基于时序关联和价值评估的技术主题演化研究*
——以生物制药产业为例

李雪[1] 王婧朱[1] 刘心蕊[1] 夏焕章[2] 袁红梅[1]
(1. 沈阳药科大学工商管理学院，沈阳 110016；2. 沈阳药科大学生命科学与生物制药学院，沈阳 110016)

摘要：为全面、准确地揭示技术发展脉络和发展态势，提供技术主题演化的整体视角，提出一种基于时序主题关联与主题价值评估的技术主题演化分析方法。基于专利数据，依据技术生命周期S曲线动态划分时间窗口，利用LDA主题模型识别各时间窗口的技术主题，从时序主题关联和主题价值评估2个视角对技术主题演化路径进行全面探析。基于生物制药产业进行实证研究，揭示生物制药相关技术的演进路径和发展状态，验证所提方法的可行性和有效性，为相关企业的技术创新提供科学的理论支持。

关键词：时序关联；主题价值评估；技术主题演化；生物制药
中图分类号：R95；G252.7　　DOI：10.3772/j.issn.1673-2286.2023.06.003
引文格式：李雪，王婧朱，刘心蕊，等. 基于时序关联和价值评估的技术主题演化研究：以生物制药产业为例[J]. 数字图书馆论坛，2023 (6)：22-30.

随着新一轮科技革命和产业变革，人工智能、生物技术等新兴技术不断涌现，技术创新正在成为各国推动经济社会转型、构筑竞争新优势的重要抓手，准确高效地识别及追溯技术创新的发展趋势和演进方向已成为技术管理领域研究人员日益关注的重要命题。专利作为最有效的创新技术信息载体，在内容上几乎涵盖了所有最新的技术创新信息，蕴含着极大的经济价值和技术价值，是理想的技术挖掘的信息来源[1]。基于此，本文利用专利数据对技术主题进行深度挖掘，刻画相关技术的发展脉络，掌握技术主题的演化规律，以快速捕捉技术发展契机，为国家制定战略决策和企业确定研发布局提供新思路。

1 文献综述

技术主题演化是技术创新动态发展的核心。技术主题演化是指技术主题在不同的、连续的、滑动的时间窗口之间表现出来的动态变化，旨在描述技术从产生、发展、成熟到消亡的发展脉络，强调从发展变化的动态角度掌握技术进步规律和演变模式。为揭示某项特定技术的演化过程，不断有学者从不同的视角对技术主题演化展开研究，逐步完善了技术主题演化研究的理论和方法。目前技术主题演化分析的研究方法主要分为三大类。①基于文献计量学的方法，该方法采用统计学对文本数据进行描述，通过构建量化指标和量化趋势来研究技术的发展现状及趋势。Carpenter等[2]通过统计学的方法研究技术与技术之间的交互关系。Zhang等[3]通过对关键词词频进行统计分析，提取出了虚拟现实技术的技术演化路径，并对结果进行人工解读，总结出了虚拟现实技术的演化情况。Yoon等[4]以生长曲线和技术生命周期为标准判断当前技术所处的阶段，通过使用专利分类号来定义技术的位置并进行技术演化分析，探索技术的延伸方向。虽然这种方法可以对技术发展趋势进行判断，但对专利技术演化情况

收稿日期：2023-05-13
*本研究得到辽宁省教育厅高等学校基本科研项目（面上项目）“基因治疗产品质量控制和药学评价技术研究”（编号：LJKR0300）、沈阳药科大学药品监管科学研究院专项基金“基因治疗产品质量控制和药学评价技术研究”（编号：2021jgkx006）资助。

的描述过于粗糙，无法深入利用专利文本所包含的大量有用信息，因此无法满足探索特定领域技术演化方向的需求。②基于专利引文分析的方法，通常依靠由被引专利指向施引专利的有向弧来反映技术之间的流动性、相似性[5]，进而识别技术演化方向，判断技术发展趋势。Batagelj[6]提出了搜索路径数算法，并将其应用于电动汽车的专利引文网络以抽取技术主题，验证了方法的有效性。Verspagen[7]、Martinelli[8]对燃料电池和电信交换器行业的专利进行引用分析，构建专利引文网络，得到较清晰的技术主题演化轨迹。Huang等[9]将专利引文共分类分析和主路径分析等分析方法相结合，并将其应用于3D打印领域，以揭示技术热点及追溯技术演化路径。基于专利引文分析的方法虽有较高的准确性，但存在时间滞后性的弊端，一些被引频次较少但包含重要技术信息的专利容易被忽略，导致结果不准确。③基于文本挖掘的方法，该方法主要对文本信息进行挖掘，可以从大量非结构化文本信息中发现潜在的数据模式和内在联系，提取有价值的信息和知识进行技术演化分析[10]。Zhu等[11]利用LDA（Latent Dirichlet Allocation）主题模型对专利文本的主题词进行提取，并通过专家咨询和参数实验的方法设置阈值，将文本相似度超过阈值的技术节点提取出来，生成技术演化路径，进而实现技术演化分析。Hu等[12]利用LDA主题模型对新能源汽车产业的技术主题进行提取，并从重要性和受欢迎程度等方面进行分析，从而实现对该产业技术发展趋势的准确认识。Liu等[13]提出了一种非参数主题模型CIHDP，该模型能够动态地检测和跟踪主题，完整映射出人工智能研究领域的技术演化路径。基于文本挖掘的方法，能够量化文本中的特征词，最大限度地发掘文本所包含的意义，可以揭示技术主题间的关联性。对这类方法的研究逐渐形成了以LDA及其扩展模型应用为主、以词嵌入及深度学习等其他文本挖掘技术为辅的研究趋势，这类方法常与前两类方法结合使用并受到广泛关注。

