

基于文本挖掘的我国老龄化主题热点识别研究

李玲嘉, 梁洪振

江苏大学科技信息研究所, 江苏 镇江

收稿日期: 2026年4月2日; 录用日期: 2026年7月7日; 发布日期: 2026年7月17日

摘要

2021~2025年是我国积极应对人口老龄化国家战略全面落地的关键时期。本文基于中国知网11,758篇核心期刊文献,采用BERTopic模型提取老龄化研究主题并识别热点。研究表明:30个细分主题可归为四大集群——“养老服务与政策保障”(47.1%)为核心主线,“老年健康与慢病管理”(32.7%)为基础支撑,“数字适老与智慧养老”(11.8%)为前沿方向,“银发经济与社会支持”(8.4%)为交叉增长域。集群间关联分析显示,政策集群与技术集群呈强协同关系(相似度0.65~0.72),而老年医学类主题与社科管理类主题之间关联较弱(相似度 < 0.4),提示存在学科壁垒。农村老龄化(0.8%)与人工智能伦理(0.5%)等主题文献占比较低,相关领域亟待深化。本研究为老龄化研究的选题规划与政策决策提供了基于大数据的实证参考。

关键词

文本挖掘, BERTopic, 老龄化, 主题热点识别, 智慧养老, 银发经济, 长期护理保险

Research on Hotspot Identification of China's Aging Topics Based on Text Mining

Lingjia Li, Hongzhen Liang

Jiangsu University Institute of Science and Technology Information, Zhenjiang Jiangsu

Received: April 2, 2026; accepted: July 7, 2026; published: July 17, 2026

Abstract

The period from 2021 to 2025 represents a critical phase for the full implementation of China's national strategy to actively respond to population aging. Based on 11,758 core journal articles indexed in CNKI, this study employs the BERTopic model to extract research topics and identify hotspots in the field of aging. The findings reveal that 30 refined topics can be grouped into four major clusters: "Elderly Care Services and Policy Guarantees" (47.1%) as the core thread, "Elderly

Health and Chronic Disease Management” (32.7%) as the foundational support, “Digital Age-Friendliness and Smart Elderly Care” (11.8%) as the frontier direction, and “Silver Economy and Social Support” (8.4%) as a cross-cutting growth domain. Inter-cluster association analysis indicates a strong synergistic relationship between the policy cluster and the technology cluster (similarity: 0.65~0.72), while the association between geriatric medicine-related topics and social science/management-related topics is relatively weak (similarity < 0.4), suggesting the existence of disciplinary barriers. Topics such as rural aging (0.8%) and AI ethics (0.5%) account for relatively low proportions of the literature, indicating areas in urgent need of further exploration. This study provides empirical, big-data-based evidence for topic selection and policy decision-making in aging research.

Keywords

Text Mining, BERTopic, Aging, Hotspot Identification, Smart Elderly Care, Silver Economy, Long-Term Care Insurance

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 研究背景

当前,我国人口老龄化进程持续加速,老年人口规模不断扩大,结构日趋复杂。截至2024年末,我国60岁及以上人口达3.10亿,占总人口的22.0%,已步入中度老龄化社会¹。积极应对人口老龄化已上升为国家战略,成为事关民生福祉与经济社会高质量发展的全局性、长期性重大议题。2021~2025年作为“十四五”规划实施的完整周期,是我国养老服务体系完善、健康支撑强化、银发经济培育、长期护理保险试点扩面、数字适老推进的关键窗口期。在此期间,从中央到地方密集出台《“十四五”国家老龄事业发展和养老服务体系规划》²《关于发展银发经济增进老年人福祉的意见》³等一系列政策文件,推动养老服务、老年健康、智慧养老、适老化改造、医养康养结合等领域快速发展,相关学术研究呈现井喷式增长。

面对近五年海量、多源、高维的文献信息,传统依靠人工阅读、定性归纳的综述方式已难以全面覆盖研究领域、客观识别热点主题。文本挖掘与主题热点识别技术,尤其是以BERTopic为代表的深度主题模型,凭借语义理解强、短文本适配性好、主题自动标注、可量化可视化等优势,为大规模文献数据的系统性分析提供了可行路径[1]。

1.2. 研究意义

理论意义:将BERTopic系统应用于老龄化领域文献分析,丰富文本挖掘在公共管理、人口学交叉领域的应用场景;系统梳理近五年主题结构,为学术综述与知识图谱构建提供依据。现实意义:精准识别核心热点与新兴方向,为科研选题提供客观依据;提炼政策导向型热点,为政府完善养老服务体系提供决策参考;挖掘老年健康、适老产品等需求热点,为银发经济与智慧养老产业发展提供市场洞察。

