

基于 K - means 聚类的老年人养老服务需求识别*

张馨月 宋泽宇 田晓伟 韩 霜 孔 杨

(滨州医学院卫生管理学院 烟台 264003)

〔摘要〕 目的/意义 精准识别老年人养老服务需求,为调整养老服务供给提供参考。方法/过程 利用八爪鱼采集器在网络平台收集养老服务需求相关文本数据 1 624 条,运用 K - means 算法进行聚类分析,确定最佳聚类数量;结合词云图分析结果,精准识别各类养老需求。结果/结论 养老服务需求分为医疗照护类、生活照料类、预防保健类、精神慰藉类、环境安全类 5 个方面。政府、社会组织、医疗机构和养老机构等应对现有资源进行分类整合,通过政策引导、专业培训、社会动员等多种途径,协同构建全方位、多层次的养老服务体系,以满足老年人养老需求。

〔关键词〕 养老服务;老年人需求;聚类分析;K - means 算法

〔中图分类号〕 R - 058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673 - 6036.2025.02.007

Identification of Elderly Care Service Needs Based on K - means Clustering

ZHANG Xinyue, SONG Zeyu, TIAN Xiaowei, HAN Shuang, KONG Yang

School of Health Management, Binzhou Medical College, Yantai 264003, China

〔Abstract〕 **Purpose/Significance** To accurately identify the needs of elderly care services, and to provide references for adjusting the supply of elderly care services. **Method/Process** The Octopus collector is used to collect 1 624 pieces of text data related to the needs of elderly care services on the network platform, and cluster analysis is carried out by K - means algorithm to determine the optimal number of clusters. Combined with the results of word cloud analysis, various needs for elderly care are accurately identified. **Result/Conclusion** The demands for elderly care services are divided into 5 categories: medical care, life care, preventive health care, spiritual comfort, and environmental safety. In order to meet the needs of the elderly, it is necessary for the government, social organizations, medical institutions and elderly care institutions to classify and integrate existing resources, and jointly build a comprehensive and multi - level elderly care service system through policy guidance, professional training, social mobilization and other means.

〔Keywords〕 elderly care services; needs of the elderly; cluster analysis; K - means algorithm

〔修回日期〕 2024 - 07 - 30

〔作者简介〕 张馨月,硕士研究生;通信作者:孔杨,博士,教授。

〔基金项目〕 国家自然科学基金项目(项目编号:72274023);国家级大学生创新创业项目(项目编号:202310440105);山东省社会科学规划研究项目(项目编号:22CGLJ01)。

1 引言

受个体健康状况、生活方式等多种因素的影响,老年人的养老需求呈现出显著的差异性和隐蔽性,导致服务提供方难以精准把握其真实需求。随着时间的推移,养老需求不断变化,养老服务供需错位等问题日益突出^[1]。因此,精准识别老年人的需求,并根据不同需求对现有多方资源进行整合分类,成为服务供给的首要任务,直接关系到服务设置的合理性、供需匹配的准确性以及养老事业的可持续发展^[2]。

在学术研究中,养老需求的识别与分类方法已趋于多元化。学者们借助人口统计学、社会学、心理学等多学科交叉研究方法,识别并评估老年人的养老需求。其中,马斯洛需求层次理论作为需求分类的基本理论框架,将老年人服务需求细分为生活需求、生理需求和精神照顾需求^[3-4]。王建云等^[5]运用 Kano 模型,区分了“需要-想要”所表现的“依赖性-期待性”,并通过年龄分类对社区居家养老的不同需求水平进行量化分析,以缓解供需错位问题。顾立等^[6]基于自理能力分类,对不同自理能力等级的老年人健康服务需求及其优先满足排序进行深入研究。

然而,既有研究多基于小范围地区的问卷调查和案例分析等,样本的局限性导致研究结果的代表性和普适性受限。如果以普查的形式大范围调查各地区老年人的养老服务需求,将耗费大量人力、物力资源,且难以实现。鉴于此,本研究借助信息化手段弥补样本来源范围的局限,将聚类分析方法引入养老领域,通过科学模型算法得出养老服务需求识别与分类结果,可以避免因个体差异和主观因素造成的偏差,为服务提供方满足老年人多样化需求提供理论支撑。