综上所述，现有研究在技术主题演化分析方面已经取得诸多成果，但在研究方法层面仍存在一定的局限性：①基于文献计量学和专利引文分析的方法本身存在局限性，以此来描述技术领域中的主题变化过程不够全面、准确，研究结果的准确性和科学性还有待提升。②应用LDA主题模型研究技术主题演化可以准确识别出技术主题，但大部分研究只关注技术主题关系和演化过程，而忽略了主题的发展状态。因此，为克服现有技术主题演化研究的不足，全面细致地解释技术主题演化过程，并提供技术主题演化的整体视角，本研究提出一种基于时序主题关联演化与主题价值评估的综合方法，全方位地揭示技术领域的主题演变动态。

2 研究方法设计

分析框架见图1，研究主要分为5个阶段：数据收集与预处理、时序主题生成、时序主题演化、主题价值评估和技术主题演化路径构建。首先，通过梳理目标领域相关文献以确定检索式，并从incoPat数据库中下载专利数据，对数据进行清洗后，构建研究数据集合。其次，根据技术生命周期S曲线，动态划分时间窗口，将研究数据根据其时间窗口划分为若干子集。利用LDA主题模型对每个数据子集进行主题分析，识别出每个时间窗口下的技术主题。再次，使用余弦相似度公式计算相邻时间窗口主题之间的相似度，挖掘出关联主题对，并判别主题间演化关系，绘制时序主题关联演化桑基图。通过设计主题活跃度和主题质量指标体系，构建战略坐标图，评估主题的价值，揭示主题的发展状态。最后，基于时序关联和主题价值2个视角，构建出技术主题演化路径并绘制可视化路径，对领域内的关键演化路径展开分析。

2.1 时序主题生成

时序主题生成是指将时间标签与主题模型相结合，展示主题随时间的变化过程。基于技术生命周期S曲线[14]，按照时间线对专利文献进行时间窗口划分，利用LDA主题模型对各个时间窗口下的主题进行识别。

主题识别是主题演化分析的基础和前提，主题模型可以对文档的隐含语义结构进行聚类和挖掘，识别出“文档-主题-词汇”的潜在语义关系。常见的主题模型包括潜在语义分析（Latent Semantic Analysis，LSA）、非负矩阵分解（Nonnegative Matrix Factorization，NMF）及LDA等[15-17]。其中，LDA主题模型已在多个领域的主题识别中得到广泛应用，可以有效地分析大规模非结构化文档集[18-19]。LDA是一种基于无监督机器学习的词袋模型，可对文本中隐含的语义维度进行挖掘，其基本思想是假设文档由具有不同概率的若干个主题组成，主题由具有不同概率的若干个词汇

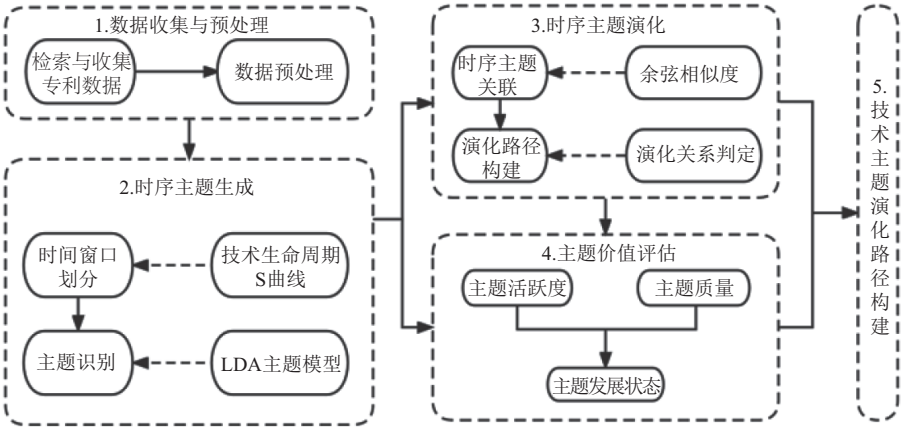

图1 研究框架图

组成，LDA模型结构见图2。图2中，z表示模型生成的主题，w表示模型生成的词汇，K表示主题数量，M表示文档数量，N_m表示文档m的词汇数量，θ表示主题分布，φ表示主题词分布，α和β分别表示主题分布θ和主题词分布φ的先验分布参数。

