

引用格式:周隽如,刘智勇.医学人工智能领域专利技术主题发展态势研究[J].世界科技研究与发展, 2024, 46(4): 497-510.

医学人工智能领域专利技术主题发展 态势研究*

周隽如 刘智勇**

(华中科技大学同济医学院医药卫生管理学院,武汉 430030)

摘要:对医学人工智能领域 1971—2022 年的专利文本进行文本挖掘,揭示其技术主题内容和演化趋势,分析技术研发热点和发展态势,有助于为科研人员梳理技术发展脉络,进而为未来研究和应用提供参考和借鉴。本文首先获取德温特专利数据库中 14184 条医学人工智能领域相关专利数据,并进行数据清洗,结合生命周期理论将其划分为三个阶段(萌芽期、发展期、快速发展期)。随后,使用 BERTopic 主题模型对专利文本进行主题识别,并通过计算技术主题对之间的语义相似度和主题过滤的方法对技术主题进行演化分析。结果显示,共识别出萌芽期主题 5 个、发展期主题 9 个、快速发展期主题 29 个,医学人工智能技术正处于快速发展阶段。其中,基础技术(信号处理与分析技术、图像处理与计算机视觉技术、数据挖掘技术)产生于萌芽期和发展期,并在发展期和快速发展期逐渐成熟,形成了分支领域技术;需求与技术的进步推动了基础技术的分化,技术之间的融合将会产生新的技术。最后,提出未来应加强对医学人工智能领域基础关键技术的研究和创新,并持续关注医学人工智能领域重要实际需求背后的技术问题,以助力技术创新。此外,还应强调医学人工智能领域的技术融合发展的重要性,鼓励多领域科研人员组建交叉学科人才团队,为促进多领域技术的深度融合和前沿多维度探索奠定基础。

关键词:医学人工智能;文本挖掘;专利分析;发展态势;BERTopic 模型

DOI: 10.16507/j.issn.1006-6055.2023.09.001

Research on the Development Trends of Patent Technology Themes in the Field of Medical Artificial Intelligence*

ZHOU Junru LIU Zhiyong**

(School of Medicine and Health Management, Tongji Medical College of Huazhong University of
Science & Technology, Wuhan 430030, China)

Abstract: The text mining of patent texts in the field of medical artificial intelligence (AI) from 1971 to 2022 reveals the content and evolution trend of its technical topics, and analyzes the hot spots and development trends of technology research and development, which is helpful to provide reference and reference for scientific researchers in the context of technological

* 中央高校基本科研业务费项目(5003516062)

** E-mail: zhiyongliu@hust.edu.cn; Tel: 130-72706735

development, and then for future research and application. In this paper, 14184 patent data related to the field of medical artificial intelligence in the Derwent patent database are obtained, and the data is cleaned, and it is divided into three stages (embryonic stage, development stage, and rapid development stage) based on the life cycle theory. Subsequently, the BERTopic topic model was used to identify the subject of patent texts, and the evolution of technical topics was analyzed by calculating the semantic similarity between technical topic pairs and topic filtering. The results showed that a total of 5 embryonic themes, 9 development themes, and 29 rapid development topics were identified, indicating that medical artificial intelligence technology is in the rapid development stage. Foundational technologies such as signal processing and analysis, image processing and computer vision, and data mining emerged during the early and development stages, gradually maturing during the development and rapid development stages, leading to the formation of specialized technological domains. Both demand and technological advancements drove the differentiation of foundational technologies. The convergence of technologies is expected to generate novel advancements. Medical artificial intelligence technology is currently in a rapid stage of development. Finally, it is proposed that the research and innovation of basic and key technologies in the field of medical artificial intelligence should be strengthened in the future, and the technical issues behind the important practical needs in the field of medical artificial intelligence should be continuously paid attention to help technological innovation. Additionally, the significance of technological convergence and development in the field of medical AI should be emphasized, encouraging multidisciplinary teams comprising researchers from various fields, to lay the foundation for promoting deep integration and multidimensional exploration of interdisciplinary technologies at the forefront.

Keywords: Medical Artificial Intelligence; Text Mining; Patent Analysis; Development Trend; BERTopic Model

医学人工智能是指将计算机视觉、语音识别、自然语言处理、机器学习等人工智能技术应用于临床和医学研究领域,为临床和研究领域的医学专业人员提供支持。随着人工智能技术的加速成熟,其在医学实践领域的应用场景不断丰富,为疾病检测、诊断及治疗带来深刻变革,在医学多领域产生了颠覆性影响。医学人工智能已成为人工智能的重要研究领域,世界主要发达国家均出台了相关国家规划和科技政策,围绕基础研究、核心技术等方面展开部署,力图在新一轮国际科技竞争中抢占发展先机并掌握主导权。

近年来,我国高度重视并积极推进人工智能在医学领域的发展,先后发布《“十三五”国家科技创新规划》《“十三五”卫生与健康科技创新专项规划》《新一代人工智能发展规划》等战略规划;制定《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》等实施方案和行动计划;通过“数字诊疗装备研发”重点专项等国家科技计划,重点聚焦医学知识图谱、影像分析、辅助诊断等科技研发方向,引领、推进我国医

学人工智能的进一步发展。

专利文献作为技术创新能力的重要表现形式,蕴含着极大的经济价值及技术价值,是理想的技术挖掘的数据来源^[1]。专利技术主题是专利文献揭露技术内容的主题和核心^[2],开展面向医学人工智能领域的专利文献主题识别与演化研究,可以了解该领域的技术热点、创新方向和发展趋势。随着时间的推移,医学人工智能技术不断演进和创新,新的主题不断涌现,而旧的主题可能逐渐淡化或融合进新的主题中。技术主题识别与演化的研究可以帮助科学家、工程师和决策者更好地了解医学人工智能领域的技术发展情况,梳理技术发展脉络,为技术创新和决策提供指导。

目前,医学人工智能领域的相关研究呈现出多样性和广泛性,主要聚焦于特定领域、特定技术和专利特征项的分析。首先,在特定领域研究中,学者已关注医学人工智能技术在医学信息^[3]、中医药^[4,5]、放射学^[6]、辅助生殖^[7]和重症医学^[8]等医学领域的应用。这些研究介绍了相关技术在医学各领域中的实际应用,及其为不同领域的医疗

