:BERTopic-LSTM模型在海量专利数据中识别出的技术主题更具细粒度,与其他模型对比,在均方根误差、平均绝对误差及 R^2 等指标上均表现更优,预测结果与现实技术迭代轨迹高度契合。研究为政府与企业复杂环境下的创新管理决策提供了科学的理论支撑与工具参考。

关键词: 新兴技术; 技术预测; BERTopic; LSTM; 人形机器人

中图分类号: G350 **DOI:** 10.3772/j.issn.1673-2286.2025.12.006

引文格式: 王珩, 杨坤, 殷涛. 基于BERTopic-LSTM模型的新兴技术主题预测与价值评估:以中国人形机器人为例[J]. 数字图书馆论坛, 2025, 21(12): 48-58.

新兴技术是新质生产力的核心驱动力,加速其产业化培育是重塑全球产业格局与科技安全的关键着力点^[1]。在此背景下,前瞻性识别高潜力新兴技术主题,助力高新技术企业把握前沿趋势、制定科学的技术投资策略,并结合技术活跃周期动态掌握领域发展动向,已成为推动新兴产业高质量发展的关键课题。

新兴技术的形成可视为知识在吸收、成长、扩散3个阶段中持续流动的动态过程^[2]。通过刻画知识吸收广度、增长速率与扩散范围等属性,可实现对新兴技术的有效识别^[3]。在知识流动驱动技术演化的过程中,技

术机会不断被激发,演进轨迹通过专利引文等载体得以显性呈现^[4-5]。此外,现有研究有必要综合考虑新兴技术对经济与社会的影响,从多维计量指标出发,以精细化方式识别具有高潜力或变革性的技术主题^[6]。专利文本作为重要的知识载体,蕴含丰富的技术语义结构与演化信息,借助机器学习方法挖掘其语义特征与发展趋势,有助于揭示技术主题的成长路径与扩展方向,为前瞻性识别技术潜力提供方法支撑^[7]。

鉴于新兴技术源于领域内知识要素的持续流动与重组,通过驱动系统演化而实现整体性涌现的特征,本

收稿日期: 2025-10-16

*本研究得到教育部人文社会科学研究规划基金“‘知识重组-场景重构’情境下数字创新空间中集群行为的演化、涌现及调控研究”(编号: 22YJA630104)资助。

研究从语义层面挖掘其潜在价值,融合专利文本语义结构与时序建模方法,构建新兴技术主题预测与价值评估模型。该模型通过多维指标刻画技术演化特征,系统评估不同技术主题的战略潜力,为前瞻性技术布局与决策提供理论依据与分析工具。

1 相关研究

1.1 新兴技术识别

“新兴技术”最早由乔治·戴等提出,定义为尚处于发展初期但具备显著颠覆性与战略潜能的技术^[8]。现有研究将新兴技术内涵划分为广义与狭义两个层次^[9]:广义上指处于发展初期、尚未形成主导市场的前沿技术;狭义上则聚焦于特定学科内部具体技术主题的演化,这一内涵界定支持了通过细化技术词汇的处理颗粒度来捕捉萌芽阶段微观动向的研究路径,为技术识别提供了理论基础。新兴技术的识别方法正经历由定性判断向定量建模、由专家主导向数据驱动的系统性演进。早期研究多依赖德尔菲法、情景分析和技术路线图等专家经验主导的前瞻方法,主观判断预测技术发展路径^[10],虽在战略层面具有一定启发意义,但由于其方法主观性强、可重复性差,难以充分适应技术系统日益复杂与快速演化的动态特征^[3]。21世纪以来,新兴技术识别逐渐形成以科技文献与专利数据为核心的数据驱动分析范式。在方法体系上,主流路径可归纳为3类^[11]:基于文献计量的统计分析方法、基于引文或合作网络的结构分析,以及基于自然语言处理与文本挖掘的主题建模。然而,上述方法仍存在明显不足:①对技术主题的语义建模多依赖浅层文本处理手段,难以捕捉专利语料中的上下文关联与细粒度差异;②在动态预测方面,对技术演化时间序列的深层建模仍显不足,难以揭示新兴技术的发展轨迹与潜在爆发点。针对既有研究在复杂语义理解与动态演化测度方面的不足,本研究结合深度语义建模与时序学习方法,兼顾语义表示与演化趋势的双重建模需求,提升主题识别的精度与自动化水平^[12-13]。

1.2 新兴技术主题预测

在新兴技术主题预测领域,已有研究基于专利和科技文献构建了多元化的模型和指标体系,涵盖技

术趋势识别、成熟度评估与潜力挖掘等多个方面^[14]。相关研究通过构建多维指标体系或引入聚类与映射方法,对技术演化方向与潜在突破领域进行了系统识别^[15]。近年来,随着研究持续深入,预测方法逐步向多源数据融合与跨领域建模拓展^[16],通过整合专利、论文及宏观信息提升预测的系统性与决策支持能力^[17],但仍存在一定局限性:一方面,多数研究侧重于静态趋势识别,缺乏对技术演化路径的动态建模与结构性刻画;另一方面,预测结果与产业战略规划及政策导向的衔接不足,难以有效支撑技术优先级排序与战略资源配置。因此,本研究在融合语义理解与动态建模的基础上引入BERTopic-LSTM模型,作为技术主题识别与趋势预测的核心工具。其中:BERTopic模型凭借深度双向语义建模能力,能够有效捕捉专利文本中的上下文关联关系,挖掘潜藏于复杂语料中的细粒度技术主题;LSTM模型在处理非线性时间序列方面具有显著优势,能够刻画技术主题在长期演化过程中的动态变化。