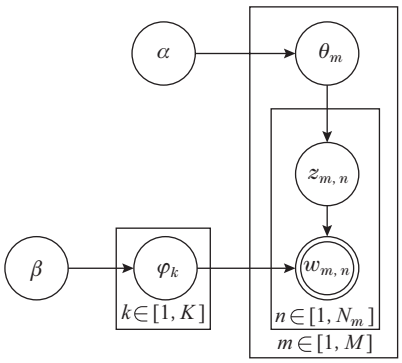

图2 LDA主题模型结构

LDA模型先在每个文档中提取一个主题，然后从提取的主题对应的词汇中随机提取一个词，重复上述过程，直到遍历整个文档中的每个词，最终生成文档-主题分布和主题-词汇分布[20-21]，模型的联合分布如式（1）所示。

$$p(w, z_m, \theta_m, \varphi|\alpha, \beta) = \prod_{n=1}^{N_m} p(w_{m,n}|\varphi_{z_{m,n}}) \cdot p(z_{m,n}|\theta_m) \cdot p(\theta_m|\alpha) \cdot p(\varphi|\beta) \quad (1)$$

在LDA模型中，主题数需要预先设定。为了得到最为合适的主题数，采用困惑度评价法选择最优主题数k[22]，困惑度计算公式如式（2）所示。

$$P(D) = \exp\left[-\frac{\sum_{m=1}^{M} \log p(w_m)}{\sum_{m=1}^{M} N_m}\right] \quad (2)$$

式中：D为包含M篇文档的语料库；$p(w_m)$为文档m中的词汇分布。困惑度表示主题的不确定性（信息熵），故而当困惑度最小时，主题数最优。一般情况下，当困惑度下降趋势不再明显或处于拐点处时，此时的k为最优主题数。

2.2 时序主题关联演化

2.2.1 主题关联建立

要构建时序主题关联演化路径，需要建立相邻时间窗口各主题间的关联关系。LDA主题模型生成的各个时间窗口下的主题都是独立存在的，为探索主题间演化关系，可通过计算相邻时间窗口主题之间的相似度来衡量主题间的延续性[23]。通过计算相邻时间窗口各主题之间的余弦相似度来衡量时序主题间的相似度，进而建立相邻时间窗口主题之间的关联关系，计算方法如式（3）所示。

$$S(T_i^{t-1}, T_j^t) = -\frac{\sum_{n_0=1}^{N_0} p(w_{n_0}|T_i^{t-1}) \times p(w_{n_0}|T_j^t)}{\sqrt{\sum_{n_0=1}^{N_0}\left[p(w_{n_0}|T_i^{t-1})\right]^2} \times \sqrt{\sum_{n_0=1}^{N_0}\left[p(w_{n_0}|T_j^t)\right]^2}} \quad (3)$$

式中：T_i^{t-1}为时间窗口$t-1$下的第i个主题，T_j^t为时间窗口t下的第j个主题；$p(w_{n_0}|T_i^{t-1})$和$p(w_{n_0}|T_j^t)$分别为时间窗口$t-1$和t下的主题-词汇分布；w_{n_0}为主题中的第n_0个词汇；N_0为主题所包含的词汇总数。余弦相似度值为0~1，$S(T_i^{t-1}, T_j^t)$的值越接近1，2个主题越相似。设定阈值ε以去除弱关联和无效关联：当$S(T_i^{t-1}, T_j^t) \geqslant \varepsilon$时，当前主题是前一主题的延续；如果$S(T_i^{t-1}, T_j^t) < \varepsilon$，则当前主题与前一主题无关。

2.2.2 演化关系判定

为解决时序主题演化中演化关系的判别问题，参

照曾江峰等[24]、李慧等[25]提出的主题演化关系模型，将演化关系分为新生、继承、消亡、融合和分裂5种（见图3），通过计算相邻时间窗口各主题之间的余弦相似度并设置阈值，提取出上述时序主题演化关系，具体判定条件如表1所示。

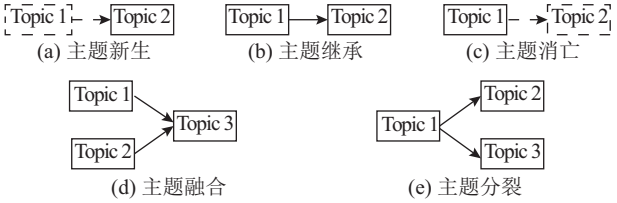

图3 主题演化关系图

表1 时序主题演化关系的判定条件

演化类型	判定条件	描 述
新生	对前一时间窗口内所有主题T^{t-1}, $S(T^{t-1}, T_j^t)<\varepsilon$	当前主题和前一时间窗口的所有主题均无关联关系，即当前主题是新出现的
继承	$S(T_i^{t-1}, T_j^t)\geqslant\varepsilon$	当前主题仅与前一时间窗口的一个主题具有关联关系，即一对一的关系
消亡	对后一时间窗口内所有主题T^{t+1}, $S(T_j^t, T^{t+1})<\varepsilon$	当前主题与后一时间窗口的所有主题均无关联关系，即当前主题在后一时间窗口中消失
融合	$S(T_i^{t-1}, T_j^t)\geqslant\varepsilon$且$S(T_k^{t-1}, T_j^t)\geqslant\varepsilon$	当前主题是由前一时间窗口的两个或多个主题合并产生的，即多对一的关系
分裂	$S(T_i^{t-1}, T_j^t)\geqslant\varepsilon$且$S(T_i^{t-1}, T_k^t)\geqslant\varepsilon$	前一时间窗口的某个主题在当前时间窗口中分裂成两个或多个主题，即一对多的关系

