

中文图书分类号: C931

UDC: 300

学 校 代 码: 10005



硕 士 学 位 论 文

MASTERAL DISSERTATION

论 文 题 目: 我国养老科技产业的技术供需匹配评价
研究

论 文 作 者: 郝亚丽

学 科: 管理科学与工程

指 导 教 师: 黄鲁成

论文提交日期: 2021 年 6 月

UDC: 300
中文图书分类号: C931

学校代码: 10005
学号: S201811031

北京工业大学管理学硕士学位论文

题目: 我国养老科技产业的技术供需匹配评价研究

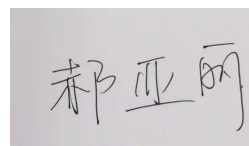
英文题目: RESEARCH ON TECHNOLOGY SUPPLY AND DEMAND
EVALUATION OF PENSION TECHNOLOGY INDUSTRY
IN CHINA

论文作者: 郝亚丽
学科专业: 管理科学与工程
研究方向: 科技与产业创新管理
申请学位: 管理学硕士
指导教师: 黄鲁成
所在单位: 经济与管理学院
答辩日期: 2021 年 5 月
授予学位单位: 北京工业大学

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京工业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：



日 期： 2021 年 06 月 25 日

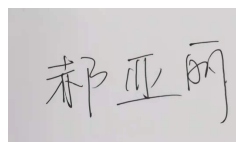
关于论文使用授权的说明

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）

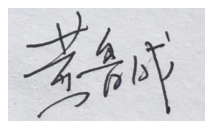
签

名：



日 期： 2021 年 06 月 25 日

导师签名：



日 期： 2021 年 06 月 25 日

摘要

老龄化社会的不断发展必然会带来巨大的挑战，因此，发展养老科技产业，以应对人口老龄化成为了重要途径。而当前养老科技产业面临着供需失衡问题，本文基于深度学习构建技术供需匹配模型，进而对养老科技产业的供需匹配情况进行深入研究，把握产业技术的供需能力，对于识别养老科技产业发展方向，为合理制定老龄化的相关政策，具有重要意义。

首先，通过对文献的梳理和回顾，界定了养老科技产业的概念。其次，根据养老科技产业的定义进行了多源数据的获取，依据文献数据源、政策、技术报告构建养老科技产业的技术词典；通过网络数据源获取用户需求，结合构建的技术词典识别技术需求点。然后，基于 BERT 和 TF-IDF 构建技术供需匹配模型，建立技术需求与技术供给的联系。再次，对识别出的技术需求点利用 K-means 算法进行聚类分析，划分为了 9 类，分别为日常生活护理、机器人技术、智能家居、可穿戴技术、移动通信类技术、检测类技术、传感器技术、辅助类技术、康复类技术。最后构建技术供需评价模型，对 9 类技术进行评价，研究发现养老科技产业技术的供需情况有六种：传感器技术、可穿戴技术和机器人技术为低供给、低需求；康复类技术为中供给、低需求；检测类技术和日常护理技术为中供给、中需求；移动通信类技术为中供给、高需求；辅助类技术为高供给、低需求；智能家居为高供给、中需求，并根据各类匹配情况，提出了技术研发主要方向，以及政策建议

关键词：养老科技产业；技术供给；技术需求；供需匹配模型；深度学习

Abstract

The continuous development of aging society will inevitably bring great challenges. Therefore, the development of pension technology industry to cope with the aging population has become an important way. At present, the pension technology industry is faced with the problem of imbalance between supply and demand. Based on deep learning, this paper constructs the matching model of technology supply and demand, and then makes an in-depth study on the matching situation of supply and demand of pension technology industry. Grasping the supply and demand ability of industrial technology is of great significance for identifying the development direction of pension technology industry and making reasonable policies on aging.

Firstly, the concept of pension technology industry is defined by combing and reviewing the literature. Secondly, according to the definition of pension technology industry, multi-source data is obtained, and the technology dictionary of pension technology industry is constructed based on literature data sources, policies and technical reports; Through the network data source to obtain user requirements, combined with the construction of technical dictionary to identify technical requirements. Then, the matching model of technology supply and demand is built based on Bert and TF-IDF, and the relationship between technology demand and technology supply is established. Thirdly, the identified technology demand points are classified into 9 categories by K-means clustering analysis, including daily life nursing, robotics, smart home, wearable technology, mobile communication technology, detection technology, sensor technology, auxiliary technology and rehabilitation technology. Finally, the evaluation model of technology supply and demand is built to evaluate nine kinds of technology. The results show that there are six kinds of technology supply and demand in the pension technology industry: sensor technology, wearable technology and robot technology are low supply and low demand; Rehabilitation technology is medium supply and low demand; Detection technology and daily nursing technology are medium supply and medium demand; Mobile communication technology is medium supply and high demand; The auxiliary technology is high supply and low demand; Smart home has high supply and medium demand. According to various matching conditions, the paper puts forward the main direction of technology research and development, as well as policy recommendations.

Keywords: pension technology industry, technology supply, technology demand, supply and demand matching model, deep learning

目录

摘要	I
Abstract	III
第 1 章 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 国内外研究现状	2
1.2.1 养老科技产业的研究现状	2
1.2.2 老年需求挖掘研究	3
1.2.3 技术供需匹配的研究现状	4
1.2.4 多源数据融合的研究现状	5
1.2.5 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	9
第 2 章 相关概念界定与研究方法	11
2.1 相关概念界定	11
2.1.1 养老产业	11
2.1.2 养老科技产业	11
2.2 网络数据及爬取方法	12
2.2.1 网络数据应用	12
2.2.2 网络爬虫技术	13
2.3 自然语言处理方法	14
2.3.1 文本预处理	14
2.3.2 文本分布式表达——词向量	16
2.3.3 文本相似度计算	17
2.4 本章小结	18
第 3 章 基于多源数据的养老科技产业技术需求挖掘	19
3.1 技术需求挖掘研究框架	19
3.2 数据来源	20
3.2.1 技术词典数据来源	20

3.2.2 技术需求挖掘数据来源	20
3.3 养老科技产业技术需求挖掘	21
3.3.1 文本数据处理	21
3.3.2 技术词典构建	22
3.3.3 技术需求挖掘	22
3.4 本章小结	24
第 4 章 基于深度学习的养老科技产业技术供给识别	25
4.1 养老科技产业技术供需匹配研究框架	25
4.2 数据来源	26
4.3 面向技术供给识别的匹配模型构建	26
4.3.1 技术供需匹配模型结构	26
4.3.2 基于 BERT 模型和 TF-IDF 的输入层	27
4.3.3 特征提取层和输出层	29
4.4 技术供需匹配模型算法实现	30
4.5 结果分析	31
4.6 本章小结	32
第 5 章 养老科技产业技术供需匹配评价	33
5.1 技术供需评价研究框架	33
5.2 技术点聚类分析	34
5.2.1 K-means 聚类流程	34
5.2.2 最优 K 值的选取	35
5.2.3 聚类结果分析	35
5.3 技术供需能力评价指标体系及模型构建	36
5.3.1 技术供需能力评价指标体系	37
5.3.2 技术供需能力评价指标细化	38
5.3.3 技术供需能力匹配评价模型构建	39
5.3.4 评价结果分析	40
5.4 对策建议	43
5.5 本章小结	44
结论	45
参考文献	47
攻读硕士学位期间所发表的学术论文	55
致谢	57

第1章 绪论

首先介绍了选题的背景与意义，接着梳理养老科技产业发展、老年需求挖掘、技术供需匹配、多源数据融合的国内外研究现状，通过对研究现状得分析提出现有研究存在得不足之处和局限性，进而根据相关结论提出本文的研究内容和用到的研究方法，并提出本文创新点。

1.1 研究背景及意义

1.1.1 研究背景

随着社会快速发展，人口老龄化问题凸现，如何应对人口老龄化问题是现阶段人类社会的一个重大课题。我国老龄化目前呈现出“规模大”和“进程快”等特征，截止到2019年，我国60岁以上人口达到2.49亿，超过总人口的17.9%，65岁以上人口达到1.66亿，超过总人口的11.9%，占世界老年人口总量的1/4。

人口老龄化程度的不断加深，其对社会各方面产生的不利影响也将逐渐突显出来，如劳动力供给不足，国民储蓄减少，经济增长潜力减小等^[1]。因此，积极应对人口老龄化成为了当前的重要任务，结合科技创新来应对这个问题是一种重要途径^[2]。国家近年来政策上也多次提出积极应对人口老龄化问题，提升科技创新能力，加大养老科技的支撑力度。

因此，经济社会发展的重要因素将包含满足老年人需求，积极改善老年生活品质。但是，随着社会发展和生活水平的不断提升，老年人的需求也逐渐呈现特殊性、新颖性、丰富性等变化，由此看来，传统“补缺式”模式下的养老服务无法满足当前多样性的养老需求^[3]。从需求角度来看，根据对北京老年群体使用科技产品情况得抽样调查结果发现，23.83%老年人认为现有科技产品不符合自己的需求，32.99%的老年人认为操作复杂且不方便使用，还有近三分之一的老年人认为现有科技产品价格高，新品种种类少；近85%的受访老人对新技术和产品有兴趣。从供给角度，全球范围内，老年用品已有6万种，仅日本就有4万中，但是我国的老年用品只有两千多种。在技术上，在incopat数据库种检索到的老年科技相关技术仅有4410件，落后于日本、美国等国家。面对当前养老供需严重不平衡问题，明确养老科技技术供给和需求的真实现状指导养老科技产业的研发方向，完善国家相关政策是十分必要的。

针对产业需求问题，用户的需求散布在海量的互联网数据中，使用单一数据源进行相关问题的研究是片面的，存在一定的不准确性。随着从工业社会向知识社会的转变，技术和信息等无形资产的重要性逐渐增加，学者发现使用多个数据集将数据进行融合可以得到更高质量的挖掘结果，从而使合并后的挖掘效果尽可能全面且满足实际需求^[4]。因此，本文将基于多源数据对养老科技产业技术进行识别分析，探索该产业在技术层面的供需差距。

1.1.2 研究意义

第一，这符合落实国家针对人口老龄化问题提出的各项政策建议，技术创新将成为缓解老龄化带来的社会压力，应对可能出现的社会问题的重要途径，同时，也促进了产业结构的优化、扩大社会供需问题，进而对经济社会的稳定发展有着重要意义。

第二，在提出新方法的支撑下，明确的为养老科技产业技术匹配差距，为企业和研发机构养老科技产业的技术创新提供了方向和新思路。

第三，为政府科技管理部门科技规划和涉老服务部门提供决策支持，促进养老健康事业与产业协调发展。

1.2 国内外研究现状

本研究涉及养老科技产业、养老需求挖掘、多源数据及技术供需匹配相关的内容，故需要对养老科技产业研究，养老需求挖掘研究，技术供需匹配研究及多源数据研究等进行综述。

1.2.1 养老科技产业的研究现状

随着社会人口老龄化程度的加深，依靠科技创新成为了应对老龄化问题的关键，众学者对养老科技产业的关注度也越来越高。各学者关注的话题主要分为以下几方面。

第一方面，研究养老科技理论的相关研究。Torre 等人^[5]指出在老年人衰老的过渡阶段存在一些问题，通过对养老科技技术的快速评估，养老科技正试图解决这些问题，越老越多的国家也在推动养老科技的发展，以及其与社会学、通信技术等的跨领域融合。Fernandez 等人^[6]认为养老科技是以为老年人提供衣食住行各方面的智能系统为目的的。此外，Chen Ke 等人^[7]通过老年人在养老科技产品和技术等方面上的态度和体验，研究老年人对于养老科技使用态度上的潜在理由。

第二方面，研究老年人对养老科技的技术接受度。Penine 等人^[8]认为老年人作为养老科技技术的主要使用者，是创新消费者，应将其意愿作为技术开发的主要参考方面，使其从技术被动接受者的角色转为技术开发的参与者。Macedo 等人^[9]利用结构方程模型（PLS）对从 278 名葡萄牙老年人样本中收集的数据进行分析，结果证实，老年人对信息通信技术的技术接受度受到老年人的行为意图和习惯的直接因素，及表现期望、努力期望、社会影响、享乐动机等间接因素。Chen 和 Chan 等人^[10]通过构建技术接受模型分析老年人对新技术接受程度的影响因素，主要包括了健康能力、生活质量等。此外，李敏等人^[11]也使用技术接受模型，探讨了老年人对科技产品态度的影响因素，其中包括了个体特征、外在的子女因素及产品因素等。

第三方面，对养老科技产业的发展研究，该部分主要研究养老科技产业的发展、重点研发领域及技术前沿等。黄鲁成等人^[12]构建了技术景观四侧面分析框架对养老科技产业进行分析，从不同层次较为全面地评估了我国养老科技产业的发展。李晋等人^[13]将网络数据和专利数据结合，对养老科技产业的重点创新领域进行分析，并分析重点创新领域的技术竞争态势。吴菲菲等人^[14]通过构建针对前沿性技术的一种识别方法来对养老科技领域的前沿性技术进行识别。娄岩等人^[15]通过构建技术融合的分析框架对养老科技技术与信息技术的融合进行研究。

1.2.2 老年需求挖掘研究

目前关于需求的研究主要从生理学、心理学、医学和社会学四个角度进行^[16]。根据马斯洛需求层次理论^[17]，将老年人的需求分为了生理需求、安全需求、情感需求、尊重需求和自我实现需求。其中，生理需求包括吃、穿、住、用；安全需求包括健康、安全；情感需求包括友情、亲情、归属感等；尊重需求包括尊严、尊重、认可；自我实现需求包括自我价值。基于这些需求，技术目的主要包括：增强和满意，预防和参与，补偿和援助以及护理和组织，老年人的主要目标是达到增强和满意，为老年人创造自主和独立的条件，同时实现自我价值^[18]。Caro 等人^[19]认为要积极应对人口老龄化，支持老年人需求的基础设施和服务，让老年人参与到科技创新的发展中来。而对于老年人的需求，Álvarez-García 等人^[20]提出在人逐渐老化的过程中，会收到多种因素的影响，这也使得老年人的需求呈现出多样性。国内外学者利用不同的数据对老年人相关需求进行了研究。

一是通过相关文献来分析需求，朱玥颖等人分析了智慧养老产业存在的问题，并针对问题，提出了养老产业的发展在经济新常态下应了解其发展规律，从供给侧进行结构性的改革，不断满足老年人对物质文化方面日益增长的需求^[21]。

二是利用调查问卷的数据分析需求，Lauriks 等人^[22]研究发现老年科技需求主要集中在健康安全、社会参与与日常生活等方面。Robinson M 等人^[23]通过电话访谈的方式对老年人的需求进行了研究，发现其需求主要集中为健康信息类需求，例如自我护理等。张二涛等人^[24]根据调查问卷结果对养老科技的需求进行了分析，主要由健康、居家与日常生活、信息沟通及社会参与四个方面。李晋等人根据调查问卷结果，就老年人对养老科技产品的显隐性需求进行了分析，包括日常生活必需品在内的显性需求、智能家居类等的隐性需求以及老年人辅助类产品等的的不确定性需求^[25]。林佳蓉等人利用调查问卷数据研究养老需求，主要对老年人对科技产品的态度、使用情况、感受及期望进行了分析^[26]。李娜利用中国健康与养老追踪调查数据库（CHARLS）数据进行了相关研究，发现在养老机构方面存在供不应求的问题，并根据发现的供需问题情况提出了相应的对策建议^[27]。张静基于 CGSS2015 和 CHARLS2015 两项全国性调查数据，将养老服务需求分为了日常照护需求和精神关怀需求^[28]。

三是利用网络数据对分析需求，钱宇星等人利用老年论坛“老年人之家”的网络数据，挖掘网络数据中存在的社会网络，研究分析得到了老年用户的健康信息需求^[29]。

此外，也有学者对养老产业供需匹配进行研究。严峻对我国目前养老供给侧及需求侧现状进行了分析^[30]。高艳等人针对海南养老服务产业发展种存在的供需问题提出了要从供需平衡视角推进相关发展^[31]。杨素等人通过实地调查了天津在失智老人照护服务方面的供需现状，对现存的供需矛盾进行了分析^[32]。张新辉等人^[33]基于 CLHLS 数据,分析了我国从 2005 年开始到 2014 年间的养老服务问题上的供需动态变化。

综上，老年市场相关需求挖掘的数据主要来源于调查访谈、专利数据和网络平台；对于供需匹配问题的研究中，主要集中在养老服务的供需失衡，且从研究角度来看，主要是分别对供给侧和需求侧现状进行分析。