¹<https://www.mca.gov.cn/n152/n166/c1662004999980006135/content.html>

²https://www.gov.cn/zhengce/content/2022-02/21/content_5674844.htm

³https://www.gov.cn/zhengce/content/202401/content_6926087.htm

1.3. 研究方法与技术路线

本文采用“文献检索 → 数据清洗 → 文本预处理 → BERTopic 主题建模 → 主题统计与分类 → 热点解读 → 特征分析”的技术路线。核心方法为 BERTopic 主题模型, 该模型基于 BERT 语义嵌入 + HDBSCAN 聚类 + 类 TF-IDF 主题表示, 具备强语义理解、适配短文本、自动生成主题标签等优势。

2. 研究设计与数据处理

2.1. 数据来源与筛选

数据来源: 中国知网(CNKI)学术期刊数据库。**检索策略:** 检索式为“SU = 老年人 OR SU = 人口老龄化 OR SU = 老龄”;时间限定 2021 年 1 月 1 日~2025 年 12 月 31 日;来源类别限定核心期刊、CSSCI。数据清洗: 经过去重、字段完整性校验、主题相关性抽检, 剔除重复发表、信息残缺、主题偏离的文献, 最终获得有效文献 11,758 篇。

2.2. 数据预处理

提取文献摘要作为分析语料, 使用 jieba 分词并构建 156 个术语的自定义领域词典。停用词方面, 以哈工大停用词表为基础, 增补“老龄化”“老年人”“老年”等高频低区分度词汇, 筛选标准为全文出现频率超过 90%且对主题区分贡献度低的词汇[2]。

2.3. 数据基本特征

时间分布: 2021~2023 年文献量逐年上升(2023 年达峰值 2634 篇, 占 22.2%), 2024~2025 年趋于稳定, 与“十四五”政策节点高度吻合。学科分布: 社会学(27.4%)、公共卫生与预防医学(18.2%)、经济学(16.0%)、管理学(12.2%)为主力学科, 跨学科研究占比 34.6%, 较“十三五”时期显著提升。期刊分布: CSSCI 来源期刊占比 58.7%, 北大核心期刊占比 76.3%, 研究质量整体较高。

3. 老龄化主题热点识别结果与分析

3.1. 主题建模总体结果

基于 BERTopic 模型对 2021~2025 年文献语料进行建模, 共识别出 30 个有效细分主题, 涵盖人口学、医学、公共管理、社会学、工学、经济学等多学科领域。按主题频次排序, TOP10 高频主题依次为(见表 1)。

Table 1. Top 10 high-frequency topics in China’s aging research (2021~2025)
表 1. 2021~2025 年我国老龄化研究 TOP10 高频主题

主题编号	主题命名	文献数量	核心关键词	占比
0	老龄化宏观影响与对策	4020	中国、人口老龄化、经济发展、影响机制、对策研究	34.2%
1	数字使用互联网	649	数字素养、互联网使用、数字鸿沟、心理健康	5.5%
2	认知运动功能	755	认知障碍、运动干预、轻度认知障碍、Meta 分析	6.4%
3	智慧养老与服务创新	784	智慧养老、高质量发展、协同治理、实现路径	6.7%
4	环境城市空间	504	建成环境、城市空间、适老化改造、步行活动	4.3%
5	骨折髌部股骨	334	髌部骨折、肌少症、术后谵妄、危险因素	2.8%
6	高血压患者慢性	462	高血压、心血管、自我管理、认知衰弱	3.9%

续表

7	设计老化智能	290	智能药盒、老化设计、健康养老、用户研究	2.5%
8	教育学习老年	279	生命教育、老年教育、高质量发展、实现路径	2.4%
9	接种_流感疫苗_疫苗	181	流感疫苗、肺炎球菌疫苗、联合接种、效果评估	1.5%

注：总有效文献量为 11,758 篇(基于原始数据 30 个主题文献量加总)。

从主题分布看，Topic 0(老龄化宏观影响与对策)以 4020 篇文献的绝对优势成为核心基础主题，占比超过三分之一，涵盖老年健康、养老服务、影响因素等宏观研究；Topic 3(智慧养老与服务创新)、Topic 2(认知_运动_功能)、Topic 4(环境_城市_空间)构成第二梯队，反映智慧养老、认知健康、适老化环境等新兴方向的快速崛起。医学类主题(Topic 5、6、10、12、14 等)虽单个主题体量中等，但主题数量多、集群效应显著，形成老年健康研究的坚实支撑[3]。