2.3 主题价值评估

在主题识别的基础上，构建主题活跃度和主题质量2个计量指标来描述不同时间窗口下的技术主题，以评估技术主题的价值，进而揭示主题发展状态。

2.3.1 主题活跃度

技术主题活跃度指标可以用同一时间窗口下与该主题相关的专利数量来表示[26]，计算公式如式（4）所示。

$$A_T(t)=M(t) \tag{4}$$

式中：$A_T(t)$为时间窗口t下的主题T的活跃度；$M(t)$为时间窗口t下该主题的文档数量。

2.3.2 主题质量

技术主题质量可从技术影响力、技术交叉性、覆盖范围和保护范围4个维度综合测量：技术影响力可从专利被引频次的角度进行衡量[27-28]；技术交叉性可通过技术宽度来分析，即涉及技术领域的数量，采用专利分类号数量测度[29-30]；覆盖范围[31]和保护范围[32]分别可以用专利的简单同族个数和权利要求数量来表示。技术主题质量指标的计算公式如式（5）所示。

$$Q_T(t)=\sum_{r=1}^{R}p_{r,T}\times V_r=\sum_{r=1}^{R}p_{r,T}\times(\omega_1\times Q_r+\omega_2\times I_r+\omega_3\times F_r+\omega_4\times C_r) \tag{5}$$

式中：$Q_T(t)$为时间窗口t下的技术主题T的主题质量；r为技术主题T下的第r个专利；R为技术主题T包含的专利数量；$p_{r,T}$为技术主题T和第r个专利之间的文档-主题分布；V_r为第r个专利的价值；Q_r为第r个专利的被引频次；I_r为第r个专利的专利分类号数量；F_r为第r个专利的简单同族个数；C_r为第r个专利的权利要求数量；ω_1、ω_2、ω_3、ω_4分别为Q_r、I_r、F_r、C_r对应的权重系数。人工分配的主观性较高，而熵权法可依据原始数据确定指标权重，计算结果客观合理[33]，故采用熵权法来确定指标权重。为了增强结果的准确性，在使用熵权法之前采用最小-最大归一化方法对数据进行标准化处理，如式（6）所示。

$$Y=\frac{X-X_{\min}}{X_{\max}-X_{\min}} \tag{6}$$

式中：Y为标准化后的数据值；X为原始数据值；$X_{\max}$、$X_{\min}$分别为数据的最大值和最小值。

为直观、清晰地反映出技术主题的发展状态，基于主题活跃度和主题质量指标计算结果，将所有技术主题映射到二维空间中，具体以主题活跃度为x轴，以主题质量为y轴，构建各个时间窗口的战略坐标图（见图4），可将技术主题划分为4种类型。①关键主题：该主题的活跃度和质量都较高，是技术的核心，对整个技术领域有很大的影响和推动作用。②潜在主题：该主题具有较高的质量，但活跃度较低，处于潜在期。③热点主题：此类主题的活跃度较高但质量较低，有望成为未来目标领域最具影响力的技术话题，具有极大的研究价值。④萌芽主题：这类主题是活跃度和质量均较低

的主题，存在一定的创新阻力。

技术主题价值评估结果可以为技术发展战略的制定提供参考，不同的战略将适用于不同类型的主题：对于关键主题，应保持长期投资；对于热点主题，要尽早加大研发投入，提高研究质量；对于潜在主题，应保持适度关注，当其相关因素发生变化时，应重新评估该主题的价值；对于萌芽主题，应避免过度投资。

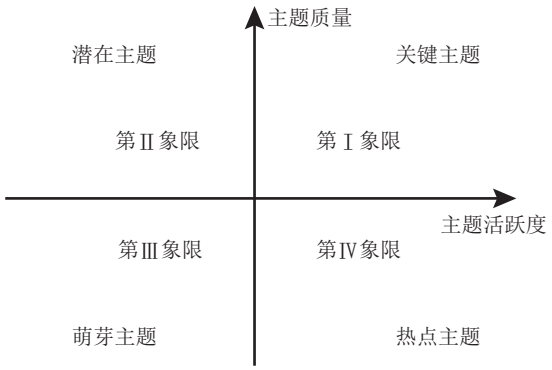

图4 主题价值评估战略坐标图

3 实证研究——以生物制药产业为例

随着生物学不断发展，生物技术逐渐融入医药研发领域，生物制药产业获得了飞速的发展，也为医药领域开拓了广阔的前景。生物制药产业作为我国战略性新兴产业之一，是我国医药产业的重大支柱，处于现代医药发展的关键位置，其发展对国民经济和人民生活质量提升具有重要意义[34]。深入了解生物制药产业技术演变规律并预测未来的趋势，对助力我国生物制药产业的进一步发展有着至关重要的作用。

3.1 数据获取与预处理

选取生物制药领域专利数据进行实证研究。从incoPat专利检索数据库中收集专利，基于文献研究和专家意见确定了检索策略。以TIAB=（gene OR nucleic acid OR nucleotide OR stem cell OR antibody OR vaccine OR hormone OR genetic engineering OR amino acid OR tissue culture OR cell culture） AND IPC=（A61K OR C12M OR C12N OR C12P OR C12Q OR C12S OR C07G OR C07K OR G01N）为检索式，将时间截至2020年12月31日，检索时间为2022年11月。将专利类型限定为被授权的中国申请专利，经数据清洗后共得到39 326个专利。以专利的标题和摘要为分析语料，使用Python软件中的分词技术进行分词、去除停用词等预处理，形成建模需要的语料库。