1.2.3 技术供需匹配的研究现状

目前，对于技术需求识别主要利用专利数据。黄鲁成等人利用专利数据对养老科技技术的主题进行识别，通过问卷调研的方式将技术和需求进行匹配，进而确立

京津冀养老科技技术的需求^[34]。He Xijun 等人^[35]利用 SAO 结构识别了新能源领域潜在的技术需求热点,并通过使用 Word2Vec 和 HowNet 计算 SAO 结构之间的语义相似性来实现技术需求的聚类。程力培等人^[36]通过对京津冀地区在 2010-2016 年间的专利转让信息进行文本处理,得到技术关键词,并通过聚类分析识别热点需求技术。

此外,也有学者将双边匹配理论引入到技术供需匹配中。Gale 和 Shapley^[37]在研究关于大学生入学和稳定婚姻匹配决策时,提出了“双边匹配”的概念。双边匹配概念由 Roth 教授在 20 世纪 80 年代进一步明确^[38],理论最先应用于医疗,教育,婚姻等领域。近年来,供需匹配问题的研究也应用到实际问题中,如医疗服务供需双方匹配问题^[39]、就业网站的职位信息及求职者评价信息的匹配问题^[40]、跨境商务问题^[41]、出租车与乘客匹配问题^[42]、供应链中供需双方合作问题^[43]、风险投资机构与创业企业的信息匹配问题^[44]等。双边匹配理论也应用于研究技术供需匹配问题,Brownlie 等人^[45]在研究中发现过于关注技术供给而忽略技术需求可能会导致产品导向的问题,故提出了供需匹配的管理思维。Klerkx 等人^[46]提出了知识中介管理模式和技术服务中心的农业知识供需匹配模式。Hung^[47]提出建立案例与知识管理结合的供需匹配机制,并提出要建立知识库。Den^[48]提出了商业模式下的技术供给要素。何喜军等人^[49]提出融合供需文本词频特征、相关性特征和语义特征的匹配模型,通过采集在线技术转移平台(技 E 网)数据,并针对京津冀技术需求进行供给匹配。张婷婷^[50]基于技术供需匹配视角下,构建了技术需求点的筛选模型进行技术筛选,并构建双层聚类模型对企业技术需求进行识别。宋志红等人^[51]通过对 1995 到 2011 年间索尼公司的技术联盟网络进行分析,研究公司的创新能力。陈希等人^[52]提出了两阶段决策分析方法,通过建立满意度评估机制对供需技术主体进行分析。张娴等人^[53]通过建立知识组织模式来解决技术研发与产业需求的信息不对等问题。

1.2.4 多源数据融合的研究现状

针对产业需求问题,用户的需求散布在海量的互联网数据种,使用单一数据源进行相关问题的研究是片面的,存在一定的不准确性。军事领域是最早应用数据融合理论的,后来在传感器^[54]、地理空间^[55]等领域中得到了广泛应用,随着数据融合的普及,情报分析领域也逐渐将数据融合作为研究的数据应用^[56]。多源数据融合是将多种渠道获得的不同数据结构的信息融合在一起,形成统一的格式,并应用于多种方式的数据集合过程^[57]。随着信息挖掘领域的发展,学者发现使用多个数

据集将数据进行融合可以得到更高质量的挖掘结果，从而使合并后的挖掘效果尽可能全面且满足实际需求^[58]。Dong^[59]发现网络数据存在质量差异且数据源不同还可能导致冗余问题。Song^[60]为改善此问题，将使用多个数据抽取工具，从不同信息源中抽取相应数据进行集成。在多源数据的应用中，周群等人^[61]通过将多源文本数据进行综合，对科技部机构用户的重点关注领域主体进行分析，并通过计算需求主题的各文本的词向量，更精准的获取机构用户的科技决策需求主题。Fang Chen等人^[62]利用多源数据改进了对情感分类的方法。

1.2.5 研究现状评述

目前该领域研究虽然取得一些进展，但还处于起步阶段，存在许多需要进一步完善的地方。

对于老年需求的研究，从数据源来看，基本采用单一数据源，需求信息的获取主要来源于调查问卷或访谈数据，具有一定的主观性；也有学者利用了网络平台的数据，仅对养老科技产业某一细分领域进行了研究，复用性小；还有学者将专利数据与调查问卷相结合，研究养老科技领域需求，专利数据反映了技术供给，调查问卷反映了产品需求，两者结合覆盖面比较窄。从研究角度来看，对于老年相关的供需问题主要是通过宏观数据来表示供需平衡，缺乏从微观角度，即老年用户需求。

对于技术供需匹配的研究，主要利用专利数据及企业技术需求来进行研究，从方法层面，采用传统相似度计算等方法识别供需，从研究角度层面，主要研究企业技术需求，而本文研究产业技术需求，缺乏通过用户需求来挖掘技术需求部分，不能全面代表整个产业的技术需求。

整体来看，对于整个产业的供需问题，现有研究主要采用定性方式分别从产业的供给侧和需求侧进行分析。从数据源上看，数据源单一，主观性强；从方法上看，研究方法的复用性小；从研究角度看，缺乏从微观角度研究老年用户需求。

1.3 主要研究工作

主要研究工作主要包括研究内容、研究框架和研究方法三部分的内容。

1.3.1 研究内容

本文的研究对象是养老科技产业的技术供需匹配，包括基于多源数据挖掘养老科技产业的技术需求点，并构建技术供需匹配模型，面向需求识别技术供给现状，

最后对技术供求匹配进行评价，依据评价结果结果，为我国养老科技产业的技术研发提供决策建议。主要内容包括如下几个方面：

第1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；并根据本文主题详细地论述了国内外的研究现状，包括养老科技产业、老年需求挖掘、技术供需匹配和多源数据融合的研究现状，最后，针对现有研究存在的不足和局限性进行总结；最后介绍了本文的研究工作，包括研究内容，研究框架和创新点。

第2章，相关概念界定与研究方法。首先对本文涉及的概念进行界定，主要包括养老产业和养老科技产业的相关概念；其次，针对本文用到的网络数据进行了论述，包括现有研究中网络数据的应用及科学性，网络爬虫技术；最后，根据本文研究数据源的特殊性，主要介绍了在研究中用到的自然语言处理相关方法。

第3章，我国养老科技产业技术需求挖掘。首先通过政策文本、技术报告、文献数据构建技术词典；然后依据技术词典，对知乎、百度新闻等网络数据训练生成词向量模型，并计算相似度，获取技术需求点；最后对获取的技术需求点进行分析。

第4章，基于深度学习的养老科技产业技术供给识别。首先提出了基于 BERT 和 TF-IDF 算法的技术供需匹配模型，构建了输入层、特征提取层和输出层三层模型结构，通过该模型运算，可依据技术需求筛选出相应的专利技术——即技术供给。

第5章，养老科技产业技术供需匹配评价。首先利用 K-means 算法对第四章筛选出的技术供给点进行聚类分析，得到分类结果；然后构建技术供需匹配评价指标体系；最后对第3章的技术需求和第4章的技术供给进行匹配评价，分析评价结果，提出相应对策建议。

1.3.2 研究框架

根据研究内容，确立如下的研究框架（图 1-1）：首先基于多源数据挖掘养老科技产业的技术需求点；其次，构建技术供需匹配模型，识别技术供给；最后进行技术供需评价研究，分析我国养老科技产业的技术供需现状。

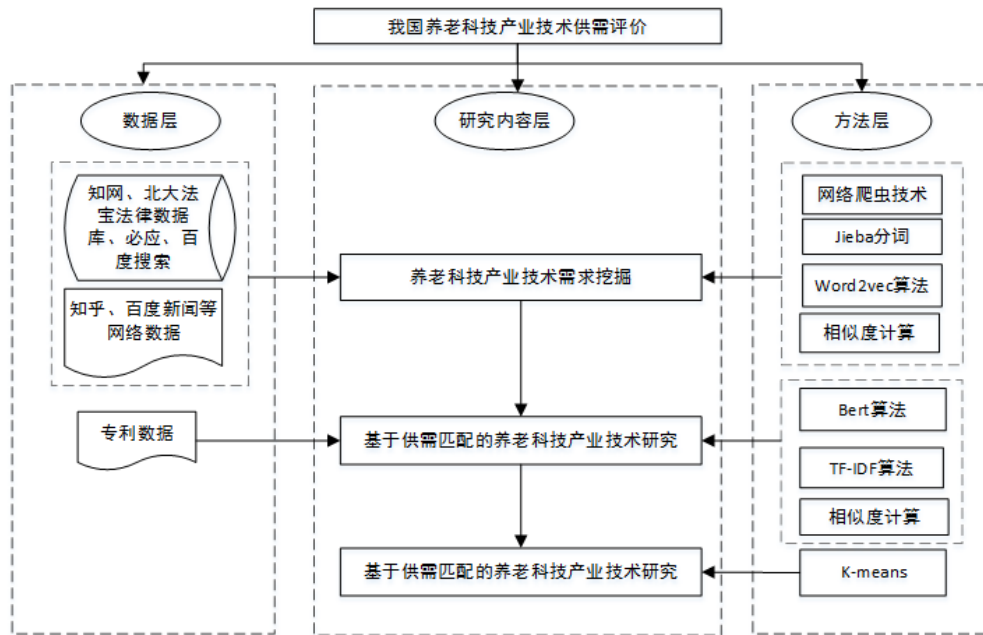


图 1-1 本文研究框架

Fig. 1-1 the research framework of this paper

本文的技术路线图如图 1-2。

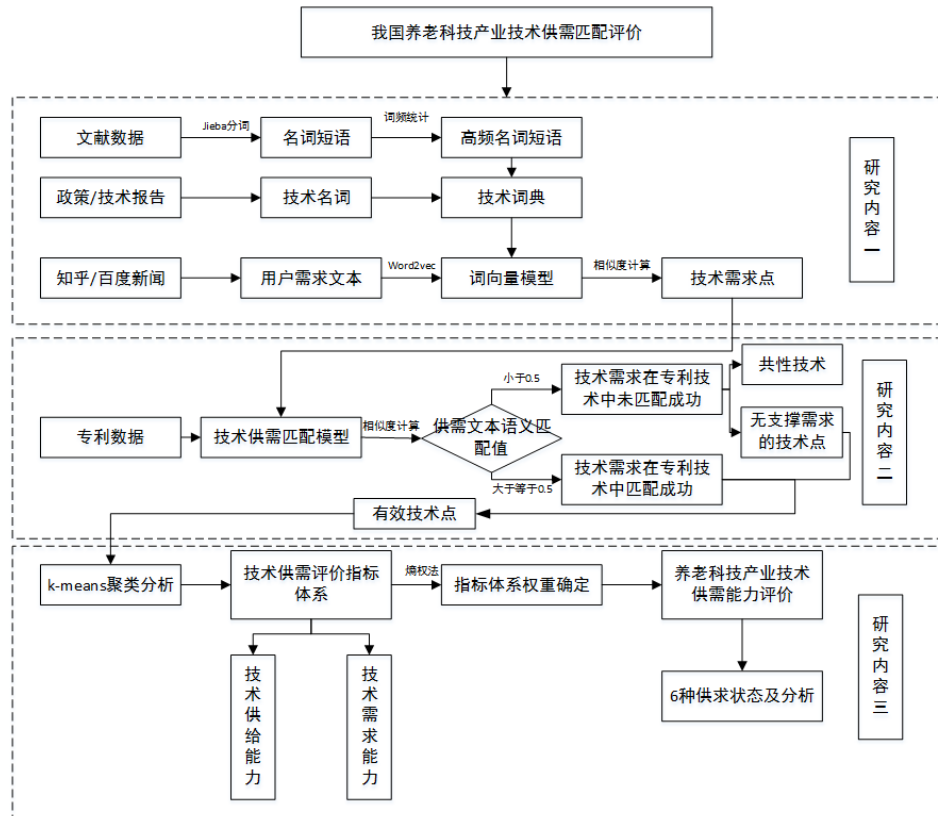


图 1-2 本文技术路线图

Fig. 1-2 the technical roadmap of this paper

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）利用多源数据挖掘养老科技产业的技术需求，避免了由于数据单一对需求挖掘的不全面性和采用调查问卷或访谈等定性方法对需求分析的主观性，通过多源数据融合，更加全面客观的识别出养老产业技术需求，且从数据源来看更为合理，能真实反映出养老科技产业的技术需求，从而对我国养老科技产业技术发展起到一定的指导作用。

（2）构建了基于 BERT 和 TF-IDF 算法的养老科技产业技术供需匹配模型，通过匹配模型建立技术需求和技术供给的联系，不仅能更加准确、快速识别出技术供给，也为研究技术供需匹配提供了新思路。

（3）构建了养老科技产业技术供需评价的指标体系，评价了我国养老科技产业的技术供求匹配现状，由此判断出养老科技产业的发展能力状态，也明确了养老科技产业技术研发方向。

第2章 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。首先，明确新兴养老科技产业的相关概念，包括养老产业和养老科技产业；其次，介绍网络数据的应用和科学性，以及网络爬虫技术方法；最后介绍本文用到的自然语言处理技术方法。

2.1 相关概念界定

2.1.1 养老产业

养老产业（Aging Industry）的服务对象为“夕阳”人群，即老年人群^[63]。老龄产业在1997年正式提出，是指为能够满足老年人衣食住行及精神文明等的产品、设施及服务^[64]。养老产业包括了第一、二、三产业，该产业是包含了关于老年人在物质和净身高层面的产品和服务，是由老年消费市场需求拉动的新兴产业^[65]。也有学者认为养老产业是以老年人为服务对象，为其提供产品和服务的产业部门总称^[66]。陈友华认为养老产业具有特殊性、综合性、市场性、复杂性、营利性等特点^[67]。杨立雄等认为养老产业包括养老服务、老年用品、老年健康、养老金融、老年旅游及地产等内容^[68]。此外，欧洲委员会在2007年提出了建设“银发经济”，指出“银发产业”是基于社会文化、基础设施和护理等方面的一种产业形态^[69]。由此，“银发经济”是以创新和服务进行的经济活动，同时还可以满足老年用户相关产品和服务需求的产出^[70]。

2.1.2 养老科技产业

王波^[71]分析了国际养老产业的发展趋势，提出了我国需要以产业化思维推动养老服务模式的创新。邵其会^[72]也提出养老模式和养老市场消费需求的变化，也促进了养老产业的融合，其中包括了养老产业和高新技术产业的融合。这也合理的解释了养老科技产业的产生，在新兴科技快速发展的背景下，对养老产业具有颠覆性影响，必将推动养老产业的结构的变化，进而形成养老科技产业。

老龄社会的发展和信息技术的进步，未来养老科技领域将成为一个很有潜力的市场，从而反过来正面影响老龄化社会的各个方面^[73]。养老科技这一术语结合了“老年医学”和“技术”，并在1991年正式提出“养老科技”的概念。养老科技作为跨学科的科技工作，是指利用技术在老年领域中的创新活动^[74]。

Thomas 等人认为养老科技（Gerontechnology）是指能够满足老龄化社会需求的

相关技术^[75]。国内马俊达等人将养老科技定义为能够满足老年人日常生活需求，以满足老年人健康和生活质量的跨学科科技工作^[76]。

随着养老科技的广泛应用，形成了具有鲜明特色和应用领域的产品及服务，这就是养老科技产业^[77]。其形成原因可以分别从供需两个角度来看。从供给角度来看，养老科技领域的快速发展奠定了理论基础，而新兴科技的快速发展奠定了技术基础。此外，社会各角色都开始关注养老科技领域，尤其是企业在各个领域的布局，促使了养老科技产业逐渐形成。从需求角度来看，老龄化程度的加深引发了一系列社会问题亟待解决，依靠科技创新应对人口老龄化成为了重要途径^[78]，为养老科技产业提供了契机。同时，老年人在物质和精神文化上的需求也越来越多，通过科技产品或服务满足其需求成为一种有效的方法。

本课题认为养老科技产业是由老年消费市场需求增长带动的，依靠养老科技来满足老年需求的新兴产业。养老科技产业不同于老龄产业，更加强调“科技”，养老科技产业是指以养老科技为基础，为老年人提供产品和服务的产业，其特征为：（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。

2.2 网络数据及爬取方法

2.2.1 网络数据应用

2020年9月29日，据CNNIC机构统计，截止到2020年6月，中国互联网网民规模已经达到9.4亿，互联网的普及率为67%，60岁及以上网民群体占比为10.3%^[79]。互联网的快速发展逐渐占据了人们的日常生活和学习，如维基百科及百度百科等各类知识共享平台，博客和Google Answer等内容社区，Twitter和知乎等社交网站，职业社交网站LinkedIn及学术社交网ResearchGate等^[80]。面向人们越来越多的需求，多样化的网络提供了更多的可能性，使得个人或组织愿意将更多的时间投入到网络中，进行社交、沟通、共享、搜寻搜集信息^[81]。在这个过程中产生了海量的网络数据，这部分数据逐渐受到科学研究者的关注，为科研工作提供了更有效的数据源。大量的社交网络数据中暗含了舆情事件的过程，基于海量的网络大数据，通过合理方式获取更有价值的信息已经成为了现阶段研究需求的重要方式^[82]。