2 资料与方法

2.1 资料来源

利用八爪鱼采集器,基于百度搜索引擎,围绕核心关键词“老年人”“养老需求”“医养结合”“养

老服务消费”进行数据收集。为了广泛获取老年人及其亲属在不同语境下对养老服务的直接或间接反馈,采用两种检索策略:一是单独使用每个关键词进行检索,以获取各关键词下的独立信息;二是将这些关键词进行不同方式的组合,以获取各关键词共同指向的养老服务需求信息。为进一步增强数据集的代表性和覆盖范围,遵循以下原则选取专注于医药卫生或养老领域的公众号作为补充数据源。一是权威性,优先选取由官方机构背书、知名医疗机构运营、行业协会指导或具有深厚行业背景的团队维护的公众号,其发布内容往往经过多重审核,信誉度高,信息来源可靠。二是专业性,确保所选公众号聚焦于医药卫生或养老领域,能够提供深入、专业的内容,包括但不限于政策解读、行业动态追踪、学术研究成果分享、健康科普等,以满足信息需求。三是活跃度,选择更新频率高、用户互动活跃、读者反馈良好的公众号,其通常能及时捕捉并传播最新资讯,确保信息的时效性和实用性。据此选取“中国老年人才”“养老先锋号”等作为补充。最终,爬取共计 1 624 条需求导向的文本数据。这些数据长度差异显著,最长包含 4 885 个字符,而最短的仅有 376 个字符。

鉴于文本内容的复杂性和多样性,原始数据中不可避免地掺杂无效信息和干扰项,这些因素可能会对后续的文本聚类分析产生不利影响。因此,借助 Python 语言的 NLTK 文本处理库,对收集到的文本数据进行精细化清理和筛选。具体筛选标准如下。一是内容相关性,排除与养老需求主题无关或相关性极低的文本。二是重复内容识别,通过文本比对技术,识别并删除高度重复的信息,避免数据冗余对分析结果的影响。三是无效信息过滤,过滤文本中的停用词(如常见的介词、连词等)、标点符号及特殊字符,以提高数据处理效率与分析准确性。四是异常值处理,针对缺失值或异常值,采取科学合理的填补或剔除策略,以确保数据的完整性和一致性。五是干扰项排除,根据文本内容的实际情况,排除广告、推广信息等非实质性内容的干扰。经过上述严格的数据预处理过程,筛选后保留 1 531 条需求文本数据。

2.2 方法

聚类算法广泛应用于文本分析、商务、生物医学等多个领域，但目前尚无通用算法能够适应所有数据应用场景。针对本研究涉及的未分类但相似的文本数据，选择 K-means 算法作为主要聚类方法。作为基于划分的聚类算法之一^[7]，该算法具有高稳定性、对数据输入顺序的不敏感性以及在处理乱序数据时结果的一致性，展现出优异的聚类效果^[8]，适用于处理本研究的文本数据，并能满足对数据聚类准确性和稳定性的需求。本研究聚类分析过程如下。第 1 步，利用手肘法确定最佳聚类数 k 值。在传统的 K-means 算法中，聚类数目的确定往往依赖于人工判断，这种主观性极大影响了聚类的准确性和可靠性^[9]。本研究利用手肘法，通过分析样本聚类误差平方和（sum of squared error, SSE）与 k 值的关系图，观察关系折线图的拐点，从而确定最佳聚类数 k 值。

$$SSE = \sum_{i=1}^k \sum_{x \in m_i} |d(X, C_i)| \tag{1}$$

式中：SSE 为所有样本的聚类误差， X 为某个样本， C_i 为第 i 个质心， $d(X, C_i)$ 为样本到质心的距离。在模拟簇数 k 小于 $k_{\text{最佳}}$ 时，增大 k 可以提高各类别的聚类程度，并降低各类别的 SSE。在模拟簇数 k 大于 $k_{\text{最佳}}$ 时，各类的集聚度会随 k 的增加而逐渐趋于平稳，之后簇的中心不再发生变化，这时“肘部”所对应的数值就是最佳聚类数（ $k = k_{\text{最佳}}$ ）^[10]。第 2 步，利用 K-means 算法完成聚类分析。Word2vec 模型是一种高效的词向量算法^[11]，能够将文本数据转换成二维或高维的数字向量，K-means 算法根据 Word2vec 模型计算出的词向量之间的距离^[12]，实现对文本信息的聚类分析。

