

基于 Topic - LLM 框架的医学信息学跨学科主题演化研究

苏中琪¹ 董 森² 崔 雷³

(¹ 中国医科大学图书馆 沈阳 110122 ² 中国医科大学财务处 沈阳 110122

³ 中国医科大学健康管理学院 沈阳 110122)

[摘要] **目的/意义** 探索医学信息学跨学科主题演化路径识别方法, 为该领域跨学科研究布局与科研管理提供参考。**方法/过程** 首先利用 BERTopic 和大语言模型 (large language model, LLM) 在获取的医学信息学文献中识别全局主题和阶段主题; 然后基于主题影响力与学科多样性二维分析框架, 确定跨学科主题; 最后分析跨学科主题的演化路径。**结果/结论** 基于 Topic - LLM 的分析方法, 发现医学信息学跨学科主题的学科多样性和主题影响力呈现稳步增长趋势。

[关键词] BERTopic; 大语言模型; 跨学科; 主题识别; 主题演化

[中图分类号] R - 058 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2026.01.008

Study on the Evolution of Interdisciplinary Topics in Medical Informatics Based on the Topic - LLM Framework

SU Zhongqi¹, DONG Miao², CUI Lei³

¹Library, China Medical University, Shenyang 110122, China; ²Department of Finance, China Medical University, Shenyang 110122, China; ³School of Health Management, China Medical University, Shenyang 110122, China

[Abstract] **Purpose/Significance** To explore the identification method of interdisciplinary topic evolution path in medical informatics, so as to provide references for the interdisciplinary research layout and scientific research management in this field. **Method/Process** Firstly, global and staged topics are identified from medical informatics literature using BERTopic and large language models (LLM). Secondly, interdisciplinary topics are determined based on a two - dimensional analysis framework integrating topic influence and disciplinary diversity. Finally, the evolution paths of these interdisciplinary topics are analyzed. **Result/Conclusion** An analysis method based on Topic - LLM is proposed, and it is found that the disciplinary diversity and topic influence of interdisciplinary topics in medical informatics exhibit a steady growth trend.

[Keywords] BERTopic; large language model (LLM); interdisciplinary; topic identification; topic evolution

1 前言

随着数字技术的快速发展与“健康中国”等国

家战略的深入推进, 医学信息学正经历显著的转型与发展, 逐步演变为重塑医疗服务模式与公共卫生治理的关键学科。作为交叉学科, 医学信息学通过高度的跨学科融合加速了知识创新, 同时也促进学

[修回日期] 2025 - 12 - 24

[作者简介] 苏中琪, 馆员, 发表论文 4 篇; 通信作者: 崔雷, 教授, 博士生导师。

科边界的动态演化与研究内容持续变迁。因此，如何在主题识别的同时量化其跨学科属性并分析其演化特征，对理解医学信息学的复杂知识体系非常重要。

本研究提出基于 Topic - LLM 的分析方法，将跨学科测量与主题识别相结合，利用大语言模型 (large language model, LLM) 增强主题的语义可解释性，构建基于主题影响力和学科多样性的二维分析框架，旨在识别医学信息学领域兼具多学科来源与影响力的成熟跨学科主题及潜在跨学科主题，并分析这些跨学科主题独特的演化路径，为该领域跨学科研究布局与科研管理提供参考。

2 相关研究

2.1 跨学科测度与传统主题识别方法

既往研究多采用多样性指数、Rao - Stirling 指数及中介中心性等宏观指标量化跨学科性^[1-3]，或者基于引文网络分析学科间的知识吸收与输出特征^[4-5]。在主题内容识别方面，基于共词网络的分析虽能揭示医学信息学中移动健康、电子病历、人工智能等热点^[6-8]，但受限于对高频词的依赖，忽略了词汇在不同上下文中的语义差异。隐含狄利克雷分布 (latent Dirichlet allocation, LDA) 主题模型常用于医学信息和健康医疗相关文本挖掘^[9-11]，但基于词袋假设的局限性导致其难以捕捉词序与深层上下文语义。

2.2 基于 BERTopic 的主题识别

BERTopic 模型结合预训练语言模型与聚类算法，在捕捉文本深层语义与上下文信息方面具有优势^[12-13]。有研究将其应用于医学人工智能^[11-14]及公共卫生政策^[15-16]等领域的主题挖掘，证实了其在高维语义识别方面的有效性。但其默认的主题表示形式多为离散的关键词列表，缺乏语义连贯性，依赖人工进行标签梳理与解读。因此，本研究基于 BERTopic 进行主题识别，利用大语言模型辅助主题语义阐释，并对成熟跨学科主题和潜在跨学科主题进行主题演化分析。