此外，截止2020年6月，我国网络新闻用户规模为7.25亿，占网络用户整体的77.1%。网络媒体颠覆了传统媒体的形式，扩大了资讯可触达的范围，更加提升了新闻的传播效果^[83]。为了吸引用户，网络媒体内容更多的满足用户

的信息需求。因此，通过对新闻报道内容的分析，可以对公众的信息需求有一定的了解^[84]。移动互联网的发展带动了社交网络的快速发展，涌现了各种网络社交平台。在这种背景下，综合社交平台不仅可以满足大众的基本社交需求，还通过更丰富的维度更多细分群体，从而能够增加其归属感和认同感。细分社交网络平台将能更凸现出用户的典型需求，用户价值的挖掘能力将更强，信息也更有效^[85]。Eric von Hippel 等人也表示在互联网环境中，人人都可以发布信息，分享想法，通过挖掘网络信息内容，可以获取已知或者未知的需求^[86]。

2.2.2 网络爬虫技术

网络数据的获取是研究用户需求的基础，从海量网络数据中获取信息仅仅通过人工是不现实的，本文利用爬虫技术对网络数据进行爬取。网络爬虫框架如图 2-1 所示，目前常用的有很多，如 Beatiful Soup、Scrapy、Selenium 等。网络爬虫流程主要有通用网络爬虫和聚焦网络爬虫两种。

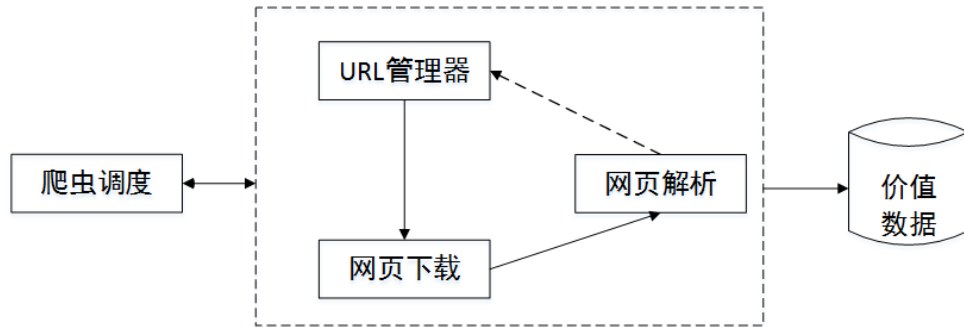


图 2-1 网络爬虫框架

Fig. 2-1 Crawler framework

通用网络爬虫得到的数据量最大，范围也是相对较广的^[87]。如图 2-2 所示，这种方式的实现原理是相对容易的，具体为：（1）确定初始 URL，加入 URL 等待序列中。读取 URL，DNS 解析，然后将提取的未经过爬取的 URL 放入等待序列种；（2）不断的从等待序列种选择 URL 重复第一步进行爬取；（3）直到强制停止或者等待序列中的 URL 全部爬取完，停止爬虫。

聚焦网络爬虫，顾名思义，这种爬虫比通用网络爬虫更具有目标性，选择性的抓取特定主题内容，因而整个实现流程较为复杂。如图 2-3 所示，具体实现流程为：（1）确定初始 URL，加入 URL 等待序列中。获取该 URL 中的内容，在获取网页内容的过程中会将与需求信息不相关的内容进行过滤、去重等操作；（2）根据爬取任务的需求构建 URL 的等待序列，并重复上述过程；（3）直到强制停止或者等待序列中的 URL 全部爬取完，停止爬虫。

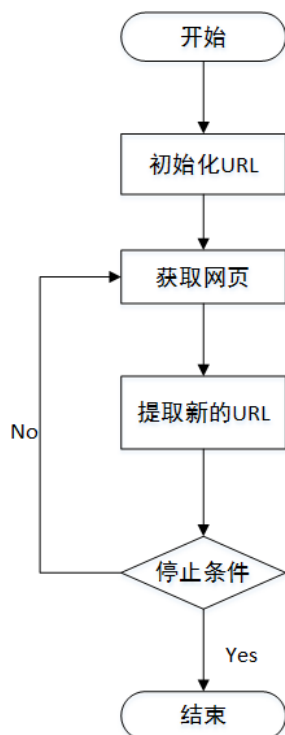


图 2-2 通用网络爬虫工作流程

Fig. 2-2 General web crawler workflow

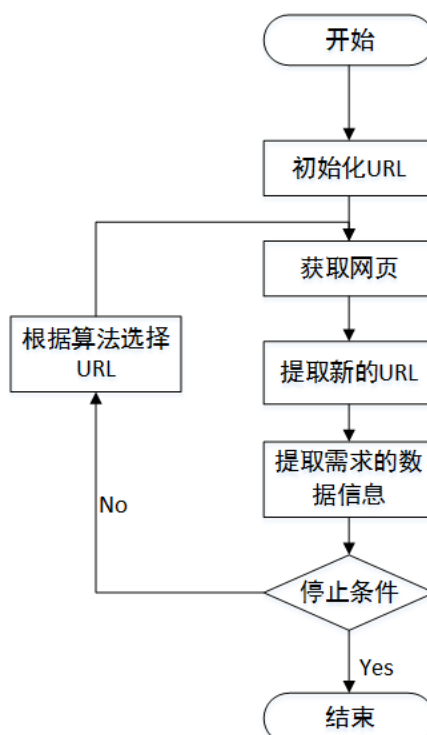


图 2-3 聚焦网络爬虫工作流程

Fig. 2-3 focus on web crawler workflow

2.3 自然语言处理方法

对于网络数据的应用，必然会涉及到自然语言处理，主要包括文本预处理，文本分布式表达及文本相似度计算，本节也将重点论述这三个部分，以便更好的理解文章内容。

2.3.1 文本预处理

本课题利用网络数据获取用户需求，是基于中长文本进行研究，文本预处理将是研究内容的第一步，影响整个自然语言处理流程的准确性。文本预处理包括文本正则化、文本分词、词性标注、去除停用词四个步骤。

2.3.1.1 文本正则化

网络数据存在比较大的噪声数据，在处理网络数据这种非结构化文本时需要进行文本清洗，避免噪声数据对文本内容的影响，主要包括表情符号、年月日、网页解析过程中带来的部分特殊符号等。本文主要利用 python 程序中的 re 正则表达式对文本清洗。

2.3.1.2 文本分词

自然语言处理中，中文跟英文有所不同，中文的基本表达单位是由字连接组成的词语，中文分词的目的就是将非结构化的句子按照既定的逻辑进行拆

分，所以中文文本分词时是以词为单位进行拆分的。目前较为成熟的技术包括基于字符串匹配、基于统计、基于语义的分词技术^[88]。

（1）基于字符串匹配的分词技术

基于字符串匹配的分词技术是根据一定的规则将文本的句子和词典进行匹配，如果发现句子中的词语与登录的词语一致时，依次将匹配成功的词组切割并记录。一般匹配规则包括最大正向匹配、最大后向匹配、最少词语切分法。该方法具有易于实现、切分速度快、使用广泛等优势。但由于需要将句子与词典进行匹配，对语料库具有较强的依赖性，需要及时更新语料库，人工维护成本大。此外，这类分词技术具有较弱的语义识别能力，不能准确识别句子中的歧义问题。因此，在实际应用中，该方法一般用于对文本的第一次分词。

（2）基于统计的分词技术

基于统计的分词技术是对相邻共现的汉字组合进行频率统计，汉字组合的共现频率越高，这个组合成为一个词语的概率越大。常用的分词模型包括最大概率、N-Gram 元、最大熵等分词模型。相比于字符串匹配的分词技术，这种方法不需要维护语料库，同时又能很好的识别新词、消除歧义。但需要大量语料来训练模型，而且有一些虽然共现频率高，但并不是一个词语。所以，通常将第一种和第二种分词方法进行结合使用，达到分词速度快、分词精度高的双优势。

（3）基于语义的分词技术

基于语义的分词技术解决了词典和歧义问题，融合了文本句法和语义分词，通过对大量语料的训练来进行分析。在三种分词技术中分词效果最佳，但由于网络用语变化太快，需要不断迭代模型，同时复杂中文语句也难以组织成计算机直接可读的形式，所以该方法仍在探索阶段。

目前，分词系统将这些方法综合起来进行分词。在国内较为成熟的几个中文分词系统包括结巴（jieba）中文分词、NLPIR 汉语分词系统、哈工大 LIP 平台等，本文采用的结巴中文分词。结巴（jieba）分词使用广泛，主要包括精确模式、全模式、搜索引擎模式、paddle 模式四种分词模式^[89-90]。文本使用的结巴分词是精确模式。

2.3.1.3 词性标注

对文本的处理还需通过一定的语法规则对分词后的每个词的词性进行标注。结巴分词中部分词性对照表如表 2-1 所示。主流的词性标注方法包括基于规则和基于统计的两种词性标注方法^[91]。

（1）基于规则的词性标注方法

该方法是提出较早的词性标注方法，基本思路是对需要进行词性标注的语料根据规定的词性进行标注，标出所有可能的词性，也称为兼类词，再结合上下文的语境和搭配关系建立词性消歧规则，最终通过规则集的筛选确定词性。最早的词性标注规则是由人工制定的，这也导致了更新不及时的问题。现在随着语料库规模的扩大，提出了机器学习的规则来自动提取方法^[92]。

（2）基于统计的词性标注方法

该方法的基本思想是：通过对输入语料的分词，给定所有可能的词性，通过概率统计的方法选出合理的词性。该方法表现出较佳的新词识别和消除歧义能力，是应用较为广泛的一种词性标注方法^[93]。

表 2-1 结巴分词部分词性对照表

符号	词性	符号	词性
a	形容词	d	副词
ad	副形词	c	连词
m	数词	e	叹词
n	名词	ng	名词素
nr	人名	ns	地名
r	代词	v	动词
w	标点符号	y	语气词

2.3.1.4 停用词

停用词一般指的是在文本中的词频很高，但是这类词单独拿出来又没有实际意思，仅仅在文本结构中起到连接和起承转合的作用，例如“地”，

“且”，“的”等，这类词对文本分类没有直接影响，且大多都是副词、介词、连词等。所以一般在自然语言处理中会剔除掉停用词，从而降低文本维度，提高文本处理效率。中文常用的停用词表有哈工大停用词、百度停用词、四川大学机器智能实验室停用词等。

2.3.2 文本分布式表达——词向量

文本表达中比较经典的模型为向量空间模型，该模型能够将长文本以最小的原子词语进行划分，因此可以很好的适应长文本。但对于标题摘要等短文本。本文中包含的词语较少，且少数几个词包含大量信息，以原子为单元没有办法将文本语义全面的表达出来，因此短文本使用分布式表达能够为文本表达提供新的思路。

分布式表达由 Hinton 在 1986 年提出：文本语言的处理应该先将文本进行符号化，以转变成机器语言^[94]。同时也引入了词向量这一概念。词向量是指通过对大量语料的模拟训练，将文本的每个词语转化为固定长度的向量值^[95]。自然语言处理最常用的词语表示方式是 One-hot Representation^[96]。One-hot Representation 虽然便于理解，使用也简单，但是存储方式不太友好，因此，在实际应用中一般配合其他算法来使用。其存储缺陷主要体现在维度灾难^[97]、词汇鸿沟两个方面。

词向量也可以作为语言模型的中间转换的产物出现，语言模型不断迭代优化也使得词向量的产生方式在不断变化。Xu^[98]提出了一种将词向量作为中间产物的单层神经网络训练语言模型。Bengio^[99]构造了三层的前馈式神经网络的经典模型。Hinton^[100]等在基于经典模型进行了优化，进一步提升了模型在分类、速度等方面的性能。Mikolov^[101]等提出了一种利用上下文信息的循环神经网络模型。Huang^[102]提出了多向量代表多义词的方法。目前，影响最为广泛的词向量模型是 word2vec 算法^[103]，其运用深度学习的思想，可以表示基于单词及其上下文之间关系向量在语义上接近的单词^[104]。第三章将对 word2vec 算法进行详细讲解。

2.3.3 文本相似度计算

自然语言处理的核心部分就是文本相似度，该方法就是评估两个文本（包括词汇、句子等形式）之间的语义相似度，常用于信息检索、复述识别、问答系统等。常见的计算相似度的算法包括了最长公共子序列、欧式距离、余弦相似度^[105]。

（1）最长公共子序列

最长公共子串可以直接反应了文本之间的相似度，其原理时利用动态规划的思想，对两个词之间的最长公约字符子串的长度进行迭代计算，最终得到词语之间的相似度^[106]。

（2）欧式距离

欧式距离在传统聚类中最常使用到，在空间向量模型下，两个词分别用向量表示，通过计算平面上两点的距离，来计算相似度^[107]。计算公式见公式（2-1）：

$$O_d = \sqrt{\sum_{i=1}^n (O_{xi} - O_{yi})^2} \quad (2-1)$$

其中， O_{xi} 和 O_{yi} 表示词 x 和词 y 在空间向量中的向量。

（3）余弦相似度

利用余弦相似度计算文本的相似度是通过对空间向量中的两个词语向量之间的余弦夹角进行计算^[108]。计算公式公式（2-2）：

$$\text{sim}(D_i, D_j) = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (2-2)$$

2.4 本章小结

本章首先对养老科技产业的相关概念进行了界定和论述；然后对网络数据源的合理性、科学性进行了论证，并阐述了涉及的网络爬虫技术方法；最后对本文涉及的自然语言处理方法进行了说明，为后续的研究工作奠定了基础。

第3章 基于多源数据的养老科技产业技术需求挖掘

依据上文的研究基础，本章主要阐述我国养老科技产业技术需求的挖掘。首先，通过政策、技术报告和文献等多数据源构建较为完整的养老科技产业技术词典；然后对知乎、百度新闻的数据爬取，并通过技术词典进行相似度计算，挖掘潜在的技术需求；最后对技术需求进行分析。

3.1 技术需求挖掘研究框架

对我国养老科技产业技术需求的挖掘，其目的是为了了解公众对养老科技技术的需求。研究框架如图 3-1 所示，首先是技术词典的构建，通过 python 脚本实现对文献数据集的 jieba 分词处理，对分词之后的词性进行分析，过滤掉不属于技术名词的短语，并提取高频名词短语，结合政策、技术报告确定最后的技术词典；然后利用爬虫代码爬取知乎、百度新闻养老科技技术相关内容的数，同样进行 jieba 分词和词性分析，过滤掉无用词后，将分词结果作为 word2vec 的训练集，以技术名词作为参考词，挖掘出技术需求点；最后对技术需求点进行可视化分析。

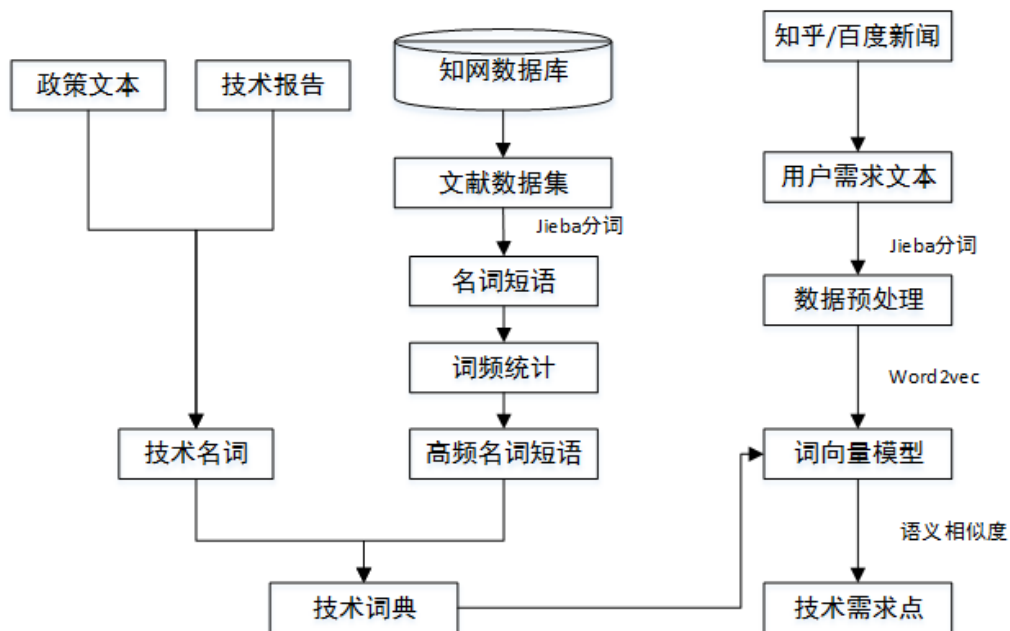


图 3-1 技术需求挖掘研究框架图

Fig. 3-1 research framework of technology requirement mining

3.2 数据来源

3.2.1 技术词典数据来源

(1) 文献数据

本文采用的文献数据的出处为 CNKI 数据库，是我国最为权威、影响力最大的平台^[109]，因此本文构建中文技术词典选取知网具有一定的合理性。根据本文的分析主题是“养老”“科技”，并对相关词进行扩充，检索表达式为：SU='养老科技'+ '养老技术'+ '养老创新'+ '养老产品'+ '智慧养老'+ '老年科技'+ '老年技术'+ '老年创新'+ '智慧老年'+ '老年产品'，通过检索表达式进行全时段检索，检索时间为 2020 年 9 月 20 日，最终共得到 4567 条数据，对数据进行二次筛选，最终共得 3787 条。