1.3 潜力价值评估

潜力价值评估是对新兴技术未来演化的可持续性、扩散影响力与战略贡献进行前瞻性判别与量化表征的分析过程。理论层面,新兴技术潜力价值主要体现在技术成长导向与生态嵌入导向两个方面:前者侧重从时间序列中描绘技术生命周期中的加速增长阶段及知识产出增长节奏^[18];后者通过技术关联网络或知识溢出机制探讨技术在跨领域、跨产业体系中的生态影响力^[19]。因此,技术成长动能与生态势能在内涵上并不割裂,二者共同揭示新兴技术潜力形成的双源机制。方法层面,尽管部分研究尝试从技术成熟度^[20]、技术覆盖范围^[21]、热点性^[22]、连贯性^[23]等角度评估新兴技术的外部表现,但存在两个方面的不足:①缺乏与语义识别或时间序列建模的融合,难以实现“识别—预测—评估”的联动分析;②评价结果多呈现为单一打分或静态排序,缺乏结构化分析框架,难以支撑技术主题的战略分类与政策响应。鉴于此,本研究从成长动能与生态势能双维视角出发构建二维评估矩阵,作为对识别结果进行战略价值赋能的结构化机制。其中:成长动能维度侧重刻画技术主题在时间序列中的趋势变化,反映其内部增长性;生态势能则衡量技术主题在技术网络中的影响程度,反映其外部系统渗透力。本研究旨在

通过评估矩阵,将技术主题映射到“成长动能×生态势能”二维矩阵中,实现对不同类型新兴技术的战略分层,为技术优先级划分及政策扶持路径制定提供决策依据。

2 研究设计

2.1 研究框架

基于上述分析,本研究融合BERTopic与LSTM模型构建“技术主题识别—新兴技术预测—潜力价值评估”分析框架,以应对新兴技术识别中的复杂性与动态性挑战。研究框架如图1所示,主要包含技术主题识别、新兴技术主题预测及预测结果的潜力价值评估3个阶段:①基于专利数据开展技术主题的结构化识别,凝

练具有战略潜质的技术核心方向;②综合新颖性指标与时序特征,刻画技术主题的成长轨迹与演化规律,预测具备持续驱动效应的新兴技术群体;③通过构建成长动能与生态势能二维评估矩阵,从双维视角评估技术主题潜力价值。

2.2 基于BERTopic模型的技术主题识别

2.2.1 数据获取及预处理

本研究以专利数据库为核心数据源,围绕研究对象设定多组检索策略,通过关键词匹配、国际专利分类号(International Patent Classification, IPC)限制及时间窗口约束,提取专利文本的标题、摘要、权利要求与技术分类等核心字段,构建原始技术文本语料。

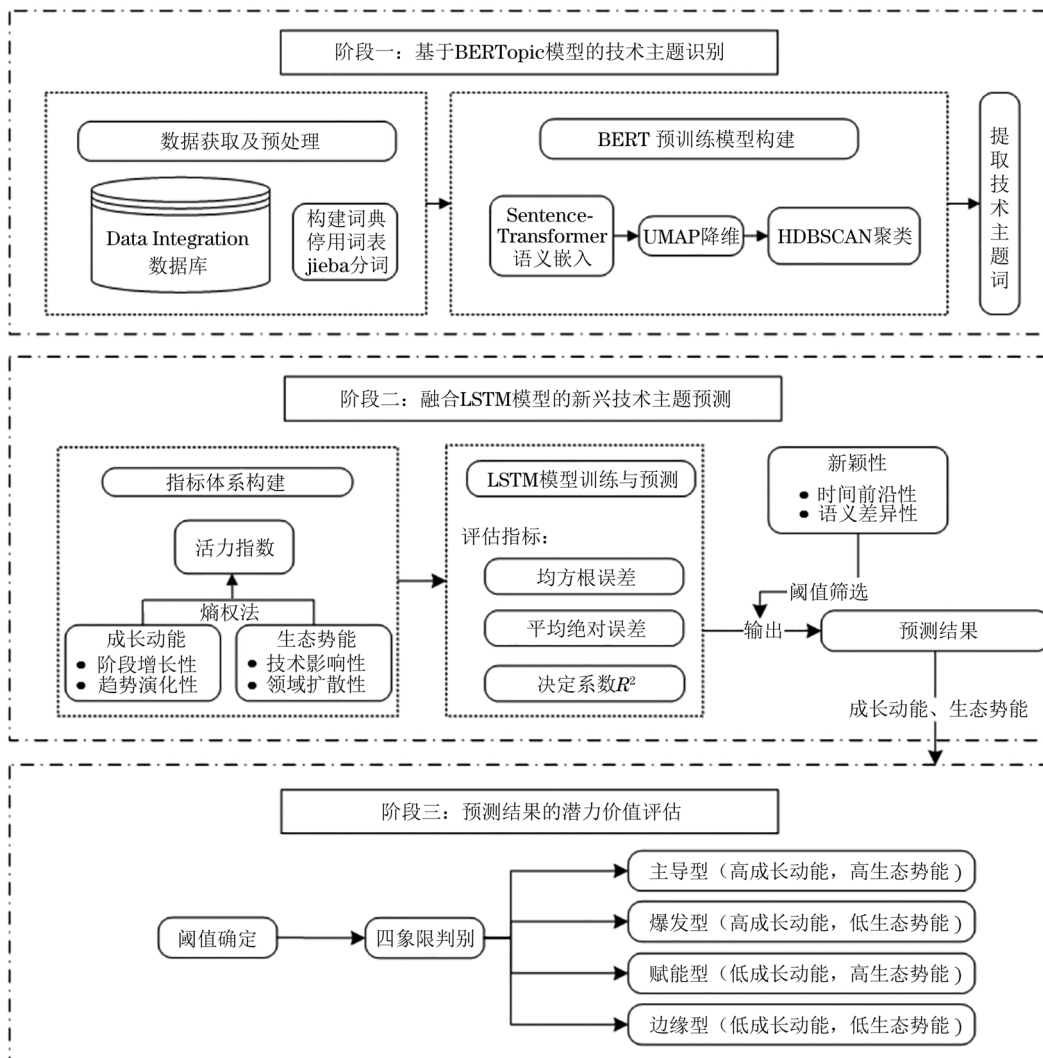


图1 研究框架

在此基础上开展数据预处理工作：首先，进行数据清洗，剔除冗余、缺失与无效记录，保障数据质量；其次，针对专利文本专业术语结构复杂、复合词高频等特点，融合WIPO Pearl术语库、IPC分类手册及行业白皮书内容，构建领域定制化词典，并结合专家意见形成术语库，同时建立停用词表，以过滤无效信息，对于长度超过512个字符的文本进行截断，以确保输入一致性；最后，使用jieba工具完成中文分词与停用词过滤，提升文本处理质量。