3 资料与方法

3.1 数据来源与处理

以 Web of Science 医学信息学研究论文作为实验数据。于 2025 年 2 月 20 日，利用检索策略 TS = (“Medical Informatics” OR “Health Informatics” OR “Biomedical Informatics” OR “Clinical Informatics” OR “Nursing Informatics” OR “Public Health Informatics” OR “Consumer Health Informatics” OR “Dental Informatics” OR “Pharmacy Informatics” OR “Imaging Informatics” OR “Translational Informatics” OR “eHealth” OR “e - Health” OR “mHealth” OR “m - Health” OR “Mobile Health” OR “Telemedicine” OR “Telehealth” OR “Telerehabilitation” OR “Digital Health” OR “Digital Medicine” OR “Connected Health” OR “Health Information Technology”) AND DT = (“Article” OR “Review”) AND PY = (2001—2024) 进行领域论文检索，共检索到 81 165 篇论文摘要。经过去重，并剔除标题、出版年、摘要、作者、期刊等字段为空的论文，最终获得 67 995 篇论文，涵盖 21 个 Web of Science 学科分类。

3.2 研究方法与分析流程

本研究整合主题模型与大语言模型，提出 Topic - LLM 分析框架，见图 1。其中 BERTopic 用于医学信息学文献主题识别，DeepSeek 用于对提取出的聚类进行语义增强和主题标签生成。

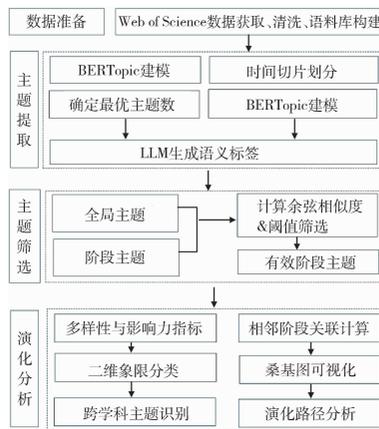


图 1 本研究技术路线

3.2.1 主题提取 利用 BERTopic 对语料库进行主题识别以获取全局主题。利用 PubMedBERT 对语料库进行文本编码, 根据语料库样本量, 将 HDBSCAN 聚类的 `min_cluster_size` 参数设置为 80, UMAP 降维过程的 `n_neighbors` 参数设置为 150, `n_components` 设置为 5。将 BERTopic 模型中 `calcu - late_probabilities` 设置为 True, `n - gram` 设置为 (1, 2), 其他参数使用默认值。基于词频统计生成的主题词可能存在语义泛化的问题, 对此, 同时增强主题特征词的专业性和区分度, 在主题表征阶段使用 BERTopic 内置的 KeyBERT 模型, 提取反映医学信息学上下文语义的主题词。利用信息熵^[17] 确定能够代表医学信息学整体发展的全局主题数量。信息熵能够量化生成主题中单词概率的不均匀分布程度。利用 BERTopic 的相关参数作为变量, 对模型进行多次迭代, 选择信息熵值最低的模型, 以此模型生成的主题作为本研究的全局主题。以 3 年为间隔, 将医学信息学语料库划分为 8 个阶段, 进而分析不同阶段之间的演化过程。结合各阶段语料库样本量, HDBSCAN 聚类过程使用默认参数, 其他参数与全局主题获取时相同。基于 DeepSeek 描述 BERTopic 生成的主题, 利用提示词引导 DeepSeek 归纳 BERTopic 生成的各主题核心信息, 最终形成对主题更为准确和合理的描述。

3.2.2 主题筛选 为筛选有效的阶段主题, 计算阶段主题和全局主题的余弦相似度^[18], 并设定阈值。如果某阶段主题与某全局主题的相似度是其与所有全局主题相似度中的最大值, 且超过阈值, 那么该阶段主题属于该全局主题。根据 Chen B 等^[19] 提出的方法, 按照低成熟度、中成熟度和高成熟度进行相似性强弱三等分, 选择高成熟度阶段主题作为有效主题。