(2) 政策数据和技术报告数据

本文的政策数据来源于北大法意网，根据本文主题，检索词为：“养老”，“老年”。技术报告数据来源于网络检索，根据本文分析主题，并进行相应的扩充，由于英文技术报告居多，因此，使用必应搜索结合百度搜索两个工具，英文检索词为：“gerontechnology”，中文检索词为：“养老科技”，“养老技术”，“养老创新”，“养老产品”，“智慧养老”，“老年科技”，“老年技术”，“老年创新”，“智慧老年”。

3.2.2 技术需求挖掘数据来源

本文选取知乎和百度新闻中的数据对其进行分析。随着互联网的普及，越来越多的人通过网络形式表现自己的需求，其中网络问答平台提供了为公众解决相关问题和需求的平台，因此，通过对相关平台中国年的“养老科技”话题进行挖掘，以获取养老科技需求的有价值信息。本文选取了国内最大的网络问答平台——知乎作为主要数据源之一，在该平台上用户可以平等的自主的进行提问或者对某些问题进行回答，进而将知识和经验可以被更多人学习和了解到，让知识共享变得更加易得^[110]。截至 2017 年 9 月，其个人注册用户总数超过 1 亿，日活跃用户量达 2600 万。人均日访问时长 1 小时，月浏览量可达 180 亿次^[111]。曹雨骋等^[112]对科普网站进行研究发现，知乎对于公众科技传播比中国科普类网站更具影响力。因此，本文选取了知乎问答数据作为一部分网络信息的数据源。

此外，新闻报道内容反映了网络媒体对相关信息得关注度，是公众接受信息得重要途径，对公众得信息需求有一定得了解^[113]。百度新闻收录了整个各大

新闻资讯网站的相关新闻评论文章，通过对百度新闻内容得分析，能在一定程度上反映公众对相关内容得关注与需求

该部分数据获取步骤为：第一步，利用关键词“养老科技，养老技术，养老创新，养老产品，智慧养老，老年科技，老年技术，老年创新，智慧老年，老年产品”等从知乎爬取评论 87703 条，从百度新闻中爬取新闻 4618 条；第二步，对数据进行二次筛选，最终共得到知乎网络数据 73621 条，百度新闻数据 3120 条。

3.3 养老科技产业技术需求挖掘

3.3.1 文本数据处理

本章利用文本数据挖掘养老科技产业技术需求，在构建技术词典和挖掘技术需求时都需对长文本数据集进行数据处理，结合各分词工具特征，选取了 jieba 分词进行对数据集进行分词，利用 python 程序实现，具体步骤如图 3-2 所示：

首先，jieba 分词对养老科技产业技术需求中网络数据集的文本句子根据词典扫描，也就是句子清洗的过程，由于处理的是网络数据，存在部分特殊字符，在清洗过程中，将这些特殊字符进行分离并标注成未知词性；接着为了实现高校的词图扫描，采用了 Trie 树结构，根据最终生成的句子成词情况，构成 DAG 词图，通过全局概率路径的计算情况得到最大的词频切分组合；其中登录词直接按照词典标识输出，未登录词识别英文和中文部分，英文、数字字典组合给予相应标注，直接输出，中文部分 HMM 模型和 viterbi 算法动态规划取得分词和标注；最后根据词性对分词结果完成相应处理，技术词典部分统计词频，提取高频技术名词，技术需求部分仅保留名词短语，作为相似度计算的训练集^[114]。

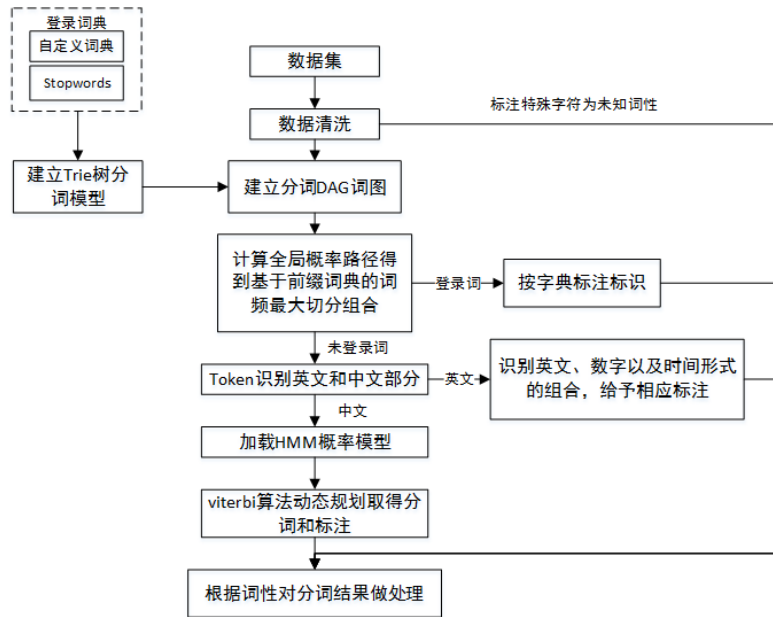


图 3-2 文本数据处理过程

Fig. 3-2 text data processing

3.3.2 技术词典构建

技术领域词典是指记录特定技术领域专有词汇或者术语的词典^[115]。本文技术领域词典中的专有词汇主要从研究文献中的摘要、政策文本和技术报告中得到。

对于文献文本，根据 jieba 分词结果，利用 python 程序对名词短语统计词频并提取高频技术名词 523 个。此外，技术词典的构建还结合一些关于智慧养老的政策文件（表 3-1）、技术报告。最后技术词典共获得 769 个技术点。

表 3-1 养老科技相关政策文件

Tab. 3-1 policy documents related to pension Technology

序号	时间	名称	制定单位
1	20170206	智慧健康养老产业发展行动计划 (2017—2020 年)	工业和信息化部、民政部、国家卫生计生委
2	20180731	智慧健康养老产品及服务推广 目录(2018 年版)	工业和信息化部、民政部、国家卫生计生委
3	2019920	民政部关于进一步扩大养老服务供给 促 进养老服务消费的实施意见	民政部

3.3.3 技术需求挖掘

技术需求的挖掘是依据 3.3.2 得到的技术词典，将爬取的知乎、百度新闻数据进行处理后的分词结果去掉词性标注作为 word2vec 算法的训练集。利用 word2vec 算法对技术需求文本和技术词典分别进行词向量训练，根据得到的词

向量计算相似度，最终得到技术需求点。如图 3-3 所示，该模型由输入层、投影层、输出层三层网络结构组成^[116]。具体分为三步：

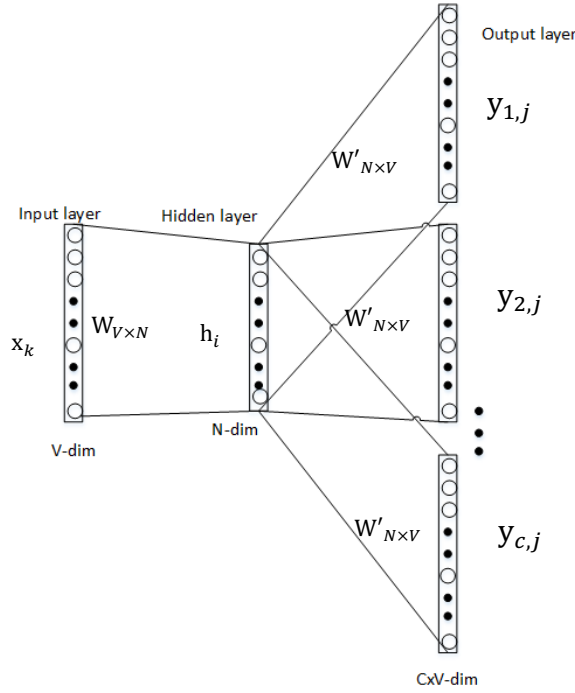


图 3-3 skip-gram 算法模型结果

Fig. 3-3 results of skip gram algorithm model

第一步，前向传播。（1）从输入层到隐藏层，公式（3-1）中 W 代表输入层到隐藏层权值向量， x 表示输入的单词向量， h 表示隐藏层向量 W^T 和 V_{wI}^T 表示输入节点为 1 所对应的向量。

$$h = W^T \cdot x = V_{wI}^T \quad (3-1)$$

（2）从隐藏层到输出层，公式（3-2）中 V_{wj}^T 表示隐藏层到输出层节点 j 的权值， $u_{c,j}$ 表示输出层第 j 个节点的输出值。

$$u_{c,j} = u_j = V_{wj}^T \cdot h, (for\ c = 1, 2, \dots, C) \quad (3-2)$$

（3）输出每一个节点的概率值， C 表示上下文 context 的数量。

$$p(w_{c,j} = w_o, c|w_I) = y_{c,j} = \frac{\exp(u_{c,j})}{\sum_{j'=1}^V \exp(u_{j'})} \quad (3-3)$$

第二步，构造损失函数，使上下文 context 所对应的多个节点最大化,输出单词组的条件概率。

$$\begin{aligned} E &= -\log p(w_{o,1}, w_{o,2}, \dots, w_{o,c}|w_I) \\ &= -\log \prod_{c=1}^C \frac{\exp(u_{c,j_c^*})}{\sum_{j'=1}^V \exp(u_{j'})} \\ &= -\sum_{c=1}^C u_{j_c^*} + C \cdot \log \sum_{j'=1}^V \exp(u_{j'}) \end{aligned} \quad (3-4)$$

第三步，反向传播更新参数，对公式（3-4）的概率求导，得到权重矩阵 W' 的更新规则为公式（3-5）：

$$w'^{(new)} = w'^{(old)} - \eta \cdot \sum_{c=1}^C (y_{c,j} - t_{c,j}) \cdot h_i \quad (3-5)$$

同理，权重 W 的更新规则为公式（3-6）。

$$w^{(new)} = w^{(old)} - \eta \cdot \sum_{j=1}^V \sum_{c=1}^C (y_{c,j} - t_{c,j}) \cdot w'_{ij} h_i \quad (3-6)$$

以上过程通过 python 程序实现，模型设定的迭代次数为 5，词频少于 1 次就被丢弃，当前词与预测词在一个句子中的设定的最大距离为 5，通过训练的词向量对技术名词进行扩展，设置语义相似度的阈值为 0.9，得到了 1145 个技术领域词典中技术名词的相似词，即为养老科技产业技术需求点。

图 3-4 为知乎数据最终处理得到的结果，出现关键词的字号越大，则表明该关键词出现的频率越高，用户的关注度越高，代表用户的需求也越大。从图中可以看出，当前中国养老科技产业技术需求主要为物联网、传感器、大数据技术、定位技术、云服务等技术，形成养老智能化、自动化。



图 3-4 养老科技技术需求词云图

Fig. 3-4 cloud chart of technology demand words for the elderly

3.4 本章小结

本章主要利用多源数据对我国养老科技产业的技术需求进行识别，首先通过政策文本、技术报告和文献数据构建技术词典；然后依据技术词典，对知乎、百度新闻等网络数据训练生成词向量模型，并计算相似度，获取技术需求点；最后对获取的技术需求点进行分析，我国养老科技产业技术需求主要为物联网、传感器、大数据等技术。

第4章 基于深度学习的养老科技产业技术供给识别

第3章挖掘出养老科技产业的技术需求，本章则是识别满足技术需求所需要的技术供给，以此为第5章的养老科技产业技术供需匹配评价奠定基础。首先，基于BERT模型和TF-IDF算法构建技术供需匹配模型，通过该模型可在专利数据中识别出支撑技术需求（第3章）的技术——技术供给点，并为后续养老科技产业的技术供需评价提供支持。

4.1 养老科技产业技术供需匹配研究框架

如图4-1所示为本章研究框架，主要通过对技术供需匹配的研究，来进一步筛选相关技术供给点。本章将基于文本特征构建了技术供需匹配模型，使用BERT和TF-IDF算法相结合，提取文本的句子特征。然后，将提取的文本特征进行相似度计算，得到相应的相似度分值。最后设定阈值，对技术供给点进行分析。

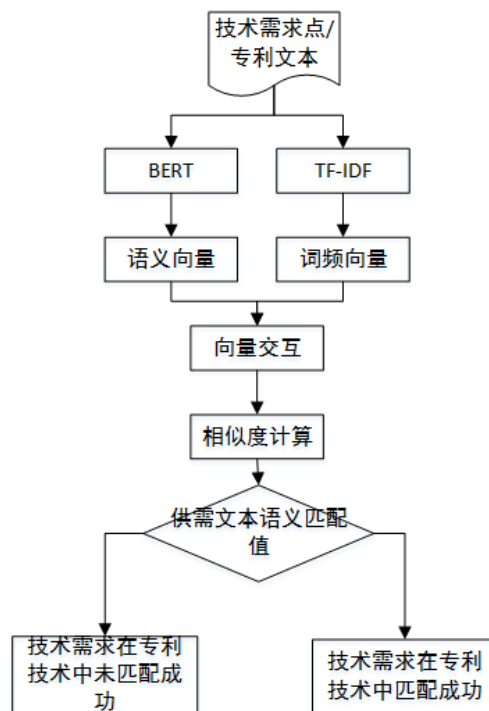


图 4-1 养老科技产业技术供需匹配研究框架

Fig. 4-1 research framework of technology supply and demand matching of pension technology industry

4.2 数据来源

本文的专利数据来源于 INCOPAT 专利数据库^[117]，该数据库提供中文的专利信息，有利于构建中文技术词典并用于分析技术需求；根据本文的分析主题是“养老”“科技”，提取检索表达式的关键词为养老、老年人、科技、创新、产品等，并对相关词进行扩充，据此得到专利检索表达式为：

TIAB=((("aged" OR "ageing" OR "aging" OR "elder" OR "older" OR "senile" OR "senior") NEAR/0 ("people" OR "human\$" OR "adult\$" OR "person\$" OR "citizen\$" OR "individual\$" OR "men" OR "women" OR "male\$" OR "female\$" OR "population" OR "patient\$" OR "societ*") OR ("population" OR "healthy" OR "societ*") NEAR/0 ("ageing" OR "aging") OR "aged care" OR "old age" OR "age*-friendly" OR "late-life" OR "later life" OR "silver economy" OR "active ageing" OR geriatric* OR geronto* OR elderly) AND (technique* OR technolog* OR innovat* OR product* OR facilit* OR device* OR apparatus* OR "educational equipment*" OR tool* OR solution* OR goods)) OR gerontechnolog*)。

由于本文研究中国养老科技领域，选定中国发明申请作为本文研究对象，最终共得到 4410 条专利数据（检索日期为 2020 年 10 月 23 日）。

4.3 面向技术供给识别的匹配模型构建

4.3.1 技术供需匹配模型结构

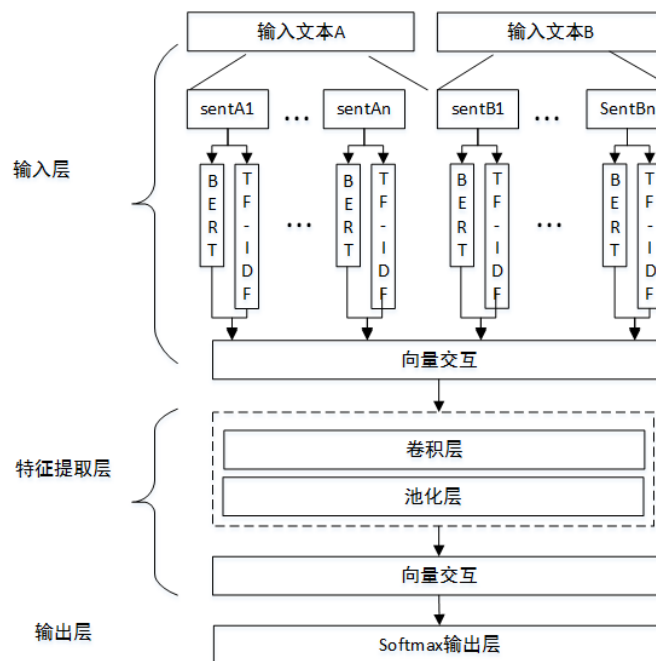


图 4-2 技术供需匹配模型结果

Fig. 4-2 results of technology supply and demand matching model

如图 4-2 所示, 将技术供需匹配模型分为了输入层、特征提取层和输出层。在输入层中应用了 BERT 模型和 TF-IDF 算法相结合的方式提取文本特征, 其主要原因为专利摘要属于长文本, 相对于短文本特征, 这类文本的特征比较模糊, 在进行文本匹配时, 需要考虑深层次的语义特征, 即需要考虑上下文关系及整体文本的逻辑关系, 故使用 BERT 模型提取语义特征, 文本采用无监督的方法训练模型, 这样可能会导致某些文本特征不明确, 因此, 加入了 TF-IDF 计算词频特征, 语义特征和词频特征相结合, 更好的表示文本特征。在特征提取层使用卷积和池化操作、向量交互提取句子特征。在输出层利用 softmax 函数输出结果。