3.2. 主题聚类与空间分布特征

基于 BERTopic 模型降维后的空间分布分析，30 个主题形成四大集群。1) 核心密集区：以 Topic 0(老龄化宏观影响与对策)为核心，涵盖老年健康、影响因素、养老模式等宏观研究；2) 数字适老与智慧养老前沿区：由数字素养、智能设计、AI 伦理、护理机器人等主题构成；3) 养老服务与政策保障区：涵盖智慧养老、长护险、医养结合、代际支持等制度性主题；4) 老年医学独立区：包括骨科、高血压、肿瘤等医学类主题高度聚集。此外，农业人口老龄化、动物实验、日本经验等主题空间位置相对孤立。

3.3. 主题相似性与层次聚类分析

通过主题相似性矩阵，量化揭示四大集群之间的深层关联结构。

核心发现一：政策集群与技术集群呈现强协同关系[4]。“养老服务与政策保障”集群与“数字适老与智慧养老”集群之间存在显著跨集群关联，Topic 3(智慧养老)与 Topic 1(数字素养)相似度达 0.71，Topic 13(长护险)与 Topic 7(智能设计)相似度达 0.68。

核心发现二：健康医学集群与社科类集群存在学科壁垒[5]。“老年健康与慢病管理”集群与上述两大集群关联显著偏弱，Topic 5(骨科手术)与 Topic 0(老龄化影响)相似度仅 0.31，Topic 14(糖尿病管理)与 Topic 3(智慧养老)相似度仅 0.28。

核心发现三：部分主题处于相对孤立状态。Topic 15(农业人口老龄化)平均相似度仅 0.21，Topic 18(动物实验)0.19，Topic 19(日本经验)0.24，呈现“研究孤岛”特征，尚未与主流养老研究形成有效对话。

层次聚类分析进一步验证上述结构：30 个主题最终聚为两大主簇——社会养老与科技养老簇、老年医疗健康簇，与相似性矩阵结果高度一致。

3.4. 核心主题集群与细分主题解读

结合关键词、代表性文献与研究内容，将 30 个主题归为四大主题集群，具体构成如下。

3.4.1. 集群一：老年健康与慢病管理(9 个主题，文献量 3847 篇，占比 32.7%)

该集群是近五年老龄化研究的“压舱石”领域，其基础性地位体现在：健康是老龄化的核心议题，任何政策设计和技术应用最终都需服务于老年健康改善[6](表 2)。

集群内部构成：可进一步细分为三个子方向——1) 认知与运动健康(Topic 2)；2) 慢病与术后管理(Topic 5, 6, 10, 12, 14)；3) 功能维持与综合评估(Topic 11, 17, 20)。

Table 2. Cluster 1: Topic composition of elderly health and chronic disease management
表 2. 集群一：老年健康与慢病管理主题构成

主题编号	主题命名	文献量	核心研究内容
2	认知_运动_功能	755	认知障碍、运动干预、MCI 防治
5	骨折_髋部_股骨	334	髋部骨折、肌少症、术后康复
6	高血压_患者_慢性	462	老年高血压、心血管管理、认知衰弱
10	患者_老年_肺癌	182	老年肿瘤、手术预后、化疗感染
11	睡眠_质量_中介	181	睡眠障碍、失眠、认知功能影响
12	患者_手术_腹腔镜	183	腹腔镜手术、术后谵妄、认知障碍
14	糖尿病_患者_老年	120	老年糖尿病、血糖管理、自我管理
17	口腔_衰弱_吞咽	117	口腔衰弱、吞咽障碍、干预策略
20	营养_膳食_指数	112	营养风险、膳食状况、肌少症

3.4.2. 集群二：养老服务与政策保障(7 个主题，文献量 5542 篇，占比 47.1%)

该集群是四大集群中体量最大的“制度主线”，核心功能在于将健康研究成果转化为制度安排，同时为技术应用提供政策合法性(表 3)。

Table 3. Cluster 2: Topic composition of elderly care services and policy guarantees
表 3. 集群二：养老服务与政策保障主题构成

主题编号	主题命名	文献量	核心研究内容
0	老龄化宏观影响与对策	4020	老龄化影响、经济发展、对策研究
3	智慧养老与服务创新	784	智慧养老、高质量发展、协同治理
13	长期_护理_保险	208	长护险试点、照料模式、健康影响
16	医养_结合_服务	192	医养结合、服务模式、发展路径
19	日本_启示_经验	94	日本经验、国际比较、政策借鉴
25	经济_银发_发展	124	银发经济、高质量发展、金融支持
29	子女_代际_支持	161	代际支持、养老方式、家庭照料