2.2.2 BERTopic预训练模型构建

BERTopic模型在语义表示与上下文建模方面表现卓越，能够有效捕捉专利文本中的隐含语义结构与主题关联，减少人工标注过程中的主观偏差。首先，采用Sentence-Transformer模型进行语义嵌入，为技术主题识别与后续时序预测奠定数据基础。其次，利用

CountVectorizer并结合停用词表对专利文本进行初步特征提取，并通过UMAP算法进行非线性降维以保留核心语义结构。在聚类阶段，基于降维后的嵌入向量采用HDBSCAN算法进行无监督密度聚类。最后，采用c-TF-IDF方法提取各主题簇的关键词，并通过人工语义解读对各主题进行标注与命名，确保结果的可解释性。在此设置下，每篇专利文本被赋予一个主要主题，未归入稳定聚类的文本则自动标记为“-1”（噪声）。

2.3 融合LSTM模型的新兴技术主题预测

2.3.1 新兴技术活力指数构建

构建新兴技术活力指数作为新兴技术的特征指标，具体由成长动能和生态势能两个部分构成（指标详情如表1所示），采用熵权法确定二者权重并融合为动态活力分量，以此为LSTM预测模型的输入特征变量。

表1 新兴技术活力评价指标体系

一级指标	二级指标	计算方式	含义
成长动能	阶段增长性	对主题在当前年份的文档频率与上一年份的文档频率进行比值换算，得到反映年度增长幅度的指数	反映该主题在相邻年份中的增长趋势，衡量主题当前处于上升、平稳或下降的阶段
	趋势演化性	对相邻年份文档频率增长率取平均值，从而得到发展速度的量化指标	衡量主题在整个研究周期内的总体增长速度，反映新兴技术扩张和扩散的速度
生态势能	技术影响性	该主题下所有文档被引次数的平均水平，即主题篇均被引次数	衡量主题内专利的平均影响力，值越高说明该主题的研究成果越受关注
	领域扩散性	该主题所含文档平均涉及的IPC数量，即主题篇均IPC数	衡量该主题所覆盖的技术领域广度，反映该主题的多元化程度与技术渗透性

(1) 成长动能。成长动能用于刻画主题在时间序列中的扩张速度与趋势变化，由阶段增长性与趋势演化性构成。阶段增长性反映主题在相邻年份的文档频率变化幅度，能够揭示其短期波动与增长趋势，计算方法如式(1)所示；趋势演化性通过计算研究周期内相邻年份文档频率增长率的平均值，刻画主题整体发展速度与规模扩张水平，计算方法如式(2)所示；成长动能计算公式如式(3)所示。

$$I_{PG} = \frac{P_t}{1 + P_{t-1}} \quad (1)$$

$$I_{TE} = \frac{\sum_T (n_{t+1} - n_t) / n_t}{T} \quad (2)$$

$$I_{GI} = a \cdot I_{PG} + b \cdot I_{TE} \quad (3)$$

式中： I_{PG} 、 I_{TE} 分别表示阶段增长性、趋势演化性； P_t 表示主题在时间 t 的文档频率； n_t 表示技术簇在时间 t 的专利数量； T 表示研究的时间跨度； I_{GI} 表示成长动

能； a 、 b 表示由熵权法确定的权重。

(2) 生态势能。生态势能用于度量主题在技术领域的影响程度与扩散范围，由技术影响性与领域扩散性构成。研究指出，专利被引次数与其知识扩散能力及在创新领域的应用广度密切相关，高被引次数意味着更高的生态影响力^[24]。因此，以篇均被引次数衡量主题文献在技术网络中的技术影响性；以篇均IPC数衡量主题技术成果在专利分类体系中的领域扩散性。生态势能计算公式如式(4)所示。

$$I_{PI} = c \cdot I_{PC} + d \cdot I_{PA} \quad (4)$$

式中： I_{PI} 表示生态势能； I_{PC} 表示主题篇均被引次数； I_{PA} 表示主题篇均IPC数； c 、 d 表示由熵权法确定的权重。

(3) 活力指数。为综合反映主题的价值潜力，对成长动能与生态势能进行组合，并对结果归一化处理，得到活力指数，计算公式如式(5)所示。

$$I_{EV} = \frac{I_{GI} \cdot \ln(1 + I_{PI}) - \text{Min}(I_{GI} \cdot \ln(1 + I_{PI}))}{\text{Max}(I_{GI} \cdot \ln(1 + I_{PI})) - \text{Min}(I_{GI} \cdot \ln(1 + I_{PI}))} \quad (5)$$

式中: I_{EV} 表示活力指数; $\text{Min}(\cdot)$ 和 $\text{Max}(\cdot)$ 分别表示所有主题集合中, 各主题 $I_{GI} \cdot \ln(1 + I_{PI})$ 计算结果的最小值和最大值。

2.3.2 LSTM模型训练

以各技术主题在连续年份中的特征指标为输入的基础数据, 构建年度时间序列数据集, 并对特征变量进行标准化处理以提升模型稳定性与泛化能力。采用滑动窗口法构建训练样本, 以连续3年的特征序列预测未来2年的值, 实现对中短期技术演化趋势的刻画。数据集按7:3划分训练集与验证集。模型结构采用两层堆叠式LSTM网络, 分别设置64个和32个神经元, 激活函数均为ReLU, 并在各层后引入Dropout层, 以缓解过拟合风险, 输出层为维度数为2的全连接层。为验证预测性能, 进一步选取BP神经网络和支持向量机(Support Vector Machine, SVM)作为基线进行对比分析。