实践提供的智能支持。其次,针对特定技术的研究对医学人工智能的发展也起到了关键作用。这些研究主要关注如何将先进的技术应用于医学领域,以提升诊断准确性和治疗效果,如:医学影像分析^[9,10]、药物发现^[11,12]、疾病预测^[13,14]、辅助诊断^[15]、智能语音识别与语义理解^[16]等技术。与此同时,研究人员还基于专利文献分析了医学人工智能领域中的专利特征项,包括专利申请人、专利申请国和IPC分类号等。研究发现西门子、飞利浦等欧美跨国大企业具备较强的竞争优势,联影科技、平安集团等具备自主研发能力的中国科技公司正迎头赶上^[17];美日荷三国引领全球发展,是医学人工智能领域主要技术来源地^[18]。这些分析有助于了解不同地区和组织在医学人工智能领域的创新贡献。此外,一些研究从文献计量学的视角,采用词共现分析、知识图谱^[8]等方法,揭示了医学人工智能领域技术开发热点,为了解领域内的前沿动态和研究趋势提供了有益信息;针对实际应用,有学者进行了系统性回顾^[19],总结了医学人工智能领域的实践活动和取得的成果^[20];部分关于未来展望的研究^[21]也对该领域可能的创新方向和挑战进行了探讨,为医学人工智能领域的发展提供了战略性的引导。

然而,现有相关研究多为针对某一子领域、某一技术主题或某一主体的研究,未能全面反映医学人工智能领域技术的历史脉络和整体态势,其整体发展历程及主题演化有待进一步梳理。BERTopic主题模型作为一种新颖且具有优势的内容分析方法,能够有效挖掘领域科技文献的潜在主题^[22,23]。基于此,本文在调研总结现有主题识别与演化分析方法的基础上,引入生命周期理论对其发展阶段进行划分,并使用BERTopic主题模型对1971—2022年医学人工智能领域的专利

文本进行技术主题识别,并利用余弦相似度计算主题对之间的语义相似度,保留强关联技术主题对进行演化分析。一方面探究了基于神经网络进行主题建模的主题识别方法在医学人工智能领域专利文本中的有效性,拓展了BERTopic的应用范围;另一方面实现了自动化的专利文本主题识别与演化分析,不再需要先验知识和人为主观判断主题数目。通过对近60年来医学人工智能领域专利的主题进行识别和演化分析,从而深入厘清医学人工智能技术主题分布特征和变化规律,为我国医学人工智能领域的发展提供一定的借鉴和参考。

1 研究理论与方法

1.1 生命周期理论

生命周期理论源于个体发展模型,是指一个生命体从出生到死亡所经历的各个阶段。经延伸和扩展后,生命周期理论被广泛应用于产品管理、行业发展、客户关系管理和信息计量学等诸多领域^[24]。生命周期理论是事物发展的基本理论,事物的发展过程是从产生、发展、成熟到消亡的过程,旧事物的消亡同时伴随着新生事物的诞生。学科领域生命周期一般分为萌芽期、发展期、成熟期和衰退期(或转型期)^[25]。技术主题的演化同样遵循这一规律,专利技术在理论上按照技术萌芽期、技术发展期、技术成熟期和技术衰退期四个阶段周期性变化。因此,本文基于生命周期理论对医学人工智能领域发展阶段进行划分,有助于梳理其演进脉络与发展趋势。

1.2 BERTopic主题模型

BERTopic模型是Grootendorst于2022年提出的一种使用了预训练模型的主题模型,它在各个领域的主题建模中取得了良好的主题识别效

果^[26-30]。该模型允许对每个术语进行语义表示,从而防止词汇不匹配问题,且无需人为设置主题数量,从而最大限度地减少了手动过程^[31]。此外, BERTopic 允许噪声主题,这可以防止将不相关的文档分配给其他主题,从而提高主题表示。综上所述, BERTopic 不仅能够自动化地有效挖掘大规模文本中的隐含主题,还能更好地保留文本语料中的语义内容,有利于更合理地揭示文本中主题及其演化规律。

使用 BERTopic 主题模型进行建模按照以下三个步骤进行:1)提取文档嵌入,2)聚类嵌入,3)使用基于类的 TF-IDF (c-TFIDF)创建主题表示。

首先使用 BERT 模型提取文档嵌入,将文本数据转换为词向量表示。然后,利用 UMAP 进行降维操作,从而避免维数灾难,并利用基于密度的 HDBSCAN 算法来生成语义相似的文档聚类。在这里, HDBSCAN 允许存在不相关的文档,即离群值。因此,其中一个主题被设置为“其他”,分配不包含在不同主题中的文档。由于 HDBSCAN 是一种基于密度的聚类算法,因此根据被视为邻居的最小距离自动选择聚类的数量。最后,计算 c-TF-IDF 生成主题表示。c-TF-IDF 计算公式如下:

$$c\text{-TF-IDF}(t, d, C) = c\text{-TF}(t, d) \times c\text{-IDF}(t, C) \quad (1)$$

其中, $c\text{-TF}(t, d)$ 是术语 t 在文档 d 中的类内词频 (term frequency within class), 即术语 t 在文档 d 中出现的次数。 $c\text{-IDF}(t, C)$ 是术语 t 的类内逆文档频率 (inverse document frequency within class), 即术语 t 在类别 C 中的文档中的逆频率。 $c\text{-TF}(t, d)$ 的计算方式与传统的 TF-ID 方法相同, 它衡量了术语在文档内的重要性, 即术语在文档中出现的频率越高, 其 $c\text{-TF}$ 值越大。 $c\text{-IDF}(t, C)$ 衡量了术语在特定类别中的重要性。具体计算公式为:

$$c\text{-IDF}(t, C) = \log(Nc / (1 + n(t, C))) \quad (2)$$

其中, Nc 是类别 C 中的文档总数; $n(t, C)$ 是类别 C 中包含术语 t 的文档数。

2 研究设计

本文基于 BERTopic 模型对医学人工智能领域的技术主题进行识别和演化分析, 研究过程包括数据获取、数据预处理、主题识别与演化分析。研究流程如图 1 所示。

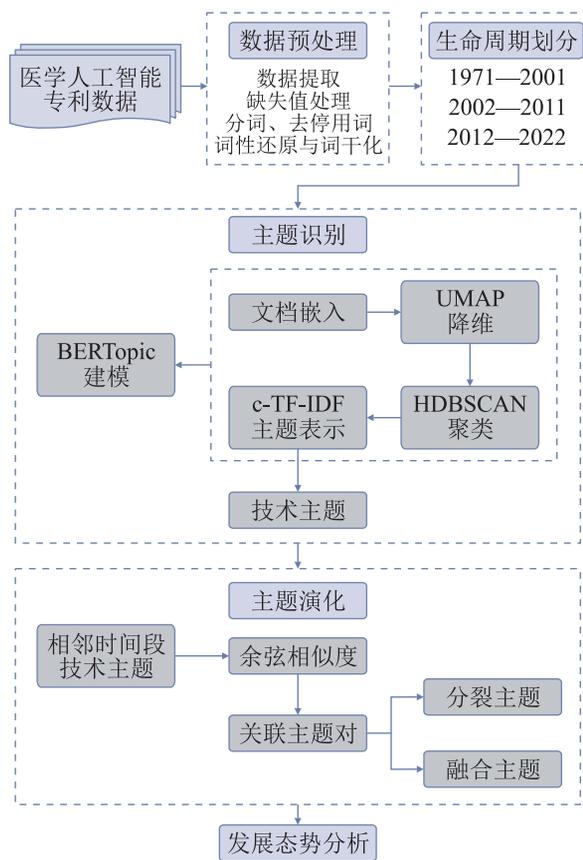


图 1 研究流程图

Fig.1 Research Flowchart

2.1 数据来源与预处理

2.1.1 数据来源

本文以德温特专利数据库 (Derwent Innovations Index, DII) 为数据来源, 通过文献调研分析构建了本文的专利检索式, $TS = (\text{“artificial intelligence*” OR “machine learning*” OR “deep learning*” OR$

“natural language processing*” OR “computer vision*” OR “neural network*” OR “pattern recognition*” OR “machine intelligence*”) AND TS = (“medicine*” OR “biomedicine*” OR “medical*”)

限定时间为2022年12月31日前,采用完整记录格式获取数据。数据检索日期为2023年5月18日,共获得专利数据14184条。

2.1.2 数据预处理

专利文本的数据预处理需要经过以下四个步骤。1)数据提取:使用Python编写正则表达式提取出txt文本格式中专利数据中的标题、摘要和时间数据,以Excel格式储存。2)缺失值处理:对包含缺失值的数据在数据库中重新检索和填充,若仍包含缺失值,则将该文本去除。3)英文分词与停用词表构建:采用Python中的NLTK库对英文文本进行分词,并在NLTK的英文停用词表的基础上根据预分词结果对停用词表进行补充,进而减少无意义的词汇对模型结果的影响。4)词性还原与词干化:使用Python中的NLTK库对英文文本进行词形还原与词干提取,经过该处理后的语料可以减少单词因时态,单复数,变形等造成噪音干扰。

2.1.3 医学人工智能领域生命周期划分

在实际研究中,生命周期的阶段通常依据年度专利数量及其增长率进行划分,结合实际情况(第一项专利的申请时间为1971年,作为起始年),本文将医学人工智能领域生命周期划分为3个阶段:第1阶段为萌芽期(1971—2001年),第2阶段为发展期(2002—2011年),第3阶段为快速发展期(2012—2022年)。从图2可见,从1971年该领域开始申请第1项专利,随后20年间,年度专利申请量保持在较低水平(50项以内)。2001年后年度专利申请量开始逐渐增长,但仍在100项以内。2012—2022年这11年时间内,医学人工智能领域从2012年的101项专利申请急剧增加为2022年的2367项专利申请,反映出医学人工智能领域正处于快速发展阶段。

2.2 BERTopic主题建模

在数据预处理的基础上,调用BERTopic、SentenceTransformer、UMAP和HDBSCAN等相关python包,使用all-MiniLM-L6-v2作为嵌入模型,利用UMAP和HDBSCAN算法进行降维和聚类。其中,n_neighbors设置为5,用于平衡局部结构与全局结构,其他参数使用默认值。

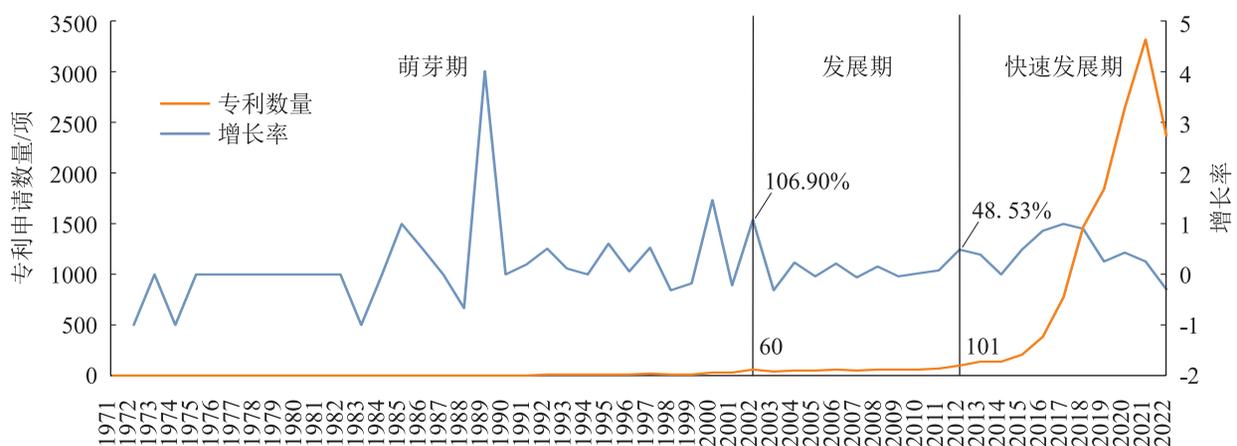


图2 专利申请数年度变化趋势

Fig.2 Annual Trend of Patent Applications

2.3 技术主题演化分析

相似度度量指标可以在一定程度上反映不同主题之间的语义关联,本文使用余弦相似度对由两个相邻时间段内的技术主题进行相似度计算,相似度大于等于所有主题对间的平均余弦相似度阈值时,则认为这两个主题之间具有相关关系。高于所有主题对平均余弦相似度的阈值设定既可以有效删除一部分弱关联,又可以最大限度地保留主题间可能存在的强关联关系。本文聚焦主题分裂和主题融合这两种主题关联模式,对医学人工智能领域的专利技术主题进行演化分析。

3 医学人工智能领域技术主题发展态势分析

3.1 技术主题识别结果分析

分别对萌芽期、发展期和快速发展期的专利数据进行 BERTopic 主题识别,共识别出萌芽期主题 5 个(表 1、图 3)、发展期主题 9 个(表 2、图 4)和快速发展期主题 29 个(表 3、图 5)。选取每个主题中选 c-TF-IDF 最大的前 10 个特征词作为主题表示,并结合领域知识,给每个主题命名。

