



毕 业 论 文

题 目 基于多层创新生态网络的

养老科技产业创新机会发现

姓 名 乔 轲

学 号 20110121

指导教师 袁 菲

日 期 2024/6/19

北京工业大学

毕业设计（论文）任务书

题目基于多层创新生态网络的养老科技产业创新机会发现

专业 信息管理与信息系统 学号 20110121 姓名 乔轲

主要内容、基本要求、主要参考资料等：

研究内容：

2020 年中国大陆 60 岁及以上老年人口总数为 2.64 亿，占总人口的 18.7%。20 世纪 60 年代第二次生育高峰形成的较大人口将逐步进入老年人口，这将使我国人口老龄化水平从近年来短暂相对缓慢的演变逆转为增长的“快车道”。另一方面，全球预期寿命从 2000 年的 66.8 岁增加到 2019 年的 73.3 岁，而健康预期寿命从 58.3 岁增加到 63.7 岁。预期寿命的增长速度略快于健康预期寿命，说明老年人残疾年限的比例变高。其中很大一部分老年人也患有慢性病，他们需要养老产业，特别是医疗产业的帮助和支持。老龄化过程不仅意味着老年人口的增加，而且还导致劳动力短缺，这进一步加剧了当前医疗保健系统所承受的负担。因而养老产业成为未来重点产业，如何搭建创新生态利用已人工智能为首的科技赋能养老产业成为重要议题。

本研究从创新生态的角度从 S(Science)-T(Technology)-B(Business) 三个层面入手，通过文献计量学，专利分析，网络挖掘的方式获取不同层面的数据，通过研究单一层面的发展状况与对比分析不同层面的转化与合作，从而获取国外不同国家与区域创新生态的情况。最后通过对比不同国家与地区养老产业创新生态情况，为中国养老产业的发展与创新生态的搭建提供政策建议。

参考资料：

[1] Office of the Leading Group for the Seventh National Population Census of the State Council. (2021). Major Figure on 2020 Population Census of China. ISBN 987-7-5053-9506-0

[2] United Nations, DESA, Population Division. (2022) Licensed under Creative Commons license CC BY 3.0 IGO. United Nations, DESA, Population Division. World Population Prospects 2022.

[3] Global health estimates 2019: Life expectancy and leading causes of death and disability, 2000–Geneva: World Health Organization; 2020

北京工业大学毕业设计（论文）

[4] Bingxin Ma, Jin Yang, Frances Kam Yuet Wong, Arkers Kwan Ching Wong, Tingting Ma, Jianan Meng, Yue Zhao, Yaogang Wang, Qi Lu (2023). Artificial intelligence in elderly healthcare: A scoping review, Ageing Research Reviews,83

[5] Xu G, Wu Y, Minshall T, et al. Exploring innovation ecosystems across science, technology, and business: A case of 3D printing in China[J]. Technological Forecasting and Social Change, 2018, 136: 208-221.

完成期限：2024 年 6 月 11 日

指导教师签章：_____

专业负责人签章：_____

2023 年 12 月 29 日

毕业设计（论文）诚信声明书

北京工业大学毕业设计（论文）

本人郑重声明：在毕业设计（论文）工作中严格遵守学校有关规定，恪守学术规范；所提交的论文是我个人在导师指导下独立研究、撰写的成果，毕业设计（论文）中所引用他人的文字、研究成果，均已在毕业设计（论文）中加以说明；在本人的毕业生设计（论文）中未剽窃、抄袭他人的学术观点、思想和成果，未篡改实验数据。

本毕业设计（论文）和资料若有不实之处，本人愿承担一切相关责任。

学生签名：_____ 日期： 2024年6月19日

关于论文使用授权的说明

本人完全了解北京工业大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

（保密的论文在解密后应遵守此规定）

签名：_____ 导师签名：_____ 日期： 2024年6月19日

摘要

全球老龄化进程正在加速，技术创新是积极应对人口老龄化的第一动力和战略支撑，养老科技产业发展的未来方向备受关注。相关研究多从技术管理的视角对于技术

与产业的发展进行分析与预测。从创新全过程管理的视角提炼现实规律，引入创新生态系统对养老科技产业进行系统分析，对形成产业创新政策、企业创新战略、科学研究方向，从而促进整个创新生态系统协同发展有着重要意义。本文将养老科技产业作为研究对象，以多层创新生态网络为框架，系统研究了养老科技产业的创新机会。

本研究通过对科学、技术与商业三层生态系统（S-T-B）的分析，探讨其协同发展对产业创新的推动作用，并划分出四类创新机会。进一步将模型扩展至投资层，提出 S-T-B-I 结构，以全面解析养老科技产业的创新生态，量化与统计其中的交互关系，发掘其中的创新机会。

首先，梳理了养老科技创新、养老科技产业发展及创新机会发现的现状。通过网络爬虫技术获取了文献、专利及公司与投资者数据，采用 KeyBERT 模型进行关键词抽取及 SciBERT 模型进行文本向量化，结合 K-means 聚类方法，分析各层生态系统的发展趋势。结果表明科学层生态方面，研究发现智能监测系统、康复技术、阿尔兹海默症研究等领域的发展迅速；技术层生态则侧重于传感器、物联网、远程医疗等方向；商业层生态则集中在人工智能与机器人、数字健康等方面。其次，通过对不同生态系统间的主题进行相关性量化与分析，构建了 S-T-B 创新生态网络。根据四象限理论，识别出四类创新机会：未产生协同发展效应、科学-技术促进发展、技术-产业促进发展和科学-技术-产业协同发展。再次，为更好地发掘产业创新机会，在 S-T-B 结构的基础上引入投资层，结合产业视角与企业生命周期的视角，构建 S-T-B-I 创新生态网络，进行层间交互的微观的分析。结果表明在投资角度与专利数量上，人工智能与机器人领域的公司都表现优异。另一方面，早期阶段的公司更容易受到投资者的青睐。但受到的投资是相对集中的，有少数几家明星公司收到了很多投资者的投资，随着生命周期的发展，投资逐渐趋于平衡。最后，通过逻辑回归肯定了拥有专利对公司获得投资的可能性有显著的正向影响。

本研究在理论上扩展了基于 S-T-B 的生态网络，加入投资者层，增加模型对于市场资本的考量，对于创新生态网络的模型构建进行了必要的补充。此外，本研究发现人工智能与机器人的技术与产品是未来养老科技产业的重点创新机会；初创企业是养老科技产业的创新主体，政策应加大对其支持力度，初创公司也应抓住技术创新机会，加大研发投入来获得市场的资本青睐，共同推动养老科技产业创新生态发展。

关键词：养老科技产业；创新生态系统；创新机会发现；S-T-B-I 结构；BERT

Abstract

As the global aging issue is alarming time, technological innovation is increasingly a primary driver and strategic support in addressing aging, therefore gerontechnology is an important solution. Existing research analyzes and forecasts the development of technology and industry from the perspective of technology management. However, it is essential to integrate the perspective of innovation process management and introduce the concept of an innovation ecosystem to systematically analyze the gerontechnology industry. This approach holds significant implications for industrial innovation policy, corporate innovation strategies, and scientific research directions, and ultimately promotes the synergistic development of the innovation ecosystem.

This thesis systematically studies the innovation opportunities in the gerontechnology industry using a multi-layer innovation ecosystem framework. By analyzing the scientific, technological, and business ecosystems (S-T-B), it explores the role of their synergistic development in promoting industry innovation. Furthermore, it extends the framework to include the investment layer, proposing the S-T-B-I structure to comprehensively analyze the innovation ecosystem of the gerontechnology industry.

Firstly, this thesis reviews the current state of research on gerontechnology innovation, industry, and innovation opportunity discovery, highlighting the deficiencies and limitations of existing studies. Using web scraping technology, literature, patents, and data on companies and investors were collected. KeyBERT was employed for keyword extraction, and SciBERT for text vectorization. Combined with the K-means clustering method, the development trends of each ecosystem were analyzed. In the scientific ecosystem, rapid advancements were observed in intelligent monitoring systems, rehabilitation technology, and Alzheimer's disease research. The technological ecosystem focused on sensors, the Internet of Things (IoT), and telemedicine, while the business ecosystem centered on artificial intelligence and robotics, as well as digital health.

By quantifying the commonalities and divergences of themes between different ecosystems, the S-T-B innovation ecosystem network was constructed. Based on the four-quadrant theory, four types of innovation opportunities were identified: those that are underdeveloped, science-technology spurred, technology-industry spurred, and science-technology-industry spurred.

To further explore industry innovation opportunities, this thesis integrates the investment layer based on the S-T-B structure, constructing the S-T-B-I innovation ecosystem network from the perspectives of industry and enterprise lifecycle, and conducting cross-layer analysis.

It was found that companies in the fields of artificial intelligence and robotics performed well in terms of both investment and patent metrics. Early-stage companies were more likely to attract investor attention. However, the investments were relatively concentrated, with a few star companies receiving most of the investors' funds. As the lifecycle progresses, investments tend to balance out. Finally, logistic regression confirmed that patent ownership has a significant positive impact on a company's likelihood of receiving funding.

In terms of theoretical contributions, this thesis expands the S-T-B ecosystem network by incorporating the investor layer, adding market capital considerations, and further supplementing the model for the innovation ecosystem. Practically, it identifies artificial intelligence and robotics as key areas for future innovation opportunities in the gerontechnology industry. Startups are the main innovators in this field, and policies should increase support for them. Startups should focus on R&D to attract market capital and jointly drive the development of the innovation ecosystem in gerontechnology.

Keywords: Gerontechnology industry; innovation ecosystem; innovation opportunity discovery; S-T-B-I structure; BERT

目录

摘要.....	IV
Abstract.....	VI
目录.....	I
1. 绪论.....	1
1.1 研究背景及意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	2
1.2 文献综述.....	2
1.2.1 养老科技创新的研究现状.....	2
1.2.2 养老科技产业的研究现状.....	3
1.2.3 创新机会发现研究现状.....	4
1.2.4 研究现状评述.....	6
1.3 主要研究工作.....	6
1.3.1 研究内容.....	6
1.3.2 研究框架.....	7
1.3.3 主要创新点.....	8
2. 相关概念界定与研究方法.....	9
2.1 相关概念界定.....	9
2.1.1 养老产业.....	9
2.1.2 养老科技产业.....	9
2.1.3 创新生态系统.....	10
2.2 网络数据获取.....	12
2.2.1 网络数据获取 Python 工具包介绍.....	12
2.2.2 静态网页内容获取.....	13
2.2.3 动态网页内容获取.....	13
2.3 文本向量化.....	13
2.3.1 基于统计的文本向量化方法.....	13
2.3.2 基于预测的文本向量化方法.....	16
2.4 本章小结.....	19
3. 基于科学、技术与商业创新生态系统的养老科技产业发展趋势研究.....	20
3.1 数据来源.....	20
3.1.1 文献数据来源.....	20
3.1.2 专利数据来源.....	21
3.1.3 投资者与公司数据来源.....	21
3.2 基于 KeyBERT 的关键词组抽取与聚类.....	22
3.2.1 KeyBERT 模型介绍.....	22
3.2.2 向量池化层策略以及聚类.....	23
3.3 养老科技产业创新生态主题发展趋势分析.....	24
3.3.1 基于科学的创新生态主题发展趋势.....	24
3.3.2 基于技术的创新生态主题发展趋势.....	25
3.3.3 基于商业的创新生态主题发展趋势.....	26

3.4 本章小结	27
4. 基于创新生态网络的养老科技产业协同发展研究	28
4.1 多层创新生态系统间主题相关性量化	28
4.1.1 基于 BERT 模型与 TF-IDF 的主题文本向量化	28
4.1.2 多层创新生态间主题相似度计算	29
4.2 基于 S-T-B 创新生态网络构建与分析框架	30
4.2.1 网络构建	30
4.2.2 S-T-B 网络分析框架	31
4.3 协同视角下养老科技产业创新机会发现	32
4.3.1 未产生协同发展效应类创新机会	32
4.3.2 科学-技术促进发展类创新机会	32
4.3.3 技术-产业促进发展类创新机会	33
4.3.4 科学-技术-产业协同发展类创新机会	33
4.4 本章小结	33
5. 基于创新生态网络的养老科技产业创新机会发现	34
5.1 产业视角下养老科技产业创新生态网络构建	34
5.2 产业视角下养老科技产业创新机会发现	36
5.2.1 投资层与商业层间创新机会分析	36
5.2.2 商业层与技术层间创新机会分析	37
5.2.3 技术层与科学层间创新机会分析	39
5.3 企业生命周期视角下养老科技产业创新生态网络构建	40
5.4 企业生命周期视角下养老科技产业创新机会发现	43
5.4.1 投资层与商业层间创新机会分析	43
5.4.2 商业层与技术层间创新机会分析	44
5.4.3 专利所有权与投资关系	46
5.5 本章小结	47
结论	48
参考文献	49
致谢	54

1. 绪论

首先解释了选题的背景与意义，之后梳理了对于养老科技创新，养老科技产业发展，创新机会发现，与创新生态系统的研究现状，通过对研究现状的分析提出现有研究存在的不足与局限，进而提出本文的研究内容与研究方法，提出本文创新点。

1.1 研究背景及意义

1.1.1 研究背景

老年人口被定义为 65 岁及以上的人，从 2015 年到 2050 年，世界上 60 岁及以上人口的比例将从 12% 扩张到 22%^[1]，反映出全球人口正在以惊人的速度老龄化。而我国是世界上人口最多的国家，也是世界上人口老龄化绝对数最大的国家。相比较而言，我国比世界上任何国家的老年人口问题压力更大。根据第七次全国人口普查结果，我国 60 岁及以上人口 26402 万人，占 18.70%（其中 65 岁及以上人口为 19064 万人，占 13.50%）。20 世纪 60 年代第二次出生高峰形成的较大人口将逐渐进入老年阶段，这将使中国的人口老龄化水平从近年来相对缓慢的演变转换为进入“快车道”的增长^[2]。另一方面，全球人均寿命从 2000 年的 66.8 岁增加到 2019 年的 73.3 岁，而健康寿命从 58.3 岁增加到 63.7 岁。寿命的增长速度略快于健康寿命，这代表了生活中有残疾、慢性疾病的老齡人口正在增加。大量老年的人患有慢性疾病，他们需要医疗保健的援助和支持^[3, 4]。但与此同时老龄化过程不仅意味着老年人口的增加，也将导致劳动力短缺。由于劳动力短缺，进一步加剧了当前医疗保健系统所承担的负担，老龄化导致老年人未满足的医疗保健需求将不可避免地增加，并逐步市场化。我们不能依赖当前的医疗保健系统，应扩大相关产业与研究在这方面的投入，美国国家技术与科技委员会^[5]在 2019 年就发布了应对老龄社会的新兴技术及产品领域(Emerging Technologies to Support an Aging Population)，细分了科技主导的产业格局，并持续运营着 Aging2.0 组织。英国则是拥有 Longevity. International 组织，由不同的企业，实验室，投资人与政府组织组成，旨在推动 R&D 与合作来面对老龄化问题，提升协同效应。与之相似的加拿大运营着 Agewell 组织，欧洲推行 AAL 计划，我国香港运营着 Innovation and Technology for Aging 计划。

在我国，2019 年 11 月，中共中央、国务院^[6]印发的《国家积极应对人口老龄化中长期规划》提出深入实施创新驱动发展战略，充分发挥科技创新引领带动作用，把技术创新作为积极应对人口老龄化的第一动力和战略支撑。2024 年 1 月国务院办公厅^[7]印发了《关于发展银发经济增进老年人福祉的意见》，这也是我国首个以“银发经济”命名的政策文件。可以说，全球在面对老龄化问题都提出了自己的产业布局，发展养老科技来增加老龄人口福祉。人口老龄化是我国社会发展的重要趋势，也是今后较长一段时期我国的基本国情，这既是挑战也带来机遇。人口老龄化从挑战方面看，将减少劳动力的供给数量、增加家庭养老负担和基本公共服务供给的压力。同时，也要看到人口老龄化能够促进“银发经济”发展，扩大老年产品和服务消费，而且有利于推动我

国科技创新和养老科技发展。因此明确养老科技产业的创新生态发展趋势，系统间协同发展现状，对发现创新机会，推动产业发展有着重大意义。

1.1.2 研究意义

第一，符合落实国家针对人口老龄化问题提出的各项政策建议，利用技术创新来解决老龄化问题符合习总书记提出的“着力发展养老服务业和老龄产业”和“要培育老龄产业新的增长点”之重任。这对养老产业的结构优化与发展，进而对经济社会的稳定发展有着重要意义。

第二，利用多层创新生态系统的方法对于养老产业中，商业生态，技术生态，科学生态的发展趋势进行研究，对于优化不同生态间协调发展，促进合作创新提供方向与思路。

第三，通过对多层创新生态系统中不同生态间的链接进行分析，为整体养老科技产业的生态提供整体的视角，并发掘其发展现状，这对政府科技管理部门与涉服务部门的规划与决策提供支撑，促进推动整体养老科技生态与产业的协调发展。

1.2 文献综述

1.2.1 养老科技创新的研究现状

养老科技—gerontechnology 由‘gerontology’与‘technology’两单词拼接而成，分别代表了老年医学的研究，与新技术与产品的研发与设计^[8]。这个词延伸出来的含义即关于面对老龄化问题的技术研究，应对老龄化社会问题的先进技术方案，是一项可以通过产业模式与商业模式去处理老龄化问题的重要举措^[8,10]。从老龄化对于科技发展方向的影响来看，老龄化对科技的发展影响路经由三条，一是“人口老龄化—技术创新—科技发展”，二是“人口老龄化—人力资本—科技发展”，三是“人口老龄化—养老支出—科技发展”，这三条影响路径共同影响通用科技向养老科技的方向发展^[11]。

对于养老科技科学发展的研究，首先是对于养老科技学科体系确立，通过 WoS 学科分类与国务院学位委员会学科分类之间的映射，构建养老科技的三层学科体系，运用社会网络分析方法识别核心学科，并对核心学科的研究态势进行分析，发现国内学科的影响力与国外发达国家存在差距^[12]。这是因为社会发展进程的差异，养老科技起源于国外，发展也相对成熟，对其借鉴与学习是加快我国技术发展的途径。因此在进一步进行新兴技术研究趋势评估中，对于国外养老科技创新研究的趋势与主题，利用聚类与战略坐标(strategic diagram) 方法，以聚类的关注度指标为横轴，新颖度指标为纵轴，建立战略坐标，发现四项新型主题有：面向老年人的智能家居科技创新研究；解决老年人孤独与社会隔离问题的科技创新研究；解决老年人环境压力的环境辅助技术创新研究^[13]。

对于养老科技技术发展的研究，刘春文^[14]构建了基于 ECT—Dim 的前沿技术识别模型，提出技术新兴性、确定性与探索性测度方式，建立前沿技术筛选机制，以养老环境辅助生活技术为例进行实证研究。在指标建立上，因为养老科技产业跨学科的特

性，吴菲菲^[15]基于新颖性与领域交叉性进行前沿技术的识别，识别出在通信、治疗和老年人行动工具方向上的专利具备知识的前沿性。还有研究将收集到的养老科技专利文档利用 TF-IDF 提取关键词，并采用 doc2bow 模型将每一个专利转换为二机制的稀疏向量，采取向量空间模型（VSM）的余弦相似度计算，得到基于专利的高维语意相似度矩阵，在进行降维处理后可以发现空白的区域为空白的研究领域，分别为：老年人大数据采集与处理；防跌倒智能控制技术与可穿戴设备；面向老年人的定制化、精准化医疗；老年智能家居。并采取了 LDA 主题抽取模型对领域专家文章进行聚类，将其与机遇相似度得到的结果相互验证^[16]。采取多元数据进行分析，可以系统的，多方面的研究养老科技整体的发展。基于文献、专利、政策文本等数据，构建了养老科技技术的领域词典，然后采用 word2vec 进行词向量模型的搭建，并用 K-means 算法对结构进行聚类。最后根据结果与专家知识相结合，搭建了四个主要类别，17 个技术主题的养老科技技术体系。并采用 logistics 曲线拟合的方法对于时间与专利数量进行技术生命周期的估算，得到移动通信类技术目前趋于成熟^[17]。

对于特定领域的前沿技术识别也是研究的重点，可以为领域内的研发提供方向与参考。在人工智能与养老科技的热点与前沿中，面向视力障碍老年人的智能服务、代际支持在老年人使用智能技术中的作用，以及老年人智能服务的安全伦理问题被认为是未来重要的研究方向^[18]。在远程医疗领域，老年人的远程看护、智能家居、跌倒检测、日常生活活动能力(ADL)、身体及精神状态的综合评估的研究被认为是该领域的前沿问题^[19]。