4.3.2 基于 BERT 模型和 TF-IDF 的输入层

输入层的目的是将文本进行预处理并进行句子向量化表示。首先是使用 BERT 提取语义特征, 将文本进行断句, 输入到 BERT 模型得到句子向量表示结果; 然后使用 TF-IDF 提取词频特征, 得到句子词频向量表示结果。最后进行向量融合并做作为特征提取层的输入。

(1) BERT 预训练得到语义向量

BERT 模型是 2018 年由谷歌提出的一种多层双向 transformer 编码器^[118], 采用的是字符编码方式, 即模型输入是 token 级别的, 这种方式避免了由于分词不准带来的误差。其结构模型如图 4-3 所示, 其中输入序列的嵌入为 E , Transformer 的 Encoder 为 Trm , 词向量的输出为 T 。BERT 将通过对文本句子的对应标记嵌入、句子嵌入和位置嵌入进行求和来构造相应的输入表示, 并将其输入到双向 transformer, 最终输出层输出深层双向语义向量, 该向量与输入长度相等且融合了句子的语义信息^[119]。

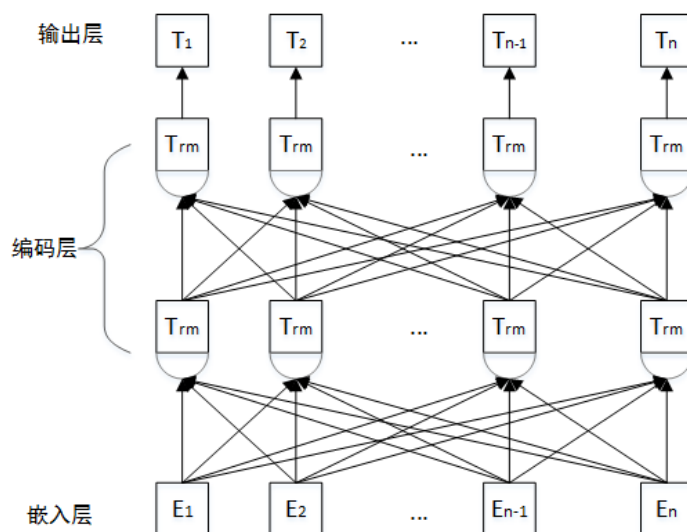


图 4-3 BERT 模型结构

Fig. 4-3 structure of Bert model

目前 BERT 模型的两模型如表所示：

表 4-1 BERT 模型

Tab. 4-1 Bert model

模型名称	隐藏层层数	隐藏层大小	注意力头数	参数总量
BERTBASE	12	768	12	110M
BERTLARGE	24	1024	16	340M

本章使用的 BERT 预训练模型是基于中文版 BERT 模型进行了微调。微调是指将经过海量语料训练过的模型应用于不同的 NLP 任务中，只需要将模型根据任务特性进行调整即可，具体模型过程为：

在嵌入层，首先要将句子的开头嵌入特殊分类标记 “[CLS]”，该标记一般用于分类任务中，因此本模型中该标记对应的 Transformer 输出被忽略掉。句子结尾嵌入特殊标记 “[SEP]”。然后根据字典将句子的文本信息转换为相应的编号，本模型的字典参照了谷歌发布的 vocab 字典，该字典是利用大量的文本数据构造的字粒度字典。对于在字典中找不到对应编号的字，标记为 “[unk]”，这样就得到了 Token Embeddings。最后根据 Token Embeddings 得到 segment Embeddings 和 Position Embeddings，对这三者进行求和，并输出。此外，该部分还另外将利用 TF-IDF 算法将词频信息也输出给下一阶段。

在经过输入层之后，进入 BERT 的编码层，每一个编码层由 Transformer 构成，又分为了注意力层、中间层和输出层三层。三层具体的结构如下：

注意力层采用了 12 heads 的 Multi-Head Attention。对于每一个 head 来讲，都是利用 query、key 和 value 的权重矩阵来得到相应的向量，并将 query、key 进行相乘并放缩，得到输出输出给下一层的矩阵。在数据处理过程中，由于各序列的长度不想等，还需要对序列进行截断操作和补齐操作。此外，在随机 mask 阶段引入了 input_mask 变量以消除该阶段可能对注意力计算的影响。因此最终注意力分数由权重矩阵与 input_mask 相加后于 value 矩阵相乘得到。

经过注意力层后，中间层将其输出接入到了全连接层，并通过 GELU 函数得到输出，GELU 激活函数具体见公式 4-1，其中， x 是变量， $P(X \leq x)$ 决定 x 中有多少信息保留下来， X 服从高斯分布。

$$\text{GELU}(x) = xP(X \leq x) \quad (4-1)$$

最后将中间层的结果传给输出层，输出层依然要经过三层，先是连接层，然后为了降低由于编码层数量增加导致的过拟合风险，要经过一层 Dropout 层，最后通过 Norm 层就能够得到整个 Transformer 的输出。

经过上述步骤，得到每个编码层的输出，最终得到 BERT 模型的输出。

(2) TF-IDF 计算词频向量

在 BERT 模型的输入层嵌入了 TF-IDF，在得到词向量的同时，计算对应的词频向量。当一个特征项在文本中出现的次数越少，表示能力越强；反之越弱。通过计算词频向量，同语义向量融合，解决由于本模型无监督学习可能会造成的特征不明确问题。

TF-IDF 权重算法^[120]由词频（TF）和逆文档频率（IDF）两部分来组成的，其中 TF 表示特征项在文档中出现的频次，IDF 表示特征项在文档中的分布，将两者结合能够同时满足特征项对文档的表示和将不同类别文档分开的需求^[121]。计算公式见公式 4-2。

$$w_{dt} = \frac{tf_{dt} * \lg(\frac{N}{n_t} + 0.01)}{\sqrt{\sum_{p=1}^K [tf_{dp} * \lg(\frac{N}{n_p} + 0.01)]^2}} \quad (4-2)$$

中各变量表示如下，t：特征项，d：文本， w_{dt} ：t 在 d 中的权重， tf_{dt} ：t 在 d 中出现的次数，N：语料库中 d 的总数， N_t ：文本语料库中包含 t 的文本数，K：文本 d 中 t 的个数。

4.3.3 特征提取层和输出层

在特征提取层主要通过卷积层和池化层对特征进行提取，然后再对得到的向量进行相似度计算。

首先经过卷积层^[122]对各序列进行编码，对句子的编码方式和 TextCNN 网络相同，使用的卷积核高度变化范围为[2,6]，宽度设置为字向量的维度，即 34*768。根据参数设置对句子进行一维卷积，得到特征图，并输入到池化层。该层使用 k-max pooling 的方法进行池化^[123]，池化窗口大小设置与特征向量是相等的，池化层中的 k 值是根据不同层数在动态变化的，具体变化函数见公式 4-3

$$k_l = \max(k_{top}, \frac{L-l}{s}) \quad (4-3)$$

各参数分别表示为： k_l 是第 l 层池化的最大值个数，s 是句子长度，L 代表总得卷积层数，l 表示当前池化是第几层， k_{top} 是最上层的卷积特征图需要通过池化保留多少个最强特征，通过公式可以看出越靠近顶层池化提取的特征最多，越到后面层提取的特征变少，这里 k_{top} 取 2。池化后得到单个句子的编码向量。

在特征提取层分别得到技术需求点和专利文本的特征向量，将两个特征向量进行拼接，最后通过 softmax 函数得到句子对相似概率。

4.4 技术供需匹配模型算法实现

技术供需匹配模型使用 python 编码实现，运行环境要求如表 4-2 所示。输入层，本文采用的 BERT 模型是谷歌预训练好的中文模型，该模型通过大规模的语料已经完成预训练阶段，通过加载该模型对文本进行训练。

表 4-2 供需匹配模型算法运行环境要求

Tab. 4-2 operation environment requirements of supply and demand matching model algorithm

工具	版本
jieba	0.42.1
Keras	2.3.1
numpy	1.18.5
scikit-learn	0.23.2
scipy	1.4.1
tensorboard	2.3.0
tensorflow	2.3.0
torch	1.6.0
transformers	3.0.2

文本提出的技术供需匹配模型具体的算法流程如下：

（1）输入的文本格式为一行为一个技术需求点或专利摘要，因此，对输入文本以一行作为一个句子的标志，得到文本集<data_sen1: x1 x2 ...xn-1 xn>和<data_sen2: x1 x2 ...xn-1 xn>。

（2）加载 BERT 预训练模型，将文本集分别输入到 BERT 训练模型中，生成一维的句子语义向量表示，集体的参数设置如表 4-3 所示。此外，根据 BERT 中的词典集字向量序列，利用 TF-IDF 计算词频向量，将一维的语义向量和词频向量分别组合成矩阵 U，然后进行相乘，最终得到的句子向量集作为文本的向量表示。

（3）将句子向量输入到卷积层和池化层学习局部特征，得到最终的特征向量。

（4）将两个文本集得到的特征向量进行点乘，并输入到 Softmax 函数中得到两个文本相似概率。

表 4-3 模型参数设置

Tab. 4-3 model parameter setting

参数	参数说明	参数设置
max_seq_length	输入文本序列的最大长度	200
train_batch_size	训练模型求梯度时的批处理大小	16
learning_rate	反向传播更新权重时，补偿大小	2e-5
num_train_epochs	训练次数	3.0

4.5 结果分析

将技术需求点与专利摘要输入模型后，得到的专利与技术需求点的匹配度如图 4-4 所示（部分展示）。

text	label0	prob0
本实用新型公开了一种智慧养老安全防护装置	安全防护装置	0.78045167
本公开涉及计算机领域，提供了一种监测仪	监测仪	0.244220872
本发明涉及移动马桶，特别是一种可抽水马桶	抽水马桶	0.998763257
本发明属于智能家居技术领域，公开智能家居	智能家居	0.904069595
本发明公开了一种基于智能家居的老智能家居	智能家居	0.887419125
本发明专利公开了一种可多功能轮椅健身椅	健身椅	0.586480436
本发明公开了一种助老座便自动伸缩抽水马桶	抽水马桶	0.097836746
本发明涉及智能养老监控领域，提出智能家居	智能家居	0.508665779
本申请提出一种信息处理方法、装置信号处理	信号处理	0.652948771
本实用新型提供一种智慧养老定位精准定位	精准定位	0.520072766

图 4-4 技术供需匹配结果

Fig. 4-4 matching results of technology supply and demand

表 4-4 筛选后的技术需求点结果

Tab. 4-4 results of technical requirement points after screening

有专利匹配的技术点	没有专利匹配的技术点
代语机	大数据技术
按摩椅	定向硬通道技术
报警装置	储柜旋转搁架
传动装置	触摸多级防盗门
温度传感器	计算机技术
信号处理器	传输技术
血糖仪	斥力弹跳器
防滑装置	测试仪

将阈值设置为 0.5，以达到与技术需求点尽可能相似的专利。阈值大于等于 0.5，即认为满足条件，需求由相应的专利进行匹配，从技术可行性的角度来看，这些筛选出的技术需求点为直接技术供给点；阈值小于 0.5，即不满足条件，没有专利与该部分技术需求相匹配，现阶段技术供给无法满足需求。

经过筛选后的技术需求点结果如表 4-4 所示，设定阈值后，阈值大于等于 0.5 的专利有 1721 个，其中有专利匹配的技术点（直接技术供给点）有 579 个。没有专利匹配的技术点主要包括两类：一是共性技术，即在各个领域都存在的基础性技术，如计算机技术、大数据技术；二是缺失技术供给点，即该类技术需求没有专利技术支撑，可以作为未来技术研发方向的参考之一，这类技术包括了产品技术和服务技术，其中服务技术的占比较高，如居家养老服务信息平台等。由于本文研究内容为养老科技产业的技术供需情况，后续研究中将重点对直接技术点和缺失技术供给点进行分析，二项合计最终得到的技术供给点为 603 个。

4.6 本章小结

本章首先建立了基于 BERT 的技术匹配模型，构造专利文本和技术需求文本的文本特征向量，基于特征向量计算两者相似度，从技术供需视角对养老科技产业技术有更深入的认识。通过设定阈值，来筛选符合条件的技术需求，得到了技术供给点 603 个。

第5章 养老科技产业技术供需匹配评价

根据第四章得到的直接技术供给点和缺失技术供给点（603个），首先利用 K-means 算法对技术点聚类，然后提出技术供给能力和技术需求能力两个评价指标体系，最后使用熵权法测算养老科技产业技术供需能力匹配进行评价。

5.1 技术供需评价研究框架

技术供需匹配评价研究目的是为了把握技术发展过程中技术研发和市场需求的关系，根据技术供需匹配情况指导研发方向，加强政策引导，提高市场技术接受度。

提出养老科技产业的技术供需匹配评价指标体系，测算技术供需匹配状态的研究过程如图 5-1 所示，首先利用 K-means 算法对第四章结果中得到的有效技术点进行聚类分析，这部分技术被认为是存在技术供给支撑的且有市场需求的或存在市场需求但是无技术供给支撑。然后分别从技术供给能力和技术需求能力构建养老科技产业技术供需匹配评价指标体系。最后使用熵权法测算养老科技产业技术供需能力，并对匹配结果进行分析。

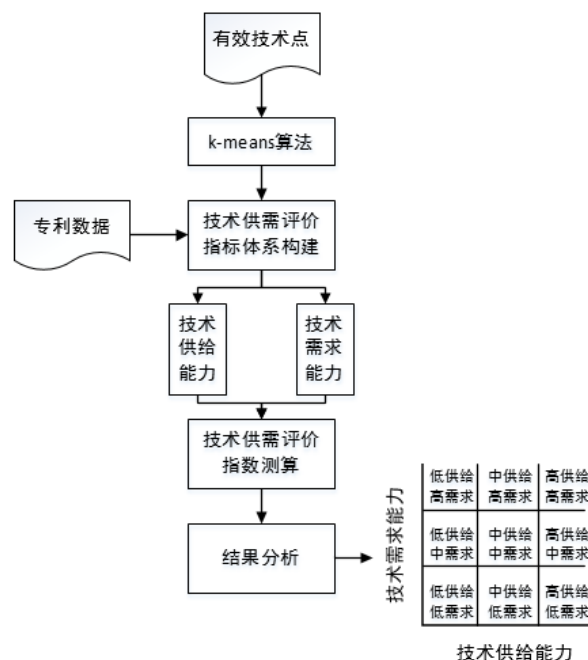


图 5-1 养老科技产业技术供需评价研究框架

Fig. 5-1 research framework for technology supply and demand evaluation of pension technology industry

5.2 技术点聚类分析

为了对养老科技产业深入分析，具体了解各技术领域的供需现状，指导研发方向和政策制定，本文首先需对识别出的技术供给进行聚类分析，得到养老科技产业的各技术领域。

5.2.1 K-means 聚类流程

本章通过 K-means 算法对有效技术点进行聚类分析。k-means 是常用的无监督学习的聚类算法，是基于形心的，其中形心是簇内所有点的均值^[124]。其主要是通过定义聚类类别 K ，聚类中心的选择是随机从样本集中选择 K 个值，通过对每个样本点和聚类中心进行计算相似度，重复过程，最终得到最为相似的 K 个类别^[125]。如图 3-2 为具体的聚类流程。