表 1 萌芽期技术主题

Tab.1 Technical Topics at the Embryonic Stage

主题标识	主题表示
T1 数据处理技术	output、data、input、network、layer、valu、neural、imag、unit、process
T2 疾病诊断与预测技术	patien、tcancer、test、time、data、ventil、individu、compris、medic、set
T3 图像处理技术	imag、use、shape、object、point、detect、segment、pattern、surfac、featur
T4 信号处理与分析技术	pattern、recognit、signal、deform、classifi、imag、speech、use、model、system
T5 信息系统	inform、system、user、data、goal、busi、word、simul、base、use

1971—2001 年,医学人工智能领域处于萌芽期,更加关注基础技术的研究,如数据处理、图像处理、模式识别等。具体的技术主题包括数据处

表 2 发展期技术主题

Tab.2 Technical Topics at the Development Stage

主题标识	主题表示
T1 图像处理与计算机视觉技术	imag、use、point、pixel、comput、object、method、draw、descript、color
T2 医疗数据挖掘	patient、medic、inform、data、treatment、method、use、system、provid、health
T3 传感器健康监测技术	sensor、signal、patient、devic、system、measur、detect、monitor、compris、use
T4 基于神经网络的预测技术	train、data、network、method、neural、set、featur、use、predict、system
T5 模式识别与数据处理系统	processor、pattern、recognit、search、system、data、devic、memori、bu、stream
T6 细胞技术	cell、compris、andor、sampl、least、detect、peptid、bind、cultur、protein
T7 文本数处理系统	system、document、text、code、input、devic、data、comput、user、process
T8 医学影像处理与分析技术	imag、medic、method、use、train、draw、set、segment、descript、comput
T9 模式分类与识别系统	classifi、pattern、class、recognit、optim、classif、system、use、set、descript

表 3 快速发展期 TOP 10 技术主题

Tab.3 Technical Topics at the Rapid Development Stage (TOP 10)

主题标识	主题表示
T1 医学信息管理系统	medic、data、method、imag、use、draw、devic、system、patient、inform
T2 癌症诊断技术	cancer、tumor、breast、imag、cell、use、method、data、patient、model
T3 心脏疾病诊断与监测	cardiac、heart、signal、electrocardiogram、salt、ecg、arrhythmia、patient、data、use
T4 手术机器人	robot、control、arm、system、connect、surgic、drive、intellig、provid、devic
T5 超声成像技术	ultrasound、imag、ultrason、probe、fetal、use、method、obtain、draw、system
T6 语音信号处理与语音识别技术	audio、signal、voic、speech、sound、acoust、recognit、speaker、nois、method
T7 生物信号分析技术	signal、ecg、wave、brain、eeg、network、puls、neural、method、characterist
T8 口腔医学影像分析	tooth、oral、dental、teeth、point、imag、jaw、cbct、centroid、detect
T9 肝脏肿瘤影像分析	liver、tumor、segment、imag、set、cancer、ct、model、featur、fibrosi
T10 骨骼影像与骨密度分析	bone、imag、osteoporosi、joint、method、age、densiti、ray、knee、human

理技术(如:数据的输入、输出和训练等)、疾病诊断与预测技术(如:癌症诊断)、图像处理技术(如:

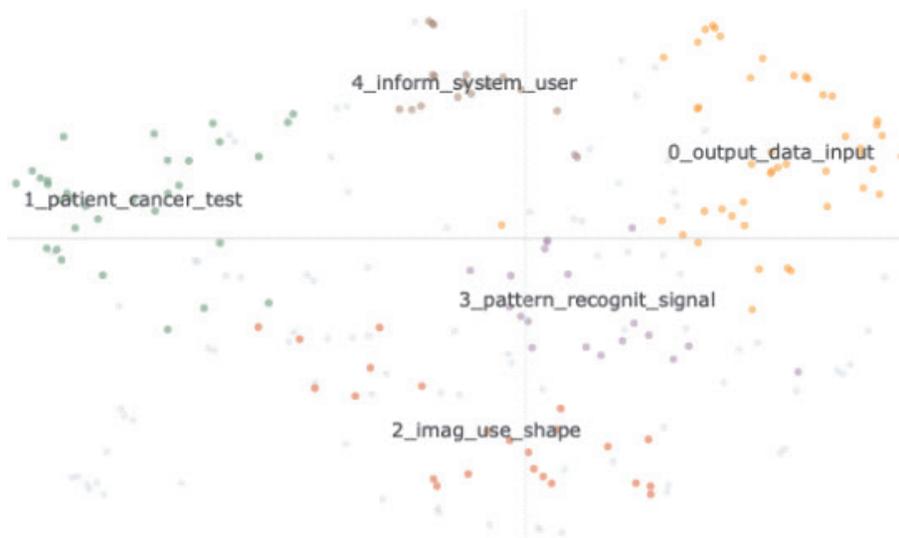


图3 萌芽期文档聚类图

Fig.3 Document Clustering Results at the Embryonic Stage

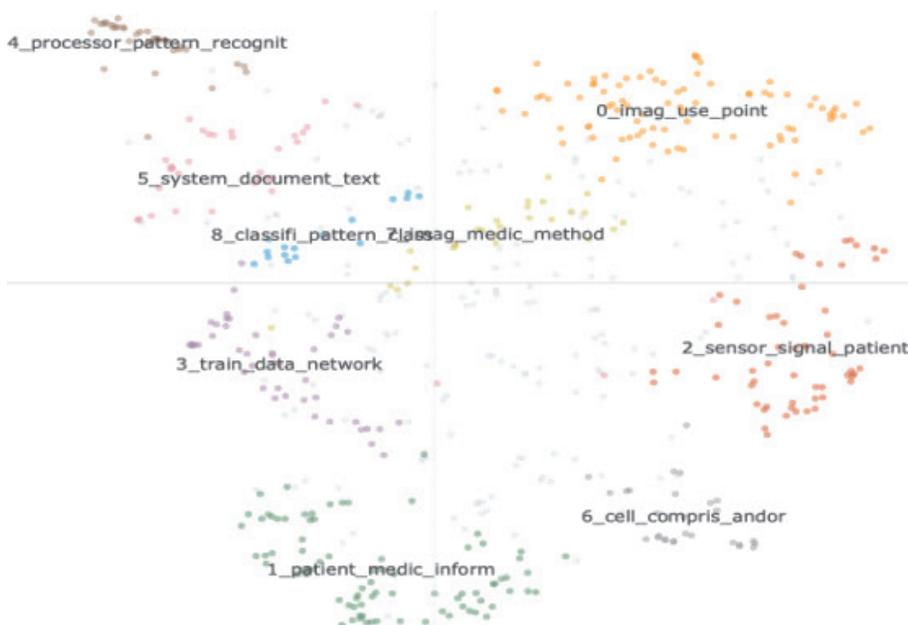


图4 发展期文档聚类图

Fig.4 Document Clustering Results at the Development Stage

图像分割、特征提取)、信号处理与分析技术(如:图像、语音等信号的识别与分类)和信息系统等。

随着时间的推移,2002—2011年,处于发展期的医学人工智能领域迎来了更多的技术发展和应用探索,涌现出了更多专利技术主题。在这一时期,图像处理技术仍是主要的技术主题之一。随