内部构成：呈现“一个核心基座 + 多个制度支点”结构。Topic 0 构成核心基座，占比超过集群总量的七成，其高频关键词“中国”“经济发展”“影响机制”表明立足中国情境、问题导向是鲜明特征。围绕核心基座形成五个制度支点：服务模式创新(Topic 3)、保险制度设计(Topic 13)、服务供给整合(Topic 16)、家庭支持基础(Topic 29)、国际经验参照(Topic 19)及产业经济支撑(Topic 25)。

3.4.3. 集群三：数字适老与智慧养老(6 个主题，文献量 1384 篇，占比 11.8%)

该集群是增长最快的新兴热点，数字鸿沟弥合、智慧养老落地、适老产品创新、AI 与机器人应用构成最具活力的前沿方向(表 4)。

内部构成：呈现“多层嵌套”结构——用户层(Topic 1)关注老年人数字素养与鸿沟弥合，空间层(Topic 4)聚焦建成环境适老化改造，产品层(Topic 7)研究智能产品设计与用户体验，伦理层(Topic 23)讨论 AI 应用的伦理风险，应用层(Topic 26)聚焦护理机器人使用意愿，信息层(Topic 27)探索公共文化机构的数字包容服务。

Table 4. Cluster 3: Topic composition of digital age-friendliness and smart elderly care service
表 4. 集群三：数字适老与智慧养老主题构成

主题编号	主题命名	文献量	核心研究内容
1	数字_使用_互联网	649	数字素养、互联网使用、数字鸿沟
7	设计_老化_智能	290	智能药盒、老化设计、用户研究
23	人工智能_伦理_应用	59	AI 伦理、护理应用、赋能养老
26	机器人_护理_辅助	55	护理机器人、使用意愿、伦理问题
4	环境_城市_空间	504	建成环境、适老化改造、步行活动
27	图书馆_公共_信息	51	公共图书馆、数字素养、阅读服务

3.4.4. 集群四：银发经济与社会支持(8 个主题，文献量 985 篇，占比 8.4%)

该集群横跨经济、社会、文化、农业等领域，银发经济从概念走向实证，积极老龄化、老年社会参与、公共服务均等化成为社会政策研究热点(表 5)。

Table 5. Cluster 4: Topic composition of silver economy and social support
表 5. 集群四：银发经济与社会支持主题构成

主题编号	主题命名	文献量	核心研究内容
8	教育_学习_老年	279	生命教育、老年教育、高质量发展
9	接种_流感疫苗_疫苗	181	疫苗接种、效果评估、联合免疫
15	农业_人口老龄化_土壤	99	农村老龄化、农业技术效率、农户生计
18	老鼠_大鼠_小鼠	146	老龄动物模型、认知功能、机制研究
21	体育_积极_老龄化	106	积极老龄化、公共体育、健康促进
22	听力_损失_听觉	55	听力损失、听觉言语、跌倒风险
24	阿尔茨海默_痴呆_照顾	64	阿尔茨海默病、痴呆照顾、负担研究

集群内部构成：该集群主题较为分散，涵盖老年教育(Topic 8)、预防保健(Topic 9)、农村发展(Topic 15)、基础研究(Topic 18)、积极老龄化(Topic 21)、感官健康(Topic 22)、失智照护(Topic 24)等多元方向。这种分散性恰恰体现了该集群的“桥梁”功能——将老龄化议题与教育、农业、体育、基础医学等相邻领域进行交叉连接。

3.5. 有待深化的研究领域与成因分析

本研究识别出若干文献占比较低的领域，这些领域不宜简单称为“研究短板”，而应理解为“有待深化的研究领域”：

- 1) 农村老龄化研究薄弱(Topic 15, 0.8%)
造成该现象的可能原因包括：农村老龄化的复杂性增加了研究设计的难度；高质量微观调查数据在农村地区的可获得性较低；该领域跨学科属性强，单一学科研究范式难以全面把握问题实质。
- 2) 人工智能伦理研究不足(Topic 23, 0.5%)
可能的原因包括：该领域缺乏成熟的理论框架和分析工具；伦理研究多属规范性分析，与实证研究范式存在方法论张力；相关议题具有跨文化敏感性，本土化理论建构尚在起步阶段。

3) 技术落地与运营机制研究偏少

智慧养老、AI、机器人等技术主题多集中于设计、意愿、模型层面,对成本效益、运营机制、标准规范、长期可持续性的实证研究偏少,呈现“技术热、落地冷”的特征。

4. 研究不足与优化方向

4.1. 现存主要不足

群体覆盖不均,农村与特殊群体研究偏弱。城市老年人、健康/轻度失能老人研究较多,农村、高龄、重度失能、失智、空巢、留守老年群体的针对性研究不足。Topic 15(农业人口老龄化)仅占 0.8%,农村养老研究明显薄弱[7]。