2.3.3 结果筛选

新颖性指标作为对预测结果的筛选工具, 旨在从已完成动态演化建模的技术主题中进一步筛选出具备原创潜力的新兴技术方向。新颖性是技术主题在知识结构与语义路径中具有突破性的重要表征, 体现为相对于既有知识体系的偏离程度与重组能力。较高新颖性的技术往往处于知识图谱的边缘, 甚至表现为“离群点”形式。此类技术虽尚未进入主流视野, 但其凭借对既有路径依赖的摆脱及异质知识的重组特征, 打破了常规技术轨迹束缚, 从而演化为孕育新技术轨道的逻辑起点^[25]。新颖性与影响力并非线性相关的, 其转化效果受到研发主体、产业适配性与知识网络位置等多重因素调节^[26]。因此, 在识别新兴技术时需要特别强调新颖性指标对未来创新突破的前瞻价值。

本研究将新颖性拆解为时间前沿性与语义差异性两个维度: 前者反映技术主题在时间维度上的近期活跃度, 体现其是否处于研究浪潮的前沿; 后者度量该主题与现有主流研究路径之间的内容差异程度, 捕捉其在语义结构上的原创性。

时间前沿性用某主题文档的篇均发表年份与该主

题首次出现的年份之差来计算, 反映该主题文档的时间集中度, 计算方法如式(6)所示, 差值越大, 表明该主题研究越集中在近年, 具有较强的前沿性。

$$I_{TN} = \bar{t}_{pub} - t_{first} \quad (6)$$

式中: \bar{t}_{pub} 表示该主题文档的篇均发表年份, t_{first} 表示主题首次出现的年份。

语义差异性用1减去某主题与其他主题的平均语义相似度, 反映该主题与既有主题的差异程度, 计算方法如式(7)所示, 差值越大, 表明该主题与既有主题差异越大, 反映其具有更高的原创性。

$$I_{SD} = 1 - \bar{S} \quad (7)$$

式中: \bar{S} 表示该主题的平均语义相似度。

两个维度等权构成新颖性指标, 作为技术主题进入后续潜力价值评估的关键筛选依据, 其计算方法如式(8)所示。

$$I_{NI} = \alpha \cdot I_{TN} + \beta \cdot I_{SD} \quad (8)$$

式中: I_{NI} 表示新颖性; α 和 β 分别表示两个指标的权重系数, 参考相关研究惯例并结合专家评估建议, 取 $\alpha=\beta=0.5$, 以综合反映各维度的同等重要性。

2.3.4 模型评估

为全面衡量LSTM模型的预测性能与拟合能力, 选取均方根误差(Root Mean Square Error, RMSE)、平均绝对误差(Mean Absolute Error, MAE)以及决定系数 R^2 这3项指标对模型进行评估, 其具体计算公式如式(9)~式(11)所示。

$$E = \sqrt{\frac{1}{n} \sum_{i=1}^n (\tilde{y}_i - y_i)^2} \quad (9)$$

$$A = \frac{1}{n} \sum_{i=1}^n |\tilde{y}_i - y_i| \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

式中: E 为RMSE; A 为MAE; \tilde{y}_i 为预测值; y_i 为实际值; \bar{y} 为实际值的均值; n 为预测次数。

2.4 预测结果的潜力价值评估

通过LSTM对成长动能与生态势能进行中短期预测, 并构建二维评估矩阵(横轴为成长动能, 纵轴为生

态势能), 实现对结果主题潜力价值评估。采用上四分位阈值进行分区: 将“高”定义为处于样本分布75%分位及以上, 将“低”定义为低于该分位数。

2.4.1 阈值定义与计算域

为避免前视偏差并保障跨期可比, 设定动态预测阈值, 具体计算逻辑如式(12)~式(13)所示。

$$I_{GI}^{0.75} = Q_{0.75} \left\{ \widetilde{I}_{GI}^{k,t} : k \in K, t \in T_{\text{predict}} \right\} \quad (12)$$

$$I_{PI}^{0.75} = Q_{0.75} \left\{ \widetilde{I}_{PI}^{k,t} : k \in K, t \in T_{\text{predict}} \right\} \quad (13)$$

式中: $\widetilde{I}_{GI}^{k,t}$ 、 $\widetilde{I}_{PI}^{k,t}$ 为模型预测得到的成长动能与生态势能的平均值; K 为主题集合; T_{predict} 为预测期年份集合。

2.4.2 四象限判别规则

对任意主题 k 在年份 t 的位置进行分类, 分类规则如式(14)所示。主导型主题在成长动能与生态势能两个维度上均处于高位, 表明其已形成成熟的技术体系并在应用场景中实现广泛部署; 爆发型主题在成长动能维度表现活跃, 具备快速扩张潜力, 但生态势能相对有限; 赋能型主题生态势能高, 但成长动能相对有限, 处于相对稳定的发展阶段; 边缘型主题在两个维度上的表现均不突出, 短期内面临较高的不确定性与演化风险。

$$C_{k,t} = \begin{cases} \widetilde{I}_{GI}^{k,t} \geq I_{GI}^{0.75} \cap \widetilde{I}_{PI}^{k,t} \geq I_{PI}^{0.75}, \text{主导型} \\ \widetilde{I}_{GI}^{k,t} \geq I_{GI}^{0.75} \cap \widetilde{I}_{PI}^{k,t} < I_{PI}^{0.75}, \text{爆发型} \\ \widetilde{I}_{GI}^{k,t} < I_{GI}^{0.75} \cap \widetilde{I}_{PI}^{k,t} \geq I_{PI}^{0.75}, \text{赋能型} \\ \widetilde{I}_{GI}^{k,t} < I_{GI}^{0.75} \cap \widetilde{I}_{PI}^{k,t} < I_{PI}^{0.75}, \text{边缘型} \end{cases} \quad (14)$$