3.2.3 学科多样性指标 使用主题词学科多样性指数 (topic terms interdisciplinarity, TI)^[20] 衡量。其中 g 表示某主题所分布的学科总数, tf 表示该主题的频率。TI 值越高, 表示该主题的学科多样性越高。

$$TI = g \times \log tf \quad (1)$$

3.2.4 主题影响力指标 使用前文获取的全局主题, 结合 Web of Science 学科分类^[21], 构建主题 - 学

科共现网络, 边的权重为对应共现论文的数量, 采用特征向量中心性 (eigenvector centrality, EIG) 作为衡量指标。EIG 通过衡量节点中心性为不同连接分配权重, 进而计算节点的影响力, 能够作为衡量跨学科研究领域重要性的指标^[22]。基于学科多样性与主题影响力, 构建二维四象限模型, 对主题进行特征分类。高 TI 值、高 EIG 值为成熟跨学科主题, 高 TI 值、低 EIG 值为潜在跨学科主题, 低 TI 值、高 EIG 值为专业化支撑主题, 低 TI 值、低 EIG 值为基础研究主题。

3.2.5 主题演化分析 为揭示跨学科主题的内容演化路径, 利用余弦相似度计算相邻阶段主题间的关联强度, 当余弦相似度超过设定阈值时, 判定两者存在演化关系。基于 Chen B 等^[19] 提出的演化框架, 将主题演化路径划分为继承、合并和分裂 3 种模式, 据此分析医学信息学全局主题在各阶段的演化路径。

4 结果

4.1 全局主题提取

BERTopic 共提取全局主题 95 个。利用 BERTopic 对语料库进行层次聚类, 见图 2。全局主题较为分散, 需要继续迭代, 结合信息熵确定最优模型的主题数量。

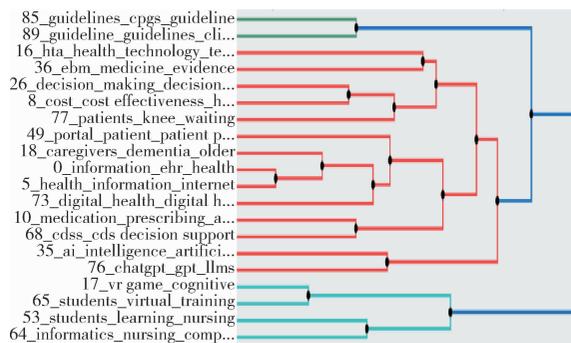


图 2 BERTopic 层次聚类结果

4.2 阶段主题提取

以 3 年为间隔划分时间窗口后, 使用 BERTopic 对各阶段论文进行主题识别, 获取阶段主题。各阶段初始主题数量, 见表 1。

表 1 各阶段主题数量 (个)

序号	阶段	初始主题数量	有效主题数量
1	2001—2003 年	68	59
2	2004—2006 年	64	56
3	2007—2009 年	83	70
4	2010—2012 年	77	66
5	2013—2015 年	91	83
6	2016—2018 年	84	72
7	2019—2021 年	91	80
8	2022—2024 年	87	79

4.3 全局主题筛选

由于初步提取的 95 个全局主题较分散, 基于信息熵继续迭代模型, 以最优模型的主题数量确定最终主题数。经计算最低信息熵值为 0.287, 最优主题数量为 25 个。前 8 个全局主题中前 10 个核心特征词及其权重, 见图 3。筛选后的 25 个全局主题可视化分布, 见图 4, 圆圈大小代表文档规模, 距离代表语义相似度。