技术的发展离不开使用者的需求，对于技术需求发掘，有学者进行实证调研分析得出老年群体在健康、居住、信息沟通、社会参与 4 个方面对福祉科技的需求^[20]。将技术供需匹配是技术与产业均衡发展的重要因素，构建供需评价模型，对目前发展的技术进行需求的判断与匹配，从需求侧为技术的研发提供方向^[21]。

1.2.2 养老科技产业的研究现状

已经有许多学者采取不同的分析方式对养老科技产业的性质、竞争态势与合作模式，研究现状与未来发展进行研究。

从产业性质入手可以为分析方法的选择提供思路，黄鲁成^[22]总结了养老科技企业的性质，分别为科技密集性，成长新兴性，交叉融合性，与顾客多类型性。其中不难看出，养老科技企业不是单一的产业类型，拥有固定的产业链结构，而是多类复合型，融合多学科与领域的新兴科技创新赛道。因为我国人口结构变化周期与国外西方国家不同步的原因，国内的养老科技产业发展速度慢于国外，因此借鉴西方发展路径来吸取经验是非常重要的方式。在总结国外新兴养老产业发展特点中，第一点是养老科技产业的发展离不开政府的支持与引导，如同之前阐述过，各国都有自己的福利性或政府组织与项目。第二点是两种发展路径，其一是已有企业去拓展自己的业务领域；第二种是以满足老年人对于科技产品需求为目标而建的企业。第三点是良好的发展生态系统，构成特点其一是建立新兴养老科技企业发展群落，初创公司、产品提供商、服

务提供商、金融支持主体、咨询机构、研发机构等都是整体生态活动中必不可少的组成部分，可以加快生态的建构与繁荣。其二是不同种群中提供特定产品与服务的企业丰富，可以相互促进。这一点可以看出，对于养老科技产业，从创新生态的视角去审视与发现是十分必要的。最后一点是科技竞争力强，拥有核心专利的企业数量多，以护理机器人技术为例，均为科技含量高的专利，同时也启发将专利分析的视角代入分析是十分必要的。

利用 ESTP-Chain 四维分析法从环境链，主体链，技术链和地位链四个维度构建指标，分析国际间养老科技产业的技术竞争格局^[23]。米兰^[24]则是从社会网络分析的角度与时间发展的角度对国际养老新兴技术合作模式演化及影响因素研究，从宏观的网络指数的角度分析了全局情况，发现时间尺度下，合作网络的规模在不断扩大，连通性在不断增强。中观个体合作深度与合作广度的角度搭建了合作模式模型，发现合作网络核心由初期的个人逐渐过渡到产学研主体。从微观演化影响因素的角度采取 ERGM 模型对影响网络图发展的内生型与外生性因素进行分析，发现基于国家协同的合作模式是当前养老科技合作网络的重要模式，而技术的创新能力始终无显著影响。

对于养老科技产业研究的发展，陆杰华^[25]梳理了养老产业的三个阶段，在起步阶段特点是是将国外研究成果的介绍与结合中国国情的理论研究有机地结合。第二阶段是发展阶段，更关注理论与现实发展的问题，提出国家产业政策的制定和完善事带动其发展的最基本和最主要的要素。第三个阶段是研究发展阶段，这一阶段中商业对于养老产业发展的兴趣以及投资者出现，并且养老产业跨学科的特征也被挖掘。徐雨森^[26]对于则是指出目前的研究视角从狭义的“技术管理”走向更为系统的“创新管理”，即不再局限于研发管理和专利分析，更多从创新全过程管理的视角提炼实践规律。提出要构建产业创新系统，其中包括考察养老科技产业生态系统的技术体制，归纳产业中创新系统的运作模式。根据创新生态系统理论去研究创新实体在系统中共同的演化转变。以及考察与识别产业生态系统中的“关键种”，来促进整体产业创新能力与政策引导的进步。

1.2.3 创新机会发现研究现状

创新机会发现（Innovation opportunity discovery）是考虑到价值创造过程中中多维度的机会发现，其囊括了技术机会发现（Technology opportunity discovery）与技术机会分析（Technology opportunity Analysis）等概念，创新机会发现关注的问题更倾向于面向商业生态的价值创造，通过生态系统中整体研发活动的推进，去推动创新，是创新管理的范畴，其视角更佳宏观更系统化。技术机会发现与技术机会分析更倾向于技术的创新，也就是技术管理的范畴，“Technological Forecasting & Social Change”与“Scientometrics”是此类型研究的重点期刊。

技术机会发现关注两个主要问题，一是识别与探索新兴的技术与商业领域，二是通过利用现有技术和产品，开拓新的领域。朱晨^[27]使用技术市场矩阵（technology-market concordance matrix）对利用多头注意力模型（Dual-attention model）获得的市场技

术数据与公司专利数据进行预期的计算，然后通过这之间的差值去发现产业与企业进一步发展的方向，回答了公司是应该利用现有技术开拓市场，还是基于现在的市场去完善技术积累的问题。

在技术机会分析在技术领域的创新机会识别与预测中，专利文本的挖掘，与技术发展的分析与预测十分重要。在专利文本的挖掘方面，徐硕基于深度学习框架 BiLSTM-CRF 和 BiGRU-HAN 模型采用标注数据集训练，有效地从专利文本中提取详细的语义信息，比传统基于 SAO 的语义信息抽取表现更精准，有助于后续的任务的精准性^[28]。在专利的分类任务上，针对新兴领域一直存在数据集小的问题，有学者通过对抗式生成网络（Generative adversarial network），在进行数据扩容，以达到在很小的数据集上训练出更好的专利分类任务的表现，其在 3D 打印产业的价值链分类任务中进行了尝试^[29]。在技术发展的分析与预测方面，岳俊举利用 Apriori 算法进行技术创新地图与关联规则的挖掘，在煤层气开采技术领域，提出了多个煤层气开采技术创新方案^[30]。还有学者关注创新想法的生成也是技术机会分析的研究重点，有研究通过利用 GTM（Generative Topographic Mapping）和链路预测，提供了集成的方法，将自动化技术与专家见解相结合，为不同领域的技术开发和创新提供了有价值的指导^[31]。

关注技术与研究之间的联系，也是识别创新的重要方式，文本挖掘在这类研究中十分重要，Shen 利用文本挖掘、ORCLUS 聚类和余弦相似度量来识别智能健康监测技术中科学和技术领域的潜在机会^[32]。更宏观的通过文本挖掘近 30 年间的专利与文献文本，发掘技术与研究的推动关系，发现以基础研究为主导去推动技术发展是发展趋势^[33]。

基于创新生态系统进行创新机会的发现也是重点的研究领域，创新生态系统主要分为知识和商业生态系统，针对图的拓扑分析在这一类研究中十分重要，许冠南利用专利与增值税数据搭建多层创新生态系统映射（MIEM）来分析知识和商业生态系统，采用 Newman 拓扑聚类识别网络群落，并采用策略角色矩阵分析群落中的角色对对数控机床领域的^[34]。利用拓扑分析去检测虚拟集群，来进行实证分析，研究在氢能源领域的社会网络中，公司所处网络的中心度是否会影响其创新能力，结果显示除了中心度，在虚拟集群中的结构洞位置也可以增加企业创新能力^[35]。后续研究借助专利数据通过机构之间科技合作网络与商业数据通过机构之间商业合作网络进行分析，对数控机床领域的知识与商业网络进行分析，研究发现知识网络与商业网络嵌入连通度和支配度均对技术创新绩效有正向影响，且商业网络与知识网络之间存在联动作用，商业网络嵌入带来的优势能够进一步强化知识网络嵌入对技术创新绩效的正向影响^[36]。

利用创新生态系统搭建网络来进行创新机会发现进一步发展后，将原本的知识生态系统进一步划分为以专利为载体的技术生态系统，与以文献为载体的科学系统，这样代表科学-技术-商业的 S-T-B 结构主要是利用价值链去分析层级之间的关系与链接，可以发现在产业尺度上创新的推动方向，以提供更多产业尺度上创新机会的洞悉。许冠南对于我国 3D 打印行业进行了建模与分析^[37]。Choi 利用此方法对智能电网领域进行

了建模与分析^[38]。本研究在原本的科学发展到技术，最终到商业进行价值创造的发展过程中，加入了投资者层，考虑到资本对商业化进程进行加速与提供必要的资源的作用。主要分为两个研究过程，其一是分析单一生态中创新主题发展的趋势，其二是统计分析不同创新生态系统中的层间交互。以此来进行进行创新机会发现。对于本文创新机会发现的定义，其不同与狭义的技术机会发现，其仅关注潜在的技术发展机会，本文关注的创新机会发现还关注于创新过程中的资源组合带来的经济价值潜力与商业化机会，具体的发现的内容是：（1）各系统的技术发展方向；（2）各系统间潜在的技术与商业创新机会；（3）各系统中重点的创新主体与技术方向。

1.2.4 研究现状评述

前人对于对于养老科技产业与创新机会发现的研究非常具有启发性，总的来说，对于养老科技产业的研究逐渐从单一国外模式的分析到想要用创新生态的视角去统筹规划发展。对于创新机会发现的研究也开始融入创新生态系统的思路，从单一的技术管理向创新管理发展。

对于养老科技产业的研究仍缺少可以将科学研究，技术开发，与产品市场的视角统一的分析的实践，本文将解决这个问题与研究空白，从S-T-B三层网络的视角，汲取创新生态系统的思路以更加全面的了解养老科技产业。

对于创新机会发现的研究，缺少考虑对于初创企业，风险投资（Venture Capital）是非常重要的资源来源，并高度的影响着企业的发展，是创新中不可或缺的一环。所以投资者也是创新生态中重要的一项主体^[39,40]。因此将商业生态网络进行进一步细分，探究其中的投资关系与资源流动是有价值的，目前这部分研究还有空缺，因此本研究将S-T-B结构进行进一步的完善，提出S-T-B-I结构来进一步解构与分析创新生态网络。

1.3 主要研究工作

主要研究工作主要包括：研究内容、研究框架、和主要创新点三部分内容。

1.3.1 研究内容

本文的研究对象是养老科技产业的创新生态系统，包括了对于其科学、技术与商业生态的发展趋势研究，并通过计算各生态网络之间的主题相似性去构建S-T-B创新生态网络，最后根据创新主体的匹配搭建产业视角下的S-T-B-I创新生态网络与企业生命周期视角下的T-B-I创新生态网络，并分析层次之间的关系。主要内容包括以下几方面：

第一章，绪论。首先对本文的研究背景和研究意义做了详细的描述，并根据本文的主题详细地论述了国内外的研究现状，包括养老科技创新，养老科技产业，创新机会发现，最后对现有研究的不足和局限性进行总结，介绍本文的研究工作，包括研究内容，研究框架和创新点。

第二章，相关概念界定与研究方法。首先对本文的研究概念进行阐述，明晰了养老产业，养老科技产业，创新生态系统的概念。之后说明了本文获取网络数据的方法，

包括使用的工具包和静态与动态网页内容获取的差异。第三部分介绍了文本向量化的方法，分文基于统计的向量化方法，与基于预测的文本向量化方法。

第三章，基于科学、技术与商业创新生态系统的养老科技产业发展趋势研究。

主要阐述对于养老科技产业创新生态系统中科学、技术与商业的发展趋势研究，首先阐述了数据的来源与获取方式，根据数据种类分为了，文献数据、专利数据、投资者与公司数据。接下来进行文本的向量化过程，首先进行了关键词抽取，之后进行向量化。最后根据聚类结果分析与描述了养老科技产业科学、技术与商业生态的发展趋势。

第四章，基于 S-T-B 创新生态网络的养老科技产业创新机会发现。通过挖掘各生态层之间的相似性，发现其产业视角下的链接，构建 S-T-B 创新生态网络，根据四象限的分析架构对产业视角下养老科技产业的创新机会进行阐述。

第五章，基于 S-T-B-I 创新生态网络的养老科技产业创新机会发现。从微观实体层面研究不同创新生态网络之间的连接，构建了产业视角下的 S-T-B-I 创新生态网络，与企业生命周期视角下的 T-B-I 创新生态网络，并研究其中的链接关系。

1.3.2 研究框架

根据研究内容确立以下研究框架图（图 1.1-1.2），首先进行养老科技产业发展趋势研究；其次进行基于 S-T-B 创新生态网络的养老科技产业创新机会发现；最后，基于 S-T-B-I 创新生态网络进行养老科技产业的创新机会发现。

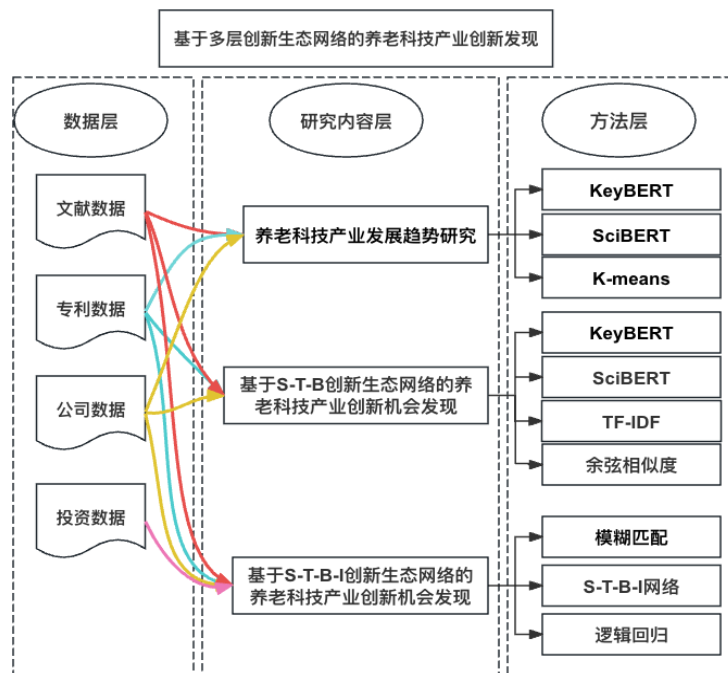


图 1.1 本文研究框架图

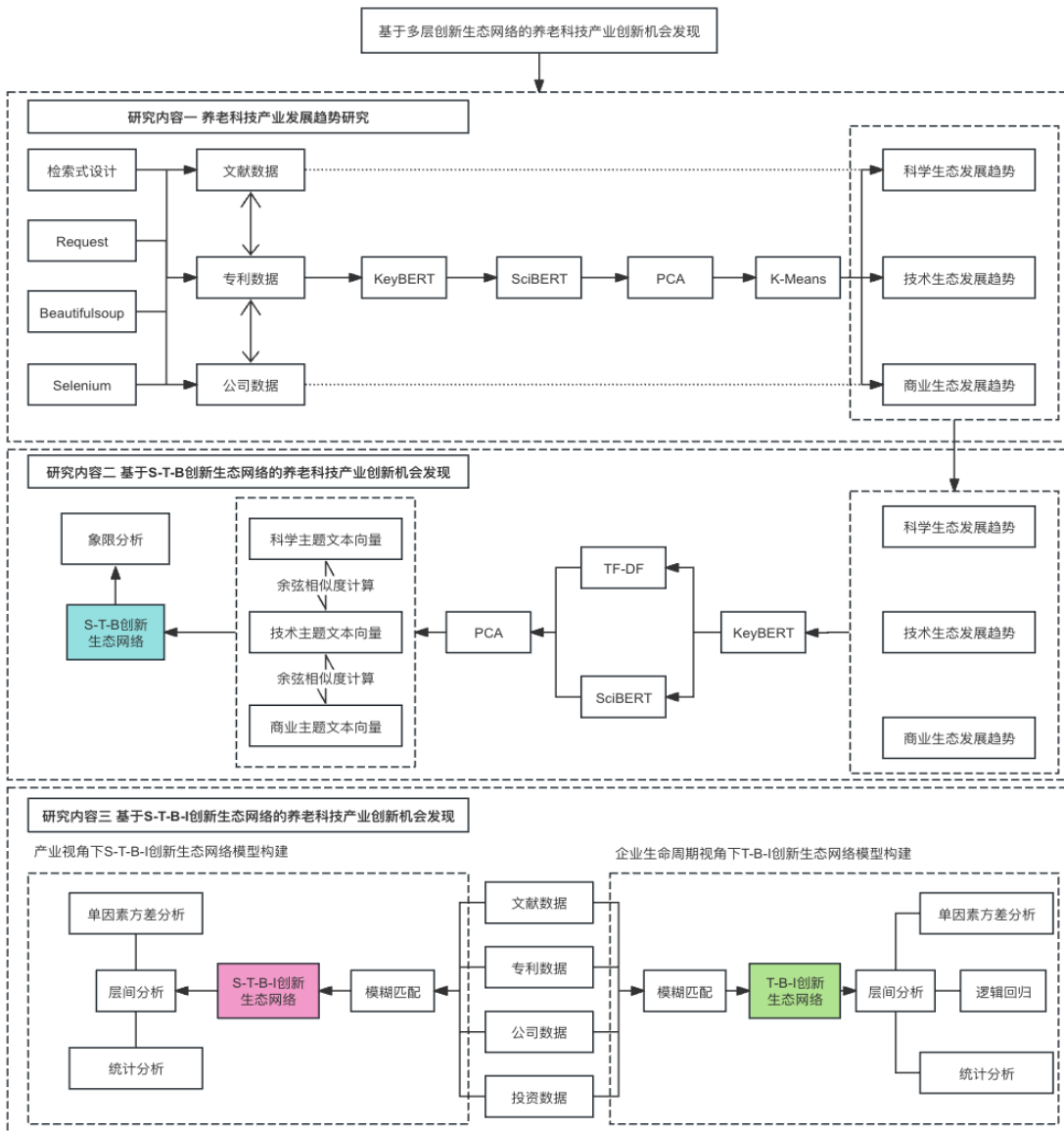


图 1.2 本文技术路线图

1.3.3 主要创新点

针对现有研究存在的不足和局限性，本研究的主要创新点如下：

- （1） 利用网页爬虫进行网页数据的获取，基于多元异构数据研究不同生态的养老科技产业发展趋势，并根据其文本相似度挖掘了产业间的发展相关性，对我国养老科技产业的产业间合作与协同发展，最终促进创新提供一定的指导作用。
- （2） 扩展 S-T-B 生态网络，加入投资者层，增加模型对于市场资本的考量，构建 S-T-B-I 创新生态网，对于创新生态网络的模型构建进行了进一步的补充。
- （3） 根据创新主体的匹配构建了产业视角与企业生命周期视角下的养老产业的创新生态网络，并分析其中的链接关系，对于养老科技产业整体的发展情况与不同生态层之间的交互关系进行了统计与挖掘。

2. 相关概念界定与研究方法

本章作为全文的支撑，主要分为三个部分。第一部分阐述相关概念，明晰了养老产业，养老科技产业，创新生态系统的概念。第二部分说明了本文获取网络数据的方法，包括使用的工具包和静态与动态网页内容获取的差异。第三部分介绍了文本向量化的方法，分文基于统计的向量化方法，与基于预测的文本向量化方法。

2.1 相关概念界定

2.1.1 养老产业

Graafmans^[41]于 1991 年首次提出了“Gerontechnology Industry”术语，学术界对其称谓不一，陆杰华^[25]教授将其称之为老龄产业，定义较为广义，泛指为老年人口服务的产业体系，包括为老年人提供产品和服务的经济实体。郑稣鹏^[42]将其称之为适老产业，将其定义为“可以实现老有所养，老有所医，老有所为，老有所学，老有所乐，老有所安等为目的，为社会公众提供各种养老以及相关产品的生产活动集合，其中既包括专门或主要为养老或老年人提供产品的活动，以及适合老年人的养老用品和相关产品制造活动。”而我国台湾与香港地区学者^[43]则采用福祉产业。黄鲁成^[22]也曾沿用过一个称谓，但考虑到传播性与政策制定相关因素，将其称之为养老产业，但无论称谓如何改变，养老产业一直作为养老科技产业的上位概念，是广义的以老年人口提供产品及服务的产业，本文更加聚焦于养老科技产业的概念。

2.1.2 养老科技产业

养老科技（Gerontechnology），黄鲁成^[12, 22]对养老科技在国外的发展进行总结，在国外学者的研究中，Gerontechnology 是科学技术（Technology）与老年（医）学（Gerontology）结合的产物，是全面满足老年人需求的科学技术，是可以通过产业和商业模式发挥作用的科学技术，是应对人口老龄化的重要举措。考虑到老龄人口的社会化需求与交叉学科的发展，认为“养老科技是以应对人口老龄化挑战、满足老年人物质与精神需求为目的，在多学科交叉基础上形成的科学技术。养老科技创新的产出形式包括产品、服务和年龄友好的空间环境。”