着深度学习技术的兴起,神经网络在医学图像分析、疾病诊断和预测等方面得到广泛应用。利用算法和模型,可以对医学影像进行分析和解读,从而辅助医生进行疾病诊断和治疗决策,如:细胞学宫颈癌筛查、肺结节检测等。

与此同时,随着医学数据的增长和谷歌公司

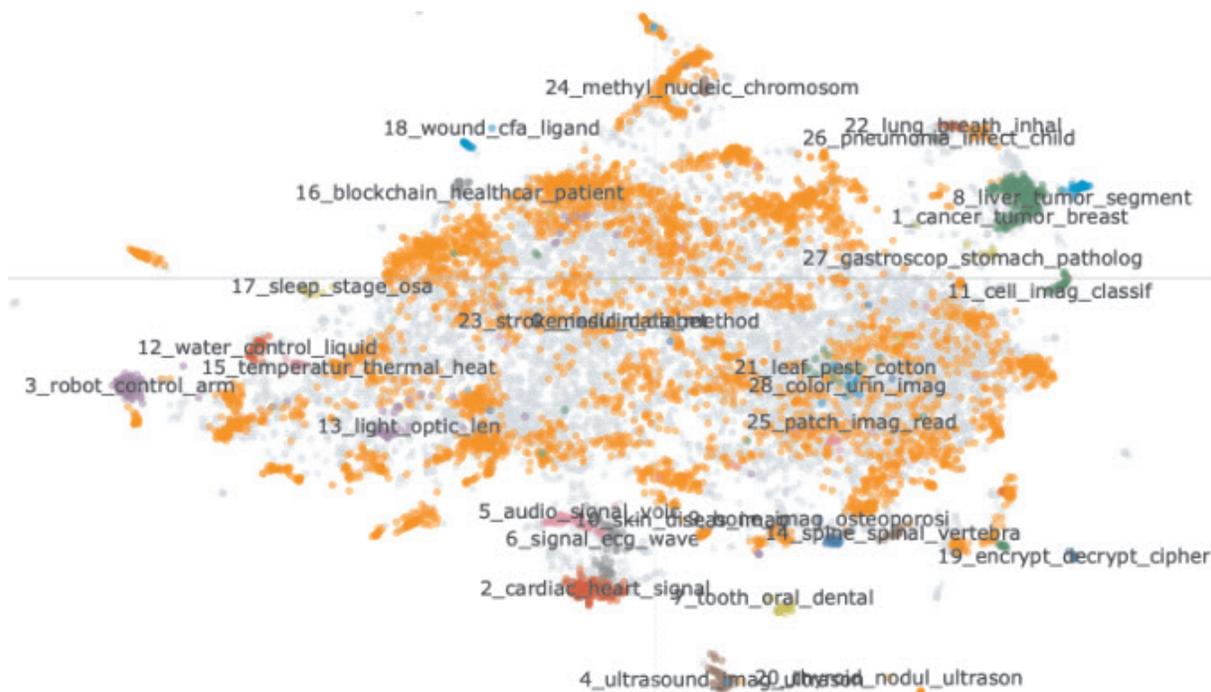


图5 快速发展期文档聚类图

Fig.5 Document Clustering Results at the Rapid Development Stage

的趋势预测模型的问世,数据挖掘和知识发现技术开始得到应用,不仅涉及基于机器学习和统计方法的数据挖掘、模式识别和预测分析,以帮助发现新的医学知识和关联,还可基于患者特征和临床数据进行个性化健康管理和治疗方案建议。数据挖掘技术通过对大量医学数据的挖掘和分析,可以发现隐藏在数据中的模式和规律,提供对疾病预测、患者风险评估和个性化治疗的支持。例如:基于电子病历数据,使用数据挖掘技术可以帮助医生有效预测死亡率、再入院率、长期住院风险和发生特定疾病的风险。

自2012—2022年,信息技术和人工智能技术的迅猛发展为医学人工智能技术发展提供了强大的助力,医学人工智能领域进入了快速发展阶段,其应用领域日益广泛,研究方向更加细化。2012年,AlexNet在ImageNet图像识别比赛中的惊人表现使得深度学习引起了广泛的关注^[32]。这一事件

标志着深度学习技术的崛起,并为医学人工智能领域的图像处理和医学影像识别提供了强大的工具和方法。2015年,美国总统奥巴马和美国国立卫生研究院(National Institutes of Health, NIH)启动了“精准医学倡议”^[33],推动了精准医疗的发展和应用。随后,基因组学和生物信息学等技术的快速发展为精准医疗提供了强有力的支持。此外,物联网技术的发展也对医学人工智能领域产生了深远影响。物联网是指通过互联网连接和交互的各种物理设备和传感器网络。它为远程监测和医疗智能设备的应用提供了新的可能性。在医学人工智能领域,物联网技术的应用推动了远程医疗、健康监测和智能医疗设备的发展,相关技术主题涵盖了与物联网相关的技术,如远程监测、数据传输和设备互联等,为医学人工智能技术的创新和应用提供了新的方向和机会。因此,在这个时期,医学人工智能技术不断演进和创新,技术主题涵

盖了医学影像识别、疾病预测、基因组学、生物信息学、精准医疗和远程医疗等多个领域。这一时期的专利技术主题的丰富性和细节性反映了医学人工智能领域在技术应用方面的快速发展。

3.2 技术主题演化情况分析

通过计算各技术主题对之间的的语义相似度,保留高于设定阈值的主题关联对作为关联主题,并通过可视化的方式展现结果,主题演化的结果如图6所示。

一些技术主题在特定时间段内分化出更具体的子领域,在医学人工智能领域的技术主题中,具有分裂演化特征的技术主题共计10个。

萌芽期的信号处理与分析技术主题在发展期分裂为模式分类与识别系统、模式识别与数据处理系统、传感器健康检测技术这三个技术主题。20世纪80年代出现的一系列基于模式匹配和特征提取的技术,如模式识别算法和机器学习方法,使得多模态信号处理与分析领域在模式分类与识