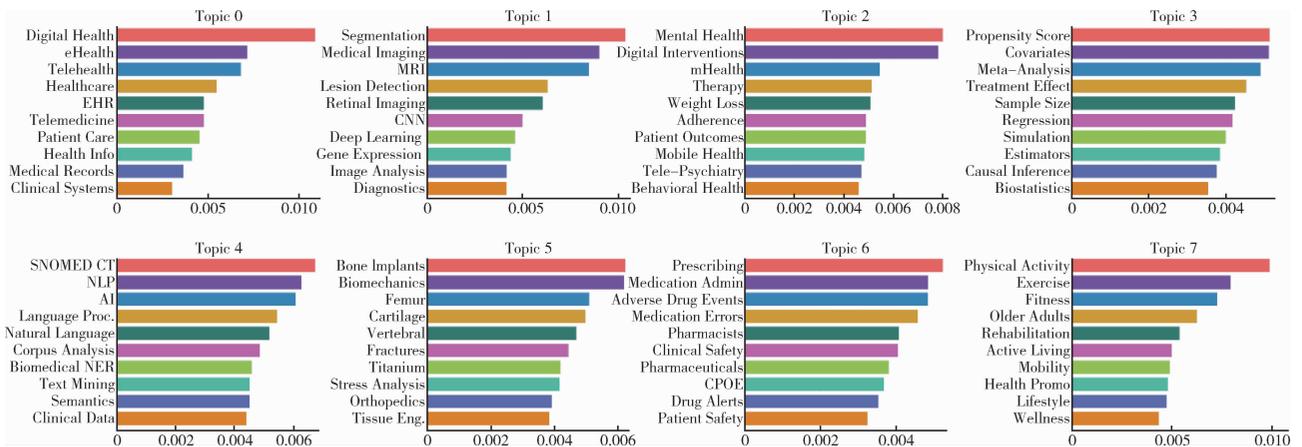


图 3 前 8 个全局主题的特征词

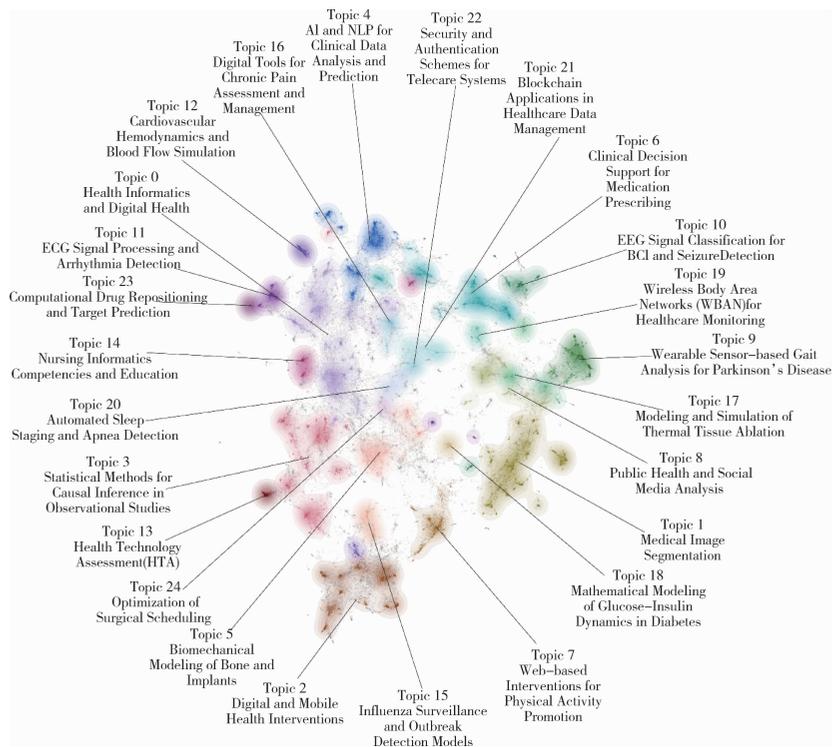


图 4 25 个全局主题的主题地图

4.4 阶段主题筛选

依据前文阶段主题筛选方法，计算每阶段主题与全局主题的余弦相似度，所有阶段主题与全局主题的相似度均低于 0.8，将相似度区间 [0, 0.8] 按低成熟度 [0, 0.27)、中成熟度 [0.27, 0.54)、高成熟度 [0.54, 0.8] 进行 3 等分，确定相似性阈值为 0.54，如果余弦相似度超过该阈值，则认为该阶段主题有效。各阶段有效主题数量见前文表 1。

4.5 学科多样性分析

25 个全局主题文献数量及对应的学科多样性指数，见表 2。高论文数和高 TI 值的主题是领域核心热点，连接多个不同学科，具备很高的跨学科性。低论文数和高 TI 值的主题，虽然尚未形成大规模研究，但其融合了多学科知识和方法解决领域问题。高论文量和低 TI 值的主题在垂直领域发展较成熟，其理论基础和研究方法较集中，具备较高的专业化程度，跨学科程度相对收敛。