老年科技产业区别于上述的老龄产业，养老产业等，是由黄鲁成^[12]提出，将科学技术与养老产业进行结合，将其定义为“以养老科技为基础，以满足老年（人）社会需求为目标，从事产品、服务和年龄友好环境的研发（设计）和生产经营活动的企业总和。”后续的研究中提出了其产业特征：（1）养老科技产业是满足老龄社会需求的新兴产业；（2）养老科技产业是涵盖第一、第二、第三产业的综合性产业。（3）养老科技产业具有营利性的特点^[21]。

2.1.3 创新生态系统

绝大多数突破性的创新不是闭门造车, 在创新战略的设计中, 引入系统的思想是必要的^[44]。引入系统的思想去研究与指导创新过程在研究进程中发生了两次重要的转变, 一是由系统向生态系统的转变。二是由单一的商业生态系统到全面的创新生态系统的转变。

系统如今被定义为“提供产品/服务的系统, 这是一个历史上自发形成或由管理层设计的多层次社会网络, 由具有不同属性、决策原则和信仰的行动者组成”^[45]。而生态系统原本是一个生物学概念, 指例如生态雨林一样的, 沿着由生命子系统组成的路径回收营养流, 这些子系统被组织成以过程为导向的角色, 连接生命和非生命子系统, 完成能量梯度推动稀有营养物质的回收^[46]。“创新系统”很早被 Niosi 提出并研究, 其基于区域经济的思想, 关注到了要对科学, 技术与创新相关的机构进行重点关注, 因为这可以为经济提供进步性^[47]。但他此时并没有采用生态系统(ecosystem)这个词汇, 而是采用系统(system), 因为其并没有想要进一步引发生态学的思考。而如今生态系统(ecosystem)已经被大量应用并替换系统(system)。并引发了相关知识原理的产生与融合, 生态学的重要思想“可持续”为例, 如今是很多对于平台生态系统研究的重要主题。共同演化是平台长期保持活力和表现力的关键。然而, 大多数观点认为平台的可持续性会不断恶化, 并最终导致赢家通吃的局面。Inoue 使用生物学中的 Lotka-Volterra 方程作为参考, 从平台生态间的关系, 平台的环境, 互补者的特征去构建模型去研究日本电子游戏市场的共生局面^[48]。不仅生态学的原理迁移是重要的研究主题, 以生态系统的思想去研究经济, 创新行为, 更是学者们关注的问题。但早期的研究主要围绕在商业生态系统, 学者基于这种思想提出通过广泛的建立链接去提升企业竞争力的观点^[49]。基于商业生态系统思想所引发的创新生态系统的概念, 让有关策略, 创新相关的研究开始涌现^[50]。也有学者将这一转变进行梳理, 并指出这一转变的根本思想是由价值获取(value capture)到价值创造(value creation)的转变^[51]。其中具有代表性的就是引发了建立平台生态的思想, 核心企业通过平台生态的搭建, 产业上下游的互补企业与供应商可以更好的利用平台进行价值创造, 提升自身与整个平台的竞争力。这种思想已经不仅限于商业的思想, 而是在向技术的管理与创新发展。学者对于 Cisco 所拥有专利的分析, 发现其成功运用并购去建立了自己的商业生态^[52]。这种思想的发展让人们意识到, 公司通过技术去获取的竞争力不仅仅依靠于自身的创新能力, 在产业链上下游的参与者所面临的创新挑战也是公司创新表现很重要的一部分。而上下游所面临的挑战对于核心企业所带来的影响也是不同的, 上游供应商遇到挑战本质上是技术发展的机遇, 可以提升核心企业的技术竞争力, 而下游服务提供商遇到的挑战与客户体验直接挂钩, 这会侵蚀核心企业的竞争优势。因此引发了垂直整合的思想, 利用技术的生命周期去进行风险的控制^[53]。创新生态系统的研究在不断的发散与细化, 如今主要的研究类型分为: (1) 开放创新生态系统; (2) 地区与国家的创新生态系统; (3) 数字创新生态系统; (4) 基于城市与创新区的创新生态系统; (5) 创新孵化者与加速

者所搭建的服务；（6）机遇大学的创新生态系统^[54]。研究主流与重点领域是：（1）产业生态；（2）商业生态系统；（3）平台生态管理；（4）多主体网络^[55]。

目前也存在对系统到生态系统的转变的反思与怀疑，主要来源于对于创新生态系统概念使用的不一致性，这会造成理论的混杂，并且学者指出即使创新生态系统引发了大量的科学贡献，但这与生态的表述（eco-prefix）并无联系，启发性的含义并不算是一个严谨定义。Oh 认为生态学启发的思想值得赞赏，但它是人工设计出来的，与自然生态系统不同。由于其目的论的意图以及承认政府的重要性，这一概念中存在经济指导主义，表明有要建立更广泛的支持结构。至于目的论，目前普遍探讨的生态系统是一个设计的系统工程，而不是演化的系统。它将一些有用的概念引入经济发展对话中，仍然只是建议，缺乏实证支持或严格的对应规则研究。与此同时，他指出与早期的科技研究、科技城、区域创新系统、科学城或创新集群的概念不同。创新生态系统文献显示六个特点：（1）更明确的系统性，更加重视众多创新行为者之间的联系。

（2）数字化，频繁利用信息技术，专注于信息和通信技术（ICT）的新产品和服务。

（3）开放创新，借用、许可、开源、众包和联盟是重要话题。（4）这一术语展示了传播价值而非研究价值。（5）更加强调组织和行业占据的不同角色或“生态位”。（6）相对于政府直接管控，更强调市场力量的重要性。提出创新生态系统未来的发展需要完善的理论与指标，将拥有广泛含义的创新生态系统细化，并给出具体的定义^[54]。

综上所述，由系统到生态系统的转变，让基于生态学的思想引发了学者对于知识原理迁移的同时，更重要的启发性的思考，以及更恰当的描述了系统的创新，是由复杂主体所共同影响的系统，以及关注价值创造过程中一切环节的参与者。而商业生态系统到创新生态系统的转变，则是由单一商业视角到全面的，关注技术创新与价值创造过程的视角。这两种转变让创新生态系统相关的研究，思想，与模型不断的系统化与全面化的发展。

“创新生态系统是一组不断发展的参与者、活动和产品，以及制度和关系，包括互补和替代关系，它们对参与者或参与者群体的创新绩效发挥着重要的作用”^[46]。定义揭示了创新生态系统也存在着圈层的差异，不同参与者、活动、产出、制度和关系区分出了不同的生态系统圈层。在创新生态系统中有两大可以独立分开的系统，分别是由基础研究支撑的知识生态系统，与市场支撑的商业生态系统^[56]，这两个独立的子系统中，知识生态系统主要负责价值创造，而商业系统负责价值获取，用来评判他们成功的因素也大不相同^[35]。去研究与了解商业生态系统与知识生态系统的独特的差异，可以为我们更好的解构与理解创新生态系统提供帮助。另外，知识生态系统中，又可以分为知识创造所构成的学术系统，与为了商业生态所构造的技术系统，其实际的载体则分别是学术期刊与专利。这两者分别代表了“科学（Science）”与“科技（Technology）”两个词汇。因为新兴科技通常起源在科学研究阶段，所以在知识生态系统中，科学生态系统与科技生态系统是两个独立又互相链接的部分，采取价值链与网络层次间的链接去分析其中的差异与空白，可以发现潜在的技术发展方向与商业化机会^[37, 38]

许冠南^[37]用科学，科技，商业数据共同搭建的 S-T-B 结构（Science-Technology-Business），解决了传统文献分析的局限性。后者只反映了学术圈的信息和状态。S-T-B 结构也解决了技术机会发现（Technology Opportunity Discovery）仅关注于关注知识到技术之间的转化^[57]。仅关注科学生态系统到技术生态系统的链接，而商业生态系统到技术生态系统中的商业化过程并没有得到研究。Choi 利用 S-T-B 结构去搭建创新生态系统来感知在层级之间的创新机会^[38]。市场侧的信息可以反映在公司的网页信息上，技术侧的数据可以反映在专利数据中^[58]。因此，目前的基于 S-T-B 结构的创新生态网络搭建，一般采用网页数据以反映商业生态的信息，来搭建依附于文献数据的科学生态系统，依附于专利数据的科技生态系统，依附于网页信息的商业生态系统。

2.2 网络数据获取

网络数据获取的技术手段是通过网络爬虫（Web Crawler）是一种用于自动化地从互联网上提取数据的技术。爬虫通过模拟浏览器行为，发送请求并解析响应，从而获取网页内容并进行数据提取。根据网页的类型，网络爬虫可以分为静态网页获取和动态网页获取，其中的本质区别就是静态网页获取的是初始加载时已经包含在 HTML 中的内容，而动态网页获取的是通过 JavaScript 在浏览器端生成或更新的内容。

2.2.1 网络数据获取 Python 工具包介绍

本文使用 Python 语言进行网页爬虫程序的编写，原因是基于 Python 语言有许多功能强大的包，本文中主要使用了：（1）Request；（2）BeautifulSoup；（3）selenium，本小节将分别介绍这三个包在工作流程中的作用。

（1）Requests

Requests 是一个 HTTP 库，广泛应用于发送 HTTP 请求和处理响应。它的主要作用包括：第一，发送 HTTP 请求：支持 GET、POST、PUT、DELETE 等各种 HTTP 请求方法，能够灵活地与 Web 服务器进行通信。第二，处理响应：自动处理 HTTP 响应的状态码、头部信息和内容，简化了处理网络请求的流程。第三，支持持久会话（Session），可以在多个请求之间保持参数，如 Cookies 等，从而模拟更复杂的用户行为。在网页爬虫工作流程中，Requests 通常用于直接获取静态网页的 HTML 内容，为后续的数据解析打下基础。

（2）BeautifulSoup

BeautifulSoup 是一个用于解析 HTML 和 XML 文档的库。它的主要作用包括：第一，解析 HTML 文档：将 HTML 文档解析为 DOM 树结构，使得文档的遍历和搜索变得非常容易。第二，提取数据：提供 API，用于查找和提取网页中的各种元素（如标签、属性、文本等）。第三，处理不完整标签：能够处理一些不规范的 HTML 文档，确保数据解析的稳定性和可靠性。在网页爬虫工作流程中，BeautifulSoup 通常用于解析从 Requests 或 Selenium 获取到的 HTML 内容，提取所需的关键信息。

（3） Selenium

Selenium 是一个功能强大的浏览器自动化工具，广泛应用于自动化测试和网页爬虫。它的主要作用包括：第一，浏览器自动化：支持多种浏览器（如 Chrome、Firefox、Safari 等），能够模拟用户在浏览器中的操作（如点击、输入、滚动等）。第二，处理动态内容：能够处理由 JavaScript 动态生成的网页内容，获取完整的网页数据。第三，复杂交互：支持与网页进行复杂的交互操作，如处理弹出框、选择下拉菜单、执行 JavaScript 脚本等。在网页爬虫工作流程中，Selenium 通常用于获取动态生成的网页内容，特别是那些无法通过静态 HTTP 请求直接获取的数据。

2.2.2 静态网页内容获取

静态网页的内容在加载时不会改变，所有数据都包含在最初的 HTML 中。静态网页获取的流程如下：

发送 HTTP 请求：使用 requests 库向目标网页发送 HTTP 请求，并接收响应。

获取网页内容：从 HTTP 响应中提取 HTML 内容。

解析网页内容：使用 BeautifulSoup 库解析 HTML 文档，构建 DOM 树。

提取所需数据：通过遍历和搜索 DOM 树，提取所需的关键信息。

这种方法适用于内容固定、不需要与网页进行复杂交互的情况。

2.2.3 动态网页内容获取

动态网页的内容是通过 JavaScript 在浏览器端生成的，无法通过简单的 HTTP 请求直接获取。动态网页获取的流程如下：

启动浏览器模拟器：使用 selenium 库启动浏览器（如 Chrome 或 Firefox）并模拟用户行为。

加载目标网页：让浏览器加载目标网页，等待所有动态内容加载完成。

获取网页内容：从浏览器中提取完整的 HTML 内容，包括动态生成的部分。

解析网页内容：使用 BeautifulSoup 库解析从浏览器获取的 HTML 文档，构建 DOM 树。

提取所需数据：通过遍历和搜索 DOM 树，提取所需的关键信息。

这种方法适用于内容动态生成、需要与网页进行复杂交互的情况。

2.3 文本向量化

2.3.1 基于统计的文本向量化方法

基于统计的方法本质上通过词频计算将文本或单词转为高维的文本向量矩阵，让语料库的每一个单词对应着嵌入矩阵的一个维度，再采取降维方式压缩信息，以更低维度的矩阵去反应文本的语义信息。进行可以分为以下几种：（1）词频-逆文档频率

TF-IDF (Term Frequency-Inverse Document Frequency) ; (2) 独热编码 (One-Hot Encoding) ; (3) 词袋模型 (Bag-of-words) ; (4) 隐含语义分析 LSA (Latent semantic analysis) ; (5) 潜在狄利克雷分布 LDA (Latent Dirichlet Allocation);

2.3.1.1 独热编码与词袋模型

独热向量 (One-Hot Encoding) 是一种词语的向量表达方式, 每个词语都被表示为一个很长的向量。这个向量的长度等于词汇表的大小, 向量中只有一个元素是 1, 其余都是 0。1 的位置对应词语在词汇表中的索引。独热编码简单明了, 但其主要缺点是向量维度高且稀疏, 无法表达词语之间的任何语义关系。

词袋模型中每个文档被表示为一个向量。这个向量的每个元素代表一个词语在文档中出现的次数, 而不考虑词语的顺序和语法结构。词袋模型的优点是实现简单, 容易理解, 但它忽略了词语的上下文信息。

引发的方法有词-文档矩阵与基于窗口的共现矩阵, 本质都是基于共现的统计方法。词-文档矩阵用于描述词语在多个文档中的分布情况。在这个矩阵中, 每一行代表一个词语, 每一列代表一个文档, 矩阵中的每个元素表示对应词语在对应文档中出现的频率。这种方法可以用于信息检索和文档相似性计算, 但同样存在高维稀疏的问题。

基于窗口的共现矩阵关注词语在特定窗口大小内的共现关系。与词-文档矩阵不同, 共现矩阵的每一行和列都代表词汇表中的词语, 矩阵的每个元素表示两个词语在所有文本的特定窗口内共同出现的次数。这种方法能够捕捉到词语之间的关系, 特别是那些经常在一起出现的词语, 从而有助于理解词语的语义关联。

2.3.1.2 TF-IDF

为了解决独热编码仅考虑增加向量维度, 而没有理由向量中的权重的缺点, 词频-逆文档频率 TF-IDF 用以评估一个字词对于一个文件集或一个语料库中的其中一份文件的重要程度。TF-IDF 值随着单词在文档中出现的次数成正比增加, 但同时会被单词在语料库中的出现频率所抵消, 这样可以帮助调整一些词汇本身出现频率较高的影响。

本文中所采用的来自于 sklearn 包中的 TfidfVectorizer 向量化方法, 从所有文档中的所有词汇构建一个词汇表。在这个过程中, 可以设定参数有 max_df, min_df, max_features 等, 以过滤掉文档频率太高或太低的词汇, 或限制词汇表的最大特征数, 之后调用 CountVectorizer 来计算每个词在每个文档中的出现次数, 这就形成了一个基本的词频 (TF) 矩阵。之后计算每一个词的逆文档频率 (IDF) 乘以 TF 矩阵中的每个元素乘以相应词的 IDF 值, 从而得到 TF-IDF 矩阵。这个矩阵表示了每个词在每个文档中的重要性。在转换后生成的 TF-IDF 矩阵中, 其结构实际上是行表示文档 (或文本), 列表示单词。这意味着每一行代表一个文档的 TF-IDF 向量, 而每一列对应于词汇表中的一个单词。

2.3.1.3 SVD 与 PCA

对于独热编码与 TF-IDF 来说最大的问题即向量过于稀疏，大部分向量存储的数值为 0，而我们想要的是更加稠密的动态的向量表达，主成分分析 PCA（Principal Components Analysis）^[59]与奇异值分解 SVD（Singular Value Decomposition）是主流的降维方法。

PCA（Principal Components Analysis）基于线性代数中的特征值和特征向量的概念。PCA 通过对数据的协方差矩阵进行特征分解来找到最能够代表数据集方差的方向，主要分为以下几个步骤：

（1）因为 PCA 对数据的尺度十分敏感，需要进行标准化，让每个特征的平均值为 0，标准差为 1，标准化公式：

$$X_{standard} = \frac{X - \mu}{\sigma} \quad (2.1)$$

其中， X 是原始数据矩阵， μ 是每个特征的平均值， σ 是每个特征的标准差。

（2）计算协方差矩阵：

$$C = \frac{1}{n-1} X_{standard}^T X_{standard} \quad (2.2)$$

其中， n 是样本数量， $X_{standard}$ 是标准化后的数据矩阵。

（3）将对称矩阵协方差矩阵进行特征分解，得到其特征值与对应的向量，特征分解公式：

$$Cv = \lambda v \quad (2.3)$$

其中 C 是协方差矩阵， λ 是特征值， v 是对应特征向量

基于线性代数中的特征值和特征向量的概念。PCA 通过对数据的协方差矩阵进行特征分解来找到最能够代表数据集方差的方向。

（4）选择主成分

选择特征值最大的几个特征向量作为主成分。这些主成分能够最大程度地保留原始数据的信息。具体选择多少个主成分通常根据特征值的累积贡献率来决定。

（5）数据投影

将原始数据投影到选定的主成分上，得到新的低维数据表示。这通过下面的矩阵乘法实现：

$$X_{new} = X_{standard} V_{selected} \quad (2.4)$$

其中, X_{new} 是投影后的数据, $V_{selected}$ 是选择的特征向量矩阵。

SVD (Singular Value Decomposition) 则是一种矩阵的因数分解, 将给定的特征矩阵分解为三个矩阵的乘积, 公式描述:

$$A = U \Sigma V^T \quad (2.5)$$

其中, A 是原始矩阵, U 是左奇异向量矩阵, Σ 是奇异值对焦矩阵, V^T 是右奇异向量矩阵的转置。

通过去除那些在对角矩阵中为 0 的列与较小的奇异值来达到特征提取与数据降维的目的。

2.3.1.4 LSA 与 LDA

基于词-文档的方法只能应用在较小的数据集中, 随着数据集的不断增长, 矩阵将维度过于大并且稀疏, 导致计算的效率低下, LSA (Latent Semantic Analysis) [60] 基于奇异值分解 (SVD) 对词-文档矩阵进行降维 [62]。

$$X \approx U_k \Sigma_k V_k^T \quad (2.6)$$

其中 X 是词-文档矩阵, U_k 和 V_k 分别是前 k 个奇异向量, Σ_k 是前 k 个奇异值组成的对角矩阵。

但 LSA 也无法应用在大规模的语料训练任务重, 因为其由矩阵分解带来的较高的运算复杂度。

LDA [61] 是一种主题模型, 用于从文档集合中发现文档内的主题。虽然 LDA 主要用于主题建模, 而不是传统意义上的词嵌入任务, 但它可以用于生成文档的主题分布表示, 从某种意义上说, 这可以被视为一种“嵌入”, 将文档映射到主题空间中。其本质是贝叶斯概率模型, 它假设文档是通过一个随机过程生成的, 其中每个文档都是从多个主题中抽取的, 每个主题又关联到特定的词汇上。所以每个文本可以依靠其包含的词汇, 在不同主题的维度上进行嵌入。

2.3.1.5 传统方法的局限

本质上, 所有以上提及的传统方法都是旨在以各种方式在潜空间中解决我们语料库的词语文本之间生成关系的问题, 即词之间未明确说明但我们可以根据数据建模方式来挖掘的关系。但在随着语料库的不断加大都会面临维度灾难与计算复杂度过高的问题。

2.3.2 基于预测的文本向量化方法

所有基于预测方法的嵌入都是基于神经网络的结构, 只不过结构在不断的发展以捕捉更多的语义, 随着结构与时间的发展可以分为以下:

2.3.2.1 word2vec

本质是单层的神经网络，基于一个假设：一个词的含义由高频率出现在它旁边的单词所决定。其一共有两种结构^[63]：

Skip-grams（SG）：利用中心词去预测外侧的词。

Continuous Bag of Words（CBOW）：利用周围的词去去预测中心词。

本文采取 Skip-grams 结构去说明其原理：

在对数似然的表示下，对于中心词的外侧词的预测目标函数为(θ 为全部需要进行优化的参数)：

$$J(\theta) = -\frac{1}{T} \log L(\theta) = -\frac{1}{T} \sum_{t=1}^T \sum_{\substack{-m \leq j \leq m \\ j \neq 0}} \log P(w_{t+j} | w_t; \theta) \quad (2.7)$$