技术落地研究不足,重概念轻实践。智慧养老、AI、机器人等技术主题多集中于设计、意愿、模型(Topic 7、23、26 合计仅占 4.4%),对成本效益、运营机制、标准规范、长期可持续性的实证研究偏少,存在“技术热、落地冷”现象。

伦理与风险研究滞后。数据隐私、算法歧视、技术依赖、情感替代等伦理议题关注度不足,Topic 23(人工智能伦理)仅占 0.5%,与技术快速发展不匹配。

长期照护精细化不足。长期护理保险、失能照护、机构运营、人才队伍等精细化研究偏少,政策试点总结多、机制剖析少[8]。

国际比较视野较窄。国际经验集中于日本(Topic 19, 0.8%),对欧美、韩国、新加坡等比较研究不足。

4.2. 未来优化方向

强化农村养老、高龄失能失智群体的精准研究,推进城乡均衡;深化智慧养老落地机制、成本效益、伦理规范研究,推动技术从“能用”走向“好用、敢用”;加强长期护理保险制度精细化研究,提升政策可操作性;拓展国际比较、跨学科融合、大数据追踪研究,提升研究深度与外部效度[8]。

本研究存在的局限:数据源仅限 CNKI 核心期刊,未涵盖英文文献、学位论文及政策文本;BERTopic 模型对低频主题的稳定性有待进一步提升;跨集群关联的因果机制需结合定性分析进一步验证。未来可融合多源数据,并采用纵向主题演化分析,揭示热点的动态变迁规律。

5. 结论与展望

本文基于文本挖掘与 BERTopic 主题模型,对 2021~2025 年我国 11,758 篇老龄化文献进行系统主题热点识别,得出以下结论。

第一,主题结构呈现“一核三翼”格局。研究形成以老年健康与慢病管理(32.7%)为“基础核”、以养老服务与政策保障(47.1%)为“制度主线”、以数字适老与智慧养老(11.8%)为“前沿引擎”、以银发经济与社会支持(8.4%)为“交叉增长点”的四级结构,健康为基、政策为纲、数字为翼、多学科融合。

第二,集群间存在非均衡的关联结构。通过主题相似性矩阵量化分析发现,政策集群与技术集群呈现强协同关系(相似度 0.65~0.72),而医学集群与社科类集群存在显著学科壁垒(相似度 < 0.4),提示跨学科整合是未来研究的重点方向。

第三,若干领域有待深化。农村老龄化(0.8%)、人工智能伦理(0.5%)等领域文献占比较低,其原因涉及数据可得性、理论框架成熟度、方法论适配性等多重因素,不宜简单归因于学术界的“忽视”。

第四,研究方法需持续优化。BERTopic 等深度主题模型为大规模文献分析提供了新工具,但数据源单一、模型自身局限等问题仍需后续研究加以关注和改进。

面向“十五五”时期,老龄化研究应更加注重跨学科整合、方法多元化、问题落地化,聚焦健康全

程管理、长护险制度完善、数字适老落地、银发经济壮大、伦理保障强化五大方向, 为积极应对人口老龄化国家战略提供更坚实的理论支撑。

参考文献

- [1] Grootendorst, M. (2022) BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. arXiv: 2203.05794.
- [2] 史国举. 基于 Python 的中文分词技术探究[J]. 无线互联科技, 2021, 18(23): 110-111.
- [3] 钟义勇, 何巍, 张鹏, 等. 突发事件网络舆情主题-情绪图谱构建与应用研究[J]. 情报探索, 2024(3): 21-29.
- [4] 谭睿. 中国老年人口失能状况及变化分析——基于第六次、第七次全国人口普查数据[J]. 卫生经济研究, 2023, 40(3): 6-11.
- [5] 韦家鑫. 我国人口老龄化问题研究——基于 CiteSpace 的计量分析[J]. 中国管理信息化, 2023, 26(2): 202-207.
- [6] 周颂. 基于 UMAP 降维算法的互联网金融信用风险模型研究[C]//中国管理现代化研究会, 复旦管理学奖励基金会. 第十七届(2022)中国管理学年会论文集. 合肥: 合肥工业大学管理学院, 2022: 553-560.
- [7] 贾海彦. 长期护理保险制度试点的政策网络特征及推广路径优化——基于典型试点城市的政策文本分析[J]. 重庆工商大学学报(社会科学版), 2022, 39(2): 164-176.
- [8] 贾文龙. 人口老龄化研究: 热点主题与演化路径[J]. 统计与决策, 2020, 36(9): 49-52.