表 2 全局主题学科多样性指数

序号	主题名称 (英文)	主题名称 (中文)	论文数 (篇)	TI 值
Topic 0	Health Informatics and Digital Health	健康信息学与数字健康	6 151	195.241
Topic 1	Medical Image Segmentation	医学图像分割	4 636	145.685
Topic 2	Digital and Mobile Health Interventions	数字与移动健康干预	4 294	184.910
Topic 3	Statistical Methods for Causal Inference in Observational Studies	观察性研究中的因果推断统计方法	3 473	164.554
Topic 4	AI and NLP for Clinical Data Analysis and Prediction	临床数据分析与预测中的人工智能与自然语言处理	2 308	140.676
Topic 5	Biomechanical Modeling of Bone and Implants	骨与植入物的生物力学建模	1 540	152.791
Topic 6	Clinical Decision Support for Medication Prescribing	药物处方临床决策支持	1 425	167.304
Topic 7	Web-based Interventions for Physical Activity Promotion	基于网络的身体活动促进干预	1 420	156.335
Topic 8	Public Health and Social Media Analysis	公共卫生与社交媒体分析	1 231	161.097
Topic 9	Wearable Sensor-based Gait Analysis for Parkinson's Disease	基于可穿戴传感器的帕金森病步态分析	1 103	156.492
Topic 10	EEG Signal Classification for BCI and Seizure Detection	用于脑机接口与癫痫检测的脑电信号分类	1 011	123.041
Topic 11	ECG Signal Processing and Arrhythmia Detection	心电信号处理与心律失常检测	941	137.364
Topic 12	Cardiovascular Hemodynamics and Blood Flow Simulation	心血管血流动力学与血流模拟	620	115.491
Topic 13	Health Technology Assessment (HTA)	卫生技术评估	529	66.292
Topic 14	Nursing Informatics Competencies and Education	护理信息学能力与教育	509	102.487
Topic 15	Influenza Surveillance and Outbreak Detection Models	流感监测与暴发检测模型	487	98.877
Topic 16	Digital Tools for Chronic Pain Assessment and Management	慢性疼痛评估与管理的数字工具	425	104.991
Topic 17	Modeling and Simulation of Thermal Tissue Ablation	热组织消融的建模与仿真	353	63.893
Topic 18	Mathematical Modeling of Glucose-Insulin Dynamics in Diabetes	糖尿病葡萄糖-胰岛素动力学数学建模	310	97.773
Topic 19	Wireless Body Area Networks (WBAN) for Healthcare Monitoring	用于医疗监测的无线体域网	303	82.252
Topic 20	Automated Sleep Staging and Apnea Detection	自动化睡眠分期与呼吸暂停检测	278	96.079
Topic 21	Blockchain Applications in Healthcare Data Management	医疗数据管理中的区块链应用	245	91.894
Topic 22	Security and Authentication Schemes for Telecare Systems	远程医疗系统的安全与认证方案	159	41.280
Topic 23	Computational Drug Repositioning and Target Prediction	计算药物重定位与靶点预测	105	58.289
Topic 24	Optimization of Surgical Scheduling	手术排程优化	98	60.424

4.6 主题影响力分析

基于学科多样性 (横轴) 与主题影响力 (纵轴) 绘制二维象限图，见图 5。根据各主题分布特征，第一象限的主题被定义为成熟跨学科主题，第四象限主题被定义为潜在跨学科主题，本研究共识别出 12 个跨学科主题。成熟跨学科主题具备高学

科多样性和高主题影响力，这类主题整合了多学科知识，在知识网络中处于核心地位。例如 Topic 0 和 Topic 2 构建了从数据基础到临床应用的知识链。潜在跨学科主题具备高学科多样性和低主题影响力，属于领域内的新兴增长点或特定应用方向。例如 Topic 8 在疫情背景下快速整合流行病学、信息科学和社交网络分析等多学科主题。