3 实证研究

3.1 研究对象

人形机器人作为典型的多学科交叉融合技术, 正逐步成为引领未来产业转型和国家科技战略布局的重要方向。其在具身智能、人机协同和自主决策等方面的突破被视为推动智能系统范式转型的关键动力, 受到全球科技强国的高度重视^[27]。因此, 基于专利数据对

人形机器人领域开展新兴技术方向预测, 对于识别潜在战略制高点具有重要意义。

3.2 数据来源及预处理

本研究数据来源为全球知识产权服务商incoPat专利数据库, 检索范围定为中国发明授权专利, 主要有以下考虑: ①研究聚焦中国技术生态内部的战略潜力识别, 基于中国区域的数据更具针对性、解释力与政策相关性; ②研究聚焦模型与方法验证, 后续可扩展至全球数据以增强外部适用性。设定检索词TIABC=(“四足机器人”OR“躯干”OR“仿人机器人”OR“仿生机器人”OR“仿人”OR“机器人头”OR“仿人智能机器人”OR“手臂舵机”OR“机器人脚”OR“躯体”OR“机器人足”OR“大腿关节”OR“小腿关节”OR“机器人腿”OR“胯关节”OR“步行机器人”OR“仿生机械”OR“六足机器人”OR“躯”OR“四肢”OR“身躯”OR“上肢”OR“手臂”OR“腿部”OR“双足机器人”OR“仿人机械手”OR“关节动作”OR“双足步行机器人”OR“关节运动”), 检索时间范围为2015—2024年, 合计检索专利数据32 816条, 形成初始专利集。依次下载专利及所属IPC类别、引证次数、被引次数、权利要求数量、发明人数量、简单同族数量等指标, 形成原始样本库。随后, 遵循前述预处理流程对数据进行清洗与分词, 最终构建出用于后续分析的结构化研究数据集。

3.3 技术主题识别结果

在语义嵌入阶段采用paraphrase-multilingual-MiniLM-L12-V2模型, 生成维度数为384的向量表示。实验过程中的主要参数配置如表2所示。

表2 关键参数配置

模块	参数	配置
UMAP降维	邻近样本数	40
	降维维度	10
	最小间距	0.05
	随机种子	30
	距离度量	Cosine
HDBSCAN聚类	最小簇大小	100
	最小样本数	15
	距离度量	Euclidean

为探究参数对主题识别结果的影响,选取BERTopic模型中的两个关键参数:UMAP的邻近样本数与HDBSCAN的最小簇大小进行敏感性测试。以3组参数组合进行对比实验,观察主题数及噪声专利数的变化情况,3组参数的敏感性实验结果如表3所示。

表3 参数敏感性实验结果

邻近样本数	最小簇大小	主题数/个	噪声专利数/个	评估结果
40	100	25	9 603	较优平衡
20	50	70	14 683	噪声回升
60	150	42	12 495	过度碎片化

由表3可知,主题识别对建模参数具有一定敏感性,邻近样本数与最小簇大小的增加虽能提升主题数,但可能引入更多离群样本并影响主题语义聚合度。综合考虑主题粒度与语义聚合度,选取表2所示的参数组合作为最终配置。为概括各主题的核心语义特征,提取每一主题中c-TF-IDF得分排名前10的关键词,并据此结合专利文本进行主题命名。命名过程由研究团队背靠背独立完成并复核,同时借助不同的人工智能大模型(GPT-4o、DeepSeek与Gemini)进行对比验证,确保语义标签具备可解释性与准确性。本次实验共识别出25个技术主题,部分结果示例如表4所示。

表4 技术主题关键词分布(部分)

主题编号	主题名称	关键词
Topic 0	机械臂安装驱动组件	装置、机构、安装、连接、手臂、固定、组件、用于、驱动、机械
Topic 1	机器人关节运动控制系统	机器人、运动、关节、控制、驱动、机构、连接、结构、实现、模块
Topic 2	下肢康复训练固定机构	训练、固定、锻炼、连接、支撑、患者、腿部、机构、装置、调节
Topic 3	人体图像识别与动作分析	图像、人体、数据、目标、识别、信息、进行、模型、获取、动作
Topic 4	纳米催化材料制备	制备、材料、催化剂、纳米、得到、溶液、反应、躯体、前驱、金属
Topic 23	多旋翼无人飞行平台	无人机、飞行器、飞行、旋翼、翅膀、控制、机械臂、扑翼、降落、机身
Topic 24	吸收材料结构设计	吸收性、区域、吸收、物品、弹性、制品、纵向、接合部、边缘、具有

3.4 新兴技术主题预测结果

3.4.1 模型评估

对已识别的技术主题进行动态评估,各指标的权

重计算结果如表5所示。基于TensorFlow平台构建两层堆叠式LSTM网络用于新兴技术预测,采用Adam优化器(学习率为0.001)与tanh激活函数,训练轮数为100,批次大小为32,并引入Early Stopping机制以防止过拟合。为验证预测性能,在相同训练数据划分与主要参数条件下与BP神经网络和SVM进行对比。结果表明,LSTM在RMSE、MAE和 R^2 等指标上均优于对比模型, R^2 达到0.549 4,能够更有效捕捉技术演化过程中的非线性特征。最终各模型预测评估结果如表6所示。

表5 各指标权重结果

指标	熵值	差异系数	权重
阶段增长性	0.817 3	0.182 7	0.410 2
趋势演化性	0.937 8	0.062 2	0.139 5
技术影响性	0.827 0	0.173 0	0.388 3
领域扩散性	0.972 4	0.027 6	0.062 0

表6 各模型评估结果

模型	RMSE	MAE	R^2
SVM	0.084 1	0.021 2	0.084 1
BP	0.034 8	0.018 2	0.451 6
LSTM	0.032 9	0.016 7	0.549 4

3.4.2 结果筛选

基于训练优化后的LSTM模型,对已识别技术主题2025—2026年的新兴技术活力指数进行预测,以刻画其演化趋势,并以新颖性指标的75%分位数为阈值,筛选出在人形机器人领域中兼具较高活力与创新潜力的新兴技术方向,具体结果如表7所示。