3.2 时序主题生成

3.2.1 时间窗口划分

在数据预处理的基础上，绘制生物制药领域的技术生命周期S曲线，见图5。根据S曲线结果并考虑专利数量分布，将生物制药专利数据的时间段划分为1985—1998年、1999—2007年、2008—2011年、2012—2014年、2015—2017年、2018—2020年6个时间窗口。

3.2.2 主题生成

对各个时间窗口下的专利数据进行LDA主题聚类以提取技术主题，通过计算主题困惑度确定这6个时间窗口的最优主题数分别为7、13、11、6、14、12个。6个时间窗口的主题生成结果见表2。

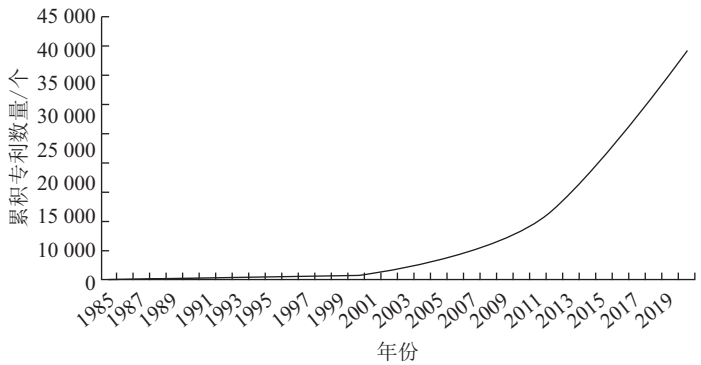

图5 生物制药领域的技术生命周期S曲线

表2 各时间窗口技术主题生成结果

时间窗口	技术主题及主题标签
1985—1998年	1_T0 基因工程技术、1_T1 疫苗的制备、1_T2 甲型肝炎活疫苗、1_T3 检测试剂盒、1_T4 外源基因表达、1_T5 抗血吸虫疫苗、1_T6 氨基酸合成
1999—2007年	2_T0 乙型肝炎病毒、2_T1 基因突变检测方法、2_T2 肿瘤基因治疗、2_T3 融合蛋白的制备、2_T4 HIV感染、2_T5 微生物技术、2_T6 单克隆抗体的制备、2_T7 基因表达载体构建、2_T8 基因及其编码的多肽、2_T9 干细胞培养、2_T10 抗菌肽制备和应用、2_T11 基因检测、2_T12 纳米药物递送系统
2008—2011年	3_T0 流感病毒、3_T1 诊断试剂盒、3_T2 细胞培养、3_T3 基因SNP检测、3_T4 单克隆抗体、3_T5 基因表达载体的构建、3_T6 肿瘤治疗、3_T7 检测试剂盒、3_T8 微生物发酵、3_T9 氨基酸修饰、3_T10 多肽及其应用
2012—2014年	4_T0 单克隆抗体的制备和应用、4_T1 肿瘤靶向治疗、4_T2 融合蛋白的制备和应用、4_T3 基因检测试剂盒、4_T4 多肽疫苗、4_T5 干细胞培养
2015—2017年	5_T0 PD-1/PD-L1抗体、5_T1 基因检测试剂盒、5_T2 抗菌肽的制备和应用、5_T3 纳米药物、5_T4 疫苗的制备、5_T5 单克隆抗体的制备和应用、5_T6 抗肿瘤联合用药、5_T7 CIK和NK细胞的培养、5_T8 基因工程菌的构建、5_T9 基因诊断标志物、5_T10 抗肿瘤多肽、5_T11 间充质干细胞培养、5_T12 嵌合抗原受体及其用途、5_T13 衍生物的制备
2018—2020年	6_T0 嵌合抗原受体、6_T1 单克隆抗体制备及应用、6_T2 基因突变体及其应用、6_T3 SNP位点检测试剂盒、6_T4 PD-1/PD-L1及其抗肿瘤作用、6_T5 核酸纳米颗粒、6_T6 SARS-CoV-2抗体、6_T7 siRNA及其应用、6_T8 多肽的制备及应用、6_T9 新型冠状病毒疫苗、6_T10 癌症诊断标志物、6_T11 间充质干细胞培养

3.3 时序主题关联演化

随着技术研究的不断深入，相邻时间窗口的技术主题之间也存在着演化变迁关系。对相邻时间窗口下主题间的余弦相似度设置阈值0.6来确定2个主题之间是否存在关联关系：若2个主题之间的相似度≥0.6，则表示2个主题之间存在演化关系，将纳入演化图。利用pyecharts图表类库将主题演化关系绘制成桑基图，结果见图6。图6中，每个方块都具有对应主题，主题间的连线表示主题的流动方向，连线的粗细表示相似度的大小：连线越粗，相似度越大，表明主题间的演化关系越紧密。