别方面取得了重要进展。随着大数据时代的到来,信号处理与分析技术需要处理海量的数据和复杂的信息,为了更好地理解、分析和利用这些数据,模式识别与数据处理系统应运而生。大数据技术的发展、云计算平台的兴起以及人工智能算法的改进都为模式识别与数据处理系统的发展提供了支持和助力。而传感器健康检测技术的出现则与医疗健康领域对健康监测和疾病预防的需求以及技术进步密切相关。传感器健康检测技术通过使用各种传感器来采集人体生理参数和环境信息,以信号处理和健康管理与医疗诊断技术为基础,从而实现对健康状况的监测和评估。

发展期的传感器健康监测技术在快速发展期又分裂为更为精细的心脏疾病诊断与检测技术与睡眠数据异常监测技术。这两个细分的技术主题的分裂与发展受到了心血管疾病和睡眠障碍等领域研究需求和应用前景的影响。随着医疗技术的进步和人们对健康的关注不断增加,对心脏疾



图6 技术主题演化

Fig.6 Evolution of Technical Topics

病和睡眠问题的监测和诊断需求逐渐凸显,进一步推动了这两个技术主题的发展。这种精细化的演化为相关领域的研究和应用提供了更为具体和针对性的解决方案,有助于改善心脏疾病和睡眠障碍的诊断与治疗效果,提升人们的健康水平。

在医学人工智能领域的技术主题中,具有融合演化特征的技术主题共计9个。其中,融合程度最高的为快速发展期中的医学信息管理系统技术主题,它由发展期中7个技术主题融合形成,将模式识别与数据处理系统、医疗数据挖掘技术、文本数据处理系统、医学影像处理与分析技术、基于神经网络的预测技术和图像处理与计算机视觉技术融合,逐渐发展成为具有代表性的医学信息管理系统。此外,癌症诊断技术的出现也是基于医疗数据挖掘技术和基于神经网络的预测技术的发展而逐渐演化形成。

4 结论与建议

本研究引入生命周期理论对医学人工智能领域近五十多年来的技术发展阶段进行划分,并使用BERTopic主题模型对其专利文本进行技术主题识别,利用余弦相似度计算主题对之间的语义相似度,保留强关联技术主题对进行演化分析。与现有研究相比,本研究较为全面地反映了医学人工智能领域技术发展的整体态势,对其技术主题的发展历程及演化情况进行了较为细致地梳理,得到的主要结论如下。

1) 医学人工智能技术萌芽早、起步慢,近年来正处于快速发展阶段。早期的医学人工智能技术受限于数据质量、计算能力以及医疗行业相对保守的特点,导致其起步较为缓慢^[34]。随着大数据技术、深度学习算法的发展^[35-37],以及医疗数据的累积,医学人工智能开始展现出巨大的潜力^[38]。

2) 医学人工智能领域技术主题围绕着信号处理与分析技术、图像处理与计算机视觉技术以及数据挖掘技术而展开,它们构成了医学人工智能的关键技术体系^[39]。信号处理与分析技术可用于提取和分析心电图和脑电图等数据中的关键信息,帮助诊断疾病、监测患者状态以及预测疾病风险;图像处理和计算机视觉技术可以实现医学影像的自动分析、疾病检测和区域分割;数据挖掘技术则有助于发现医疗数据中的隐藏模式和关联关系,从而预测疾病的发展趋势、治疗效果和相关风险。这些关键技术为疾病诊断、预防和治疗提供了有力支持^[40]。

3) 医学人工智能领域基础技术在临床实践中的应用正受到不断增长的需求推动。随着医疗技术的进步和人们对健康的关注不断增加^[41],未来癌症诊断技术、心脏疾病诊断与监测技术、睡眠数据异常监测技术等为相关领域的研究和应用提供更为具体和针对性的解决方案的技术将成为未来技术研发与创新的主要方向。

4) 医学人工智能领域基础技术逐渐成熟并不断分化,不同技术间的融合也将不断促进技术创新,从而推动了分支领域技术的进一步发展。从萌芽期关注数据处理、图像处理、模式识别等基础技术到快速发展期聚焦医学影像识别、疾病预测、基因组学、生物信息学、精准医疗和远程医疗等多个基础技术在医学领域的实践场景,这体现了医学人工智能领域技术主题的不断拓深与细化。发展期中医疗数据挖掘、医学影像处理与分析等技术的融合推动了医学信息管理系统逐步发展。

结合研究结论,本文就医学人工智能技术战略发展提出如下建议。

1) 继续加强对医学人工智能领域基础关键技术的研究和创新。在未来的发展中,持续加强对

基础关键技术的研究和创新是至关重要的^[42]。随着医学人工智能领域的快速发展,我们需要不断探索新的方法和算法来进一步提升信号处理与分析技术、图像处理与计算机视觉技术以及数据挖掘技术的性能和应用效果。通过加强基础理论研究和优化,可以持续推动医学人工智能的发展,加速医疗技术的革新和医疗质量的改善,为医疗领域提供更多的创新解决方案和技术进步。

2) **重视实际需求,推动技术创新。**需求与技术的进步是医学人工智能领域基础技术逐渐分化的推动因素。需求导向意味着我们需要深入了解医学领域的需求^[43],从医疗实践和临床应用的角度出发,明确医学人工智能所需解决的实际问题^[20,21]。这包括改善疾病诊断和治疗过程、提高医疗效率、减少误诊和病例漏诊等方面。通过深入了解需求,我们可以有针对性地开展技术攻关研究,确保技术的应用符合实际需求。问题导向则强调解决医学人工智能领域面临的挑战和难题。在技术研究中,我们需要面对各种问题,如数据隐私与安全、模型可解释性、算法的鲁棒性等^[38]。通过解决这些问题,我们能够提高技术的可靠性、稳定性和可行性,进一步推动医学人工智能的发展。效果导向是指技术创新应注重实际效果的验证和评估。我们应该重视医学人工智能技术的实际应用效果,进行充分的实验和评估,确保技术的有效性和可行性。这包括在真实世界中的挖掘证据、与医生和患者的合作反馈等。只有通过验证和评估,才能真正了解技术的效果,并对其进行改进和优化^[37]。结合需求和实际应用进行研究努力推动技术的发展和优化,进一步提升医学人工智能在医疗领域的应用效果和价值。

3) **鼓励技术融合与跨学科合作。**人工智能领域涉及众多技术领域的交叉与融合,如机器学习、

自然语言处理、计算机视觉等。这些技术相互作用和相互补充,可以带来更加强大和创新的解决方案。通过关注人工智能领域的关键共性技术,能够推动不同技术之间的融合,从而产生新的技术突破。鼓励多领域科研人员组建交叉学科的团队也是非常重要的。不同背景和专业知识的科研人员组成的团队,可以促进知识和经验的交流,加速技术创新的发展。通过跨学科的合作,可以整合各个领域的专业知识,提供更全面和综合的解决方案^[44]。跨学科的合作也有助于推动多领域技术的深度融合。不同领域的专业知识和技术在结合时,可以相互借鉴和迭代,产生新的理念和方法,有助于拓宽技术应用的边界,推动技术的突破和应用的创新,带来更多的机遇和发展前景。