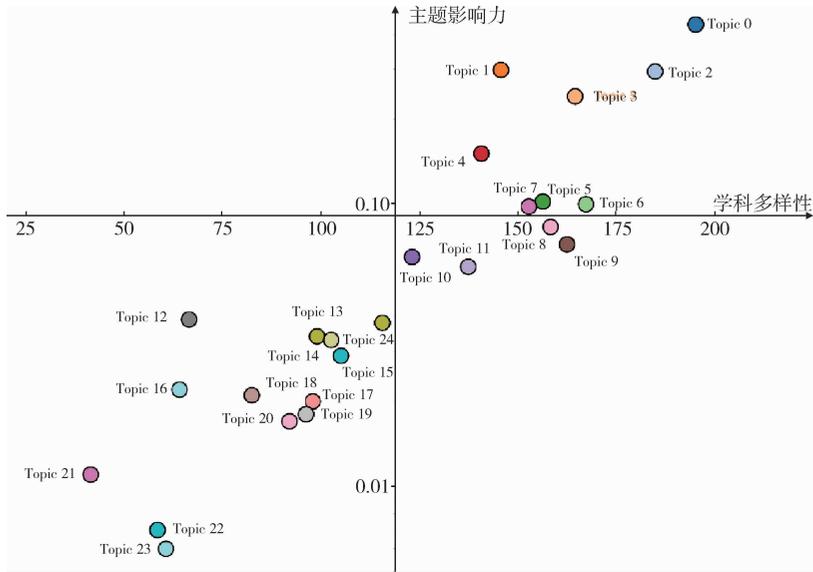


图 5 全局主题象限分析结果

4.7 主题演化分析

以成熟跨学科主题 Topic1 和潜在跨学科主题 Topic10 为例，进行主题演化分析。

4.7.1 Topic1 医学图像分割 医学图像分割的主题演化分析结果，见图 6。

(S8 - T65)、3D 建模 (S8 - T40) 及多器官分析 (S8 - T16) 等前沿方向，反映了医学图像研究向精准医学的演进。

4.7.2 Topic10 用于脑机接口与癫痫检测的脑电信号分类 主题演化分析结果，见图 7。该主题整体以继承和合并为主，处于吸收学科知识并逐步发展的阶段。S1—S3 阶段，研究主要围绕信号去噪 (S1 - T15) 与特性分析 (S2 - T4、S3 - T17) 展开，路径相对单一。S4—S8 阶段，以 S4 - T50、S5 - T30、S6 - T30 为代表的脑机接口技术进一步发展，并吸收癫痫监测 (S7 - T68) 中的特征提取方法，两者在 S8 阶段融合为脑机接口分类算法 (S8 - T33)。疾病诊断应用持续深化，以 S6 - S81、S7 - T31、S8 - T36 主题为代表的阿尔茨海默病的 EEG 研究逐步演化，形成独立的特征识别与诊断体系。

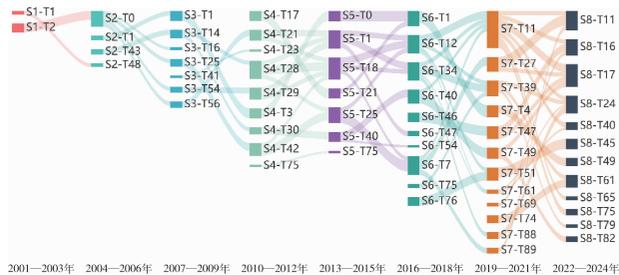


图 6 Topic1 主题演化分析

该主题呈现继承 - 分裂 - 合并 - 再分裂的演化路径，研究重点从图像分割逐步演变为图像分类与影像组学。S1—S3 阶段，通用图像分割方法 (S2 - T0) 成为研究核心，并应用于眼科 (S3 - T25) 与超声 (S3 - T56) 等分支。S4—S7 阶段的发展主要包括分裂和合并，特定器官如大脑 (S4 - T29)、颈动脉 (S4 - T21) 的图像分析需求促进通用图像算法 (S5 - T1) 与图像分类方法 (S7 - T11) 融合。至 S8 阶段，主题呈现高度分化，图像分类方法进一步演化为癌症图像分类 (S8 - T17)、影像组学

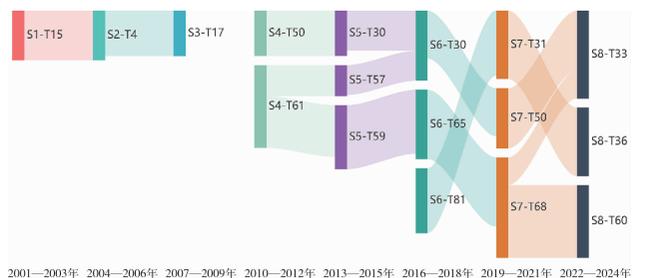


图 7 Topic10 主题演化分析

5 讨论

5.1 针对成熟跨学科主题建立持续发展机制

成熟跨学科主题形成了融合理论、技术和应用的发展路径,应针对其建立持续发展机制。通过制定跨学科的技术规范与临床指南,降低技术落地门槛。改革科研评价与人才培养体系,形成跨学科研究成果评价机制,培养复合型人才。推动多中心真实世界研究,持续验证成熟跨学科主题研究成果对专科领域的支撑能力。

5.2 针对潜在跨学科主题优化资源配置

潜在跨学科主题由于外部技术与学科的融合程度不足,应用场景未普及。应优化其资源配置,提升其影响力。政府科研管理部门应设立专项长周期资助,支持跨学科前沿探索。高校和科研院所应打破学科界限,组建以医疗场景需求为导向的跨学科研究团队。医疗机构与产业界应加强应用场景相关研究,促进跨学科研究成果在细分领域落地。