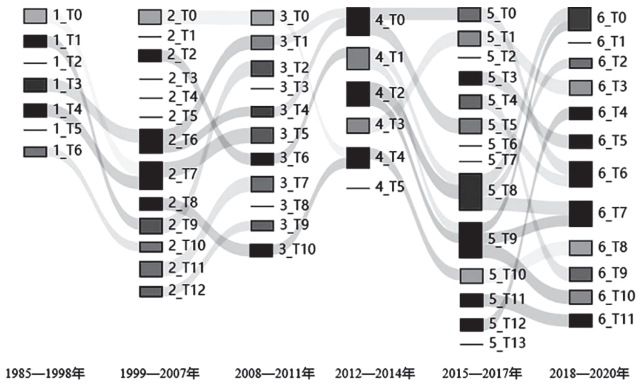

图6 生物制药领域时序主题演化路径图

由图6可知，生物制药领域的技术主题存在新生、继承、消亡、融合和分裂5种演化关系。

（1）新生：4_T2 融合蛋白的制备和应用在2012—2014年成为新的技术主题，5_T3 纳米药物、5_T12 嵌合抗原受体及其用途主题为近些年生物制药领域的新生主题，这些主题在未来具有很大的发展空间。

（2）继承：1985—1998年的1_T3 检测试剂盒、1999—2007年的2_T6 单克隆抗体的制备、2008—2011年的3_T4 单克隆抗体、2012—2014年的4_T0 单克隆抗体的制备和应用、2015—2017年的5_T0 PD-1/PD-L1抗体、2018—2020年的6_T6 SARS-CoV-2抗体主题属于继承主题，具有高连续性。

（3）消亡：2008—2011年的3_T5 基因表达载体的构建在下一时间窗口消失，该主题研究弱化。

（4）融合：2012—2014年的4_T1 肿瘤靶向治疗、4_T2 融合蛋白的制备和应用2个主题融合生成5_T8 基因工程菌的构建和5_T9 基因诊断标志物2个主题，合并后的主题再次发生融合，并入6_T0 嵌合抗原受体。

（5）分裂：5_T9 基因诊断标志物在2018—2020年分裂为6_T0 嵌合抗原受体、6_T7 siRNA及其应用和6_T10 癌症诊断标志物3个主题，分化出新的研究方向。

3.4 基于时序关联和价值评估的动态演化路径构建

在时序主题演化路径的基础上，结合主题活跃度和主题质量2个技术主题价值评估指标，构建生物制药领域技术主题动态演化路径。动态演化路径图的横向角度展示了每条路径上的主题演化类型，纵向角度清晰直观地展示各时间窗口的主题发展状态和战略坐标分布。生物制药领域的相关研究随着时间的推移不断丰富，主题的演化形式也更加多样化，各个不同演化路径之间存在着许多交叉，演化过程较为复杂。根据时序主题的演化关系，发现在所有时间窗口中有4条主题演化路径具有高连续演化（继承、融合或分裂）关系，见图7。这4条演化路径共包含8个技术主题节点，其对应的主题词和主题描述见表3。

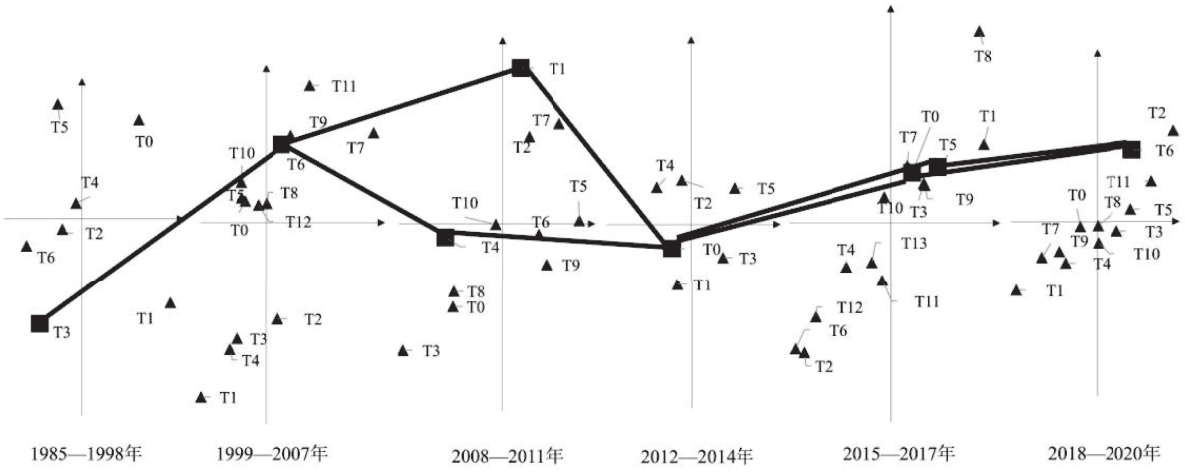

图7 生物制药领域主题动态演化路径图

表3 高连续演化主题内容

1985—1998年	1999—2007年	2008—2011年	2012—2014年	2015—2017年	2018—2020年
T3：检测、抗体、特异性、抗原、试剂、试剂盒、单克隆抗体、产物、血清、诊断	T6：抗体、检测、单克隆抗体、抗原、试剂盒、诊断、特异性、肝癌、血清、试剂	T1：检测、抗体、试剂盒、诊断、抗原、试剂、特异性、igG、血清、胶体金； T4：单克隆抗体、抗体、血管、细胞株、杂交瘤、重链、氨基酸序列、特异性、轻链、细胞	T0：抗体、试剂盒、单克隆抗体、抗原、细胞株、杂交瘤、人乳头瘤病毒、诊断、igG、血清	T0：抗体、融合蛋白、单克隆抗体、PD-1、重链、轻链、氨基酸序列、IL、PD-L1、纳米抗体； T5：抗体、试剂盒、单克隆抗体、抗原、细胞株、杂交瘤、诊断、血清、ELISA、生物素	T6：抗体、抗原、氨基酸序列、纳米抗体、重链、轻链、Cov、SARS、CDR、毒素