数据可用性声明

支撑本研究的科学数据已在中国科学院科学数据银行(Science Data Bank) ScienceDB 平台公开发布,访问地址为 <https://www.doi.org/31253.11.sciencedb.j00053.00055> 或 <http://resolve.pid21.cn/10.57760/sciencedb.j00053.00055>。

参考文献

- [1] 杨恒,王曰芬,张露.基于核心专利技术主题识别与演化分析的技术预测[J].情报杂志,2022,41(7):49-56. (YANG Heng, WANG Yuefen, ZHANG Lu. Technical Prediction Based on Core Patented Technology Topics Recognition and Evolution Analysis [J]. Journal of Information, 2022, 41(7): 49-56.)
- [2] 魏玲,许海云,刘春江,等.技术领域主题发现研究——以基因工程疫苗领域为例[J].数字图书馆论坛,2017(1):37-45. (WEI Ling, XU Haiyun, LIU Chunjiang, et al. Study on the Technological Domain Recognition: A Case Study of Gene

- Engineered Vaccine Technology [J]. Digital Library Forum, 2017(1): 37-45.)
- [3] 徐璐璐, 杨嘉乐, 康乐乐. 医学信息领域人工智能技术的主题漂移与未来展望——基于JCR 26本医学信息期刊文本的命名实体识别[J]. 现代情报, 2022, 42(10): 163-176. (XU Lulu, YANG Jiale, KANG Lele. Topic Drift and Future Prospect of Artificial Intelligence Technology in Medical Information Field: Named Entity Recognition of 26 JCR Medical Informatics Journals Based on JCR [J]. Journal of Modern Information, 2022, 42(10): 163-176.)
- [4] 陈琦, 张君冬, 郑婉婷, 等. 基于LDA模型的中医药人工智能领域主题演化分析[J]. 世界科学技术—中医药现代化, 2022, 24(9): 3315-3324. (CHEN Qi, ZHANG Jundong, ZHENG Wanting, et al. Analysis of Topic Evolution in the Field of Artificial Intelligence in Traditional Chinese Medicine based on LDA model [J]. Modernization of Traditional Chinese Medicine and Material Medica-World Science and Technology, 2022, 24(9): 3315-3324.)
- [5] 孟晓媛, 张艳, 陈智慧. 人工智能在中医药领域的应用与发展[J]. 吉林中医药, 2023, 43(5): 618-620. (MENG Xiaoyuan, ZHANG Yan, CHEN Zhihui. Application and Development of Artificial Intelligence in t TCM [J]. Jilin Journal of Chinese Medicine, 2023, 43(5): 618-620.)
- [6] HOSNY A, PARMAR C, QUACKENBUSH J, et al. Artificial Intelligence in Radiology [J]. Nature Reviews. Cancer, 2018, 18(8): 500-510.
- [7] 谢燕鹏, 杨欢, 刘建荣. 人工智能在辅助生殖领域中应用的研究进展[J]. 中国医药导报, 2023, 20(19): 53-56. (XIE Yanpeng, YANG Huan, LIU Jianrong. Research Progress on the Application of Artificial Intelligence in the Assisted Reproduction Field [J]. China Medical Herald, 2023, 20(19): 53-56.)
- [8] 张开友, 王思佳. 基于WOS的重症医学领域人工智能研究的可视化分析[J]. 检验医学与临床, 2023, 20(9): 1287-1293. (ZHANG Kaiyou, WANG Sijia. Visual Analysis of Artificial Intelligence Research in Critical Care Medicine Based on WOS [J]. Laboratory Medicine and Clinical, 2023, 20(9): 1287-1293.)
- [9] CHAN H-P, SAMALA R K, HADJIISKI L M, et al. Deep Learning in Medical Image Analysis [J]. Advances in Experimental Medicine and Biology, 2020, 1213: 3-21.
- [10] SAVADJIEV P, CHONG J, DOHAN A, et al. Demystification of AI-driven Medical Image Interpretation: Past, Present and Future [J]. European Radiology, 2019, 29(3): 1616-1624.
- [11] YOO J, KIM T Y, JOUNG I, et al. Industrializing AI/ML During the End-to-end Drug Discovery Process [J]. Current Opinion in Structural Biology, 2023, 79: 102528.
- [12] ZHONG F, XING J, LI X, et al. Artificial Intelligence in Drug Design [J]. Science China. Life Sciences, 2018, 61(10): 1191-1204.
- [13] WASSAN J T, ZHENG H, WANG H. Role of Deep Learning in Predicting Aging-Related Diseases: A Scoping Review [J]. Cells, 2021, 10(11): 2924.
- [14] DE VELASCO ORIOL J, VALLEJO E, et al. Benchmarking Machine Learning Models for Late-onset Alzheimer's Disease Prediction from Genomic Data [J]. BMC Bioinformatics, 2019, 20(1): 709.
- [15] ZHANG Y, KONG M, ZHAO T, et al. Auxiliary Diagnostic System for ADHD in Children Based on AI Technology [J]. Frontiers of Information Technology & Electronic Engineering, 2021, 22(3): 400-414.
- [16] ZHENG J, ZHANG Q, XU S, et al. Cognition-Based Context-Aware Cloud Computing for Intelligent Robotic Systems in Mobile Education [J]. IEEE Access, 2018, 6: 49103-49111.
- [17] 胡子琦, 胡彩燕, 马镛. 医疗人工智能技术的专