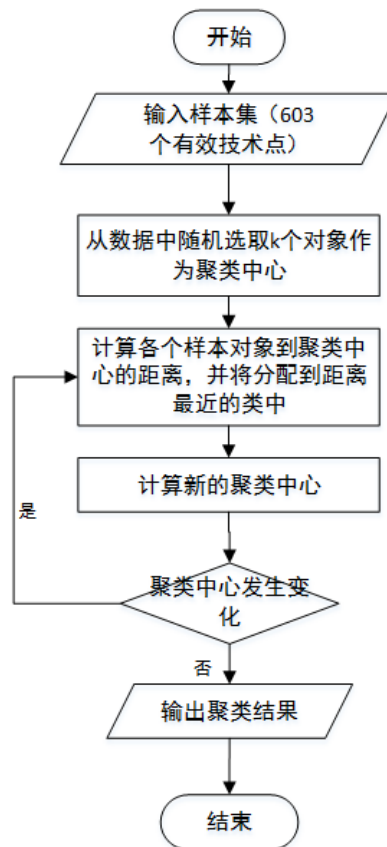


图 5-2 K-means 聚类流程

Fig. 5-2 K-means clustering process

(1) 确定 K 值，即需要将养老科技产业分为几个类别，然后从样本集中随机选择 K 个对象作为 K 个类别的聚类中心。

(2) 将第四章中得到的 603 个有技术供给点作为样本集，对技术供给点样本的每个技术供给点与各个聚类中心计算距离，最终将每个技术供给点划分到最近的类别中。

(3) 完成对有技术供给点的划分后，计算每个聚类结果的聚类中心。

(4) 将计算得到的新聚类中心同原聚类中心进行比较，如果发生了变化，返回第(2)步，否则继续聚类过程。

与前一次计算得到的 k 个聚类中心比较，如果聚类中心发生变化，转(2)，否则转(5)。

(5) 直到 K 个聚类中心不发生变化时，输出聚类结果，结束聚类流程。

5.2.2 最优 K 值的选取

K-means 算法在聚类前需要确定聚类类别 K ，常用的方法有轮廓系数法^[126]和手肘法^[127]。本文利用轮廓系数完成对 K 值的选取，由公式(5-1)计算得到。

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5-1)$$

公式(5-1)中， $a(i)$ 表示每个类别中技术供给点 i 与同类别中的有效技术点不相似程度的平均值； $b(i)$ 表示每个类别中有效技术点 i 到其他类别中的有效技术点的不相似程度的最小值。

轮廓系数最终的范围在-1 和 1 之间，得到的系数越接近 1 表示聚类效果越好^[128]。本文计算得到聚类类别 K 与平均轮廓系数曲线图如图 5-3 所示。当聚类类别 $K=9$ 时，平均轮廓系数最大，聚类效果最好，因此将聚类类别最终确定为 9。

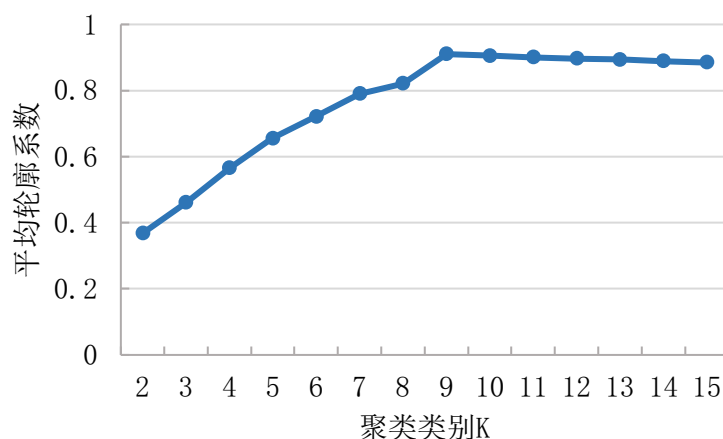


图 5-3 聚类结果与平均轮廓系数曲线图

Fig. 5-3 curve of clustering result and average contour coefficient

5.2.3 聚类结果分析

对第三章得到的 603 个有效技术点进行 K-Means 聚类分析，设置聚类类别 $K=9$ ，最终将技术供给分为了 9 类，分别为日常生活护理、机器人技术、智能

家居、可穿戴技术、移动通信类技术、检测类技术、传感器技术、辅助类技术、康复类技术。如表 5-1 所示为分类结果。

表 5-1 聚类结果

Tab. 5-1 clustering results

topic1: 日常生活护理	topic 2: 机器人技术	topic 3: 智能家居
老年病护理装置	机器人控制技术	安全防震床
智能护理床	服务机器人	牙齿治疗用漱口装置
自动排便护理机	老人看护机器人	智能感知助老看护系统
电子吃药提醒器	智能语音机器人	养老介卫浴装置
高龄人群睡眠急救护理装置	陪护机器人	家庭监控装置
topic 4: 可穿戴技术	topic 5: 移动通信类技术	topic 6: 检测类技术
便携式安全监护仪	基于蜂窝移动通讯定位便携装置	保健综合检测仪
可穿戴式监视器	即时通讯技术	无损检测技术
老年人穿戴式记录告警装置	图像实时放大设备	防跌倒检测系统
腕表式通讯呼叫器	信息通信技术	温湿度检测
穿戴式呼吸装置	语音信息通信装置	神经生理检测
topic 7: 传感器技术	topic 8: 辅助类技术	topic 9: 康复类技术
安全传感器	步行辅助车	颈椎病康复器
探测传感器	辅助喂药器	康复排痰技术
无线传感器网络	拐杖式购物车	快速康复护理技术
心率传感器	坐站辅助器	虚拟现实康复训练设备
重力传感器	轻携式爬楼装置	上肢有氧运动康复器

日常生活护理类技术主要包括老年病护理装置、智能护理床、电子吃药提醒器等老年人医疗护理相关技术；机器人技术主要包括了像服务机器人、陪护机器人的功能型机器人，机器人相关技术及用于特定功能的机械等；智能家居类技术主要包括安全防震床、养老卫浴装置等以家为平台，集成各类技术的家居生活设施；可穿戴技术主要包括腕表式通讯呼叫器、穿戴式呼吸装置等穿戴式智能老年设备；移动通信类技术主要包括即时通讯技术、语音信息通信装置等技术及移动电话、智能手机等技术；检测类技术主要扩防跌倒检测、神经生理检测等检测相关技术，还包括了血糖仪等设备；传感器技术主要包括重力测试、心率测试等各种功能型传感器技术；辅助类技术主要包括步行辅助器、拐杖式购物车等辅助老年人出行的技术；康复类技术主要包括康复排痰技术、上肢有氧运动康复器等老年康复医疗技术。

5.3 技术供需能力评价指标体系及模型构建

技术供需匹配评价的基础是技术供需能力的测算，这需要对二者的含义及构成（指标体系）进行界定。

5.3.1 技术供需能力评价指标体系

（1）新兴养老科技产业发展能力评价指标选取原则

养老科技产业技术供需评价指标体系的选取应该严格遵循科学性、系统性、可比性和可操作性等原则，具体说明如下：

1) 科学性原则。评价指标体系的构建综合考虑养老科技产业与其他一般产业的共有特征以及其自身的特殊性，选取的指标应与养老科技产业发展密切相关，可以全面反映养老科技产业的供需情况。

2) 系统性原则。指标选取不仅要考虑宏观环境对养老科技产业的影响，还应该考虑养老科技产业其组成成分以及各组成成分之间的相互关系，尽量避免指标之间的交叉。

3) 可比性原则。在指标选取的过程中，应该重视养老科技产业发展之间的差异，在选择具体指标时使其具备可比性。

4) 可操作性原则。充分考虑选取的各个指标必须是可以获得的，保证数据来源的真实性和可靠性。

（2）技术供需能力含义及构成

养老科技产业技术供需指标是指可以用来反映养老科技产业技术供需水平，由技术供给能力和技术需求能力。

1) 技术供给能力。技术供给是指一定时期内，系统为满足不同的技术需求而提供的技术成果^[129]。技术供给能力是指为了满足经济、社会、科技发展的特殊需求，在一定的创新环境下，通过科技资源优化配置，生产出具有高附加值的，且能够转化为实现生产力的技术成果的能力。因此，将技术供给能力进行量化主要从最终实现的技术成果上进行研究。专利数据能够作为技术发展水平的衡量标准之一。联合国教科文组织曾在 1996 年发表的世界科技报告中提出，技术活动可以通过专利局公布的专利数量来描述^[130]。此外，实力比较强大的国家或组织会利用专利来抢占技术制高点，并且通过寻求技术保护形成对某一领域市场的合法垄断，保证专利成果更好的转换。专利数量是技术成果的一种形式，而有效的技术成果则体现为专利质量。

因此，本文定义的技术供给能力是指，面向养老科技产业的技术需求，能够提供的专利技术的能力。

2) 技术需求能力。技术需求是指养老科技产业发展对技术产生的需求。技术需求能力是技术需求的种类和数量大小,它决定了拉动养老科技产业创新的能力(由于将对每个具体细分领域进行研究,故仅采用“数量”指标,而不考虑“种类”指标),包括有技术支撑的技术需求和无技术支撑的技术需求。前者是指在专利数据库中可以识别出满足技术需求的技术,后者是指专利数据库缺失满足技术需求的技术。

5.3.2 技术供需能力评价指标细化

文章 5.2 部分对有效技术点进行了分类,为了更好的了解各类技术特征,本章将利用专利数据源、技术需求数据源两种数据源构建技术供需评价指标体系。根据数据源及评价对象特征提出了从技术供给能力和技术需求能力进行详细的指标构建,如表 5-2 为技术供需评价的指标体系。

表 5-2 养老科技产业技术供需评价指标体系

Tab. 5-2 evaluation index system of technology supply and demand of pension technology industry

一级指标	二级指标	三级指标	计算含义
技术供给能力	技术专利数量	专利申请数量	相关技术点专利的申请数量
	技术专利质量	专利申请人数	相关技术点的平均专利申请人数
		专利许可量	相关技术点的平均专利许可量
	技术专利影响	专利平均被引次数	相关技术点的平均专利被引次数
		当前影响指数	平均被引次数/领域中所有专利平均被引次数
		技术强度	相关技术点的专利数量*当前影响指数
		科学关联性	相关技术点的专利平均引用次数
		科技强度	专利数量*科学关联性
技术需求能力		有支撑的技术需求量	有专利支撑的技术需求的点数
		无支撑的技术需求量	无专利支撑的技术需求的点数

技术供给能力评价指标及计算:不少专家构建了专利技术评价指标,其中在 1980 年,CHI 研究人员^[131]提出了专利量化指标。本文主要参考 CHI 指标,并结合本文评价对象构建了养老科技产业技术供给能力评价指标,包括专利数量、专利质量和专利影响三个指标。技术专利数量^[132],可以用来衡量产业的专利活动的活跃程度,通过专利数量进行计算;技术专利质量^[50],可以用来衡量产业的技术水平程度,利用专利申请人数和专利许可量进行表征,专利申请人数,可以反映出该专利的投入成本,从而反映专利质量。专利许可量,指该技

术的专利所有人或其授权人许可他人在一定期限、一定地区、以一定方式实施其所拥有的专利，并向他人收取使用费用，一定程度上反映了专利质量；技术专利影响^[132]，可以用来衡量产业的技术实际影响力和技术总体水平，利用专利平均被引次数、当前影响指数、技术强度、科学关联性和科技强度来进行计算。

技术需求能力及指标计算：本文根据养老科技产业技术需求能力的定义构建了评价指标，包括可支撑需求的技术数量和未支撑需求的技术数量两个指标。可支撑需求的技术数量，可以用来衡量将技术需求的潜在价值转化为实际价值的能力，表征技术需求的转化能力。未支撑需求的技术数量，可以用来衡量现有技术的覆盖率情况，进而反映技术市场化的发展状况。

5.3.3 技术供需能力匹配评价模型构建

技术供需能力通过技术供给能力和技术需求能力来计算、监测和评价，其依据上述构建的技术供需能力评价指标体系，并进行计算分析。首先是对数据标准化处理，然后利用熵权法确定权重，最后得到各类别的技术供需能力。

(1) 原始数据标准化处理

为了能够使得构成养老科技产业技术供需能力的 10 个指标之间进行比较和加权，需要消除这 10 个指标的原始数据存在的单位不同和量级上的问题，即需要进行标准化处理，具体计算方式见公式（5-2）。

$$x'_{ij} = \frac{x_{ij} - x_{ijmin}}{x_{ijmax} - x_{ijmin}} \quad (5-2)$$

其中， x_{ij} 表示第*i*个类别在第*j*个指标上的值， x'_{ij} 为经过标准化后的值。

(2) 熵权法确定权重

熵权法是给各指标赋权，计算权重依据是各指标反映的信息量，权重越大，表示这个指标提供越多的信息^[133]。具体的计算步骤如下所示：

步骤一，根据公式（5-3）计算第*j*项指标的输出熵，其中 $p_{ij} = \frac{b_{ij}}{\sum_{i=1}^m b_{ij}}$ ， $j = 1, 2, \dots, n$ 。

$$E_j = -(\ln m)^{-1} \sum_{i=1}^m p_{ij} \ln p_{ij} \quad (5-3)$$

步骤二，根据公式（5-4）计算第*j*项指标的变异程度。

$$D_j = 1 - E_j ; j = 1, 2, \dots, n \quad (5-4)$$

步骤三，根据公式（5-5）计算第*j*项指标的权重值。

$$w_j = \frac{D_j}{\sum_{j=1}^n D_j} ; j = 1, 2, \dots, n \quad (5-5)$$

5.3.4 评价结果分析

根据前文构建的养老科技产业技术供需评价模型，对 9 个技术类别分别计算得到三个一级指标，首先是对数据进行去除量纲化，然后计算每个一级指标得分，得到如表 5-3 所示。

从技术供给能力来看，养老科技产业在辅助类技术的供给能力最大，传感器技术的供给能力最小，主要原因为传感器技术有别于其他类别技术，更多为共性技术，因此整体来看，该类技术供给能力相对较低；辅助类技术涉及到老年人出行问题，越来越多的技术应用于传统产品中。从技术需求能力来看，移动通信类技术的需求能力最大，机器人技术的需求能力最小，主要原因为随着互联网的普及，越来越多的老年用户开始使用移动互联网，因此带动了对移动通信类技术的需求量；机器人技术相对来讲，这类技术还处于技术研发阶段，市场上的关注度主要是企业，对个人来讲这类需求的技术接受度相对较低，因此对机器人技术的需求量较小。

表 5-3 技术供需能力指标得分

Tab. 5-3 index score of technology supply and demand capacity

类别 \ 指标	技术供给能力	技术需求能力
日常生活护理	1.10252287	0.364412476
机器人技术	0.340471464	0.025943668
智能家居	1.291657806	0.304937257
可穿戴技术	0.328935397	0.185176976
移动通信类技术	0.622369797	0.698910833
检测类技术	0.642708146	0.306108141
传感器技术	0.103861	0.154950338
辅助类技术	1.814600432	0.104292252
康复类技术	0.6505716	0.030108917

黄文馨^[134]指出，从产品供需角度来看，可以将产品分为三个阶段，分别为供大于求、供求平衡和供小于求，当产品进入市场初期，市场中的需求量较高，但是供给方较少，就会出现供小于求的现象。随着产品市场的成熟，同一类产品的供给方看到市场需求，抢占市场，供给能力加大，会达到供需平衡的状态。后续市场饱和，部分产品供给方退出市场，会进入供大于求的阶段。因此研究供需能力，可以看到此类技术当前所处阶段，更好的指导未来研发方向。

为了更直观的对各类别的两个指标进行横向分析，将各指标按百分位划分为“高、中、低”，将指标处于全领域前 1/3 的指标值设为“高”，中间 1/3 的指标值设为“中”，最后 1/3 的指标值设为“低”，从而将指标变量转化为哑变量进行分析。

如图 5-4 所示，9 个技术类别表现出了六种特征，即低供给、低需求；中供给、低需求；中供给、中需求；中供给、高需求；高供给、中需求；高供给、高需求。为了更深入了解不同特征的变量，将对各级指标进行详细分析，如表 5-4 所示。

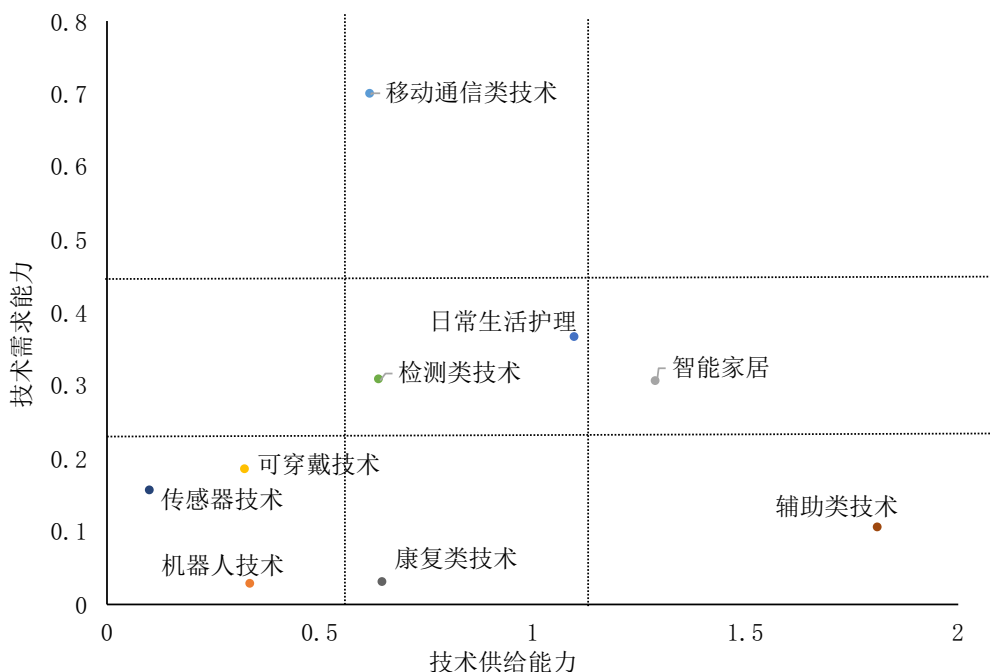


图 5-4 技术供需能力评价图

Fig. 5-4 evaluation chart of technology supply and demand capacity

(1) 低供给、低需求类型技术

低供给、低需求类型技术处于未被市场所关注的状态，包括了传感器技术、可穿戴技术和机器人技术。其中传感器技术从具体指标来看都属于低水平状态，该技术多属于共性技术，在养老科技产业中的特征不突出，由此可见，养老科技产业是跨学科的科技工作，政府应对该领域的基础科学（老年生理与医学、信息与数据科学）与共性技术给予重点支持；可穿戴技术各项指标中处于领域低水平，处于起步阶段，是导入型技术；机器人技术当前影响指数占总领域的 23.6%，代表了该类技术具有较高的科技价值及影响力，14.3%的技术强度反映了该类技术具有较强的科学性。