5.3 推动建立统一的数据体系

数据是跨学科研究的重要保障,应持续推动数据体系建设发展。加快推进国家级医疗大数据标准建设,推动构建专病队列与多模态数据集。医疗机构应强化数据治理,利用隐私计算和先进信息技术打破数据孤岛,构建跨区域、跨机构的数据融合平台,为跨学科研究提供底层支撑。

5.4 加强成果转化体系建设

跨学科研究以解决临床问题、满足健康需求为核心目的,相关学术成果的落地依赖完善的成果转化体系。建立科研成果与临床实践双向反馈机制,促进科研成果在真实场景中快速迭代。企业应基于先进信息技术构建个人健康管理体系,实现智慧医疗的普惠落地。

6 结语

跨学科研究是应对当代复杂社会挑战和推动科

学创新的关键。本研究应用 BERTopic 从医学信息学论文中识别主题,然后通过主题影响力和学科多样性二维分析框架确定跨学科主题,系统分析了成熟跨学科主题和潜在跨学科主题的演化路径。本研究仍存在局限性,数据来源主要为期刊文献,而未包含会议论文、技术报告和临床指南等其他知识载体。未来将整合多元文献类型,以提升该跨学科主题识别与演化分析框架的全面性。

作者贡献: 苏中琪负责数据处理、实验实施、论文撰写;董森负责数据收集与清洗;崔雷负责研究设计、论文修订。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 LEYDESDORFF L, WAGNER C S, BORNMAN L. Interdisciplinary as diversity in citation patterns among journals: rao-stirling diversity, relative variety, and the gini coefficient [J]. *Journal of informetrics*, 2019, 13 (1): 255-269.
- 2 LEYDESDORFF L, HELLSTEN I. Measuring the meaning of words in contexts: an automated analysis of controversies about "monarch butterflies," "frankenfoods," and "stem cells" [J]. *Scientometrics*, 2006, 67 (2): 231-258.
- 3 STIRLING A. A general framework for analysing diversity in science, technology and society [J]. *Journal of the royal society interface*, 2007, 4 (15): 707-719.
- 4 陈小清,李炜超,刘丽,等. 医学信息学跨学科测度指标与论文影响力关系 [J]. *中华医学图书情报杂志*, 2022, 31 (2): 42-49.
- 5 李点,文庭孝,许林勇. 引文施引双角度医学信息学交叉测度分析 [J]. *医学信息学杂志*, 2024, 45 (5): 46-52, 64.
- 6 刘玉婷. 国际医学信息学领域研究现状与前沿分析 [J]. *医学信息学杂志*, 2023, 44 (7): 38-44, 67.
- 7 HE X, PENG C, XU Y, et al. Global scientific research landscape on medical informatics from 2011 to 2020: bibliometric analysis [J]. *JMIR medical informatics*, 2022, 10 (4): e33842.
- 8 HU J, LI C, GE Y, et al. Mapping the evolution of digital health research: bibliometric overview of research hotspots, trends, and collaboration of publications in JMIR (1999-2024) [J]. *Journal of medical internet research*, 2024, 26 (10): e58987.
- 9 荣越,商宪丽,谢冰冰. 生命周期视角下医学信息学科的主题识别与演化研究 [J]. *情报科学*, 2023, 41 (5): 169-180.

(下转第 69 页)