如图7所示，1985—1998年的1_T3 检测试剂盒为萌芽主题，在1999—2007年演变为关键主题2_T6 单克隆抗体的制备，受到了极大关注，在成长的过程中不断分化，最终分裂为关键主题3_T1 诊断试剂盒和萌芽主题3_T4 单克隆抗体。2012—2014年的萌芽主题4_T0 单克隆抗体的制备和应用由3_T1 诊断试剂盒和3_T4 单克隆抗体合并而来，并在下一时期发生分裂，形成2015—2017年的关键主题5_T0 PD-1/PD-L1抗体、5_T5 单克隆抗体的制备和应用。5_T0 PD-1/PD-L1抗体、5_T5 单克隆抗体的制备和应用的主题质量和主题活跃度均较高，二者发生融合，合并为关键主题6_T6 SARS-CoV-2抗体。6_T6 SARS-CoV-2抗体具有较强的主题连续性，又具有较高的主题活跃度和主题质量，是当前的技术核心。SARS-CoV-2引起的COVID-19给人类健康带来了紧急危

机，以抗体为基础的SARS-CoV-2治疗药物正在加速进入临床前和临床研究[35-36]，SARS-CoV-2抗体的研究是当前生物制药领域的重点。

4 结论

提出了一种结合时序主题关联和价值评估的主题演化分析方法，采用LDA主题模型提取专利文本技术主题并识别时序主题演化路径，研究技术主题演化的过程，通过计算主题活跃度和主题质量来衡量主题价值，揭示主题的发展状态，以可视化方式绘制技术主题动态演化路径，并通过生物制药领域的实证研究证明了方法的可行性和准确性。本研究的贡献如下：①基于技术生命周期S曲线，利用专利增长规律划分多个专利时间窗口，克服了传统固定时间窗口划分方法不合理的问题；②设计主题活跃度和主题质量2个指标，开展主题价值分析，以评估每个主题的潜在价值，揭示主题的发展状态；③基于时序主题关联演化和价值评估2个维度探寻技术主题演化路径，融合桑基图和主题战略坐标图动态地展示技术主题的发展和演化情况，获得了更全面的演化路径。

尽管如此，本研究还存在局限性，需要进行完善：①仅利用了专利数据，实际研究中，技术相关信息不仅存在于专利数据上，也体现在学术论文等数据上，后续可使用多种数据源多角度地对技术主题演化路径进行研究；②LDA模型只识别每个主题下的词汇，无法对主题自动命名，主题标签的确定具有主观性，这可能会导致主题演化分析产生偏差，未来的研究可以探索新方法来进行主题标注。

参考文献

[1] 历妍. 基于专利的技术发展趋势研究[D]. 北京：北京工业大学，2011.

[2] CARPENTER M P，NARIN F. Validation study：patent citations as indicators of science and foreign dependence[J]. World Patent Information，1983，5（3）：180-185.

[3] ZHANG Z G，XU M L，HUANG J M. Proposals to promote the development of virtual reality in China—based on patent econometric analysis[J]. Engineering，2018，10（5）：291-304.

[4] YOON J，JEONG B，LEE W H，et al. Tracing the evolving trends in electronic skin（e-skin） technology using growth curve and technology position-based patent bibliometrics[J]. IEEE Access，2018，6：26530-26542.

[5] LIN Y，CHEN J，CHEN Y. Backbone of technology evolution in the modern era automobile industry：an analysis by the patents citation network[J]. Journal of Systems Science and Systems Engineering，2011，20（4）：416-442.

[6] BATAGELJ V. Efficient algorithms for citation network analysis[EB/OL]. [2023-04-01]. https://arxiv.org/abs/cs/0309023.

[7] VERSPAGEN B. Mapping technological trajectories as patent citation networks：a study on the history of fuel cell research[J]. Advances in Complex Systems，2007，10（1）：93-115.

[8] MARTINELLI A. An emerging paradigm or just another trajectory? understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry[J]. Research Policy，2012，41（2）：414-429.

[9] HUANG Y，ZHU D H，QIAN Y，et al. A hybrid method to trace technology evolution pathways：a case study of 3D printing[J]. Scientometrics，2017，111（1）：185-204.

[10] 谌志群，张国煊. 文本挖掘研究进展[J]. 模式识别与人工智能，2005，18（1）：65-74.

[11] ZHU M R，ZHANG X P，WANG H W. A LDA based model for topic evolution：evidence from information science journals[C]//Proceedings of 2016 International Conference on Modeling，Simulation and Optimization Technologies and Applications（MSOTA2016）. 2016：49-54.

[12] HU R J，MA W C，LIN W Q，et al. Technology topic identification and trend prediction of new energy vehicle using LDA modeling[J]. Complexity，2022，2022：1-20.

[13] LIU H L，CHEN Z W，TANG J，et al. Mapping the technology evolution path：a novel model for dynamic topic detection and tracking[J]. Scientometrics，2020，125（3）：2043-2090.