表7 新兴技术主题预测结果

主题编号	主题名称	新颖性	活力指数	活力指数排序
Topic 2	下肢康复训练固定机构	2.364 1	0.045 4	6
Topic 3	人体图像识别与动作分析	2.540 9	0.055 0	1
Topic 7	自动采血与止血装置	2.517 1	0.046 4	4
Topic 8	实验动物固定与操作装置	2.452 9	0.044 9	7
Topic 11	智能车辆驾驶辅助系统	2.544 4	0.051 2	2
Topic 13	VR用户交互与动作捕捉	2.687 2	0.049 7	3
Topic 16	机械臂轨迹控制系统	2.684 0	0.046 0	5

通过对比全样本数据发现, 新颖性位于75%分位数以上的技术主题, 其活力指数均位列全体样本的前50%序列, 验证了新颖性作为识别潜在新兴技术前置性指标的有效性。高新颖性主题在成长动能与生态势能两个维度上呈现协同提升特征, 表明其不仅具备初期原创性优势, 在后续技术演化中也展现出较强的发展韧性与扩散能力。也就是说, 具备较高语义异质性与结构突破性的技术主题往往能够沿新的知识路径持续演化, 实现从概念创新向系统嵌入的动态跃迁。

3.5 潜力价值评估及分析

3.5.1 潜力价值评估

基于对预测期(2025—2026年)技术主题演化趋势的分析, 结合新颖性指标, 最终筛选出7个代表性新兴技术主题并划分至不同象限(见图2): Topic 3属于主导型(高成长动能-高生态势能), Topic 2、Topic 11和Topic 13为爆发型(高成长动能-低生态势能), Topic 7、Topic 8与Topic 16为边缘型(低成长动能-低生态势能)。

3.5.2 结果验证

上述技术主题在2025—2026年均保持较高活力水

平, 表明其在短期内具备持续演化与应用外溢的潜力。为进一步增强识别结果的可信度与稳健性, 本研究从政策导向与产业发展趋势两个维度开展交叉验证。

从政策导向来看, 国家及地方政府在康复医疗、人工智能、智能制造等重点领域已相继出台多项前瞻性政策。例如: 《关于调整〈浙江省残疾人货币补贴辅助器具目录〉部分产品和补贴标准的通知》将外骨骼康复训练器纳入智能辅助器具补贴目录, 明确支持康复设备的个性化与智能化转型(Topic 2); 国家药监局发布《医疗器械技术审评中心创新医疗器械特别审查申请审查实施细则》, 开辟高端医疗器械技术审查绿色通道(Topic 7); 《上海市推进养老科技创新发展行动方案(2024—2027年)》中明确提出发展动作识别、人机交互与具身智能模型(Topic 3、Topic 13); 《新能源汽车产业发展规划(2021—2035年)》从顶层设计层面强化相关代表领域的战略地位(Topic 11、Topic 16)。

从产业趋势来看, 结合权威机构发布的行业研究报告, 康复医疗、动作识别、VR/AR与智能制造等行业均进入高速发展阶段。例如, 2024年我国康复医疗服务市场规模已达1 200亿元, 康复机器人、VR康复训练系统等智能设备渗透率将从2024年的8%提升至2030年的25%^[28], 其中下肢康复领域增长最为显著(Topic 2、Topic 3、Topic 13、Topic 16)。

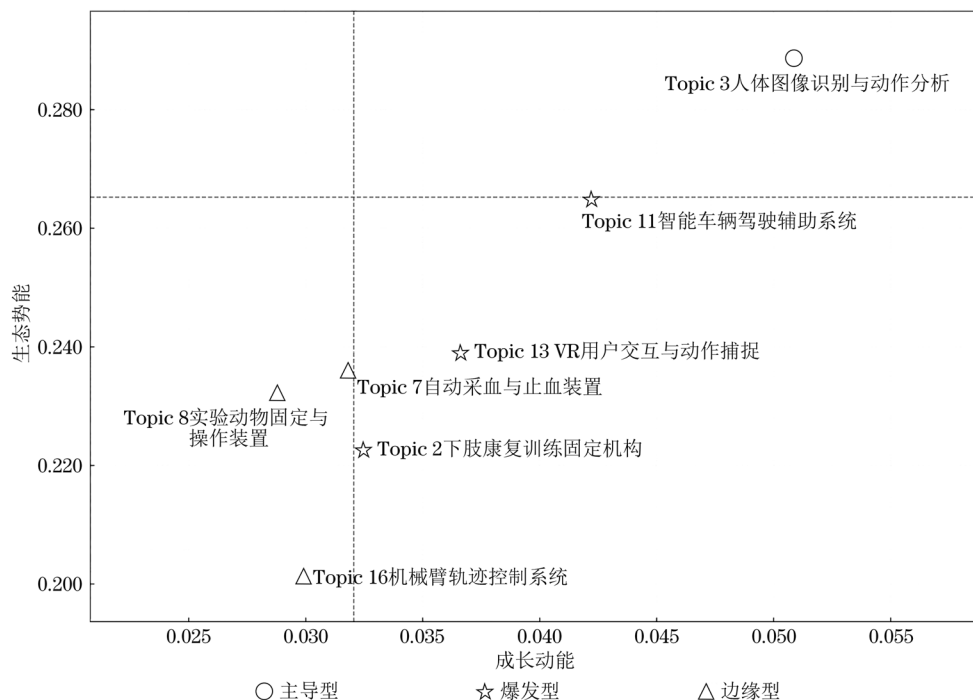


图2 技术主题象限分布

此外,结合中国信息通信研究院于2024年发布的《人形机器人产业发展研究报告》^[29],人形机器人正由高动态发展阶段向智能化阶段演进,技术重心逐渐从平衡控制转向感知交互与精细协同,模型识别出的技术主题(Topic 2、Topic 3)正好与这一智能化阶段的关键技术趋势相吻合。在产业应用层面,人形机器人技术正实现跨领域跃迁:工业生产和智慧物流领域,特斯拉、优必选等厂商正尝试推动其在汽车制造中的批量应用,中国电科21所等厂商试图利用人形机器人解决传统物流难题(Topic 11、Topic 16);医疗健康与家庭服务领域,人形机器人正由起步阶段的辅助医护向具备语音交流、情感分析功能的陪伴服务演进(Topic 7、Topic 8、Topic 13);在特殊领域,如危险作业、灾害救援及安防巡逻,人形机器人凭借精准行走与四肢灵敏运动(Topic 2、Topic 3)展现出显著的战略应用价值。综上,模型所识别的智能化方向具有明显的前沿性和技术共识基础。