其中对于中心词出现的概率 P 可以利用 SoftMax 函数进行计算

$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)} \quad (2.8)$$

优化方法是随机梯度下降（Stochastic Gradient Descent），最小化目标函数，最大化对于中心词的预测概率。

但在实际模型计算时因为计算复杂度的原因，并不会采用这种方式，最多采用的是基于负采样（Negative Sampling）的目标函数：

$$J_{neg-sample}(u_0; v_c; U) = -\log \sigma(u_0^T v_c) - \sum_{k \in \{K \text{ sampled indices}\}} \log(-u_k^T v_c) \quad (2.9)$$

其中 σ 是 logistic/sigmoid 函数：

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (2.10)$$

目标函数的第一部分表达了最大化抽样的词是中心词的外侧词的概率；第二部分表达了最小化抽样的词不是中心词的外侧词的概率。在这一阶段的发展中还有 glove 模型^[64]，其综合了矩阵分解和基于预测的方法的优点。

上述方法所训练出来的向量是静态的，在训练过后所有词的向量化表达就固定了，并且并没有这会导致对于一词多义问题没有解决。后续模型想要获取基于语境的词嵌入。然后可以根据输入上下文调整单词的表示向量。

2.3.2.2 递归神经网络的发展

上述方法所训练出来的向量是静态的，在训练过后所有词的向量化表达就固定了，无法实时更新，推理之发生在给定的窗口中，并且这些会导致对于一词多义问题没有得到解决。这是因为其只是一个单层的前馈神经网络，递归神经网络 RNN（Recurrent Neural Networks）^[65]通过在每一个神经元接受输入信息的同时传递编码后的前面状态的所有信息，解决了这个问题，但因为其在反向传播（Backpropagation）进行梯度下降过程中会出现梯度消失的问题，使优化在收敛之前就进入局部最优解。长短期网络 LSTM（Long-short term memory network）^[66]通过控制递归过程中数据的信息传递解决了这个问题。EMLO（Embedding from Language Model）^[67]基于双向长短期网络（bi-directional LSTM）组成，其根据给定的一句话去预测单词，在隐藏层的最后一层就会得到考虑了整个语境的单词语义表征。但也存在计算复杂与无法并行运算的问题。

2.3.2.3 Transformer 的出现

后续模型优化主要是 Transformer 架构的注意力机制（attention mechanism）^[68]的出现。GPT（Generative Pre-training）模型在同样也是通过文本下一个词的预测的优化进行训练，不过其采用了 Transformer 架构，其中的自注意力层（self-attention）可以让其生成文本时更加有效地理解和使用上下文信息，最主要是解决了先前模型因为梯度爆炸与消失所导致的对于长距离信息的丢失，且可以进行并行训练，使得采用巨量文本进行预训练，来获得明显优于 LSTM 的任务性能成为现实。

2.3.2.4 BERT

BERT（Bidirectional Encoder Representations from Transformers）^[69]是 Google 于 2018 年发布的一种预训练语言模型，其本质仍然是 12 层 Transformer 网络，BERT 的核心创新在于它的双向训练方式，这使得模型在处理每个单词时不仅考虑之前的词还考虑之后的词，从而获得了更为丰富的语境信息。其训练任务分别为（1）Masked Language Model (MLM)：随机遮盖（masking）输入句子中的词语，并训练模型预测这些被遮盖的词，这种方法可以使模型在理解句子语义的同时学习到词语之间的关系。

（2）Next Sentence Prediction (NSP)：给定两个句子 A 和 B，模型需要预测 B 是否是 A 的下一个句子。这有助于模型学习理解两个句子之间的关系。这种方法可以使模型更好地理解上下文之间的关系。

BERT 是自然语言处理的一个里程碑，其贡献不仅是深层的语义嵌入，而是其出现让采用大模型的预训练与微调的工作流取代了小模型，对于绝大部分自然语言处理任务，只需在原模型上对下游任务进行微调就可以达到优异的表现。

在嵌入层面，BERT 模型的输入包含三个嵌入（embedding）：token 嵌入（token embedding）、位置嵌入（position embedding）和段落嵌入（segment embedding）。这三个嵌入是分别计算得到的，然后在输入层直接相加起来。这样做的原因是因为 BERT 模型需要同时考虑输入的 token、位置和段落信息。这三个嵌入分别对应了输入 token

在句子中的位置和所处的段落，以及输入 token 本身的信息。因此，在将它们相加之后，BERT 模型可以同时获得这些信息。在获取嵌入向量时，会在隐藏层的最后一层获取其考虑了最多语境信息的向量。对于嵌入方法的选择有：（1）从新训练模型，数据量过小不足以让模型学习到词间关系；（2）采取预训练模型，可以让模型学习到词间关系，但对于后续任务没有进行优化；（3）初始化预训练模型，并进行微调，可以让模型学习到词间关系，并对后续任务进行优化。

因为 BERT 所需要巨量的计算资源与数据，所以有针对于不同任务的采取不同文本与参数训练的模型可以选择，本文选择的 SciBERT 使用 1.14M 全文文献, 3.1B 字符运算，对于科学文本的文本向量化任务会有更好的表现。

2.4 本章小结

本章节的目标是阐述相关的概念与研究方法，首先对于养老产业、养老科技产业和创新生态系统的概念和发展进行梳理，接下来并阐述了涉及的网络爬虫技术方法；最后对本文涉及的文本向量化处理方法进行了说明，为后续的研究工作奠定了基础。

3. 基于科学、技术与商业创新生态系统的养老科技产业发展趋势研究

根据上文的研究基础，本章主要阐述对于养老科技产业创新生态系统中科学、技术与商业的发展趋势研究，首先阐述了数据的来源与获取方式，根据数据种类分为了，文献数据、专利数据、投资者与公司数据。接下来进行文本的向量化过程，首先进行了关键词抽取，之后进行向量化。最后根据聚类结果分析与描述了养老科技产业科学、技术与商业生态的发展趋势。

3.1 数据来源

3.1.1 文献数据来源

本文文献数据来源于 Web of Science 核心数据库，利用过往研究中的技术词典^[17,21]与已获取公司数据的产业关键词进行检索式的设计，以技术层，应用层，与目标层的设计原则，对相关的单词与概念进行扩充，最终设计的检索式如下：

TS = (("AI" OR "artificial intelligence" OR "machine learning" OR "deep learning" OR "neural networks" OR "robotic*" OR "machine learning algorithms" OR "natural language processing" OR "iot" OR "wireless sensor*" OR "wireless network*" OR "computer vision" OR "robotic bionics" OR "exoskeleton robots" OR "wearable" OR "detect*" OR "biology" OR "geriatric gerontology")

AND("mobile device*" OR "mobile phone*" OR "smartphone*" OR "wireless communication*" OR "telecare*" OR "telehealth*" OR "telemedicine*" OR "telerehabilitation*" OR "remote monitoring" OR "telemonitoring*" OR "diagnosis*" OR "treatment*" OR "caregiver*" OR "digital health platform*" OR "remote healthcare" OR "rehabilitation" OR "mobility aids" OR "smart home" OR "health monitoring systems" OR "chronic disease management" OR "medical" OR "molecular therapy" OR "oncogenomics" OR "mental health interventions" OR "e-prescriptions" OR "data science" OR "Alzheimer*" OR "Parkinson*" OR "dementia" OR "diabetes" OR "software" OR "customized" OR "personalized" OR "digital therapeutics" OR "virtual reality" OR "augmented reality" OR "neuroscience" OR "cognitive" OR "bioscience and genetic engineering" OR "assistive technology" OR "intelligent monitoring devices" OR "smart caregiving devices" OR "fintech" OR "community" OR "entertainment")

AND ("elderly" OR "older adults" OR "senior citizens" OR "geriatric population" OR "aging adults" OR "senile individuals" OR "senior adults" OR "elder care" OR "old age care"

OR "aging society" OR "Silver Age" OR "Silver Market" OR "Active aging" OR "age-friendly technology"))

检索时间是近五年数据，最终获取数据 5970 条。

3.1.2 专利数据来源

专利数据来源主要分为两部分，第一部分来源是德文特创新数据库，用于获取全面的养老科技产业专利数据。采取和文献检索相同的检索式，检索时间为近五年数据，最终获取数据 1215 条。

第二部分数据来源为 USPTO 数据库与 Google Patents 数据库，用于获取公司所拥有的专利数据，采取的方式是，首先将公司名称进行拼接处理，用逻辑或进行连接，检索的字段是 assignee，在 USPTO 检索获得 PN（Patent Number）。将每一个专利号与 Google patent 的固定 URL 进行拼接，对应网页存储的就是对应的专利信息，采取网页爬虫获取其对应标签数据，共获取数据 1859 条。

最后将两部分数据进行合并，去重，去掉外观型专利，共获取数据 2515 条。

3.1.3 投资者与公司数据来源

本文投资者与公司数据来自于网页信息，采取网页爬虫技术获取，数据源分为两部分。

（1）Longevity.International 数据获取

Longevity.International 是首个开放获取的非营利性分散式养老科技产业知识与协作平台，旨在促进养老科技产业的参与者和利益相关者(包括公司、投资者、非营利组织、学术实验室和研发中心、政府机构和政策制定者)之间更大程度的协同作用、高效合作和讨论。网站上开放了大量的投资者与公司信息，本文重点获取了养老科技（AgeTech）相关的投资者与公司信息。

其网页结构并不复杂，没有将信息存放于需要 Javascript 运行的地方，所以只采用了 Request 进行信息获取，与 BeautifulSoup 进行基于制定标签的 HTML 数据的解析。对于公司数据最终获取 686 条，对于投资者信息最终共获取 721 条，与公司数据中的投资者信息进行合并去重后，投资者数据共 1417 条。

（2）Aging 2.0 数据获取

Aging 2.0 是一家专注于推动老龄化领域创新的组织，成立于 2012 年，由 Katy Fike 和 Stephen Johnston 创立。该组织致力于解决与老龄化相关的重大挑战和机遇，已经发展成一个全球网络，拥有超过 40,000 名创新者，遍布 31 个国家的 130 个分支。Aging 2.0 支持包括创新者、护理人员、医疗保健公司等在内的多种利益相关者，推动能够简化老年人生活过程并改善他们生活质量的创新。此外，该组织在全球举办众多活动，促进老龄产业专业人士之间的对话与合作。其拥有一款应用 Konexons，是一个平台，

旨在帮助创新者、企业家和投资者通过提供必要资源将突破性技术带入市场，有创新机会与意愿的公司会将信息放在这个平台之上，本文获取的数据就来自这部分数据。

其网页设计较为复杂，首先，脚本通过 Selenium 启动 Chrome 浏览器，访问登录页面并自动填充用户名和密码来登录。登录后，脚本遍历一个包含不同社区的表格，对每个社区的特定 URL 进行迭代访问。对于每个页面，使用 Request 获取页面元数据，利用 BeautifulSoup 解析 HTML 内容，从中提取和处理制定的标签数据，并将这些数据写入 CSV 文件。脚本还处理分页以访问多个页面的数据，并在所有相关数据被提取后关闭浏览器。最终获取公司数据 521 条。

3.2 基于 KeyBERT 的关键词组抽取与聚类

本节主要介绍对文本的关键词抽取操作以及向量化的策略。

3.2.1 KeyBERT 模型介绍

本质是利用分词将文本分割为词组或单词，采用指定的 Bert 模型对全文与词语进行向量化，利用余弦相似度分别计算词语与全文之间的相似度，根据参数 top_n 指定相似度最高的词语并返回，根据余弦相似度的理论，词语与全文的相似度越高，词语越是可以代表全文信息，越是关键词。具体结构可见图 3.1：

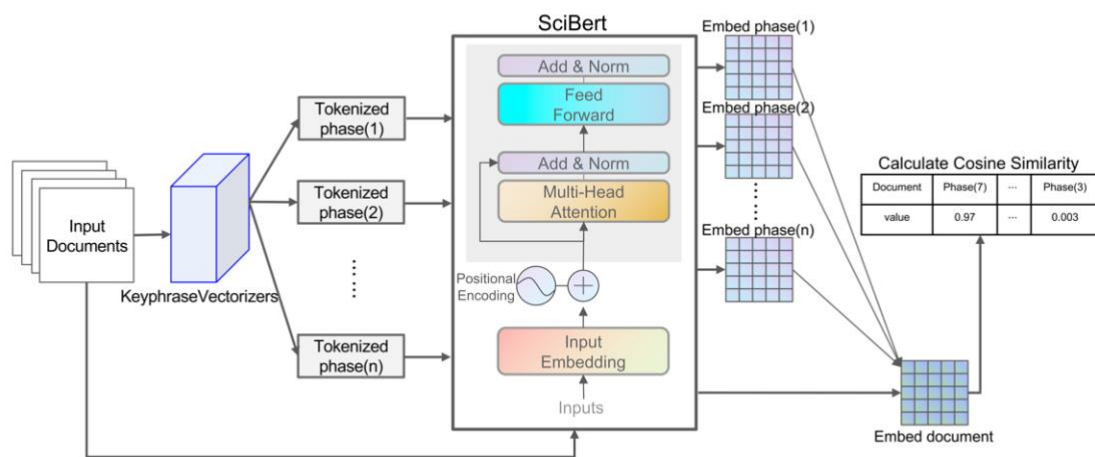


图 3.1 KeyBERT 模型架构

本文采取的分词器是 KeyphraseVectorizers，其分词过程不需要指定词组长度，而是首先将文本进行词性（part of speech）打标，然后根据完整的语义结构进行抽取。采用英文停词，提取三至五个个关键词。

3.2.2 向量池化层策略以及聚类

在文本预处理工作中，使用了多种技术来清洗数据。因此定义了一个预处理函数，该函数将文本转化为小写，并分词、利用 Wordnet 进行词性标注，然后进行词形还原，并且移除了停用词、标点符号、数字和单字词，并使用正则表达式移除关于公司名称与老年人表达相关的词汇。

之后将文本使用 KeyBert 进行关键词抽取。之后采用 SciBert 进行词嵌入，其遵循 BERT 的基本架构，嵌入维度与原始的 BERT 模型（BERT-base）相同，它的词嵌入维度是 768 维。在将词组进行分词后，每一个词元单独进行嵌入，词组的每一个词元会被转化为 768 维，然后在采取平均的操作，即每一个词元对于词组的向量化特征的贡献是一致的，每一个词组会转化为 768 维度的向量。因为每一行数据由 3-5 个词组构成，所以在表征文本时，这为不同类型的文本（公司数据、文献数据和专利文本）可能包含不同的信息密度和结构特性，在池化层中是采取拼接策略还是平均策略是依据后续的 K-means 聚类结果表现来决定的，经过实验后对于公司与文献数据采取的是拼接操作，对于专利文本采取的是平均操作。

这种策略选择的理论分析是，在拼接操作中，各个词向量按顺序连接在一起，形成一个长向量。这种方法能够保留关键词的独立性，充分利用关键词的全部信息，强调平均贡献，对于结构较为丰富的数据（公司与文献数据）非常有效，从而在聚类中获得更精细的区分。而平均操作将所有词向量的元素取平均，生成一个单一的固定长度向量。这种方法对于信息较为均匀分布的文本效果较好。专利文本往往具有标准化的语言和格式，关键信息分布相对均匀，因此采用平均操作可以有效地捕捉整体的语义特征，而不会过分强调某些特定的词汇或短语。

在进行聚类前采用 PCA（Principle Components Analysis）进行降维操作，这可以有效提升聚类任务的表现。PCA 降维的维度设定依靠累积的方差解释率，本文设置参数为 0.95，意为降维后的特征向量可以体现原本高维向量 95% 的信息。之后将数据输入 K-means 模型进行聚类，聚类数量的选择依据多轮迭代得到的轮廓系数，轮廓系数越高，越代表聚类的效果越好，因此选择相对轮廓系数更高的聚类数量，轮廓系数图如下。最终公司数据被分为了 5 个聚类，专利数据被分为了 6 个聚类，文献数据被分为了 10 个聚类，但在后续人工将文献数据的相似主题聚类进一步合并，最终文献数据被分为了 7 个聚类。

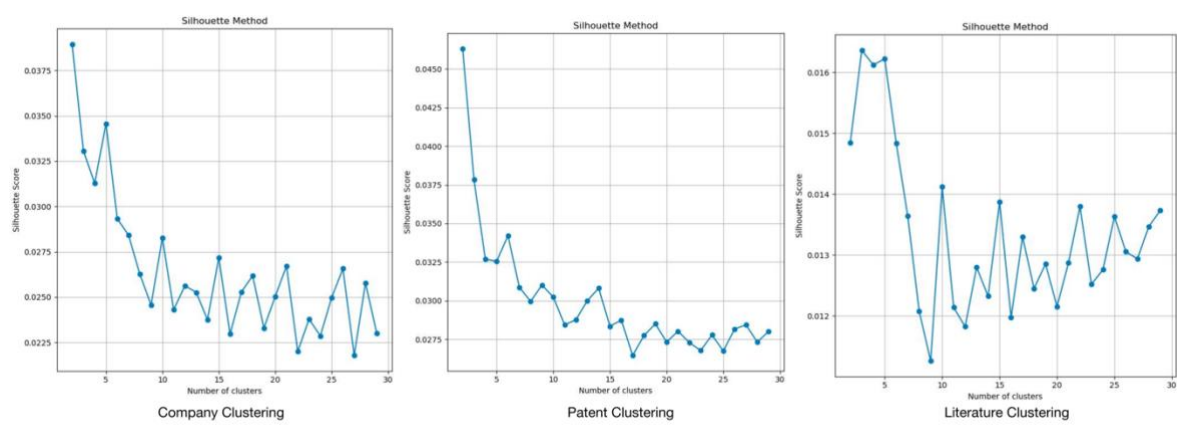


图 3.2 公司、专利与文献的聚类轮廓系数折线图

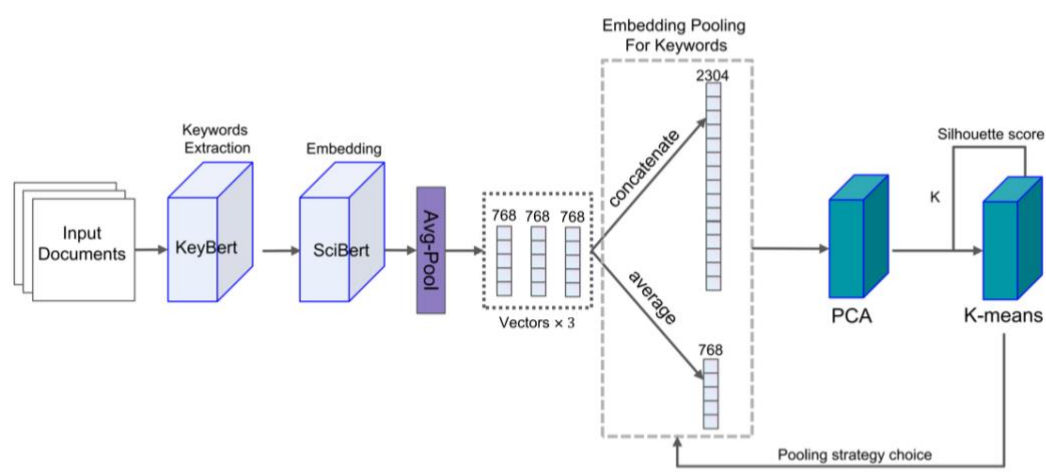


图 3.3 文本处理流程

3.3 养老科技创新生态主题发展趋势分析

3.3.1 基于科学的创新生态主题发展趋势

文献数据是科学生态发展的重要反映，在上节的数据分析过后文献数据被分为了 7 个聚类，分为了以下研究方向：

- (a) 智能监测系统
- (b) 康复
- (c) 阿尔兹海默症：包括阿尔兹海默症与心理健康；阿尔兹海默症与认知修复；阿尔兹海默症与脑神经科学三个子主题
- (d) 重大疾病：包括疾病诊断与术后康复；癌症两个子主题
- (e) 流行病学研究
- (f) 人工智能在疾病诊断与预测方向的研究