[14] FOSTER R N. Working the S-curve：assessing technological threats[J]. Research Management，1986，29（4）：17-20.

[15] HOFMANN T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine Learning，2001，42（1）：177-196.

[16] LEE D D，SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature，1999，401（6755）：788-791.

[17] BLEI D M，NG A Y，JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research，2003，3（4/5）：993-1022.

[18] TAZIBT A A，AOUGHLIS F. Latent dirichlet allocation-based temporal summarization[J]. International Journal of Web Information Systems，2019，15（1）：83-102.

[19] JENSEN S，LIU X Z，YU Y Y，et al. Generation of topic evolution trees from heterogeneous bibliographic networks[J]. Journal of Informetrics，2016，10（2）：606-621.

[20] WU H，YI H F，LI C. An integrated approach for detecting and quantifying the topic evolutions of patent technology：a case study on graphene field[J]. Scientometrics，2021，126（8）：6301-6321.

[21] 廖列法，勒孚刚，朱亚兰. LDA模型在专利文本分类中的应用[J]. 现代情报，2017，37（3）：35-39.

[22] BLEI D M. Probabilistic topic models[J]. Communications of the ACM，2012，55（4）：77-84.

[23] 秦晓慧，乐小虬. 基于LDA主题关联过滤的领域主题演化研究[J]. 现代图书情报技术，2015（3）：18-25.

[24] 曾江峰，黄泳潼，陈静怡，等. 面向融合出版技术的主题演化研究[J]. 数字图书馆论坛，2023，19（4）：9-18.

[25] 李慧，胡吉霞，佟志颖. 面向多源数据的学科主题挖掘与演化分析[J]. 数据分析与知识发现，2022，6（7）：44-55.

[26] FABRY B，ERNST H，LANGHOLZ J，et al. Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry[J]. World Patent Information，2006，28（3）：215-225.

[27] 于超，王涛，苏信宁，等. 引证关系视角下知识特质对“明星”专利形成的影响研究[J]. 软科学，2021，35（2）：19-25.

[28] 罗家豪，孙巍. 基于专利价值的技术成熟度测度与分析方法研究[J]. 数字图书馆论坛，2022（1）：17-25.

[29] 陈稳，陈伟. 科学与技术对比视角下的前沿主题识别与演化分析[J]. 情报杂志，2022，41（1）：67-73，163.

[30] 许珂，陈向东. 基于专利技术宽度测度的专利价值研究[J]. 科学学研究，2010，28（2）：202-210.

[31] 寇翠翠，金金，吴亚坤，等. 基于技术相似性和技术能力分析的中小型企业技术机会识别[J]. 情报理论与实践，2019，42（8）：104-109.

[32] 乔永忠，肖冰. 基于权利要求数的专利维持时间影响因素研究[J]. 科学学研究，2016，34（5）：678-683.

[33] 夏亮，杨江平，邓斌，等. 基于熵权法和改进灰色聚类模型的雷达软件安全性需求分析[J]. 兵器装备工程学报，2019，40（9）：94-98，168.

[34] 何方，刘杰. 基于专利分析的生物制药产业发展趋势研究[J]. 中国发明与专利，2021，18（7）：50-55，72.

[35] 饶亚岚，于爱平，王延琳，等. 抗SARS-CoV-2抗体产品治疗COVID-19的使用建议及临床试验证据[J]. 中国临床药理学杂志，2021，37（14）：1947-1952.

[36] 张辉，毛群颖，梁争论，等. SARS-CoV-2抗体标准物质的研制及应用进展[J]. 中国生物制品学杂志，2023，36（2）：230-234.

作者简介

李雪，女，硕士研究生，研究方向：药品知识产权。
王婧朱，女，博士研究生，研究方向：药品知识产权。
刘心蕊，女，博士研究生，研究方向：药品知识产权。
夏焕章，男，博士，教授，研究方向：组合生物合成研究。
袁红梅，女，博士，教授，通信作者，研究方向：药品知识产权，E-mail：yuanhm612@163.com。

Evolution of Technological Topic Based on Temporal Correlation and Value Evaluation: A Case Study of Bio-Pharmaceutical Industry

LI Xue[1] WANG JingZhu[1] LIU XinRui[1] XIA HuanZhang[2] YUAN HongMei[1]
(1. School of Business Administration, Shenyang Pharmaceutical University, Shenyang 110016, P. R. China; 2. School of Life Sciences and Biopharmaceuticals, Shenyang Pharmaceutical University, Shenyang 110016, P. R. China)

Abstract: To comprehensively and accurately reveal the context and trend of technological development, and provide the overall perspective of technological topic evolution, a technological topic evolution analysis method based on temporal topic correlation and topic value evaluation is proposed. Based on patent data, the study dynamically divides the time windows according to the S-curve of the technology life cycle, uses the LDA topic model to identify the technological topics in each time window, and makes a comprehensive analysis on the evolution path of the technological topics from the perspectives of temporal topic correlation and topic value evaluation. Bio-pharmaceutical industry is selected for the empirical study. The evolution path and development state of bio-pharmaceutical related technologies are revealed, the feasibility and effectiveness of the proposed method are verified, and scientific theoretical support is provided for the technological innovation of related enterprises.

Keywords: Temporal Correlation; Topic Value Evaluation; Technological Topic Evolution; Bio-Pharmaceutical

（责任编辑：王玮）