3 结果与分析

3.1 手肘法确定最佳聚类数 k 值

通过迭代获取不同 k 值下的聚类结果，并绘制样本聚类与 k 值的关系图，见图 1。当 $k > 5$ 时，折线的斜率基本保持不变且呈现出递减趋势。因此，为了达到最好的聚类效果，选择最佳聚类 k 值为 5，

即可将养老服务需求类型划分为 5 类，以识别不同老年群体的具体需求。

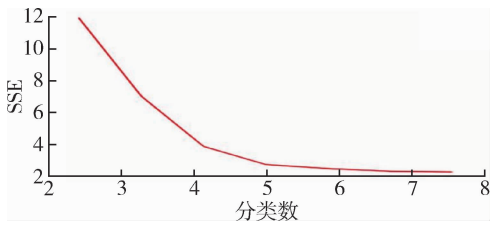


图 1 手肘法确定最佳聚类数 k 值

3.2 聚类结果

借助 Python 语言的 sklearn 库实现 K-means 算法，获得 5 类养老服务需求的聚类结果。为了验证聚类效果的合理性，结合 Matplotlib 绘图工具对聚类结果可视化处理，见图 2，5 个不同颜色的散点分别代表 5 类文本，同类颜色间的散点分布呈集中趋势，不同类颜色间的散点分布不存在明显偏差，即不同类别之间清晰分离，而同类别内的数据点则紧密聚集。结果表明，本研究所采用的聚类方法不仅有效区分了不同类型的养老服务需求，而且同类内部的相似度也达到较高水平，可认为聚类效果满足研究需求。

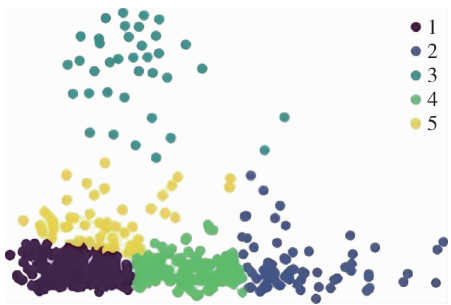


图 2 Matplotlib 工具聚类效果

进一步采用词云图可视化工具直观展示不同类别的文本内容，不仅能够揭示文本之间的共同点和差异点，还能更直观地观察和比较文本数据，见图 3。其中图 3（a）中“医疗卫生”“医疗机构”“医护人员”等高频词汇显著，突显了老年群体对医疗卫生方面养老服务的高度关注。该群体期望获得由专业医疗人员提供的全面且高质量的医疗照护服务。图 3（b）中“日常生活”“上门服务”“饮食

起居”等关键词尤为突出，反映了老年群体的主要服务需求集中在家政服务、老年助餐、代购物品等生活照料服务方面。随着老年人身体功能的衰退，其越来越依赖家庭和社会力量满足这些基本生活需求。图 3（c）中老年人对疾病预防方面的关注尤为明显。在健康中国战略的大背景下，老年人愈发重视预防保健，期望所在社区卫生服务中心能够基于自身的健康数据，制定个性化的卫生干预方案，为疾病的预防和治疗提供科学依据。图 3（d）中越来

越多老年人具有心理服务需求意愿，主要体现在对子女和亲人陪伴与关注的渴望，希望获得来自家庭的温暖与支持，同时也希望在参与社区活动的过程中，不仅能够获得社交满足，还能在自身能力范围内培养兴趣爱好，丰富精神生活。图 3（e）中“安全隐患”“基础设施”“消防安全”“发生意外”等关键词比较突出，老年人希望相关部门可以定期进行安全排查，对公共场所的基础设施升级改造，减少摔倒等意外伤害发生的可能。