3.5.3 结论分析

主导型主题的典型代表是Topic 3(人体图像识别与动作分析),其在成长动能与生态势能两个维度均表现突出,反映其已处于由“新兴”向“主导”转化的阶段。应用层面,该类技术已在医疗影像、智能安防、运动康复等领域实现系统化落地。例如,基于动作捕捉的步态分析系统被广泛应用于神经疾病康复评估,基于深度学习的人体图像识别技术已在智慧医院中用于自动化病灶检测与手术辅助。上述案例验证了其不仅具备强劲的知识增长动能,在产业与社会应用层面也已经形成较为成熟的渗透体系。

爆发型主题的代表包括Topic 2(下肢康复训练固定机构)、Topic 11(智能车辆驾驶辅助系统)以及Topic 13(VR用户交互与动作捕捉)。这类主题普遍呈现出高成长动能、低生态势能的特征,表明其仍处于快速演化的技术跃迁临界点。例如,Topic 2在康复医疗和老龄照护场景中的需求增长显著,但受制于成本与技术适配性,尚未实现规模化推广。Topic 11在智能驾驶辅助领域被广泛关注,当前L2+/L3级功能已进入商业化应用阶段,但仍受到政策、安全与公众接受度等因素影响。Topic 13受元宇宙概念和虚拟交互技术的推动,在Quest Pro、PICO等产品中已实现多模态交互探索,但仍面临硬件门槛与内容生态瓶颈等影

响。因此,该类主题是未来值得重点监测与投入的核心方向。

边缘型主题包括Topic 7(自动采血与止血装置)、Topic 8(实验动物固定与操作装置)以及Topic 16(机械臂轨迹控制系统)。这些主题在语义结构上表现出一定的创新性,但活力指数相对较低,尚未进入快速演化区间。Topic 7尽管已在少数三甲医院试点引入自动采血封管设备,但仍受限于严格的医疗监管与标准门槛;Topic 8主要服务于特定生物实验场景,受场景单一性与规模限制明显;Topic 16多为工业机器人子模块或实验自动化平台的嵌入式模块,其独立性与系统原创性相对有限。整体而言,边缘型主题处于低成长动能、低生态势能的观察窗口,后续发展仍须依赖政策驱动、产业拉动或新场景开拓等外部变量的激活。

4 结语

本研究融合BERTopic主题识别与LSTM时序预测,构成长动能与生态势能二维评估矩阵,形成从语义特征挖掘到趋势预测,再到内部类型评估的多维闭环分析流程。研究在实证层面验证了模型的有效性,证明LSTM模型在预测精度与泛化能力上优于传统模型,并进一步通过政策与产业报告的交叉验证,为新兴技术的优先级排序与战略选择提供了决策参考,确认了识别结果具备科学性与前瞻价值。以人形机器人领域为例,研究揭示了不同技术主题在二维评估矩阵中的分布特征与演化潜能:首先,主导型主题应强化其底层感知范式的规范性引导作用,以提升核心算法对全产业链的赋能价值;其次,爆发型主题展现出向医疗康复、智能驾驶及VR交互等场景跨域扩散的显著特征,构成了当前产业红利释放的关键窗口;最后,边缘型主题展现出人形机器人在精密作业等特定场景中的突破潜力,具备长期的战略监测与关注价值。基于此,本研究建议依托二维评估矩阵建立技术分层管理与资源配置机制,推动高潜力技术与现实场景对接,促进新兴技术向实际应用跃迁。

需要指出的是,本研究利用政策文件与产业报告进行定性佐证,缺乏对识别结果与市场实际需求匹配度的量化评估。未来研究将着重构建更完善的外部验证体系,包括建立专家评价机制、实现全球专利数据覆盖、引入产业实际应用指标等,进一步提升识别结果的可靠性与实践指导价值。