表 5-4 技术供需评价各指标占比

Tab. 5-4 proportion of each index of technology supply and demand evaluation

指标 类别	技术供给能力								技术需求能力	
	专利数量 (%)	专利平均申请人数 (%)	专利许可量 (%)	专利平均被引次数 (%)	当前影响指数 (%)	技术强度 (%)	科技关联性 (%)	科技强度 (%)	有支撑的技术需求量	无支撑的技术需求量
传感器技术	0.9	11.4	0.0	0.5	5.5	0.0	0.7	0.0	10.7	0.0
辅助类技术	24.9	10.8	34.8	23.3	10.0	21.6	24.3	38.0	7.4	0.0
机器人技术	4.1	10.9	0.0	9.0	23.6	14.3	8.2	2.1	2.2	0.0
检测类技术	9.7	10.7	17.9	8.7	9.5	7.6	11.1	6.8	18.7	3.6
康复类技术	9.6	11.8	0.0	8.5	9.5	7.5	9.8	5.9	0.4	3.6
可穿戴技术	6.7	10.8	0.0	5.5	8.8	4.2	4.7	2.0	8.5	0.0
日常生活护理	17.8	10.8	9.7	17.5	10.5	17.1	19.2	21.5	0.7	42.9
移动通信类技术	5.9	11.5	29.2	6.7	12.0	7.5	5.3	2.0	42.8	7.1
智能家居	20.6	11.3	8.4	20.4	10.6	20.3	16.8	21.7	8.7	21.4

(2) 中供给、低需求类型技术

中供给、低需求类型技术表现出供大于求的特征，该类技术处于起步阶段，市场上较多可提供的专利技术需求，但是人们对此类技术的需求较少，包括了康复类技术。该类技术在供给专利申请人数上多，占总领域的 11.8%，表明了研发单位对康复类技术的重视。未来随着人们对于健康问题的关注，加之国家政策的大力扶持，更多老年人将加入到康复医疗中，这类技术的需求将会逐步加大。

(3) 中供给、中需求类型技术

中供给、中需求类型技术从供给和需求能力上都未达到顶峰处于中间水平，是稳定型技术，包括了检测类技术和日常生活护理类技术。具体来看，检测类技术的可支撑的技术量占总领域的 18.7%，表明市场上对该类技术的需求量较大并现有技术供给基本满足该技术领域；日常生活护理类技术科技关联性占总领域的 19.2%，科技强度占 21.5%，反映了这类技术具有较强的科学性，但是未支撑需求的技术量占 42.9%，表明市场需求与技术研发存在异步性，即当前技术研发方向未能很好的满足市场基本需求。

(4) 中供给、高需求类型技术

中供给、高需求类型技术表现出供不应求的特征，包括了移动通信类技术。移动通信类技术移植性较强，该类技术相对于其他技术来讲，老年相关的特征性不太明显，因此技术供给能力较弱。但是随着移动互联网的发展，老年人对该类技术的需求变大，占总领域的42.8%，这部分需求带动的是移动通信类产品，该类技术属于交易活跃型技术。因此，该类技术的研发方向应该注重同老年特征相结合，同时注重产品上的便利性、操作简单等特点，从而能够满足老年人社会参与和娱乐的需求。

（5）高供给、低需求类型技术

高供给、高需求类型技术属于活跃性技术，包括了辅助类技术。该类技术在各项指标上都表现出较高水平，从技术供给能力来看，专利数量占总领域24.9%，专利许可量占34.8%，专利平均被引次数占23.3%，当前影响指数10%，技术强度占21.6%，科技关联性占24.3%，科技强度占38%，表明了相关技术多、质量高、影响力高且科学性强。从技术需求能力来看，技术需求能力处于中间水平。因此，未来该类技术的重点研发方向应注重产品的便利性、安全性、操作简单等特点，能够支持老年人能够独立行动，进行社会参与、娱乐与训练等，以吸引更多的需求。

（6）高供给、中需求类型技术

高供给、中需求类型技术表现出供大于求的特征，人们对于该类技术的认知不足，此类技术的接受度低，传统观念中的产品形态难以一时代替，因此，此类技术的需求较小，但是由于技术发展，越来越多的产品在做技术突破，这类技术的供给必然是呈现增长趋势。智能家居类技术表现出高供给、中需求特征，属于成长型技术，从技术需求能力来看，未支撑需求的技术量占21.4%，表明当前智能家居类技术的研发方向同市场需求存在一定的差异，未来该类技术的重点研发方向应该注重产品的服务性。

5.4 对策建议

根据对养老科技产业技术供需能力评价的研究分析结果，指出了我国养老科技产业在发展过程中存在的主要问题，因此，本文针对如何提升我国养老科技产业发展能力提出对策建议。

（1）传感器、计算机、大数据等共性技术在养老科技产业中普遍应用，因此，可在国家重点研发计划、国家自然科学基金立项中，对养老科技基础科学与共性技术给予重点支持。通过技术融合，为养老科技创新提供新思路、新方法。

(2) 康复类技术、辅助类技术的切实影响着老年人的生活，引导老年人接受并使用此类技术，是当前解决市场供需不平衡问题的关键。开展老年人需求调研，加大养老科技的宣传，根据老年人需求为养老科技企业提供技术研发方向，为政府政策提供决策信息，保障养老科技市场的稳定性。

(3) 移动通信类技术逐渐成熟，市场需求量也比较大，把控产品质量是当前养老科技市场中必不可缺的部分。制定养老科技产业与企业的标准，提高养老政策的针对性和有效性。

(4) 日常护理类技术、智能家居类技术的市场需求的技术覆盖率较低，市场化程度较低，这两类技术领域中存在大量居家养老相关系统、平台需求，未来如何将平台与产品相结合，推动相关技术产品的应用是当前该技术领域的重点方向。

(5) 大力培养养老科技相关的创新性人才，培养多学科交叉的养老科技相关的专业型人才，鼓励养老科技的技术创新，为解决我国当前老龄化问题创造良好的学术环境。

(6) 结合当代老年人需求特点，建议养老科技未来重点研发方向为：老年产品创新，融合信息化、智能化技术更好地满足老年人的需求；注重产品的便利性、安全性、操作简单等特点，能够支持老年人能够独立行动，进行社会参与、娱乐与训练等；注重养老科技产业的服务化技术，规范各类养老平台、系统的应用。

5.5 本章小结

本章基于有效技术点对我国养老科技产业的技术供需进行评价，首先根据轮廓系数确定将技术点分为 9 类，并利用 K-means 算法进行对技术点进行聚类，最终得到日常生活护理、机器人技术、智能家居、可穿戴技术、移动通信类技术、检测类技术、传感器技术、辅助类技术、康复类技术等 9 类技术，然后构建技术供需评价模型，对各类别进行指标测算，最后将 9 类根据技术供需能力和技术需求能力分为了 6 个特征表现，分别为低供给、低需求；中供给、低需求；中供给、中需求；中供给、高需求；高供给、低需求；高供给、中需求。

结论

随着科学技术的发展，技术创新成为了产业发展中重要的影响因素，也是国家或组织核心竞争力的重要来源。因此，对产业技术的供需能力进行识别，可以把握产业技术的发展现状，并形成以“需求”为导向的技术研发良性发展。

本文首先对文献进行梳理和回顾，界定了养老科技产业的概念，并根据养老科技产业的定义进行了多源数据的获取，通过文献数据源、政策、技术报告构建养老科技产业的技术词典，通过网络数据源获取用户需求，结合构建的技术词典识别用户需求中的技术需求点；然后基于 BERT 和 TF-IDF 构建技术供需匹配模型，识别出有效技术点；接着对识别出的有效技术点利用 K-means 算法进行聚类分析，划分为了 9 类，分别为日常生活护理、机器人技术、智能家居、可穿戴技术、移动通信类技术、检测类技术、传感器技术、辅助类技术、康复类技术；最后构建技术供需评价模型，从技术供给能力和技术需求能力对 9 类技术进行评价，将养老科技产业技术现有的供需情况分为了六种：低供给、低需求；中供给、低需求；中供给、中需求；中供给、高需求；高供给、中需求；高供给、高需求，并根据各类型情况对技术类别进行分析，进而引导技术研发方向。

根据研究结果，文本也提出了相应的建议，比如：针对供大于求的移动通信类技术，未来技术研发应该结合老年特征，注重产品的便捷性和操作简单等特性；针对供不应求的康复类技术，国家应该给予相关技术重点支持，引导企业或者组织加大相关技术的研发；针对供需相对平衡的技术，如日常护理类技术，未来应重点关注产品创新，提高老年技术接受度等。

本文为研究产业技术提供了一种新思路，但是研究过程中，仍存在一些问题进行改进和完善：本文纵向得分析了养老科技产业技术供需发展，但是在技术发展中可能会由于时间问题存在时间维度上的差异，但是本文并没有从时间上对技术供需问题进行分析。望今后可以从多角度对问题进行分析，增加时间维度的模型计算，研究问题更具有严谨性。

参考文献

- [1] 吴玉.中国老龄产业发展报告[M]. 北京:社会科学文献出版社, 2014.
- [2] 黄鲁成,刘春文,苗红,吴菲菲.开展依靠科技创新应对人口老龄化研究的思考[J].中国软科学,2019(05):1-10.
- [3] 田新朝. 以供给侧改革助推养老服务业发展[N]. 中国人口报,2019-10-14(003).
- [4] Wang S, Zhang H, Wang H.Object co-segmentation via weakly supervised data fusion[M]. America:Elsevier Science Inc, 2017.
- [5] De la Torre, F; Morales, D; Quiroz, C P.Gerontechnology rapid review and global trends[J]. EEGmagic:Revista mexicana de ingenieríabiomédica.2015,36(3):171-179.
- [6] Fernandez,Caballero,Antonio;Gonzalez,Pascual;Navarro Elena.Cognitively Inspired Computing for Gerontechnology[J].COGNITIVE COMPUTATION.2016, 8(2):297-298.
- [7] Chen,Ke;Chan Alan Hoi-shou.Use or Non-Use of Gerontechnology-A Qualitative Study[J].INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH.2013,10(10):4645-4666.
- [8] Peine A , Rollwagen I , Neven L . The rise of the “innosumer”-Rethinking older technology users[J]. Technological Forecasting and Social Change, 2014, 82:199-214.
- [9] Macedo I M . Predicting the acceptance and use of information and communication technology by older adults: An empirical examination of the revised UTAUT2[J]. Computers in Human Behavior, 2017, 75(10):935-948.
- [10] Chen K , Chan A H S . Predictors of gerontechnology acceptance by older Hong Kong Chinese[J]. Technovation, 2014, 34(2):126-135.
- [11] 李敏,王振振,王立剑.居家老年人使用养老科技的影响因素分析——基于北京、南京、咸阳三市的调查[J].人口与发展,2017,23(03):84-92.
- [12] 黄鲁成,常兰兰.基于专利的技术景观四侧面分析框架——以老年福祉技术为例[J].科技管理研究,2016,36(21):34-40.
- [13] 黄鲁成,李晋,吴菲菲.老年科技重点创新领域的竞争态势分析[J].情报志,2019,38(0 6):57-63+71.
- [14] 吴菲菲,栾静静,黄鲁成,张亚茹.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.
- [15] 娄岩,杨嘉林,黄鲁成,苗红,杨培培.基于专利共类的技术融合分析框架研究——以老年福祉技术与信息技术的融合为例[J].现代情报,2019,39(09):41-53.
- [16] Phang C W, Sutanto J, Kankanhalli A, et al. Senior citizens acceptance of information systems: A study in the context of e-government services[J]. IEEE Transactions on Engineering Management, 2006, 53(4): 555-569.

- [17] Wang K H, Chen H G, Chen G. On the priority of needs for home care services of older adults in Taiwan, as based on Maslow's hierarchy of needs[J]. Gerontechnology, 2016, 14(4): 239-243.
- [18] Bouma H, Fozard J L. Gerontechnology as a field of endeavour[J]. Gerontechnology, 2009, 8(2):68-75.
- [19] Fitzgerald K G, Caro F G. International Perspectives on Age-Friendly Cities[M]. Cincinnati: University of Cincinnati, 2015.
- [20] Álvarez-García, José, Durán-Sánchez, Amador, del Río-Rama, María, et al. Active Ageing: Mapping of Scientific Coverage[J]. International Journal of Environmental Research and Public Health, 2018, 15(12):22-45.
- [21] 朱玥颖,刘治彦.我国智慧养老产业发展的新思路[J].江苏科技信息,2018,35(04):68-70+74.
- [22] Lauriks S, Reinersmann A, Van Der Roest H G, et al. Review of ICT based Services for Identified Unmet Needs in People with Dementia[J]. Ageing Research Reviews, 2007, 6(3):223-246.
- [23] Robinson M, Robertson S. Health Information Needs of Men[J]. Health Education Journal, 2014, 73 (2) : 150-158.
- [24] 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,2016,26(13):102-105.
- [25] 黄鲁成,李晋,苗红,吴菲菲.基于老年人科技产品需求的研发[J].技术经济,2019,38(01):19-27.
- [26] 林佳蓉,周玮,洪紫静,张林.智能养老视角下中国城市老年人科技产品的使用[J].科技创新导报,2019,16(08):224-225.
- [27] 李娜.养老服务需求与供给分析[J].山西农经,2019(16):19-20+22.
- [28] 张静,任振兴,范叶超.养老服务需求与发展对策研究——基于 CGSS 和 CHARLS 两项全国老年人口调查数据的实证分析[J].老龄科学研究,2019,7(03):16-36.
- [29] 钱宇星,周华阳,周利琴,任美伶,李浩.老年在线社区用户健康信息需求挖掘研究[J].现代情报,2019,39(06):59-69.
- [30] 严俊.老有所依:公众养老的需求侧与供给侧[J].国家治理,2016(09):32-41.
- [31] 高艳,黄怡,陈泰完.基于供需视角下海口市养老服务产业发展研究[J].当代经济,2017(34):105-107.
- [32] 杨素,敬德芳.供需视角下失智老人照护服务现状研究——基于天津市的调查[J].劳动保障世界,2019(20):20-22.
- [33] 张新辉,李建新.社区老年服务供需动态变化与平衡性研究——基于 CLHLS 2005—2014 的数据[J].社会保障评论,2019,3(02):122-136.
- [34] 黄鲁成,陈安奇,吴菲菲,李欣,黄斌.京津冀老年福祉技术需求对比分析[J].软科学,2018,32(03):55-60+101.
- [35] Xi-jun He, Xue Meng, Yan-bo Dong, Yu-ying Wu. Demand identification model of potential technology based on SAO structure semantic analysis: The case of new energy and energy saving fields[J]. Technology in Society, 2019(9):58-58.