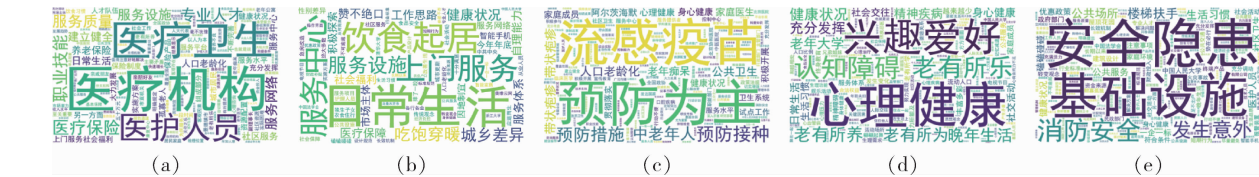


图 3 高频词汇在词云可视化中的分布

根据词云图分析结果，结合前文图 2 聚类效果图，系统归纳和总结养老服务需求，见表 1。当前老年人的养老服务需求主要聚焦于与日常生活相关的传统型服务。相比之下，对于心理健康支持以及环境适老化改造等深层次、个性化的需求，关注度较低。

表 1 养老需求类型数量及图像表示

类型序号	文本数量（条）	需求文本特征
1	514	医疗照护
2	436	生活照料
3	258	预防保健
4	129	精神慰藉
5	194	环境安全

4 讨论及建议

4.1 构建医养结合养老体系，精准满足老年人医疗照护需求

医疗照护类需求尤为关键，涵盖了定期健康检查、专业护理服务等诸多层面。应积极推进医疗机构与养老机构紧密合作，切实满足老年人的基础医疗需求。同时，建立完善的老年人健康档案系统，定期开展健康检查和医疗咨询活动^[13]，

为行动不便、失能、患有慢性病以及高龄的老年人提供便捷、有针对性的家庭病床服务和上门巡诊服务^[14]。这样的综合照护体系不仅能够有效提升老年人的生活质量，也是实现健康老龄化社会的重要途径。

4.2 提供多元化生活照料服务，全面提升老年人生活质量

生活照料类需求聚焦于老年人的日常起居，包括饮食、起居等基础性生活照料。随着年龄增长，许多老年人因身体功能逐渐衰退而难以独立应对日常活动，需要得到他人的协助。然而，目前社区提供的生活照料主要集中于饮食服务，对家政服务、代购等多元化需求涉足较少。应在政策层面发挥引导作用，丰富日间照料服务类型，加强社区服务功能建设，鼓励志愿者参与老年照料服务^[15]；引导物业、企业、零售服务商等多元化服务机构拓展服务范围，为老年人提供生活用品代购、代收代缴等基本生活照料服务，确保老年人的日常生活需求得到全方位、多层次的满足。

4.3 强化预防保健教育，助力老年人健康自我管理

老年人对自我健康的关注度日益提高，预防保健类需求不仅体现了老年群体对健康生活的追求，

更是其积极应对老龄化挑战的具体体现。《“十四五”国家老龄事业发展和养老服务体系规划》^[16]明确指出,加强老年健康教育和预防保健等举措是完善老年健康支撑体系的关键。服务供给方应充分利用多元化的传播渠道,广泛开展预防保健活动^[17],如计划免疫、营养指导等,对老年人及其看护人员进行有效宣传,包括慢性病预防、就医注意事项等。通过系统的健康教育、个性化的疾病预防指导等服务,使老年人更全面地了解自身健康状况,提升健康素养。