(g) 慢性病
具体的代表性研究为：

表 3.1 科学生态各主题代表性研究

智能监测系统	康复	阿尔兹海默症	重大疾病	流行病学	人工智能	慢性病
活动识别	穿戴设备	认知修复	术后并发症	呼吸道合胞病毒（RSV）	老年痴呆/阿尔兹海默症	轻度认知障碍（MCI）
穿戴式传感器	运动康复科技/理疗	老年护理	骨折和骨质疏松症	新冠病毒（COVID-19）	机器学习算法	阿尔兹海默症
智能医疗保健系统	VR	心理健康	心血管疾病	疫苗	卒中预测与康复	心力衰竭
跌倒监测系统	肌肉力量与行动能力评估	轻度认知障碍（MCI）	围手术期护理	抗生素抗药性	轻度认知障碍（MCI）	慢性肾脏疾病
远程健康监控系统	步态分析	神经心理学测试	老年护理	流行病管理	跌倒监测	心血管疾病
人工智能	机器人外骨骼辅助康复系统	蒙特利尔认知评估	疾病诊断	流行病学	卷积神经网络在影像医学中的应用	糖尿病
智能家居	卒中康复	脑生物标记物（早期检测）	癌症与癌症扩散	抗菌与抗病毒治疗	预测模型	营养状况
物联网	认知修复	精神认知障碍	急性髓系白血病（AML）	病毒与细胞病原体	神经退行性疾病	骨密度
		帕金森症	机器人手术	传染病检测		风险预测模型
		脑炎症	多学科肿瘤护理			
		神经退行性疾病	慢性淋巴细胞白血病（CLL）			
		氧化应激与衰老	预测性生物标记物			
		β -淀粉样蛋白（早期检测）	遗传与分子诊断			
		分子生物学	化疗与放疗			

3.3.2 基于技术的创新生态主题发展趋势

专利数据是技术生态发展的重要反映，在上节的数据分析过后专利数据被分为了 6 个聚类，分为了以下研发方向：传感器，物联网，听觉辅助，医疗科技，远程医疗，人工智能。具体的代表性技术为：

北京工业大学毕业设计（论文）

表 3.2 技术生态各主题代表性技术

传感器	物联网	听觉辅助	医疗科技	远程医疗	人工智能
处理器	穿戴设备	语音处理与增强	疾病诊断与监控	持续血糖监测系统（CGM）	神经网络与深度学习
皮肤传感器	移动机器人	信号处理算法（多频段信号）	治疗干预与药物输送	远程医疗系统	阿尔兹海默症诊断
环境传感器	远程监控	近 场 磁 感 应（NFMI） 通信	Tau 磷酸肽免疫疗法（治疗精神退行性疾病）	穿戴系统	糖尿病预测
NPU	智慧家居系统	无线技术	B- 咔 琳 化 合 物（治疗精神退行性疾病）	远程患者监护	安全与应急管理（跌倒监测）
数据处理与分析	云服务器	电池技术	脂质体疫苗组合物（治疗精神退行性疾病）	AI 诊疗	穿戴设备与监测系统
葡萄糖传感器	无线通信模块	自适应声学反馈消除（降噪，滤波器）	用于植入式器械的生物相容性涂层	个人化医疗应用程序	语音识别与合成
机器人与穿戴设备	智慧系统	环境适应（声学）	用于核酸分析的微流控系统	VR 疗愈	自动语音识别
分析物电极/电级	远程呈现（智真系统）	耳鸣掩蔽器装置	生物可吸收支架	移动端的健康应用程序	机器人与自动驾驶
连续分析物传感器		双耳波束成形技术	B 淀粉样蛋白抑制剂 / β 淀粉样蛋白 PET 成像化合物（检测）	智能医疗设备	社交机器人
通信收发器		自适应定向麦克风	植入式止痛装置		AR/VR
			心电图电极定位		大数据
			经皮分析物传感器（非侵入性监测设备）		人机交互
			用于慢性疼痛的植入式神经刺激器		智慧交通系统
					数据安全与隐私

3.3.3 基于商业的创新生态主题发展趋势

公司网页数据是商业生态发展的重要反映，在上节的数据分析过后公司网页数据被分为了5个聚类，分为了以下发展方向：人工智能与机器人，远程医疗，医疗与疾病诊断，数字健康，生活辅助。具体的代表性技术与服务为：

北京工业大学毕业设计（论文）

表 3.3 商业生态各主题代表性技术与服务

人工智能与机器人	远程医疗	医疗与疾病诊断	数字健康	生活辅助
自然语言处理	提升远程医疗可达性与效率	分子治疗、肿瘤基因组学	智能家居	数字健康平台
数据科学	数字健康平台	医疗与康复器械	可穿戴健康检测设备	智能监测设备
智能机器人软件	护理者服务(提供支撑, 人力资源, 与教育)	电子化处方与病例	个人化医疗	智能看护设备(赋能养老护理工作)
VR/AR	康复技术	老年学/老年病学研究	P4 药物	社交, 社区, 老年教育服务
金融科技	智能家居及健康监控设备	针对于阿尔兹海默症与帕金森的前端医学	线上健康平台	移动性以及安全保护类产品
机器人仿生科技	慢性病管理	定制化诊疗	健康生活指南服务	个性推荐, 健康生活指南类服务
	心理健康干预	脑神经科学与认知修复	B2B 医疗服务	护理者服务(提供支撑, 人力资源, 与教育)
仿生科技与外骨骼机器人	智能看护设备(赋能养老护理工作)	生物科学与基因工程	移动健康应用	健康监测设备
物联网		p4 药物	助残服务	跌倒监测
机器人			远程患者管理	
计算机视觉			制造与能源	

3.4 本章小结

本章主要描述了数据获取到发现各创新生态发展趋势的过程, 首先通过网络爬虫获取了文献、专利、公司与投资者数据。接下来对文本进行关键词的抽取, 并介绍了向量化过程中的策略与步骤, 之后进行聚类, 最后解释了各生态系统的发展趋势。科学生态分为: 智能监测系统、康复、阿尔兹海默症、重大疾病、流行病学研究、人工智能在疾病诊断与预测方向的研究、慢性病。科技生态分为: 传感器、物联网、听觉辅助、医疗科技、远程医疗、人工智能。商业生态分为: 人工智能与机器人、远程医疗、医疗与疾病诊断、数字健康、生活辅助。并详细描述了其代表性研究。

4. 基于创新生态网络的养老科技产业协同发展研究

根据第三章聚类出的各生态的发展方向，本章节则是基于已经分类好的发展趋势去挖掘各生态层之间的关系，发现其产业视角下的链接，不同发展主题之间的关系可以是发展的差异与相似，为后续第五章针对于生态之间创新实体链接研究提供支持。

4.1 多层创新生态系统间主题相关性量化

4.1.1 基于 BERT 模型与 TF-IDF 的主题文本向量化

基于第三章聚类的结果，本节将每个主题的全部文本进行嵌入。由于向量化的数据增加，使用上一节相同的方法，则会导致向量所表示文本的特征没有凸显，所以进行了流程的改进以增加向量对与文本中特征信息的表征能力，以下将阐述其中的策略：

前期流程依旧是采用 KeyBert 模型进行关键词的提取，每个文本都要提取三个关键词。对于每一个关键词进行分词分为单一词元，之后采取 SciBert 模型进行词嵌入，这里依旧是提取隐藏层最后一层包含最复杂语义关系的向量，对于每一个关键词的每一个词元都将转化为一个 768 维度的向量。

之后引入注意力机制，通过计算输入序列中每个词元的重要性得分来对其进行加权，从而在最终的嵌入表示中突出关键的词元特征。首先初始化一个可训练的注意力权重向量，并将其与每个词元的嵌入向量进行点乘，计算每个词元的重要性得分。然后，使用 softmax 函数将这些得分归一化，得到每个词元的权重。接下来，对每个词元的嵌入向量进行加权，生成加权后的嵌入向量。最后，对所有加权后的嵌入向量求和，得到一个综合的嵌入表示。这样的好处是在多个词元向量合并的过程中可以强调重要的特征，让最终的关键词可以代表词元的特征。之后采取拼接（concatenate）的方式将三个关键词的向量首尾连接，得到基于 BERT 模型的文本关键词向量化结果。

与此同时采用 TF-IDF 对提取出的关键词进行向量化，TfidfVectorizer 的参数设定是采用英文听词，最大文档频率阈值为 0.5，过滤掉出现在超过 50% 文档中的词。最小文档频率阈值等于 2，过滤掉出现在少于 2 个文档中的词。使用次线性词频缩放，将词频取对数。使用 1 到 3 元组，考虑单个词、两个词的组合和三个词的组合。将采用 BERT 模型的向量与采用 TF-IDF 模型的向量首尾连接，得到最终的代表一行数据的向量。并对于每一个聚类中的所有行进行平均，最终得到可以代表聚类的向量。具体过程表示为下图：

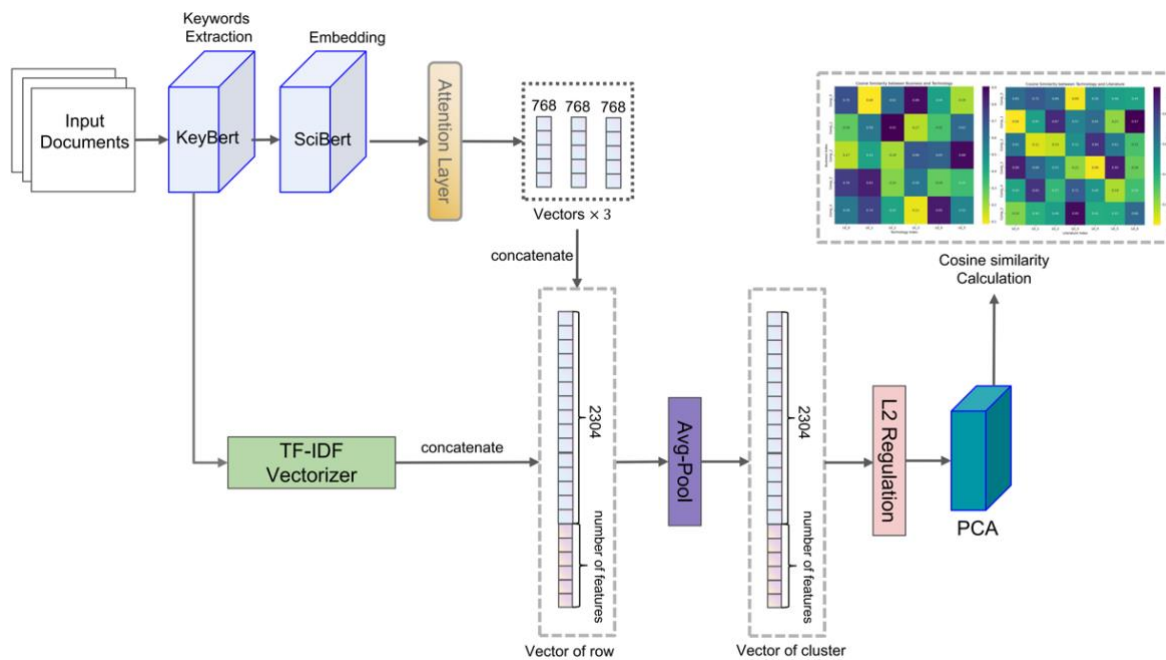


图 4.1 文本处理流程

4.1.2 多层创新生态间主题相似度计算

获得可以代表主题的向量后，需要进一步进行归一化，这里采用的是 L2 范数（欧几里得范数）等于一。将一个向量归一化（normalization）是指将该向量转换为一个单位向量（长度为 1 的向量）。具体操作是将原向量的每个分量除以该向量的 L2 范数。这对后续的余弦相似度计算非常重要，使得余弦相似度只关注向量之间的角度，而不是向量的长度。

之后进行 PCA 降维，参数设置与上一章节中的相同，之后进行余弦相似度的计算，得到以下结果：

表 4.1 商业生态与技术生态各主题相似度

	1 传感器	2 物联网	3 听觉辅助	4 医疗科技	5 远程医疗	6 人工智能
1 人工智能	0.70*	0.08	0.62	0.88***	0.44	0.19
2 远程医疗	0.29	0.58	0.91***	0.17	0.32	0.63*
3 医疗	0.17	0.43	0.19	0.68*	0.65*	0.88***
4 数字健康	0.76**	0.83**	0.24	0.58	0.28	0.33
5 生活辅助	0.58	0.74*	0.42	0.11	0.85*	0.53

表 4.2 技术生态与科学生态各主题相似度

	a 智能监测系统	b 康复	c 阿尔兹海默症	d 重大疾病	e 流行病学	f 人工智能	g 慢性病
1 传感器	0.65*	0.72**	0.66*	0.08	0.56	0.46	0.44
2 物联网	0.06	0.45	0.87***	0.57	0.65*	0.21	0.97***
3 听觉辅助	0.61*	0.11	0.14	0.52	0.84**	0.61*	0.33
4 医疗科技	0.89***	0.68*	0.55	0.21	0.08	0.90***	0.28
5 远程医疗	0.43	0.83**	0.36	0.70**	0.49	0.18	0.34
6 人工智能	0.23	0.39	0.46	0.95***	0.41	0.42	0.66*

Note:

- 一个星号 (*) 相似度大于 0.6
- 两个星号 (**) 相似度大于 0.7
- 三个星号 (***) 相似度大于 0.85

4.2 基于 S-T-B 创新生态网络构建与分析框架

4.2.1 网络构建

设置相似度阈值为 0.6，主题相似度高于阈值的主题对将会被链接，并认为是高度相关的发展趋势，构建出下图：

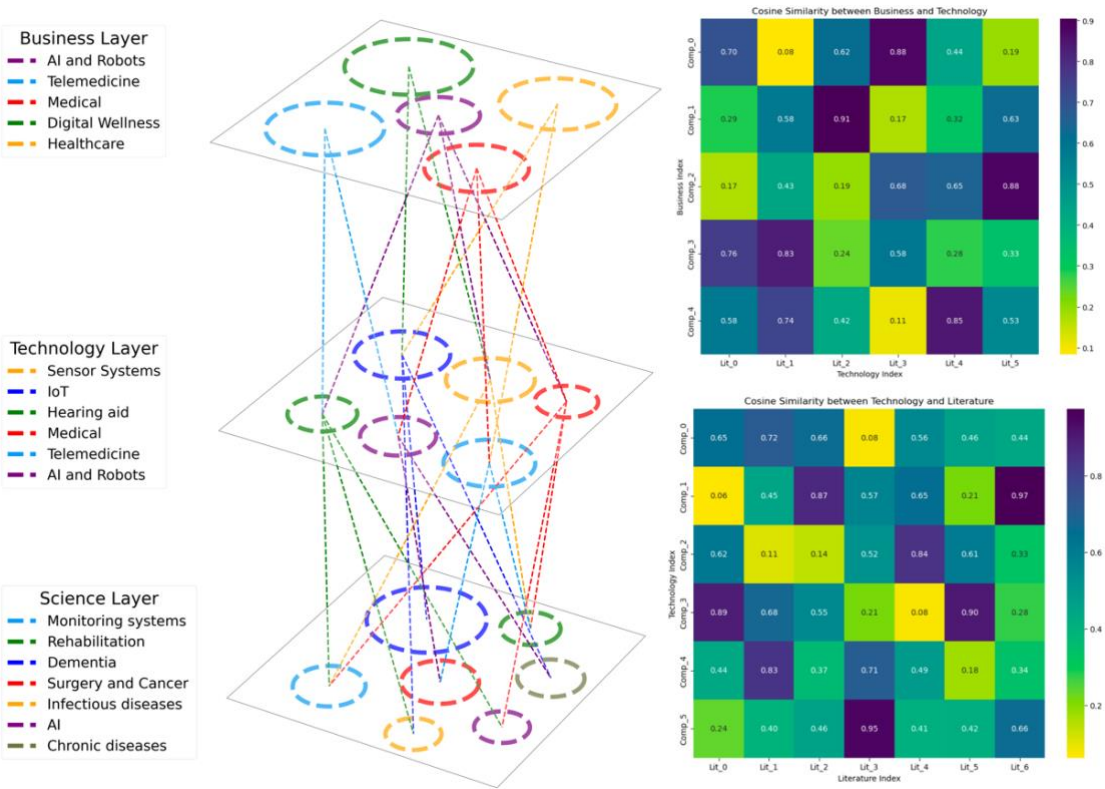


图 4.2 S-T-B 创新生态网络

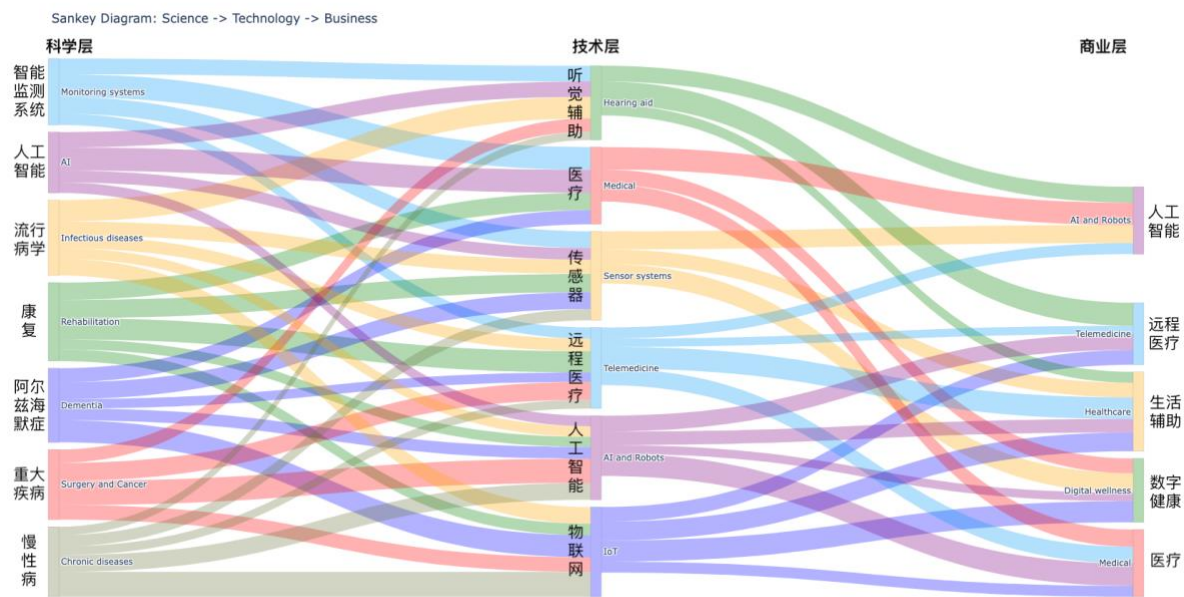


图 4.3 生态间各主题相似度桑基图

4.2.2 S-T-B 网络分析框架

本节根据许冠南^[35]根据生态间链接不同而设计的四象限分析框架，对上一节的结果进行分析。

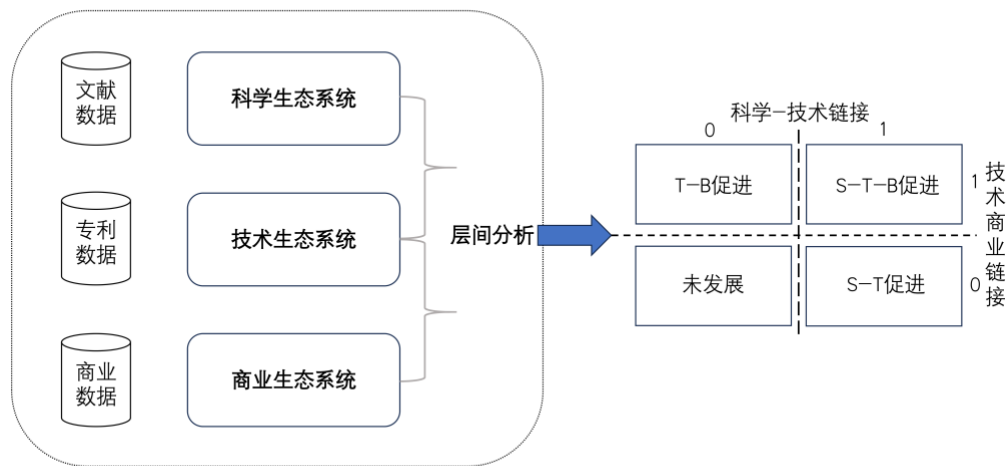


图 4.4 S-T-B 网络分析框架

基于 4 种相互作用类型象限，去分析其中的创新机会：将其可以分为：

（1）未产生协同发展效应

象限中的细分市场在科学、技术或商业生态系统中是孤立的，各层之间没有联系。

（2） 科学-技术协同发展

在 S-T 促进象限中，这些细分市场拥有成熟的科学和技术研究，但尚未商业化。

（3） 技术-产业协同发展

在 T-B 关联象限中，这些细分市场更面向市场，并已发展出相关的行业知识优势，但缺乏必要的基础研究。

（4） 研究-技术-产业协同发展

在 S-T-B 开发象限中，各个部分在所有三个层面上都相互关联，并实现了平衡的运行状态。

4.3 协同视角下养老科技产业创新机会发现

4.3.1 未产生协同发展效应类创新机会

即使在 S-T-B 领域中相似度较低的主题，在结果中可以看到对于每一个单一领域，都至少有一个发展积极相关的主题存在，但也存在大量的没有协同效应的领域，原因可能是发展的应用方向与技术基础不同。