参考文献

- [1] 周文, 许凌云. 论新质生产力: 内涵特征与重要着力点[J]. 改革, 2023 (10): 1-13.
- [2] JIANG M, YANG S L, GAO Q. Multidimensional indicators to identify emerging technologies: perspective of technological knowledge flow[J]. Journal of Informetrics, 2024, 18 (1): 101483.
- [3] 赵又霖, 崔淦, 王嘉杰, 等. 不连续创新视角下技术演进-知识流动的新兴技术识别方法[J]. 情报杂志, 2025, 44 (10): 144-154.
- [4] 杨坤, 殷涛, 汪万, 等. 面向未来场景的新兴技术产业化机制: 基于知识组织视角的探索性案例分析[J/OL]. 科技进步与对策: 1-13[2025-12-17]. <https://link.cnki.net/urlid/42.1224.G3.20251203.0835.002>.
- [5] 王金凤, 杨慧琳, 赵伟宇, 等. 知识迁移下基于融合知识网络与链路预测的技术机会识别[J]. 情报杂志, 2025, 44 (5): 165-173, 198.
- [6] XU H Y, WINNINK J, YUE Z H, et al. Multidimensional scientometric indicators for the detection of emerging research topics[J]. Technological Forecasting and Social Change, 2021, 163: 120490.
- [7] 周莉, 陈荣, 严素梅, 等. 面向技术发展脉络分析的专利引证类时序模型及图谱研究[J]. 数字图书馆论坛, 2023, 19 (12): 44-55.
- [8] 乔治·戴, 保罗·休梅克. 沃顿论新兴技术管理[M]. 石莹, 译. 北京: 华夏出版社, 2002.
- [9] WANG Q. A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science and Technology, 2018, 69 (2): 290-304.
- [10] KELLER J, VON DER GRACHT H A. The influence of information and communication technology (ICT) on future foresight processes: results from a Delphi survey[J]. Technological Forecasting and Social Change, 2014, 85: 81-92.
- [11] 奉国和, 陈恩琪, 邓伟伟. 语义与演化视角下新兴技术识别: 以“工业机器人”领域专利为例[J/OL]. 科研管理: 1-19[2025-12-17]. <https://link.cnki.net/urlid/11.1567.G3.20250122.1659.013>.
- [12] 张甜, 陈进东, 周晓纪, 等. 基于潜在影响力预测和多源信息融合的新兴技术识别方法[J]. 情报杂志, 2025, 44 (9): 134-142, 133.
- [13] 赵雪峰, 吴德林, 吴伟伟, 等. 基于深度学习与多分类轮询机制的高质量“卡脖子”技术专利识别模型: 以专利申请文件为研究主体[J]. 数据分析与知识发现, 2023, 7 (8): 30-45.
- [14] 侯剑华, 朱晓清. 基于专利的技术预测评价指标体系及其实证研究[J]. 图书情报工作, 2014, 58 (18): 77-82, 116.
- [15] JUN S, SUNG PARK S, SIK JANG D. Technology forecasting using matrix map and patent clustering[J]. Industrial Management & Data Systems, 2012, 112 (5): 786-807.
- [16] 孙笑明, 袁思懿, 彭珍珍, 等. 基于专利分析与TRIZ的新兴技术预测模型研究: 以新能源汽车动力电池为例[J]. 科技进步与对策, 2025, 42 (17): 101-112.
- [17] 徐璐, 潘禹辰. 基于多维动态大数据的新兴产业技术预测分析方法[J]. 情报资料工作, 2023, 44 (4): 13-23.
- [18] 梁瑞文, 毛进, 芦昆, 等. 生物医学领域技术方法的知识增长模式研究[J]. 情报理论与实践, 2023, 46 (8): 104-112.
- [19] 余壮雄, 董洁妙, 程嘉嘉. 技术“卡脖子”的影响与对策: 基于融合技术关联的生产网络分析框架[J]. 中国工业经济, 2025 (6): 101-120.
- [20] 辛竹琳, 魏凤, 邓阿妹, 等. 基于技术成熟度的技术评价方法研究[J]. 科技管理研究, 2024, 44 (11): 80-89.
- [21] 赵琦智, 莫赞. 基于多源数据及多维指标的新兴技术识别方法研究: 以新能源汽车技术为例[J]. 广西大学学报(哲学社会科学版), 2025, 47 (2): 187-196.
- [22] 李尽法, 许灿, 张珂, 等. 动态视域下融合深度学习与专利地图的新兴技术识别研究[J]. 情报理论与实践, 2025, 48 (7): 160-169.
- [23] 杨思洛, 江曼. 新兴技术内涵特征和识别方法研究进展[J]. 情报科学, 2023, 41 (5): 181-190.
- [24] HAN S Q, HUANG H L, HUANG X H, et al. Core patent forecasting based on graph neural networks with an application in stock markets[J]. Technology Analysis & Strategic Management, 2024, 36 (8): 1680-1694.
- [25] 韩令晖, 陈劲, 李习保. 科学家作为发明人: 离群创新的来源、新颖性及扩散[J]. 科研管理, 2025, 46 (3): 38-47.
- [26] 王巍, 刘彦娇, 陈劲, 等. 关系嵌入性动态演化对创新新颖性与影响力的作用[J]. 系统管理学报, 2025, 34 (1): 168-186.
- [27] 刘伟. 人形机器人发展面临的挑战及未来趋势研判[J]. 人民论坛, 2025 (4): 42-47.
- [28] 中研普华产业研究院. 2025—2030年康复医疗产业深度调研及未来发展现状趋势预测报告[R]. 深圳: 中研普华产业研究院, 2025.
- [29] 中国信息通信研究院. 人形机器人产业发展研究报告(2024年)[R]. 北京: 中国信息通信研究院, 2024.

作者简介

王珩, 女, 硕士研究生, 研究方向: 知识管理。

杨坤, 女, 博士, 教授, 通信作者, 研究方向: 知识管理、技术创新管理, E-mail: yangkun12345688@126.com。

殷涛, 男, 硕士研究生, 研究方向: 知识管理、技术创新管理。

Topic Prediction and Value Evaluation of Emerging Technologies Based on BERTopic-LSTM Model:
Taking Chinese Humanoid Robot as an Example

WANG Bei YANG Kun YIN Tao

(School of Management, Shanghai University of Engineering Science, Shanghai 201620, P. R. China)

Abstract: Scientifically judging the evolution path of emerging technologies has important forward-looking enlightenment value for the national strategic layout of science and technology. In view of the limitations of existing research in fine-grained topic capture ability, lack of temporal feature fusion, and single evaluation dimension, this study constructs an analysis framework of “technology topic identification-emerging technology prediction-potential value evaluation”, aiming to provide more accurate and interpretable decision-making basis for technology priority ranking and resource allocation. Firstly, a BERTopic-LSTM technology prediction model considering semantic evolution and temporal fusion is proposed to overcome the problem that traditional models are difficult to take into account the deep semantics and evolution continuity of text. Secondly, a two-dimensional evaluation matrix is designed to see through the growth impetus and ecological potential inclination of emerging technologies, and systematically identify their potential strategic value. Finally, an empirical analysis is carried out in the field of humanoid robot. The results show that the technical topics identified by the BERTopic-LSTM model in massive patent data are more fine-grained. Compared with other models, it performs better in RMSE, MAE, and R^2 , and the prediction results are highly consistent with the actual technology iteration trajectory. The research provides scientific theoretical support and tool reference for the innovation management decision-making of government and enterprises in complex environment.

Keywords: Emerging Technology; Technology Prediction; BERTopic; LSTM; Humanoid Robot

(责任编辑: 常春)