4.4 多维度强化精神慰藉,提升老年人心理关怀水平

随着年龄增长,老年人更易感受到孤独与寂寞,这种心理状态对其生活质量和整体幸福感可能产生不利影响。为改善老年群体的心理状况,政府应加大宣传教育力度,强化家庭成员对老年人心理关爱的认识,使老年人感受到家庭的温暖;同时,社区和养老机构应发挥桥梁作用,积极与高等院校中的心理学、社会工作等相关专业合作^[18],通过志愿服务、公益有偿服务等多种形式,为老年人尤其是有特殊困难的老年人提供心理咨询、情绪疏导等心理关爱服务;此外,还可以结合老年人的兴趣爱好和身体状况,设计多元化的文娱活动,以帮助老年人缓解孤独感和焦虑情绪^[19]。

4.5 构建安全舒适环境,全面满足老年人安全需求

环境安全类需求主要体现在老年人对居住环境安全性的关注。为有效降低老年人在日常生活中可能遭遇的安全风险,确保其生命安全和身体健康,社会组织应积极建设适老化服务设施,同时依据地方实际条件,合理新建社区卫生服务站和适合老年人的锻炼场所^[20],促进社区健康服务资源的全面共享和高效利用。此外,还应积极推广和普及老年人居家适老化服务技术,组织专业的社区护理人员细致评估老年人的居住环境,及时发现并消除安全隐患。

综上所述,为了全面且精准地满足养老服务需求,需要政府、社会组织、医疗机构和养老机构等充分发挥各方资源优势,提高资源利用效率,通过

政策引导、资金投入、专业培训、社会动员等多种手段,推动养老服务从“有”向“优”的转变,实现老有所养、老有所乐。

5 结语

本研究通过聚类分析法,将大量养老需求数据细化为医疗照护类、生活照料类、预防保健类、精神慰藉类、环境安全类 5 个方面,识别并剖析老年群体复杂多元的服务需求,进而提出分类指导策略,旨在帮助社区和相关服务提供者更精准地识别并满足老年人的多样化需求,从而全面提升养老服务供给的针对性。本研究未明确限定时间范围,所收集数据跨越不同时间段。老年人养老服务需求具有动态性与复杂性,其可能随时间的推移及社会环境的变化而动态调整,未来将进一步探索不同时间范围内养老需求的变化趋势,以期揭示养老需求随时间变化的动态特征,增强研究成果的全面性和可靠性。为持续优化养老服务策略、完善养老服务体系提供理论基础与数据支撑。

作者贡献: 张馨月负责研究设计、数据分析、论文撰写;宋泽宇、田晓伟、韩霜负责数据收集;孔杨负责研究设计、论文修订。

利益声明: 所有作者均声明不存在利益冲突。

参考文献

- 1 张帆,魏川林,刘小春.农村养老服务精准供给的理论逻辑、制约因素及实现路径[J].安徽农学通报,2022,28(5):38-44.
- 2 陈帅,梁昌勇,赵树平.服务供应链模式下居家养老服务提供商组合选择研究[J].科技促进发展,2019,15(10):1106-1112.
- 3 崔树义,杨素雯,田杨.供需视角下社区居家养老服务提质增效研究——基于山东省 1200 名老年人的调查[J].山东社会科学,2020(9):127-133.
- 4 孙霞,于兆丽,薛雅卓,等.互联网+居家医养结合养老服务现状与服务需求研究[J].护理研究,2020,34(2):318-321.
- 5 王建云,钟仁耀.基于年龄分类的社区居家养老服务需求层次及供给优先序研究——以上海市 J 街道为例