值得关注的，对于科学生态中‘人工智能在疾病诊断与预测方向的应用’主题与技术生态中‘人工智能’的相似度并不显著，凸显了科学研究领域与技术领域对于人工智能的发展方向是不相同的，这是一种技术创新机会，对于人工智能在疾病诊断与预测方向的应用相关的研究进行应用技术的发展与转化。

还有技术领域‘远程医疗’与商业领域‘远程医疗’的相似度也不显著，这反映出远程医疗相关技术的商业化率是很低的，很多技术并没有成功转化为产品，这是一种商业创新机会。

对于未产生协同发展效应的各主题，是拥有最大协同发展潜力的，特别是领域相近的主题，应促进科学研究和技术应用之间的协同发展，挖掘潜在的技术和商业创新机会，推动技术成果的快速转化和应用。同时，跨学科合作、基础研究投入、市场导向和政策支持的综合措施。

4.3.2 科学-技术促进发展类创新机会

这些细分市场拥有成熟的科学和技术研究，但尚未商业化，所示创新机会集中在商业机会。一方面代表科学研究与技术之间的转化效率高，另一方面也是启示在商业领域中寻找创新机会，将技术积累转化为商业机会。

科学生态的智能检测系统与技术生态中医疗科技之间促进作用；科学生态的针对阿尔兹海默症的研究与技术生态中物联网与机器人技术之间的促进作用；科学生态的对于重大疾病的研究与技术生态中人工智能的技术之间的促进作用；科学生态的对于人工智能在疾病诊断与预测方向的应用的研究与技术生态中医疗科技之间的促进作用；科学生态的对于慢性病的研究与技术生态中物联网与机器人之间的促进作用，这些链

接告诉我们这些领域在科学与技术生态的创新要素具有相似性，一方面技术转化与迁移是简单的，更容易产生创新。也启示这些领域应需要积极寻找商业机会，比如对于科学生态的智能检测系统与技术生态中医疗科技，可以积极发展商业生态中的远程医疗，让领域间知识与技术的流通更加通畅，达到科学-技术-产业协同发展的平衡。

4.3.3 技术-产业促进发展类创新机会

这些细分市场更面向市场，并已发展出相关的行业知识优势，但缺乏必要的基础研究，所以应在基础研究上进行技术与知识的积累。

比如商业生态的诊疗科技与技术生态的人工智能联系显著，但技术生态的人工智能与科学生态的‘人工智能在疾病诊断与预测方向的应用’联系并不显著，这并不一定是缺少相关的基础研究，而是相关研究与产业间的联系度不够。

面对这种情况，研究机构可以通过与医院和医疗机构的合作，开发智能诊断系统 and 健康监测设备，推动人工智能技术在医疗领域的广泛应用。或建立专项研究基金，支持与产业需求相关的基础研究项目，弥补当前市场导向细分市场的基础研究短板。推动科研机构与企业合作，共同开展基础研究，通过产学研结合的方式，解决实际产业问题。

4.3.4 科学-技术-产业协同发展类创新机会

科学-技术-产业可以协同发展，创新要素间相似度高，应保持这个发展趋势与状态，利用系统协同发展的优势。例如科学生态的智能检测系统与技术生态中的传感器科技与商业生态中的人工智能与机器人，科学生态的重大疾病与技术生态中的人工智能与商业生态中的诊疗科技等等，都反映出科学研究、技术开发和商业应用之间的高度协同效应。

这会带来很多优势与创新机会，例如因为其协同研发与资源共享的优势，可以提升研发的效率与科研成果的转化率。还可以推动跨领域的融合，例如人工智能在医疗诊断中的应用，将传感器技术与智能检测系统相结合，从而开发出更先进的诊疗设备。与前沿技术的突破，例如科学研究中的前沿技术能够迅速应用于产业，推动技术进步，如利用人工智能算法提高重大疾病诊断的准确性和效率。

4.4 本章小结

本章节基于上一章对于科学、科技与商业创新生态系统的发展趋势，利用其不同系统产业间的相似度进行 S-T-B 创新生态网络的搭建。首先介绍了将产业文本进行向量化的策略，之后利用余弦相似度计算了不同系统间产业的相似度，设定阈值，搭建了 S-T-B 创新生态网络并阐述了分析框架，最后分为四类阐述了产业视角下养老科技产业的创新机会。

5. 基于创新生态网络的养老科技创新机会发现

在上一章节中，已经在产业层面对不同生态中不同产业的发展趋势的相关性进行了分析，本章节承接上文，从微观实体层面研究不同创新生态网络之间的连接，构建了产业视角下的 S-T-B-I 创新生态网络，与企业生命周期视角下的 T-B-I 创新生态网络，并研究其中的链接关系。

5.1 产业视角下养老科技产业创新生态网络构建

基于 S-T-B-I 模型进行多层创新生态的构建，其构建方式如下图所示。其中投资层，每一个点代表一个投资者；商业层，每一个点代表一家公司；技术层，每一个点代表一个专利；科学层，每一个点代表一片文献。投资层与商业层之间的联系是公司受到投资；技术层与商业层之间的联系是专利的委托方为这家公司；技术层与科学层之间的联系是为同一家机构。其中采用的匹配方式是模糊匹配。

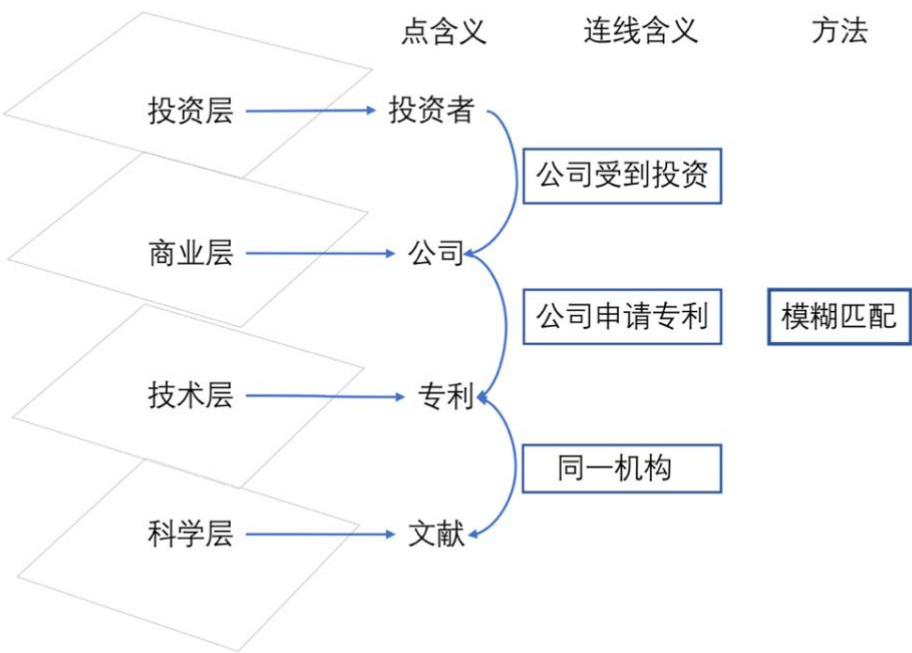


图 5.1 模型搭建图

模糊的匹配逻辑为：首先，标准化匹配字段的名称，将名称转换为小写并去除空格，根据词汇将 Univ 等词汇映射为 university。然后根据忽略列表筛选出不需要匹配的名称，这些名称由人工进行筛选，多数为名字过于简单影响匹配结果的名称。标转化结束后，使用字符串匹配算法，fuzzywuzzy 库的 process.extractOne，在标准化公司名称列表中寻找最佳匹配公司名称，并计算匹配分数。这里设定阈值为 90，如果匹配分数

高于设定阈值且最佳匹配公司名称不在忽略列表中，则认为匹配成功。最终搭建出以下结构：

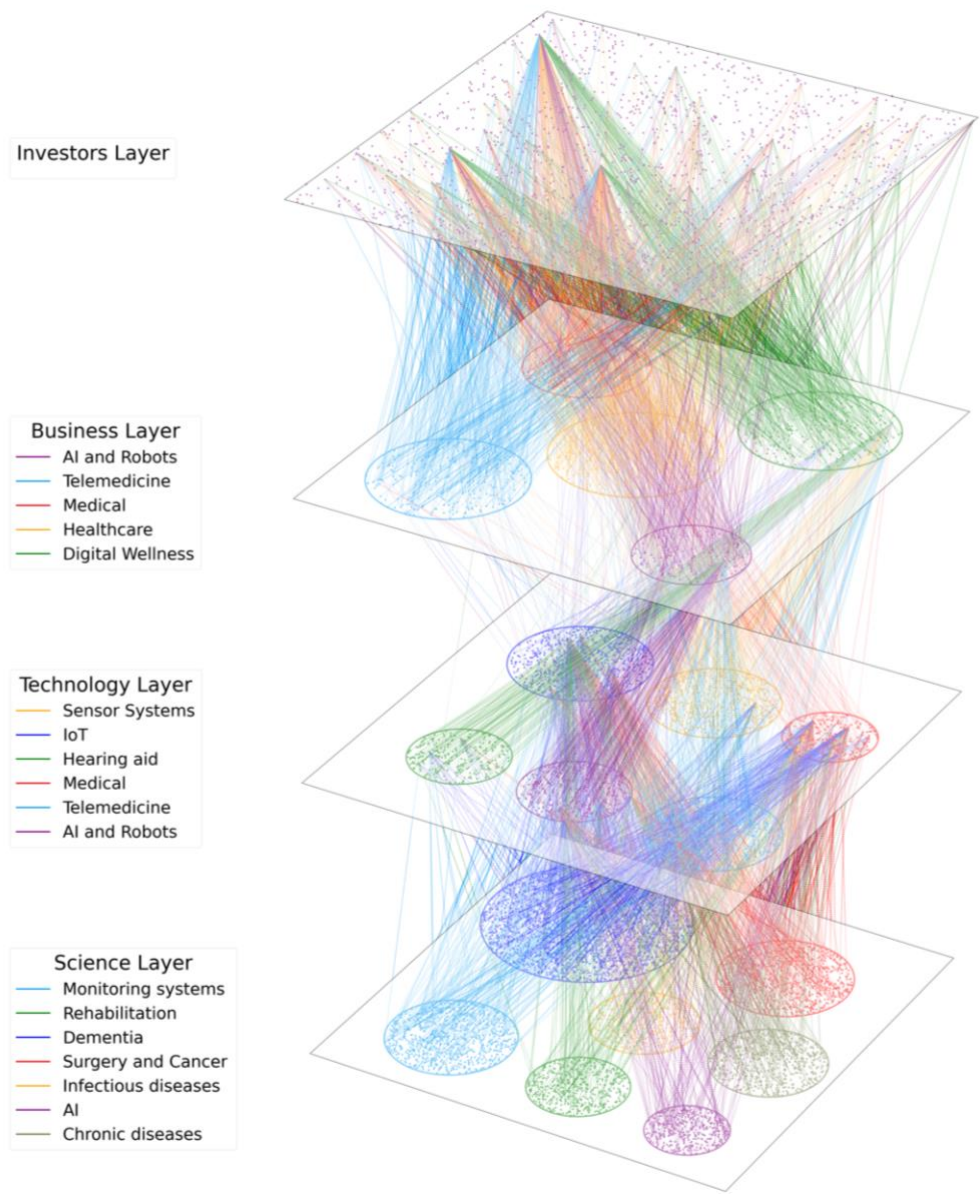


图 5.2 产业分类视角下 S-B-T-I 创新生态网络

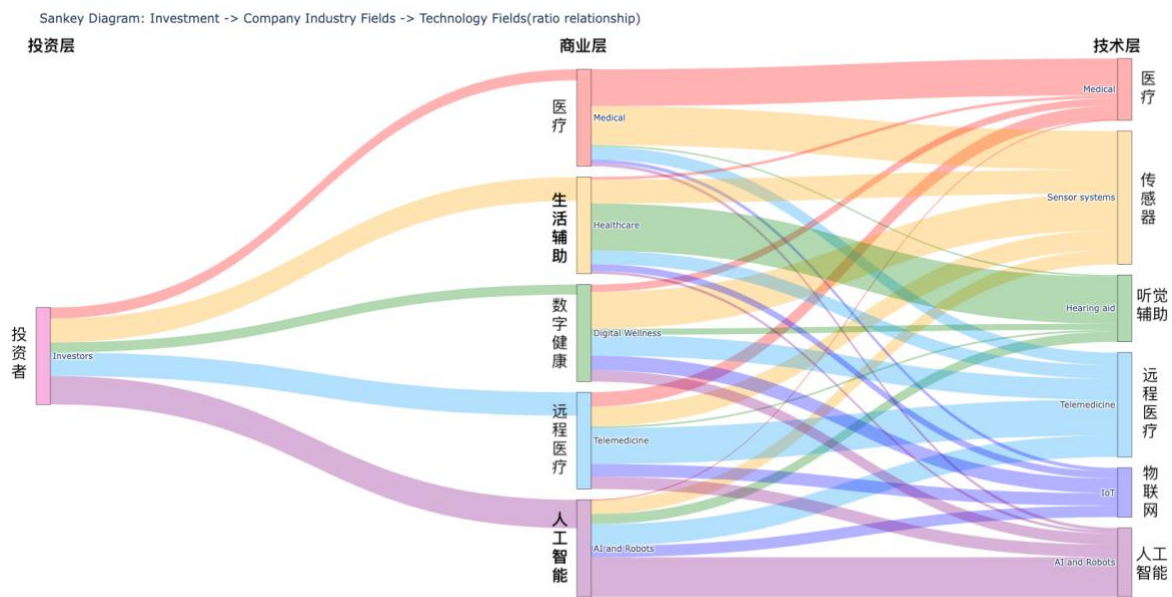


图 5.3 产业视角下的桑基图

本章节之后部分将会对其进行不同层次之间的分析。

5.2 产业视角下养老科技产业创新机会发现

5.2.1 投资层与商业层间创新机会分析

对投资者层与商业层之间的投资关系进行统计，将投资者对不同产业投资数量进行单因素方差分析，得到以下结果，P 值远小于显著性水平（0.05），表示不同组别之间在统计学上存在显著差异。表示在养老科技产业领域，投资者对于不同产业的发展前景看法与投资选择是有差异的。

表 5.1 投资者对不同产业投资数量的单因素方差分析结果

产业名称	受投资数量	均值	方差
人工智能与机器人	930	0.68	3.80
远程医疗	769	0.56	5.10
医疗与疾病诊断	362	0.26	1.76
数字健康	329	0.24	1.34
生活辅助	780	0.57	6.29
F 值	14.67	P 值	6.1274E-12

如图 5.3 所示，人工智能与机器人产业的受投资数量是最多的，并且投资份额较为平均，意味着更多投资者对于人工智能与机器人在养老科技产业的发展前景较为看好，选择加大对其的投资力度。生活辅助产业与远程医疗产业的受投资数量相近，都受到

了投资者的关注，但相比之下生活辅助产业的方差较大，证明其投资较为集中，有少数投资者更加看好这个产业的发展前景。

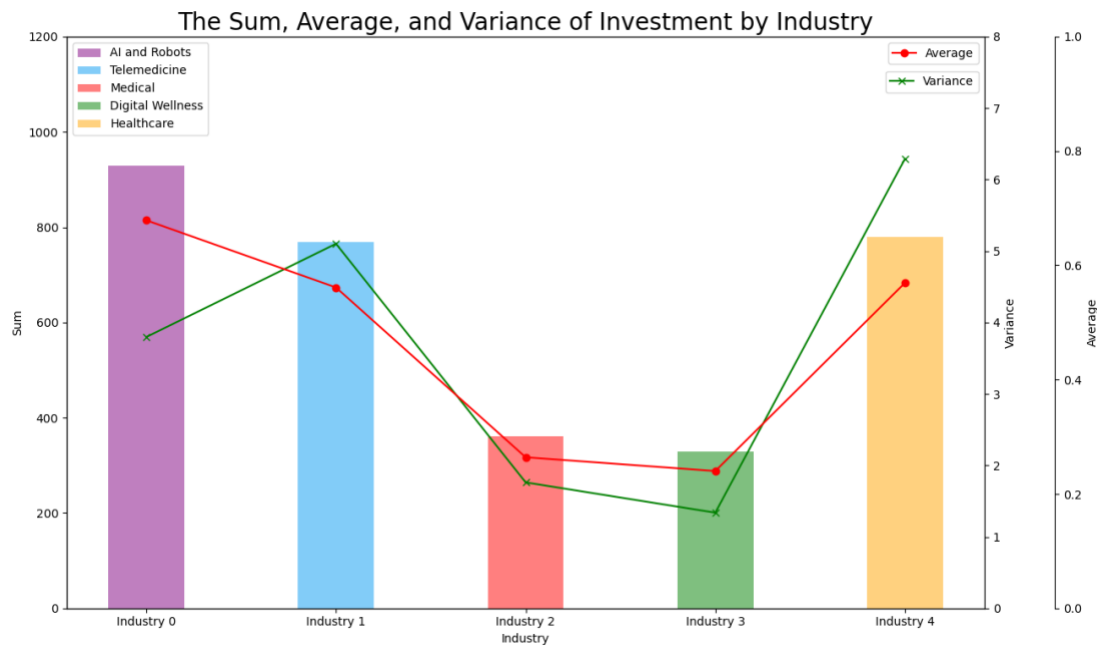


图 5.4 投资者对不同产业投资数量

5.2.2 商业层与技术层间创新机会分析

对商业层与技术之间的专利申请关系进行统计，对不同产业所申请的专利数量进行单因素方差分析，得到以下结果，P 值大于显著性水平（0.05），意味着不同组在专利拥有情况上的平均值没有显著的统计学差异。表示在养老科技领域，不同产业对于专利的申请并没有显著的差异。

表 5.2 不同产业的公司专利申请数量的单因素方差分析结果

产业名称	拥有专利数量	均值	标准差
人工智能与机器人	400	2.88	19.02
远程医疗	54	0.21	0.73
医疗与疾病诊断	313	1.64	10.77
数字健康	79	0.26	1.28
生活辅助	514	1.98	17.38
F 值	2.10	P 值	0.08

如图 5.4 所示，生活辅助类产业的公司拥有更多数量的专利，其中大部分由听力辅助类型的专利构成，这是因为其中大部分企业是主营业务是助听器产品。并且其标准差较大，说明专利的所有权是较为集中的。人工智能与机器人产业的专利拥有数量其次，其中人工智能与机器人领域的专利占比较高，且绝大多数的人工智能与机器人领