- [36] 程力培,董策玮,张兆一,郭静雯,何喜军.基于专利技术交易数据的京津冀区域热点需求技术识别研究[J].农业图书情报学刊,2018,30(07):5-9.
- [37] Gale D , Shapley L S . College Admissions and the Stability of Marriage[J]. The American Mathematical Monthly, 2013, 120(5):386-391.
- [38] Roth, Alvin E. Common and conflicting interests in two-sided matching markets[J].European Economic Review, 1985, 27(1):75-96.
- [39] 陈希,王娟.智能平台下考虑主体心理行为的医疗服务供需匹配方法[J].运筹与管理,2018,27(10):125-132.
- [40] 曹帅. 考虑个性化指标的双边匹配决策问题研究[D].江苏:南京航空航天大学,2017.
- [41] Yumeng Miao, Rong Du, Jin Li, J. Christopher Westland. A two-sided matching model in the context of B2B export cross-border e-commerce[J]. Electronic Commerce Research, 2019, 19(4):77-98.
- [42] 姜宝胜,张健建,乔亚娟.基于“互联网+”时代下的出租车资源优化配置研究[J].数学学习与研究,2016(20):139.
- [43] 单晓红,王非,何喜军,蒋国瑞.基于稳定双边匹配的供应链产销合作研究[J].计算机工程与应用,2016,52(23):260-265+270.
- [44] 邱科.风险投资机构与创业企业双边匹配决策研究[J].重庆文理学院学报(社会科学版),2019,38(05):45-53+60.
- [45] Brownlie DT, Macbeth DK.The Strategic Management of Technology:Integrating technology supply and demand perspective[J].European Management Journal, 1989, 71(1):71-83.
- [46] Klerkx L, Leeuwis C. Matching demand and supply in the agricultural knowledge infrastructure experiences with innovation intermediaries [J]. Food Policy, 2008, 33 (3) : 260-276.
- [47] Hung Sheng-Hao, Lin Chia-Hung, Hong Jen-Shin.Web mining for event-based commonsense knowledge using lexico-syntactic pattern matching and semantic role labeling[J].Expert Systems with Applications, 2010, 37(1): 341-347.
- [48] Den P . KNOWLEDGE-INTENSIVE BUSINESS SERVICES AS CO-PRODUCERS OF INNOVATION[J]. International Journal of Innovation Management, 2000, 04(04):491-528.
- [49] 何喜军,马珊,武玉英,蒋国瑞.多特征融合下在线技术转移平台供需匹配研究——以京津冀区域数据为例[J].情报杂志,2019,38(06):174-181.
- [50] 张婷婷. 基于多源数据的企业潜在技术需求点识别[D].北京:北京工业大学,2018.
- [51] 宋志红,李常洪,李冬梅.技术联盟网络与知识管理动机的匹配性——基于1995-2011年索尼公司的案例研究[J].科学学研究,2013,31(01):104-114.
- [52] 陈希,樊治平,李玉花.技术知识供需双边匹配的两阶段决策分析方法[J].工业工程与管理,2010,15(06):90-94.
- [53] 张娴,胡正银,茹丽洁,许海云,肖国华,方曙.专利技术供需信息关联知识组织模式研究[J].图书情报工作,2016,60(08):118-125.
- [54] Hall D L , Llinas J . An Introduction to Multisensor Data Fusion[J]. Proceedings of the IEEE, 1997, 85(1):6-23.

- [55] Shen, Runping, Huang, Anqi, Li, Bolun, et al. Construction of a drought monitoring model using deep learning based on multi-source remote sensing data[J]. International Journal of Applied Earth Observation and Geoinformation, 2020(79):48-57.
- [56] 李广建,杨林.大数据视角下的情报研究与情报研究技术[J].图书与情报,2012(6):1-8.
- [57] 李广建,化柏林.大数据分析情报分析关系辨析[J].中国图书馆学报,2014(5):14-22.
- [58] Si L, Callan J. A semisupervised learning method to merge search engine results[J]. ACM Transactions on Information Systems, 2003, 21(4):457-491.
- [59] Dong X L, Gabrilovich E, Murphy K, et al. Knowledge-based trust: estimating the trustworthiness of web source[J]. Proceedings of the Vldb Endowment, 2015, 8(9):938-949.
- [60] Q. Song, Y. Wu, P. Lin, et al. Sun. Mining Summaries for Knowledge Graph Search[C]. IEEE Transactions on Knowledge and Data Engineering. IEEE. 2018: 1887-1900.
- [61] 周群,化柏林.基于多源数据融合的科技决策需求主题识别研究[J].情报理论与实践,2019,42(03):107-113.
- [62] Fang Chen, Zhigang Yuan, Yongfeng Huang. Multi-source data fusion for aspect-level sentiment classification[J]. Knowledge-Based Systems, 2020(7):187-187.
- [63] 李晓梅,郭正模,刘金华.老龄产业的跨行业特征与统计规范探讨[J].人口与经济,2016(01):29-38.
- [64] 董昕,刘强,周婧玥.我国老龄产业发展现状与对策:一个文献综述[J].西部论坛,2014,24(04):78-86.
- [65] 吴玉韶.对老龄产业几个基本问题的认识[J].老龄科学研究,2014,2(01):3-12.
- [66] 吴玉韶. 中国老龄产业发展报告[M]. 北京:社会科学文献出版社, 2014.
- [67] 陈友华.中国老龄产业的发展规模及其影响因素研究[J].现代经济探讨,2015(11):44-49.
- [68] 杨立雄.北京市老龄产业发展研究[J].中国软科学,2017(03):74-83.
- [69] Klimczu A. Supporting the development of gerontechnology as part of silver economy building[J]. Journal of Interdisciplinary Research, 2012, 2(2): 52-56.
- [70] Cornet G. Europe's 'Silver Economy': A potential source for economic growth[J]. Gerontechnology, 2015, 13(3): 319-321.
- [71] 王波.论养老模式与现代服务业的融合[J].华东理工大学学报(社会科学版),2008(03):16-20.
- [72] 邵其会.基于产业融合的养老服务业发展模式与对策[J].现代商业,2014(12):272-273.
- [73] Chen K, Chan A H S. The ageing population of China and a review of gerontechnology[J]. Gerontechnology, 2011, 10(2): 63-71.
- [74] Petermans J. Gerontechnology : don't, miss the train, but which is the right carriage[J]. European Geriatric Medicine, 2017, 8(4):12-23.
- [75] Thomas L H, Marcia K H. Gerontechnology Why and How[M]. Eindhoven, the Netherlands: Herman Bouma Foundation for Geron-technology, 2000:1-2.
- [76] 马俊达,刘冠男,沈晓军.社会福利视野下我国老年福祉科技及其发展路径探析[J].中国科技论坛,2014(05):130-136.

- [77] 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.
- [78] Sixsmith A, Mihailidis A, Simeonov D. Aging and Technology: Taking the Research into the Real World[J]. Public Policy & Aging Report, 2017, 27(2): 74-78.
- [79] 中国互联网信息中心.第 46 次《中国互联网络发展状况统计报告》[DB/OL].http://www.cnnic.net.cn/hlwfyj/hlwzxbg/hlwjbg/202009/t20200929_71257.htm, 2020.
- [80] 刘娜,余光胜,毛荐其.基于网络数据的经管学科研究趋势测度[J].中国软科学,2016(4):133-142.
- [81] 欧阳逸,郭斌,何萌,等.微博事件感知与脉络呈现系统[J].浙江大学学报(工学版),2016,50(6):1176-1182.
- [82] 中国互联网信息中心.第 47 次《中国互联网络发展状况统计报告》[DB/OL].<https://news.znds.com/article/52203.html>, 2020.
- [83] 谢婉莹.档案网络信息的公众关注度与需求探析——基于百度指数的数据分析[J].档案学研究,2017(03):59-64.
- [84] 2016 年中国网络社交市场趋势预测,易观智库 [DB/OL].<https://www.chyxx.com/industry/201607/430649.html>, 2016.
- [85] Von Hippel E A , Kaulartz S . Next-Generation Consumer Innovation Search: Identifying Early-Stage Need-Solution Pairs on the Web[J]. Social Science Electronic Publishing, 2010(1):2-9.
- [86] 佟怡.基于网络爬虫技术的非上市商业银行知识管理评价研究[D].天津:天津工业大学,2019.
- [87] 冯俐.中文分词技术综述[J].现代计算机(专业版),2018(34):17-20.
- [88] 杨凡.基于深度学习的微博情感分析[D].西安:西安建筑科技大学,2020.
- [89] 王辉,潘俊辉,王浩畅,张强,张岩.基于深度学习的中文语法错误诊断方法研究[J].计算机技术与发展,2020,30(11):69-73.
- [90] 梁喜涛,顾磊.中文分词与词性标注研究[J].计算机技术与发展,2015,25(02):175-180.
- [91] 古丽拉.基于规则和统计的哈萨克语词法分析和短语识别方法研究[D].哈尔滨:哈尔滨工业大学,2017.
- [92] 杨荣根,杨忠.基于 HMM 中文词性标注研究[J].金陵科技学院学报,2017,33(01):20-23.
- [93] Hinton G E . Learning distributed representations of concepts[C]// Eighth Conference of the Cognitive Science Society. 1989(2):1-9.
- [94] 路欣远.面向智能问答系统的答案选择与自然语言推理模型[D].江苏:南京邮电大学,2020.
- [95] 许文华.基于深度神经网络的长文本分类算法的设计与实现[D].江苏:南京邮电大学,2020.
- [96] Yoshua1 Bengio, Réjean Ducharme, Pascal Vincent, Christian Janvin. A Neural Probabilistic Language Model[J]. Journal of Machine Learning Research, 2003(3):1137-1155,

- [97] AlexRudnicky. Can Artificial Neural Networks Learn Language Models?[C]// Sixth International Conference on Spoken Language Processing, ICSLP 2000 / INTERSPEECH 2000, Beijing, China, October 16-20, 2000.
- [98] Bengio Y , Schwenk H , Jean-Sébastien Senécal, et al. Neural Probabilistic Language Models[J]. The Journal of Machine Learning Research, 2003, 3(6):1137-1155.
- [99] Andriy Mnih, Geoffrey E. Hinton. Three new graphical models for statistical language modelling[C]. International Conference on Machine Learning, 2007:641-648.
- [100] Mikolov T , Martin Karafiát, Burget L , et al. Recurrent neural network based language model[C]// Interspeech, Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September. DBLP, 2010:1045-1048.
- [101] Huang E H , Socher R , Manning C D , et al. Improving word representations via global context and multiple word prototypes[C]// Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1. Association for Computational Linguistics, 2012.
- [102] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer Science, 2013(1):8-10.
- [103] Yoon B, Magee C L. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction[J]. Technological Forecasting and Social Change, 2018, 132: 105-117.
- [104] 姜佳. 文本表示模型及相似度计算算法研究与应用[D].西安:西安科技大学,2020.
- [105] 翟璐莎. 最长公共子序列查询算法研究[D].河北:燕山大学,2018.
- [106] 赵潇安. 基于欧式距离角点提取的工业机器人视觉算法 [J]. 自动化与仪表, 2020, 35(08):31-36.
- [107] 周昌,李向利,李俏霖,朱丹丹,陈世莲,蒋丽榕.基于余弦相似度的稀疏非负矩阵分解算法 [J]. 计算机科学, 2020, 47(10):108-113.
- [108] 桂晶晶,王祥至.我国“就业质量”领域研究特点和研究趋势探究——基于中国知网的数据统计分析[J].大众标准化,2020(15):22-24.
- [109] 刘素娜. 知乎的经营与管理模式分析[J]. 智库时代, 2019(13): 278-279.
- [110] 知乎: 截至 9 月份, 知乎个人注册用户总数超过了 1 亿[EB/OL]. [2018-12-24]. http://www.sohu.com/a/193351816_812860.
- [111] 曹雨骋, 李浩鸣. 科普网站的社交功能对科技传播的影响研究——以知乎网为例[J]. 出版广角, 2015(6): 79-81.
- [112] 李世霞, 田至美. 基于百度指数的旅游目的地网络关注度影响因素分析——以青岛为例[J]. 首都师范大学学报(自然科学版), 2014(01):56-59.
- [113] 郝光兆. Python 知识自动问答系统的研究与实现[D].河北:河北工程大学,2019.
- [114] 王仁武, 陈川宝, 孟现茹. 基于词向量扩展的学术资源语义检索技术[J]. 图书情报工作, 2018, 62(19):1-8.

- [115] Carretero S . Mapping of effective technology-based services for independent living for older people at home[J]. JRC Working Papers, 2015(8):21-22.
- [116] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer Science, 2013(6):1-9.
- [117] 李昊. 基于 Incopat 数据库的合肥工业大学专利分析[J]. 内蒙古科技与经济, 2019(7): 143-146.
- [118] Su Jing, Dai Qingyun, Guerin Frank, Zhou Mian. BERT-hLSTMs: BERT and hierarchical LSTMs for visual storytelling[J]. Computer Speech & Language, 2021(1):67-67.
- [119] Devlin J , Chang M W , Lee K , et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. BMC genomics, 2018(1):2-7.
- [120] Moussa Marmar, Măndoiu Ion I. Single cell RNA-seq data clustering using TF-IDF based methods[J]. BMC genomics, 2018, 19(6)1-9.
- [121] 汪静,罗浪,王德强.基于 Word2Vec 的中文短文本分类问题研究[J].计算机系统应用,2018,27(05):209-215.
- [122] Aparajit Narayan, Elio Tuci, Frederic Labrosse, Muhanad H. Road detection using convolutional neural networks[J]. ECAL 2017: The Fourteenth European Conference on Artificial Life, 2017(29):1-9.
- [123] Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, Yanbo Gao. A fully trainable network with RNN-based pooling[J]. Neurocomputing, 2019(2):338-338.
- [124] 阮光册, 夏磊. 基于主题模型的检索结果聚类应用研究[J]. 情报杂志, 2017, 36(03) : 179-184.
- [125] Ni Tianjiao, Qiao Minghao, Chen Zhili, Zhang Shun, Zhong Hong. Utility-efficient differentially private K-means clustering based on cluster merging[J]. Neurocomputing, 2020(5):1-9.
- [126] Rousseeuw P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis[J]. Journal of Computational and Applied Mathematics, 1987, 20: 53-65.
- [127] 吴广建,章剑林,袁丁.基于 K-means 的手肘法自动获取 K 值方法研究[J].软件,2019,40(05):167-170.
- [128] 吴江, 侯绍新, 靳萌萌, 等. 基于 LDA 模型特征选择的在线医疗社区文本分类及用户聚类研究[J]. 情报学报, 2017, 36(11) : 1183-1191.
- [129] 邹蔚. 技术有效需求与供给的机理研究[D].湖北:武汉理工大学,2006.
- [130] 佚名. 联合国教科文组织发布 1996 年世界科学报告[J].中国科技信息,1996(11):38-40.
- [131] PRIEMUS H.System innovation in spatial development:current dutch approaches[J].European Planning Studies, 2007, 15 (8) :991-1006.
- [132] 王在花,景丽.CHI 专利评价指标对我国构建企业知识产权综合评价体系的启示[J].知识经济,2020(01):14+16.

- [133] Weiyang Yu, Songjia Zhang, Huitao Shen, Shaodong Yu, Teng Chen. Air Quality Evaluation of Bo Hai Coastal Region Based on Entropy Weight Method[J]. Journal of Coastal Research, 2019, 94(1):10-12.
- [134] 黄文馨.产品生命周期的研究[J].商业研究,2003(17):13-15.

攻读硕士学位期间所发表的学术论文

- [1] 黄鲁成,郝亚丽,苗红,袁菲.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. (见刊)
- [2] 黄鲁成,郝亚丽,李晋,苗红.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. (见刊)
- [3] 黄鲁成,郝亚丽,苗红.我国养老科技领域发展研究[J].科技导报,2019. (录用)

致谢

时光飞逝，三年的研究生生涯即将结束，回顾三年的校园生活，仍历历在目，收获了知识，无论是专业技能、语言表达还是科学研究方面都成长很快，这也将是我未来人生路上不可或缺的宝贵财富。

首先，我要感谢我的导师黄鲁成，在求学期间，不管从学位论文和期刊论文的选题、撰写、修改到完成，都倾注了黄老师的心血。此外，黄老师对学术严谨认真的态度也在影响着我，让我在未来的学习、工作和生活中都养成了良好的习惯。感谢黄老师对我们学业上的指导，在生活上的关心。

同时，也要感谢团队中的吴菲菲老师、苗红老师、娄岩老师、李欣老师、杨早立老师、嵇灵老师和袁菲老师，感谢老师们在课堂上专业技能的传授，在团队汇报中对我论文提出的宝贵意见，在写论文遇到困难时，对我的指导和帮助。感谢我的舍友，在生活中遇到困难，可以第一时间帮助、鼓励和支持我，还愿意聆听我的埋怨、牢骚，与你们一起学习、娱乐是研究生中最美好的时光。感谢我的同门张家欣和葛豪，不管是学习还是生活中，在我需要帮忙的时候，都会帮助我。感谢同门李晋师兄，在学习上愿意给我传授自己的经验和知识，也会耐心的给我提出宝贵的意见。感谢团队的所有师兄师姐们、18级的管科同学们对我的帮助和支持，让我感受到了温暖，愿各位以后所得皆所期。

我要特别感谢我的家人，感谢您们对我所做的一切决定给予的无条件支持与鼓励，感谢您们成就了现在的我。

最后北京工业大学，感谢经济与管理学院，“不息为体，日新为道”，未来我将怀着感恩之心继续前进。