- [J]. 东北大学学报 (社会科学版), 2019, 21 (6): 607 – 615.
- 6 顾立, 刘晓楚, 陈婷, 等. 基于自理能力分类的四川省养老机构老年人健康服务需求研究 [J]. 医学与社会, 2022, 35 (3): 30 – 35.
- 7 方自建. 基于爬虫技术和文本聚类的 COVID – 19 舆情分析 [D]. 湘潭: 湘潭大学, 2021.
- 8 李鹏, 李强, 马味敏, 等. 基于 K – means 聚类的路面裂缝分割算法 [J]. 计算机工程与设计, 2020, 41 (11): 3143 – 3147.
- 9 高海淑, 张玉敏, 吉兴全, 等. 基于场景聚类的主动配电网分布鲁棒综合优化 [J]. 电力系统自动化, 2020, 44 (21): 32 – 41.
- 10 周善添, 苏凌云, 张衡, 等. 基于 K – means 算法的低碳疫苗配送中心选址研究 [J]. 物流工程与管理, 2022, 44 (5): 96 – 100.
- 11 苏梦泽, 耿骞, 刘志刚, 等. 基于属性分析及文本聚类的我国养老政策演化研究 [J]. 人口与发展, 2023, 29 (5): 133 – 144.
- 12 宋赫, 严李强, 肖杨, 等. 基于 K – means 的时参模型温室控制模式分析方法研究 [J]. 新型工业化, 2022, 12 (1): 41 – 44, 58.
- 13 葛延风, 王列军, 冯文猛, 等. 我国健康老龄化的挑战与策略选择 [J]. 管理世界, 2020, 36 (4): 86 – 96.
- 14 程雅娟, 王尚才. 高龄老人居家养老健康管理需求及制约因素研究 [J]. 中国医药, 2021, 16 (3): 448 – 451.
- 15 罗梅. 志愿服务下“以老助老”模式及其对策分析——以 G 别院为例 [J]. 老龄化研究, 2023, 10 (3): 777 – 783.
- 16 “十四五”国家老龄事业发展和养老服务体系规划 [EB/OL]. [2024 – 02 – 21]. https://www.gov.cn/gongbao/content/2022/content_5678066.htm.
- 17 沈纪川, 王鸣. 重点关注老年人群助力“全民健康”行动 [J]. 疾病监测, 2020, 35 (8): 687 – 692.
- 18 黄祖良. 基于嵌入性治理视角的社区养老服务模式研究 [D]. 北京: 中央财经大学, 2022.
- 19 王智莉, 代俏俏. 小组工作介入半失能老人精神慰藉研究——以“耆乐融融”项目为例 [J]. 理论观察, 2023 (12): 90 – 94.
- 20 王安琪, 陈欣, 薛宇. 城市社区医养结合养老场所建设改造策略探讨 [J]. 南京医科大学学报 (社会科学版), 2024, 24 (2): 150 – 154.

(上接第 41 页)

- 41 HICKEY B A, CHALMERS T, NEWTON P, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review [J]. Sensors, 2021, 21 (10): 3461.
- 42 SCHMITTER – EDGECOMBE M, PARSEY C, LAMB R. Development and psychometric properties of the instrumental activities of daily living: compensation scale [J]. Archives of clinical neuropsychology, 2014, 29 (8): 776 – 792.
- 43 OUCHI Y, AKANUMA K, MEGURO M, et al. Impaired instrumental activities of daily living affect conversion from mild cognitive impairment to dementia: the osaki – tajiri project [J]. Psychogeriatrics, 2012, 12 (1): 34 – 42.
- 44 AKL A, TAATI B, MIHAILIDIS A. Autonomous unobtrusive detection of mild cognitive impairment in older adults [J]. IEEE transactions on bio – medical engineering, 2015, 62 (5): 1383 – 1394.
- 45 DAWADI P N, COOK D J, SCHMITTER – EDGECOMBE M. Automated clinical assessment from smart home – based behavior data [J]. IEEE journal of biomedical and health informatics, 2016, 20 (4): 1188 – 1194.
- 46 GORDON M L, GATYS L, GUESTIN C, et al. App usage predicts cognitive ability in older adults [C]. New York: The 2019 CHI Conference on Human Factors in Computing Systems Association for Computing Machinery, 2019.
- 47 RAUBER J, FOX E, GATYS L A. Modeling patterns of smartphone usage and their relationship to cognitive health [EB/OL]. [2024 – 01 – 03]. <https://www.semantic-scholar.org/paper/Modeling – patterns – of – smartphone – usage – and – their – to – Rauber – Fox/1f9c2e166ebaef0cb200c5f5e70f10aa4e050bc>.
- 48 CHEN R, JANKOVIC F, MARINSEK N, et al. Developing measures of cognitive impairment in the real world from consumer – grade multimodal sensor streams [C]. Anchorage: The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.