域的专利都分布在这个产业，说明人工智能与机器人领域对于本领域专利的申请是十分重视的。

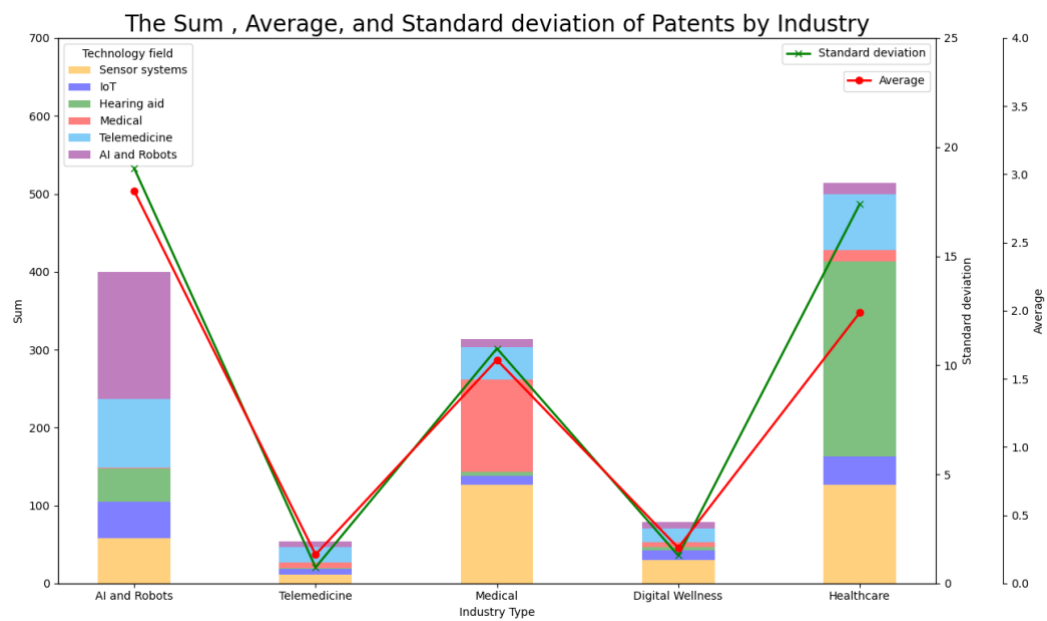


图 5.5 不同产业拥有专利数量

将企业申请的专利数据进行汇总与统计，筛选出专利拥有数量最高的 10 家企业，如下表所示，其中一半的公司是人工智能与机器人领域的公司。

表 5.3 专利拥有数量最高的十家企业

编号	企业名称	产业领域	传感器	物联网	听觉辅助	医疗	远程医疗	人工智能	总数
1	Nuance Communications	人工智能与机器人	11	15	35	0	35	113	219
2	Sivantos	生活辅助	37	3	148	0	14	2	204
3	Starkey	生活辅助	58	10	101	2	14	3	188
4	Dexcom	医疗与疾病诊断	29	8	0	13	22	3	75
5	Intuition Robotics	人工智能与机器人	3	0	0	0	22	6	31
6	Starship Technologies	人工智能与机器人	14	5	0	0	7	3	29
7	Teladoc	生活辅助	0	17	0	0	5	2	24
8	Kneron	人工智能与机器人	3	2	1	0	2	14	22
9	Koko	人工智能与机器人	2	3	4	0	4	4	17
10	AC Immune	医疗与疾病诊断	0	0	0	15	0	0	15

5.2.3 技术层与科学层间创新机会分析

知识生态系统可以进一步分为文献为载体的科学层，与专利为载体的技术层，这两者共同推进科学技术的创新，是价值的创造的重要过程。而科学研究更着眼于基础知识的研究，而专利更着眼于围绕着基础知识的应用技术。科学知识的创造为应用技术的生产者带来了知识支撑，科学知识本身是一种公共产品，具有外部性，而应用技术通过知识产权制度获得了占有性^[37, 70, 71]。Motohashi 对于 30 年内的专利与文献信息进行相关性分析，发现知识生态的传播逐渐向单向发展，即发现科学研究对于技术的依赖程度在下降，技术发展越来越依赖以大学与公共研究机构为主体的科学研究的发展，更为基础的科学研究越来越成为创新的主要驱动力^[71]。

高校的研究工作可以分为基础研究，应用研究和试验发展三类。基础研究是以文献为载体的科学研究范畴，应用研究是专利为载体以达到预定的目标，为解决实际问题提供科学依据。从专利管理的角度，知识生产力分为专利产出和专利转化两个动态阶段，专利产出即高校申请专利，而技术转化是以市场为导向，是科技成果的商品化，产业化。高校提升应用研究对于基础研究的投入比例，有利于专利的转化即高校的研究成果可以授权与售卖给企业，进行价值创造^[72]。

在本文构造的科学与技术生态中因为没有考虑商业生态的联系，主要关注专利产出的过程。在技术生态中的专利可以看作高校对于应用研究的投入，而科学生态的文献可以看作是高校对于基础研究的投入。因此分别统计文献发表数量与专利申请数量并分别排名列于下表，是在养老科技产业中对于基础研究与应用研究做出突出贡献的机构，可以看出榜单重合性并不高，可见不同机构对于研究的侧重点的不同。进一步去衡量高校基础研究与应用研究之间的关系，即高校的专利申请与发表文献之间的关系。去评价哪些机构在科学研究最终转化到专利申请上有优势，本质是衡量应用研究与基础研究的相关性与转化效率，因此本文设以下公式 5.1 的指标去衡量研究转化效率，并根据最终得分进行排名列于下表，其中领域相关性采用的第四章计算的领域之间的相似度。

$$Score = \sum_i \sum_j \frac{Patent\ number_i}{Literature\ number_j} \times Similarity_{ij} \quad (5.1)$$

其中*i*代表专利的领域，*j*代表文献的领域

北京工业大学毕业设计（论文）

表 5.4 机构排名

编号	文献发表数量排名	数量	专利申请数量排名	数量	研究转化排名	得分
1	北京大学	60	拉夫里科技大学	6	山东大学	8.837
2	中山大学	51	四川大学	5	郑州大学	5.067
3	浙江大学	50	浙江大学	4	四川大学	5.056
4	四川大学	49	山东大学	4	重庆大学	4.024
5	法国巴黎公立医院	41	郑州大学	4	华南理工大学	3.785
6	香港理工大学	38	三星电子	4	合肥工业大学	3.042
7	荷语天主教鲁汶大学	30	富士通	4	松下电器	2.759
8	山东大学	26	安徽工业大学	3	长治医学院	2.457
9	郑州大学	24	逢甲大学	3	温州医科大学	2.352
10	浙江中医药大学	19	江南大学	2	江南大学	2.287
11	台北医学大学	19	香港理工大学	2	深圳大学	2.257
12	温州医科大学	17	重庆大学	2	浙江大学	2.231
13	南京大学	16	中国科学院深圳先进 技术研究院	2	浙江中医药大学	1.989
14	日本东北大学	15	合肥工业大学	2	富士通	1.766
15	江南大学	14	长治医学院	2	宁夏医科大学	1.764

5.3 企业生命周期视角下养老科技产业创新生态网络构建

公司的发展阶段可以分为四阶段，投资在发展阶段中也起到不同的作用：（1）种子阶段：早期的投资帮助其孵化早期的商业概念；（2）研发（R&D）阶段：投资扶持其研发活动；（3）初始收入阶段：投资帮助其进行初始阶段的创收；（4）创收增长阶段：投资帮助其扩大规模^[73]。这四个阶段依据其活动的不同，也可以划分为：（1）初创阶段：确立商业模式；（2）过渡阶段：为增长奠定坚实的基础，包括产品/服务的开发、组织和管理团队的建设，以及开发必要的流程和基础设施；（3）扩展阶段：添加资源并利用流程快速发展业务；（4）退市阶段：通过首次公开募股、私募、合并或收购来收获企业价值^[74]。

因此本研究综合先前研究结合研究目的，将初创公司的发展阶段与商业活动分为：（1）种子阶段：验证创意，开发原型，并建立市场契合度；（2）早期阶段：初步获得市场牵引力和产品市场契合度；（3）成长期：大规模扩展运营，扩大市场影响力，并努力实现盈利；（4）成熟与退出阶段：扩大市场份额，实现或保持盈利，并探索退出策略。并根据不同的投资阶段信号，将发展阶段与投资信号进行对应，如下表所示：

表 5.5 公司发展阶段与投资信号表

编号	发展阶段	投资信号
1	种子阶段	天使投资；前种子轮；资助（Grants）；非股权援助；股权众筹；种子轮；创意阶段；产品众筹
2	早期阶段	A 轮融资；可转换债券；企业轮；债务融资；早期风险投资；未知轮次的风险投资
3	成长阶段	B 轮融资；C 轮融资；后期风险投资；私募股权；二级市场；债务融资；初创公司扩展
4	成熟与退出阶段	D 轮融资；E 轮融资；F 轮融资；首次公开募股（IPO）；并购（M&A）；上市后股权；上市后债务

依靠公司的受投资阶段，将拥有投资信息的 825 家公司划分为不同的发展阶段。

表 5.6 公司发展阶段数量

编号	发展阶段	公司数量
1	种子阶段	521
2	早期阶段	151
3	成长阶段	112
4	成熟与退出阶段	41

与上一节的创新生态系统搭建方式类似，因为科学层与技术层的链接关系已经在上一节进行了分析，因此本节仅构建由投资者层，商业层，与技术层构成的三层创新生态系统。链接关系与上一节也保持一致，投资者层与商业层是投资关系，商业层与技术层是专利申请的关系。

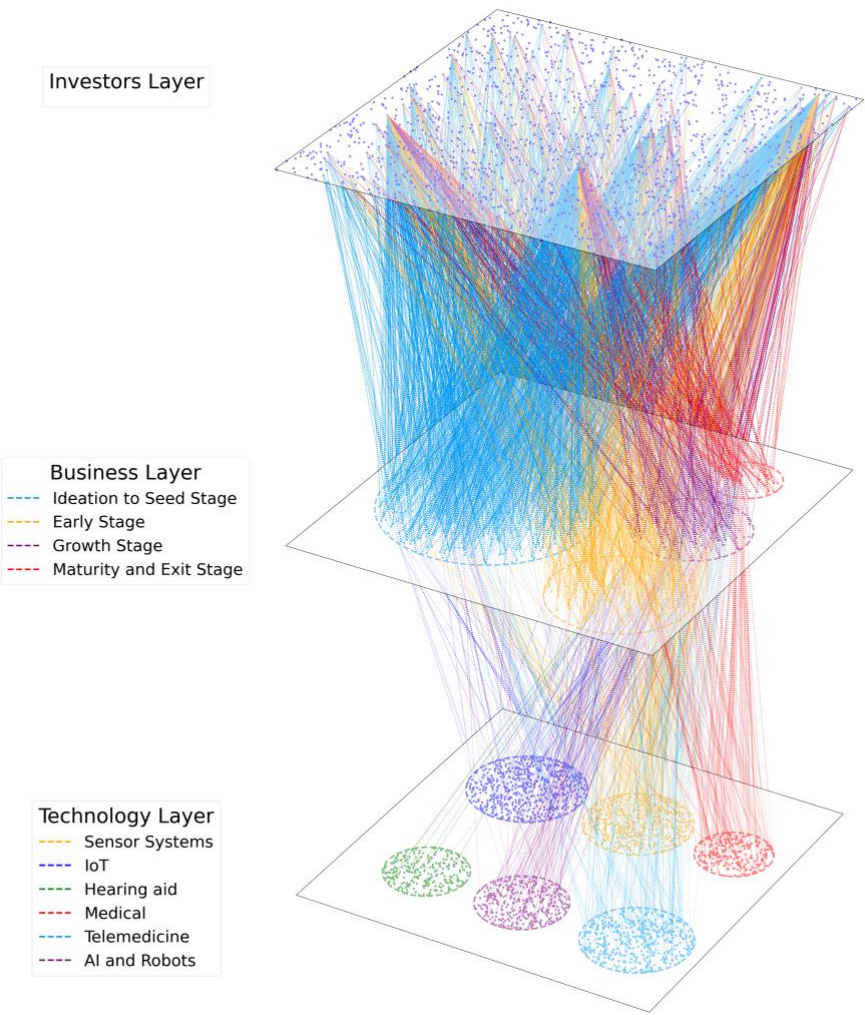


图 5.6 基于企业生命周期的 T-B-I 创新生态网络

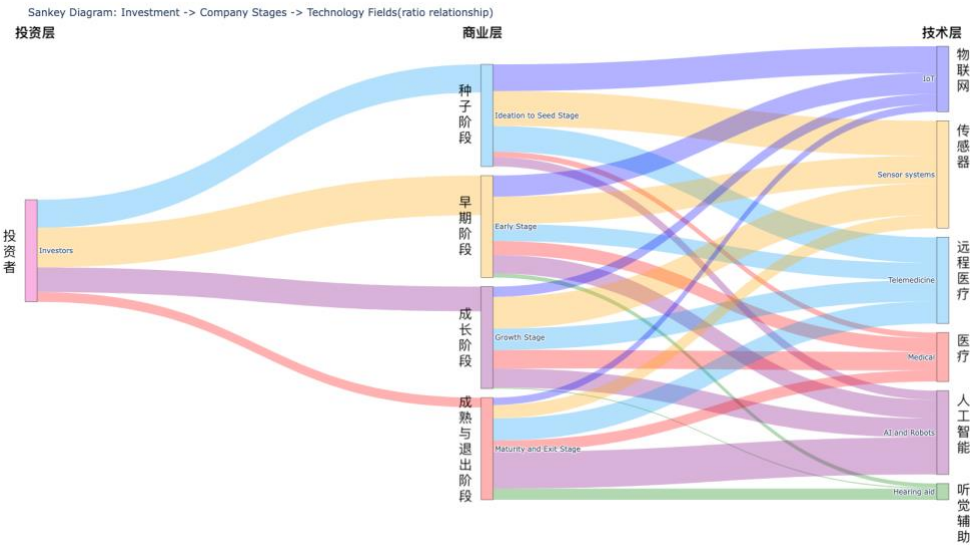


图 5.7 企业生命周期视角下的桑基图

5.4 企业生命周期视角下养老科技产业创新机会发现

5.4.1 投资层与商业层间创新机会分析

对投资者层与商业之间的投资关系进行统计，对不同发展阶段公司受到的投资数量进行单因素方差分析，得到以下结果，P 值远小于显著性水平（0.05），表示不同组别之间在统计学上存在显著差异。投资者对于不同发展阶段的公司的投资意愿与发展前景是存在显著差异的。

表 5.7 投资者对不同生命周期的公司投资数量的单因素方差分析结果

发展阶段	受投资数量	均值	方差
种子阶段	769	0.59	6.33
早期阶段	1088	0.82	11.77
成长阶段	679	0.52	3.10
成熟与退出阶段	266	0.20	0.54
F 值	12.08	P 值	9.0058E-10

如图 5.6 所示，早期阶段的公司受投资数量是最多的，之后随着生命周期的发展逐渐减少，可见在养老科技产业，早期阶段的公司更容易受到投资者的青睐。同时从方差的变动我们可以知道，早期阶段的投资是相对集中的，有少数几家明星公司收到了很多投资者的投资，随着生命周期的发展，投资逐渐趋于平衡，可以在竞争激烈的市场中存活，本身受到投资就成为了一种普遍现象。

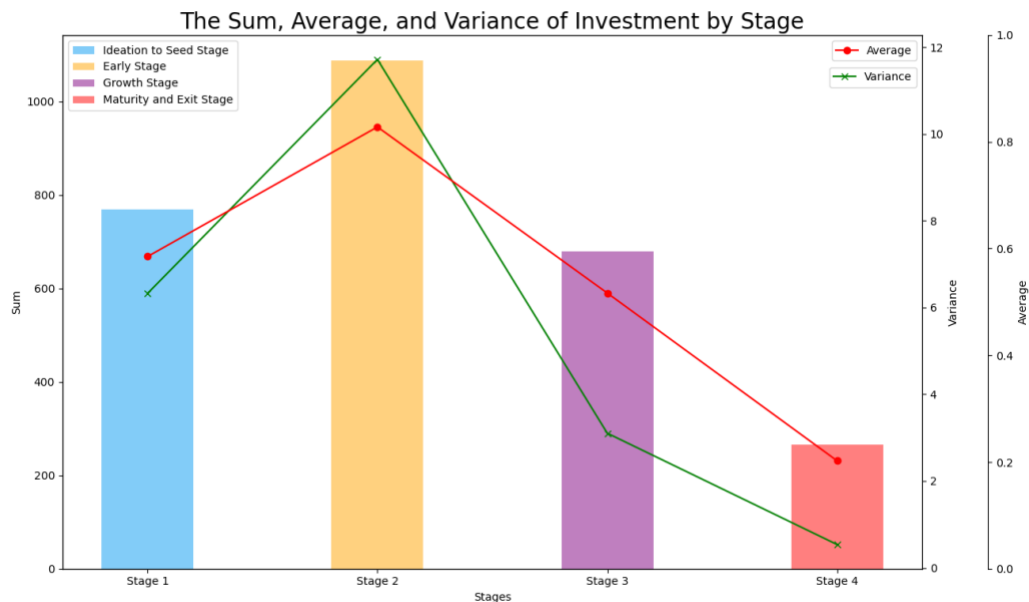


图 5.8 投资者对不同发展阶段的投资数量

风险投资（Venture Capital）与私募股权公司（Private Equity）的主要区别在于它们的投资阶段、投资目标和策略、资金规模以及管理方式。(1)投资阶段：风险投资通常关注早期和成长阶段的初创公司，帮助他们从创意到市场的早期发展。相反，私募股权公司则偏向于投资成熟企业，通常在这些企业已经具备稳定收入和盈利能力的情况下进行重组或扩展。(2)投资目标和策略：风险投资寻求高风险、高回报的机会，重点支持技术创新和市场潜力巨大的新兴企业。私募股权公司则注重通过运营改善、财务重组等方式提升企业价值，目标是通过稳定和长期的收益实现退出。(3)资金规模：风险投资的单笔投资金额相对较小，通常在数十万到数百万美元之间。而私募股权公司的投资金额较大，往往达到数千万甚至数亿美元。(4)管理方式：风险投资通常通过少数股权的形式提供资本，并在战略层面提供支持。私募股权公司则倾向于获得企业的多数股权或完全控股，以便进行深度的管理和运营干预^[75, 76]。

对于投资者的投资信息进行统计与汇总，得到对于养老科技产业投资最多的 10 家投资者，并按照起投资者类型进行归类，汇总下表。

表 5.8 投资数量最多的 10 家投资者

编号	投资者名称	投资者类型	种子阶段	早期阶段	成长阶段	成熟阶段	总数
1	Qiming Venture Partners USA	风险投资 VC	37	53	25	8	121
2	4DX Ventures	风险投资 VC	36	45	24	9	114
3	Domain Associates	风险投资 VC	35	44	23	9	111
4	MKaNN	风险投资 VC	36	43	24	6	108
5	Mercia Fund Managers	风险投资 VC	24	32	15	6	76
6	500 Startups Canada	风险投资 VC	17	32	14	6	69
7	EPIC Ventures	风险投资 VC	17	30	14	7	68
8	HPE Growth	私募 PE	17	29	13	6	65
9	Docor International Management	私募 PE	9	27	13	4	53
10	Frontier Equities VC	风险投资 VC	9	23	12	3	47

从中可以看到，在老年科技领域，风险投资（Venture Capital）是投资的主体，贡献了大量投资。但也不乏一些私募股权公司（Private Equity）公司。

5.4.2 商业层与技术层间创新机会分析

对商业层与技术之间的专利申请关系进行统计，对不同发展阶段的公司所申请的专利数量进行单因素方差分析，得到以下结果，P 值远小于显著性水平（0.05），表示

不同组别之间存在显著的统计学差异。表示在养老科技领域，不同发展阶段的公司对于专利的申请存在显著的差异。

表 5.9 不同发展阶段的公司专利申请数量的单因素方差分析结果

发展阶段	拥有专利数量	均值	标准差
种子阶段	111	0.21	1.31
早期阶段	126	0.83	2.52
成长期	187	1.67	5.05
成熟与退出阶段	321	7.83	35.82
F 值	11.15	P 值	3.5527E-07

从图 5.7 可以看到随着公司在不断发展，专利数量的申请是逐渐变多的，但与此同时专利所有权的集中性在变大，这也符合公司发展的情况，随着公司的不断发展，无论是对于研发的重视还是对于自身知识产权的保护都会逐渐加大。特别需要指出的是从成长阶段到成熟阶段专利所有权的集中程度快速增大，表示在养老科技产业较为成熟的市场中，技术的所有权是相对集中的。

另一方面需要指出的是人工智能与机器人科技在发展阶段中占比不断的加大，可见其在公司发展中的重要程度。

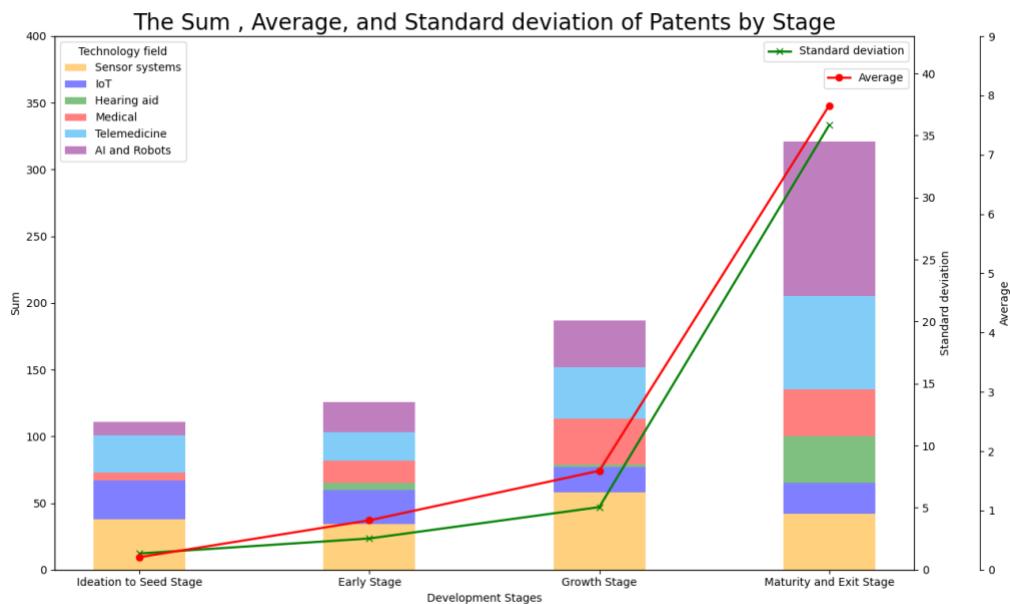


图 5.9 不同发展阶段的公司拥有专利的数量

5.4.3 专利所有权与投资关系

专利申请是打破创业投资过程中信息差的方式，可以凸显初创公司的创新能力，让其更容易受到投资^[73,74]。养老科技产业中初创公司占有很大的比例，因此本章节想要探究在养老科技产业中，专利申请是否有助于企业获得投资，以帮助其快速发展。