- 利情报研究 [J]. 中国发明与专利, 2022, 19(7): 29-35, 43. (HU Ziqi, HU Caiyan, MA Bang. Analysis of Patent Information of Medical Artificial Intelligence Technology [J]. China Inventions & Patent, 2022, 19(7): 29-35, 43.)
- [18] 张婷, 卢岩, 陈娟, 等. 基于三方专利的医学人工智能领域技术竞争态势研究 [J]. 中国医疗设备, 2020, 35(11): 133-137. (ZHANG Ting, LU Yan, CHEN Juan, et al. Technology Competition of Medical Artificial Intelligence Based on the Analysis of Triadic Patent Family [J]. China Medical Device, 2020, 35(11): 133-137.)
- [19] TRAN B X, NGHIEM S, SAHIN O, et al. Modeling Research Topics for Artificial Intelligence Applications in Medicine: Latent Dirichlet Allocation Application Study [J]. Journal of Medical Internet Research, 2019, 21(11): e15511.
- [20] AMISHA, MALIK P, PATHANIA M, et al. Overview of Artificial Intelligence in Medicine [J]. Journal of Family Medicine and Primary Care, 2019, 8(7): 2328-2331.
- [21] BRIGANTI G, LE MOINE O. Artificial Intelligence in Medicine: Today and Tomorrow [J]. Frontiers in Medicine, 2020, 7: 27.
- [22] GROOTENDORST M. BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure [J]. arXiv, 2022. DOI: 10.48550/arXiv.2203.05794.
- [23] JEON E, YOON N, SOHN S Y. Exploring New Digital Therapeutics Technologies for Psychiatric Disorders Using BERTopic and PatentSBERTa [J]. Technological Forecasting and Social Change, 2023, 186: 122130.
- [24] 朱光, 刘蕾, 李凤景. 基于LDA和LSTM模型的研究主题关联与预测研究——以隐私研究为例 [J]. 现代情报, 2020, 40(08): 38-50. (ZHU Guang, LIU Lei, LI Fengjing. Research on Topic Relation and Prediction Based on LDA and LSTM: A Case Study of Privacy Research [J]. Journal of Modern Information, 2020, 40(8): 38-50.)
- [25] 关鹏, 王曰芬. 基于LDA主题模型和生命周期理论的科学文献主题挖掘 [J]. 情报学报, 2015, 34(3): 286-299. (GUAN Peng, WANG Yuefen. Topic Mining in Scientific Literature Based on LDA Topic Model and Life Cycle Theory [J]. Journal of the China Society for Science and Technical Information, 2015, 34(3): 286-299.)
- [26] EGGER R, YU J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts [J]. Frontiers in Sociology, 2022(7): 886498.
- [27] GUIZZARDI S, COLANGELO M T, MIRANDOLA P, et al. Modeling New Trends in Bone Regeneration, Using the BERTopic Approach [J]. Regenerative Medicine, Future Medicine, 2023, 18(9): 719-734.
- [28] CONTRERAS K, VERBEL G, SANCHEZ J, et al. Using Topic Modelling for Analyzing Panamanian Parliamentary Proceedings with Neural and Statistical Methods [C]. 2022 IEEE 40th Central America and Panama Convention (CONCAPAN). 2022: 1-6.
- [29] ZHANG D, WU X, LIU P, et al. Identification of Product Innovation Path Incorporating the FOS and BERTopic Model from the Perspective of Invalid Patents [J]. Applied Sciences, MDPI AG, 2023, 13(13): 7987.
- [30] WANG Z, CHEN J, CHEN J, et al. Identifying Interdisciplinary Topics and Their Evolution Based on BERTopic [J/OL]. Scientometrics, (2023-07-23). <https://doi.org/10.1007/s11192-023-04776-5>.
- [31] 张清慧, 陈谊, 武彩霞. 基于词表示模型的领域文献数据可视分析方法 [J]. 图学学报, 2022, 43(4): 685-694. (ZHANG Qinghui, CHEN Yi, WU Caixia. A Visual Analysis Approach for Literature Data Based on Word Representation Model [J]. Journal of Graphics, 2022, 43(4): 685-694.)
- [32] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional

- Neural Networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [33] COLLINS F S, VARMUS H. A New Initiative on Precision Medicine [J]. New England Journal of Medicine, Massachusetts Medical Society, 2015, 372(9): 793-795.
- [34] 申喜凤, 李美婷, 南嘉乐, 等. 医学人工智能发展态势分析及问题浅析[J]. 科技管理研究, 2023, 43(7): 193-198. (SHEN Xifeng, LI Meiting, NAN Jiale, et al. Analysis on the Development Trend and Problems of Medical Artificial Intelligence [J]. Science and Technology Management Research, 2023, 43(7): 193-198.)
- [35] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [36] MURDOCH T B, DETSKY A S. The Inevitable Application of Big Data to Health Care [J]. JAMA, 2013, 309(13): 1351-1352.
- [37] BEAM A L, KOHANE I S. Big Data and Machine Learning in Health Care [J]. JAMA, 2018, 319(13): 1317-1318.
- [38] TOPOL E J. High-performance Medicine: the Convergence of Human and Artificial Intelligence[J]. Nature Medicine, 2019, 25(1): 44-56.
- [39] 言方荣. 人工智能在生物医药领域中的应用和进展[J]. 中国药科大学学报, 2023, 54(3):263-268. (YAN Fangrong. Application and Advance of Artificial Intelligence in biomedicine Field [J]. Journal of China Pharmaceutical University, 2023, 54(3): 263-268.)
- [40] 赵阳光. 医疗人工智能技术与应用研究[J]. 信息技术, 2018, 12(3): 32-36. (Zhao Yangguang. Medical Artificial Intelligence Application and Technology [J]. Information and Communication Technology, 2018, 12(3): 32-36.)
- [41] 辛艳姣, 项莉, 金新政. 智慧健康行业发展路径分析[J]. 智慧健康, 2016, 2(9): 29-32. (XIN Yanjiao, XIANG Li, JIN Xinzheng. Analysis on the Development Path of Smart Health Industry [J]. Smart Healthcare, 2016, 2(9): 29-32.)
- [42] 史湘宁, 房超, 李辉, 等. 我国人工智能关键技术体系演化路径及发展研究——基于科学计量视角[J]. 科学管理研究, 2022, 40(3): 2-11. (SHI Xiangning, FANG Chao, LI Hui, et al. Research on Evolution Path and Development of Key Technology System of Artificial Intelligence in China: From the Metrological Perspective [J]. Scientific Management Research, 2022, 40(3): 2-11.)
- [43] 方丽, 崔雷. 需求驱动的医疗健康大数据挖掘模型构建[J]. 中华医学图书情报杂志, 2021, 30(7): 17-23. (FANG Li, CUI Lei. Demand-driven Establishment of Medical and Health Big Data Mining Model [J]. Chinese Journal of Medical Library and Information, 2021, 30(7): 17-23.)
- [44] 袁毅, 季泽豪. 人工智能跨学科特性研究——文献计量分析的视角[J]. 图书馆杂志, 2022, 41(6): 46-52. (YUAN Yi, JI Zehao. A Study of the Interdisciplinary Characteristics of Artificial Intelligence: From the Bibliometric Perspective [J]. Library Journal, 2022, 41(6): 46-52.)

作者贡献说明

周隽如:数据收集与处理, 论文撰写;

刘智勇:设计论文框架, 提出论文修改意见。