- (社会科学版), 2022, 24 (6): 125 - 133.
- 9 翟运开, 田远航, 赵栋祥. 政策供需匹配视角下养老产业数字化政策效果评估及提升路径 [J]. 中国卫生事业管理, 2023, 40 (7): 481 - 485, 490.
 - 10 王方意, 刘春华, 王兵兵, 等. 基于 LDA 与社会网络分析的我国省级体育产业政策主题热点识别与演进分析 [J]. 山东体育学院学报, 2024, 40 (6): 97 - 107.
 - 11 魏萍, 蔡润雨, 颜家琴. 数字经济政策扩散的网络特征与动力机制研究——基于 533 份政策文本的社会网络分析 [J]. 管理学刊, 2025, 38 (1): 97 - 113.
 - 12 田园宏, 林江洋, 姚前, 等. 长三角城市群水污染防治府际合作网络特征及影响因素 [J]. 资源科学, 2025, 47 (2): 322 - 331.
 - 13 周正柱, 沈思含. 三维分析框架下长三角地区科技创新政策协同演进研究 [J]. 软科学, 2024, 38 (10): 31 - 39, 62.
 - 14 郭磊, 黄春柳, 王辰. 智慧养老政策的量化评价与优化路径——基于 PMC 指数模型的分析 [J]. 南京理工大学学报 (社会科学版), 2025, 38 (2): 18 - 30.
 - 15 王新建, 胡广伟, 司文峰, 等. 数字环境下社区养老精准化服务价值网络构建——基于政策文本内容的分析 [J]. 情报科学, 2024, 42 (7): 10 - 22.
 - 16 吴飞, 郭怡格. “政策工具—利益相关者”二维框架下公共服务政策注意力配置研究——基于央地 68 份医养结合政策文本分析 [J]. 领导科学, 2025 (2): 144 - 151.
 - 17 高歌. 城镇职工基本养老保险政策文本研究——基于政策工具、政策目标、政策主体的三维分析框架 [J]. 北京航空航天大学学报 (社会科学版), 2023, 36 (3): 133 - 142.
 - 18 CASTRO - LEAL F, DAYTON J, DEMERY L, et al. Public spending on health care in Africa: do the poor benefit [J]. Bulletin of the world health organization, 2000, 78 (1): 66 - 74.
 - 19 叶阿忠, 田小梅, 李田田, 等. 长江经济带新质生产力空间关联网络结构特征及效应分析 [J]. 华东经济管理, 2025, 39 (4): 21 - 32.
 - 20 ROTHWELL R, ZEGVELD W. An assessment of government innovation policies [M] // ROESSNER J D. Government innovation policy. London: Palgrave Macmillan UK, 1988.
 - 21 CALLON M. The sociology of an sector - network: the case of the electric vehicle [M] // CALLON M, LAW J, RIP A. Mapping the dynamics of science and technology: sociology of science in the real world. London: Palgrave Macmillan UK, 1986.
 - 22 胡世文, 祁志伟. 政策工具视角下数字政府建设政策文本研究——基于省级政策文本 (2019—2021) 的分析 [J]. 西南民族大学学报 (人文社会科学版), 2023, 44 (1): 188 - 200.

(上接第 61 页)

- 10 陈琼, 朱庆华, 闵华, 等. 基于领域主题的学科交叉特征识别方法研究——以医学信息学为例 [J]. 现代情报, 2022, 42 (4): 11 - 24.
- 11 袁永旭, 王涟, 殷彩明, 等. 基于 BERTopic 模型的我国医学人工智能领域的主题识别与内容分析 [J]. 中国数字医学, 2025, 20 (4): 89 - 96.
- 12 GROOTENDORST M. BERTopic: neural topic modeling with a class - based TF - IDF procedure [EB/OL]. [2025 - 09 - 26]. <https://arxiv.org/abs/2203.05794>.
- 13 WANG Z, CHEN J, CHEN J, et al. Identifying interdisciplinary topics and their evolution based on BERTopic [J]. Scientometrics, 2024, 129 (11): 7359 - 7384.
- 14 牛振东, 和晓峰, 刘晓琦, 等. 基于 BERTopic 模型的医学人工智能研究主题挖掘及演化特征分析 [J]. 医学信息学杂志, 2024, 46 (2): 54 - 61.
- 15 谢逸钧, 邵瑛琦, 姜成华, 等. 基于 BERTopic 的残疾儿童康复救助政策主题识别及演化分析 [J]. 医学信息学杂志, 2025, 46 (10): 51 - 57, 75.
- 16 夏苏迪, 谢靖, 是沁, 等. 基于 BERTopic 模型的我国中医药老年康养政策量化研究 [J]. 医学信息学杂志, 2025, 46 (6): 43 - 49.
- 17 MACKAY D. Information theory, inference and learning algorithms [M]. Cambridge: Cambridge University Press, 2003.
- 18 AMAN V. A new bibliometric approach to measure knowledge transfer of internationally mobile scientists [J]. Scientometrics, 2018, 117 (1): 227 - 247.
- 19 CHEN B, TSUTSUI S, DING Y, et al. Understanding the topic evolution in a scientific domain: an exploratory study for the field of information retrieval [J]. Journal of informetrics, 2017, 11 (4): 1175 - 1189.
- 20 XU H, GUO T, YUE Z, et al. Interdisciplinary topics of information science: a study based on the terms interdisciplinarity index series [J]. Scientometrics, 2016, 106 (2): 583 - 601.
- 21 LEYDESDORFF L, CARLEY S, RAFOLS I. Global maps of science based on the new web - of - science categories [J]. Scientometrics, 2013, 94 (2): 589 - 593.
- 22 HEO P S, LEE D H. Evolution patterns and network structural characteristics of industry convergence [J]. Structural change and economic dynamics, 2019, 51 (12): 405 - 426.