为此建立了逻辑回归模型去探究专利所有权与投资的关系，模型为：

$$\text{logit}(P(\text{fund} = 1)) = \beta_0 + \beta_1 \text{patent} + \beta_2 \text{stage} + \beta_3 \text{stage}^2 + \beta_4 \text{Industry}_1 + \beta_5 \text{Industry}_2 + \beta_6 \text{Industry}_3 + \beta_7 \text{Industry}_4 + \beta_8 \text{Industry}_5 \quad (5.2)$$

其中变量的含义为：

表 5.10 逻辑回归变量意义与赋值

编号	变量	变量的意义	变量赋值
1	<i>patent</i>	判断公司是否拥有专利	0：是 1：不是
2	<i>stage</i>	公司的发展阶段	连续变量，范围 1 到 4，数值越大代表公司越成熟
3	<i>Industry</i> ₁	公司属于人工智能与机器人产业	0：是 1：不是
4	<i>Industry</i> ₂	公司属于远程医疗产业	0：是 1：不是
5	<i>Industry</i> ₃	公司属于医疗与疾病诊断产业	0：是 1：不是
6	<i>Industry</i> ₄	公司属于数字健康产业	0：是 1：不是
7	<i>Industry</i> ₅	公司属于生活辅助产业	0：是 1：不是

表 5.11 逻辑回归结果

变量	模型 1	模型 2	模型 3
<i>const</i>	-3.1228	-3.7527***	-0.7372***
<i>patent</i>	0.8837***	0.9170***	1.1668***
<i>stage</i>	3.3078***	0.9170***	
<i>stage</i> ²	-0.6676***	-0.6502***	
<i>Industry</i> ₁	-0.2761		
<i>Industry</i> ₂	-0.5394		
<i>Industry</i> ₃	-0.7693		
<i>Industry</i> ₄	-1.1537		
<i>Industry</i> ₅	-0.3843		
Pseudo R-squared	0.1050	0.0896	0.0279
Log-likelihood	-481.5435	-489.8300	-523.0270
Converged	TRUE	TRUE	TRUE
LL-Null	-538.0473	-538.0473	-538.0473
LLR p-value	9.16E-21	9.08E-21	9.08E-21

Note:

一个星号 (*) : p 值小于 0.05 或 z 值绝对值大于 1.96

两个星号 (**) : p 值小于 0.01 或 z 值绝对值大于 2.576

三个星号 (***) : p 值小于 0.001 或 z 值绝对值大于 3.291

逻辑回归模型的结果表明, 是否拥有专利对公司获得资金的可能性有显著正向影响。虽然模型的伪 R 平方值较低, 说明模型的整体解释力有限, 有很多影响因素并未纳入进来, 但本研究想要挖掘专利申请对于公司获取资金的影响, 通过多个模型的一致性结果增加了此结果的可信度。

同时公司的发展阶段对获得资金具有倒 U 型的影响趋势, 这说明了公司在发展过程中, 在初期会有利于公司获得投资, 但发展到后期会不利于其获得投资。

5.5 本章小结

本章从微观创新实体的视角去构架创新生态网络, 首先构建了产业视角下的 S-T-B-I 创新生态网络, 其分为投资者层、商业层、技术层与科学层, 分别分析了层间的链接。在投资层与商业层间创新机会分析中, 发现投资者对于不同产业的发展前景看法与投资选择是有差异的, 投资者对于人工智能与机器人在养老科技产业的发展前景较为看好。在商业层与技术层间创新机会分析中, 发现不同产业对于专利的申请并没有显著的差异。接下来分析了各产业的专利数量与占比, 统计了专利数量最多的 10 家公司, 发现其中一半的公司是人工智能与机器人领域的公司。

之后在科学层与技术层间分析中分为文献发表、专利申请与研究转化进行了机构的统计与排名。在此之后构建了企业生命周期视角下的 T-B-I 创新生态网络, 分为投资者层、商业层、技术层, 分别分析了其层间链接, 在投资层与商业层间的创新机会分析中, 发现投资者对于不同发展阶段的公司的投资意愿与发展前景是存在显著差异的, 早期阶段的公司更容易受到投资者的青睐。但受到的投资是相对集中的, 有少数几家明星公司收到了很多投资者的投资, 随着生命周期的发展, 投资逐渐趋于平衡。之后统计了投资最多的 10 家投资者。在商业层与技术层间创新机会分析中, 发现不同发展阶段的公司对于专利的申请存在显著的差异, 随着公司在不断发展, 专利数量的申请是逐渐变多的, 但与此同时专利所有权的集中性也在变大, 最后通过逻辑回归肯定了拥有专利对公司获得资金的可能性有显著的正向影响

结论

本文通过对养老科技创新生态系统的发展趋势、生态系统间主题的相关性以及产业与企业生命周期视角下的创新网络进行了深入研究，得出以下结论：

（1）创新生态发展趋势：使用 KeyBERT、SciBERT、PCA 和 K-means 的等数据分析方法，对科学生态、科技生态和商业生态三个系统的主题发展趋势进行聚类分析。在科学生态系统中，发现了智能监测系统、康复、阿尔兹海默症等关键研究领域；科技生态系统则聚焦于传感器、物联网和医疗科技等技术研发；商业生态系统则关注人工智能与机器人、远程医疗等商业应用。

（2）生态系统间的主题相关性与 S-T-B 创新生态网络：通过量化分析，构建了一个基于产业间相关性的 S-T-B 创新生态网络，并根据四象限理论，将系统间产业发展关系分为四类创新机会，包括未产生协同发展效应、科学-技术促进发展、技术-产业促进发展以及科学-技术-产业协同发展。

（3）产业视角下的 S-T-B-I 创新生态网络：投资者对人工智能与机器人在养老科技产业的发展前景与创新机会持乐观态度。专利分析显示，专利数量最多的公司多来自人工智能与机器人领域，且绝大多数人工智能领域的专利来自于从事人工智能与机器人业务的公司，所以可以预见的是人工智能与机器人的技术与产品是未来养老科技产业的重点创新机会。

（4）企业生命周期视角下的 T-B-I 创新生态网络：投资者对不同发展阶段公司的投资意愿和前景存在显著差异，早期公司更易吸引投资。随着公司发展，投资趋于平衡，专利申请数量和专利所有权集中性均随公司发展而增加。此外通过逻辑回归分析证明，拥有专利显著提高了公司获得资金的可能性，而公司发展阶段对资金获取有倒 U 型影响。从创新机会的角度，初创企业是养老科技产业的创新主体，政策应加大对其支持力度，初创公司也应进行技术的研发来获得市场的资本青睐，共同推动创新生态发展。

本研究以创新生态系统的视角去研究养老科技产业的发展趋势与创新机会。然而，研究存在局限性，包括聚类方法的主观性以及数据集范围的限制，这可能影响了研究结果的客观性与全面性。以及没有在具体的技术创新层面进行更进一步的研究。未来的研究可以进一步深入分析养老科技的创新生态发展机理，促进生态系统间创新要素的流通和发展。

参考文献

- [1]World Health Organization. Definition of an older or elderly person. 2013.
- [2]国务院第七次全国人口普查领导小组办公室.（2021）. 2020 年中国人口普查主要数据.ISBN: 987-7-5053-9506-0
- [3]United Nations, DESA, Population Division. (2022) Licensed under Creative Commons license CC BY 3.0 IGO. United Nations, DESA, Population Division. World Population Prospects 2022.
- [4] Global health estimates 2019: Life expectancy and leading causes of death and disability, 2000–Geneva: World Health Organization; 2020.
- [5]WhiteHouse. Emerging technologies to support an aging population[J]. 2019.
- [6]国务院. 国务院关于印发《国家积极应对人口老龄化中长期规划》的通知[EB/OL]. (2019-11-21) [2024-05-23].
- [7]国务院办公厅. 国务院办公厅关于发展银发经济增进老年人福祉的意见[EB/OL]. (2024-01) [2024-01-15].
- [8] Micera S, Bonato P, Tamura T. Gerontechnology[J]. IEEE Engineering in Medicine and Biology Magazine, 2008, 27(4): 10-14.
- [9]Harrington T L, Harrington M K. Gerontechnology: Why and How[J]. MShaker Publishing B.V.2000
- [10]Bronswijk, Johanna EMH, et al. Defining gerontechnology for R&D purposes.[J] Gerontechnology 8.1 (2009): 3.
- [11]黄鲁成,郝亚丽,苗红,等.老龄化对科技发展的影响研究路径及结果分析[J].创新科技,2020,20(01):18-29.
- [12]黄鲁成,李晋,吴菲菲.基于文献的养老科技学科体系研究[J].情报杂志,2020,39(02):63-71.
- [13]黄鲁成,李晓宇,苗红,等.国外养老科技创新研究:趋势、主题与展望[J].科学学研究,2020,38(07):1294-1303.
- [14]刘春文,黄鲁成,苗红,等.基于 ECT-Dim 的前沿技术识别方法——以养老环境辅助生活技术为例[J].情报杂志,2023,42(08):77-82+76.
- [15]吴菲菲,栾静静,黄鲁成,等.基于新颖性和领域交叉性的知识前沿性专利识别——以老年福祉技术为例[J].情报杂志,2016,35(05):85-90.
- [16]黄鲁成,米兰,吴菲菲.面向老龄社会的新兴技术预测与评价研究[J].科研管理,2022,43(10):51-60
- [17]黄鲁成,郝亚丽,李晋,等.基于多源数据的养老科技技术体系识别研究[J].世界科技研究与发展,2019,41(06):585-595.
- [18]顾洁,郑稣鹏.AI+老龄研究的热点分析与前沿展望——国外文献知识图谱分析[J].国外社会科学前沿,2020,(03):53-63+84.

- [19]黄鲁成,刘玉敏,苗红,等.面向老年福祉技术创新的远程医疗领域前沿识别[J].情报杂志,2016,35(02):63-68+138.
- [20]黄鲁成,张二涛,苗红.老人福祉科技需求实证研究[J].科技管理研究,2016,36(13):102-105.
- [21]郝亚丽.我国养老科技产业的技术供需匹配评价研究[D].北京工业大学.2021.
- [22]黄鲁成,李晋,苗红.新兴养老科技产业及区域发展评价研究[J].科研管理,2020,41(02):51-59.
- [23]黄鲁成,常兰兰,苗红,等.基于 ESTP-Chain 四维分析法的老年福祉技术竞争态势分析[J].科技管理研究,2016,36(12):213-219.
- [24]米兰,黄鲁成,苗红,等.国际养老新兴技术合作模式演化及影响因素研究[J].科研管理,2021,42(10):28-39.
- [25]陆杰华.我国老龄产业研究评述及展望[J].北京大学学报(哲学社会科学版),2002,(01):135-140.
- [26]徐雨森,郑稣鹏,刘雨梦.适老技术创新研究述评与展望[J].技术经济,2017,36(11):24-32.
- [27]Motohashi K, Zhu C. Identifying technology opportunity using dual-attention model and technology-market concordance matrix[J]. Technological Forecasting and Social Change, 2023, 197: 122916.
- [28]Chen L, Xu S, Zhu L, et al. A deep learning based method for extracting semantic information from patent documents[J]. Scientometrics, 2020, 125: 289-312
- [29]Xu G, Dong F, Feng J. Mapping the technological landscape of emerging industry value chain through a patent lens: An integrated framework with deep learning[J]. IEEE transactions on engineering management, 2020, 69(6): 3367-3378.
- [30]岳俊举,冯立杰,冯奕程,等.基于多维技术创新地图与关联规则挖掘的技术机会识别方法研究[J].情报学报,2017,36(08):798-808.
- [31]Liu Z, Feng J, Uden L. From technology opportunities to ideas generation via cross-cutting patent analysis: Application of generative topographic mapping and link prediction[J]. Technological Forecasting and Social Change, 2023, 192: 122565.
- [32]Shen Y C, Wang M Y, Yang Y C. Discovering the potential opportunities of scientific advancement and technological innovation: A case study of smart health monitoring technology[J]. Technological Forecasting and Social Change, 2020, 160: 120225.
- [33]Motohashi K, Koshiba H, Ikeuchi K. Measuring science and innovation linkage using text mining of research papers and patent information[J]. Scientometrics, 2024: 1-21.
- [34]Kang N, Xu G, Mu X, et al. How virtual clusters affect innovation performance: Evidence from global hydropower industry[J]. Journal of Cleaner Production, 2022, 352: 131554.
- [35]Xu G, Hu W, Qiao Y, et al. Mapping an innovation ecosystem using network clustering and community identification: a multi-layered framework[J]. Scientometrics, 2020, 124(3): 2057-2081.

- [36]许冠南,胡伟婕,周源,等.创新生态系统双重网络嵌入对企业创新的影响机制[J].管理科学,2022,35(03):73-86.
- [37]Xu G, Wu Y, Minshall T, et al. Exploring innovation ecosystems across science, technology, and business: A case of 3D printing in China[J]. Technological Forecasting and Social Change, 2018, 136: 208-221.
- [38]Choi K H, Kwon G H. Strategies for sensing innovation opportunities in smart grids: In the perspective of interactive relationships between science, technology, and business[J]. Technological Forecasting and Social Change, 2023, 187: 122210.
- [39]Alexy O T, Block J H, Sandner P, et al. Social capital of venture capitalists and start-up funding[J]. Small Business Economics, 2012, 39: 835-851.
- [40]Davila A, Foster G, Gupta M. Venture capital financing and the growth of startup firms[J]. Journal of business venturing, 2003, 18(6): 689-708.
- [41]Graafmans J A M. TIDE proposal gerontechnology[J]. 1991.
- [42]郑稣鹏.适老企业创新机会形成与创新影响机理研究[D].大连理工大学,2021.
- [43]徐業良. 老人福祉科技產業的機會和挑戰[J]. 福祉科技與服務管理學刊, 2014, 2(1).
- [44]Adner R. Match your innovation strategy to your innovation ecosystem[J]. Harvard Business Review, 2006, 84(4): 98.
- [45]Tsujiimoto M, Kajikawa Y, Tomita J, et al. A review of the ecosystem concept—Towards coherent ecosystem design[J]. Technological forecasting and social change, 2018, 136: 49-58.
- [46] Granstrand O, Holgersson M. Innovation ecosystems: A conceptual review and a new definition[J]. Technovation, 2020, 90: 102098.
- [47] Niosi J. Building national and regional innovation systems[J]. Books, 2010.
- [48]Inoue Y. Winner-takes-all or co-evolution among platform ecosystems: A look at the competitive and symbiotic actions of complementors[J]. Sustainability, 2019, 11(3): 726.
- [49]Moore J F. Predators and prey: a new ecology of competition[J]. Harvard business review, 1993, 71(3): 75-86.
- [50]Dias Sant´ Ana T, de Souza Bermejo P H, Moreira M F, et al. The structure of an innovation ecosystem: foundations for future research[J]. Management Decision, 2020, 58(12): 2725-2742.
- [51]de Vasconcelos Gomes L A, Facin A L F, Salerno M S, et al. Unpacking the innovation ecosystem construct: Evolution, gaps and trends[J]. Technological forecasting and social change, 2018, 136: 30-48.
- [52]Li Y R. The technological roadmap of Cisco's business ecosystem[J]. Technovation, 2009, 29(5): 379-386.
- [53]Adner R, Kapoor R. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations[J]. Strategic management journal, 2010, 31(3): 306-333.

- [54]Oh D S, Phillips F, Park S, et al. Innovation ecosystems: A critical examination[J]. Technovation, 2016, 54: 1-6.
- [55] Tsujimoto M, Kajikawa Y, Tomita J, et al. A review of the ecosystem concept—Towards coherent ecosystem design[J]. Technological forecasting and social change, 2018, 136: 49-58.
- [56]Deborah J J. What is an innovation Ecosystem[J]. By National Science Foundation, Arlington.
- [57]Wang M Y, Fang S C, Chang Y H. Exploring technological opportunities by mining the gaps between science and technology: Microalgal biofuels[J]. Technological Forecasting and Social Change, 2015, 92: 182-195.
- [58] Motohashi K, Koshiba H, Ikeuchi K. New Indicator of Science and Technology Inter-relationship by Using Text Information of Research Articles and Patents in Japan[R]. Research Institute of Economy, Trade and Industry (RIETI), 2021.
- [59]Lebret R, Collobert R. Word emdeddings through hellinger pca[J]. arXiv preprint arXiv:1312.5542, 2013.
- [60]Deerwester S, Dumais S T, Furnas G W, et al. Indexing by latent semantic analysis[J]. Journal of the American society for information science, 1990, 41(6): 391-407.
- [61]Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.
- [62]Cvitanic T, Lee B, Song H I, et al. LDA v. LSA: A comparison of two computational text analysis tools for the functional categorization of patents[C].International Conference on Case-Based Reasoning. 2016.
- [63]Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in neural information processing systems, 2013, 26.
- [64]Pennington J, Socher R, Manning C D. Glove: Global vectors for word representation[C]. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014: 1532-1543.
- [65]Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.
- [66]Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [67]M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contextualized word representations[C]. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,, volume 1, pages 2227–2237, 2018.

- [68]Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [69]Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [70]Shibata N, Kajikawa Y, Sakata I. Extracting the commercialization gap between science and technology—Case study of a solar cell[J]. Technological Forecasting and Social Change, 2010, 77(7): 1147-1155.
- [71]Motohashi K, Koshiha H, Ikeuchi K. Measuring science and innovation linkage using text mining of research papers and patent information[J]. Scientometrics, 2024: 1-21.
- [72]周杨. 科技成果转化视角的高校知识生产力研究[D].浙江大学,2012.
- [73]Conti A, Thursby J, Thursby M. Patents as signals for startup financing[J]. The Journal of Industrial Economics, 2013, 61(3): 592-622.
- [74]Picken J C. From startup to scalable enterprise: Laying the foundation[J]. Business Horizons, 2017, 60(5): 587-595.
- [75]PitchBook. Private equity vs. venture capital: What's the difference? [EB/OL]. [2024-05-20]. <https://pitchbook.com/blog/private-equity-vs-venture-capital-whats-the-difference>.
- [76]Leland. Venture Capital Vs. Private Equity: A Career Guide for Aspiring Investors [EB/OL]. [2024-05-20].<https://www.joinleland.com/library/a/venture-capital-vs-private-equity-a-career-guide-for-aspiring-investors>

致谢

在北京工业大学的四年，是迄今为止成长最为迅速的四年，我从一个并未具有自我意识的小孩，变为大体了解并接纳自身的大人。在此我想向那些曾对我好的人，曾经一起度过一段美好时光的人，谈起心地善良时脑海中总会浮现的人，表达自己的感谢。

首先感谢学校的栽培，提供资源与机会，让我可以发挥自己的才能。其次感谢为我授课与生活指导的老师，将知识与经验传达给我。在此更要感谢袁菲老师与课题组中的两位学姐，共事的过程中提供的帮助与鼓励，让我学术道路的伊始充满力量。

之后，我要感谢我的朋友们，B105 的舍友们，你们是也一直是会是我人生中不可或缺的坐标，跨越时间，在迷茫时指引我，让我可以奔赴自己的美好前程。

感恩父母与爷爷的支持，感谢奶奶，感谢我的狗狗汤圆，感谢没有放弃的自己。

我不太确定，未来我是否还能拥有这样的情感与羁绊，但唯一可以确信的是我会生活在更多的厨房，图书馆与洗衣房。把握自己的内心与力量，坚强的成长，成为我想成为的人。

我无法永久的保留这些珍贵的美好记忆，因为所有的美都并非美的本体，它只是美的派生或摹本，因此必然包括杂质并不断磨损。但我坚信真正的美是坚不可摧的，一如苏格拉底所言：花虽凋零，但它的美永在。