

李萍, 王丽丽. 国内多模态技术的研究现状与发展趋势: 基于 CiteSpace 的可视化分析[J]. 智能计算机与应用, 2025, 15(1): 194-202. DOI: 10.20169/j.issn.2095-2163.24062803

国内多模态技术的研究现状与发展趋势: 基于 CiteSpace 的可视化分析

李萍^{1,2}, 王丽丽²

(1 吉林化工学院 信息与控制工程学院, 吉林 吉林 132022; 2 德州学院 计算机与信息学院, 山东 德州 253023)

摘要: 随着深度学习和自然语言处理方法的发展, 多模态技术在人们的日常生活中得到广泛应用。本文利用 CiteSpace 软件检索中国知网数据库, 分析近十年多模态技术研究现状, 为相关科研人员提供了研究思路。本文筛选了 1 027 篇相关期刊文献进行分析, 发现多模态技术的文献数量呈上升趋势, 反映了对该技术的持续关注。其研究热点为图像融合、视觉问答等, 在医学图像以及日常社交中都有应用; 未来多模态技术可能的发展趋势在“生理信号”与“自动驾驶”等方面。

关键词: 多模态; CiteSpace 软件; 文献计量; 可视化分析; 研究现状; 发展趋势

中图分类号: TP391 **文献标志码:** A **文章编号:** 2095-2163(2025)01-0194-09

Research status and development trend of multimodal technology in China: Visual analysis based on CiteSpace

LI Ping^{1,2}, WANG Lili²

(1 School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, Jilin, China;

2 School of Computer and Information, Dezhou University, Dezhou 253023, Shandong, China)

Abstract: With the advancement of deep learning and natural language processing techniques, multimodal technology has found widespread applications in daily life. This paper employs CiteSpace software to examine the research trends of multimodal technology over the past decade, based on the China National Knowledge Infrastructure (CNKI) database, providing valuable research insights for scholars in the field. This paper selected and analyzed 1 027 relevant journal articles. The analysis reveals that the number of publications on multimodal technology has been increasing, reflecting sustained interest in this field. Additionally, research hotspots include image fusion and visual question answering, with applications in medical imaging and social interaction. Future developments in multimodal technology are likely to focus on areas such as physiological signal processing and autonomous driving.

Key words: multimodal; CiteSpace software; bibliometric; visual analysis; research status; development trend

0 引言

在数据领域中, 数据的表达方式有文本、图片、音频、视频以及混合数据, 数据的一种存在方式或其来源称为模态, 模态是一种社会性、文化性资源^[1]。多模态数据用来表示不同形态的数据形式, 是指对同一个数据对象在不同的空间维度下进行描述^[2]。多模态学习理论由 Scherer^[3] 提出, 研究的是将不同表达形式的数据进行融合, 在实际生活中具有广泛的应用。在医学方面, 可以利用多模态医学图像相

关技术来进行病情诊断; 在物品识别方面, 多个短视频以及购物平台应用多模态技术进行视频、图片以及文本的分类, 方便消费者更好地进行选购; 在智能机器人方面, 应用多模态技术可以实现人机交互、视觉问答、视觉常识推理等; 在影视作品中, 应用多模态技术可以将音频或视频进行字幕的生成。

随着人工智能的发展与深度学习相关方法的不断更新, 多模态领域的研究热点也不断更新。基于现有的研究成果, 本文主要采用由美国德雷克塞尔大学信息科学与技术学院的陈超美教授与大连理工

基金项目: 国家自然科学基金(11903008); 山东省德州市市级研发计划项目(2022dzkj098)。

作者简介: 李萍(1999—), 女, 硕士研究生, 主要研究方向: 人工智能技术及应用。

通信作者: 王丽丽(1981—), 女, 博士, 教授, 主要研究方向: 大数据挖掘, 机器学习。Email: jsjxwl@126.com。

收稿日期: 2024-06-28

大学 WISE 实验室联合开发的 CiteSpace 软件^[4],对多模态研究领域的相关期刊文献进行可视化分析和系统性研究。以时间线为研究依据,对近十年所发表的文献期刊进行关键词的提取,然后再通过图谱的展示与分析,进一步对多模态研究领域的热点变化趋势以及未来发展趋势进行探究与总结。

1 研究设计

1.1 数据来源

知网是中国最大的学术文库,汇集了各领域的学术资源。本文研究数据来源于中国知网数据库,以“多模态深度学习”、“多模态学习”以及“多模态融合”为主题词进行检索,检索时间跨度为 2012 年 1 月至 2022 年 12 月,检索得到 1 605 篇期刊文献。通过手动筛选,剔除会议、图书等无关文献,并且以 Refworks 的格式进行导出,经过 CiteSpace 软件的数据转换和除重,最终选取其中相关性较强的 1 027 篇期刊文献作为研究数据,方便后期利用 CiteSpace 软件进行相关数据的处理与分析。

1.2 研究方法

CiteSpace 是一款以文献系统以及信息计量学为研究对象的可视化工具,能够密切跟踪一个领域的发展,对该领域的文献期刊进行系统的分析与研究^[5],帮助初学者对研究热点有更清晰的认识。

本文借助 CiteSpace 6.2. R4 软件对 2012 年~2022 年中国知网数据库中多模态领域相关的期刊文献进行定量分析,从关键词、作者以及机构三方面进行图谱的绘制。其中,关键词图谱包括关键词共现、聚类、时间线、突现等 4 个方面,并且以时间线为依据,对图谱进行分析,获取多模态研究领域的研究热点、变化以及发展趋势进行总结。关键词共现是一种研究学术论文的方法,通过分析某一领域相关关键词的共同出现频率来揭示研究主题;关键词聚类是将行为相似度最高的关键词聚集在一起,从而揭示总体特征;时间线图是将聚类包含的关键词按照首次出现的年份进行分类,在所属的聚类中铺展开来^[6];关键词突现则反映了研究热点的变化及其持续时间。

本文按各个时间段对不同关键词关注度的多少来对多模态领域的发展现状进行分析,并进行发展趋势的推断。此外,对期刊文献的作者以及机构也进行了共现分析,来推测该领域作者与机构各自之间的合作关系。

2 结果分析

2.1 发文量统计

本文共选取 1 027 篇期刊文献进行发文量统计(见图 1)。由图 1 可知,2012 年到 2022 年多模态领域的发文量总体呈上升趋势,说明国内对多模态领域的关注度以及研究程度随时间推移不断上升。2012 年~2017 年多模态领域发文量较少,均在 50 篇以下,年均发文量约为 21 篇,发文频率较低;2018 年与 2019 年呈现小幅度上升趋势;2020 年后发文量大大幅度持续上升,表示现阶段中国多模态研究领域受到高度关注。

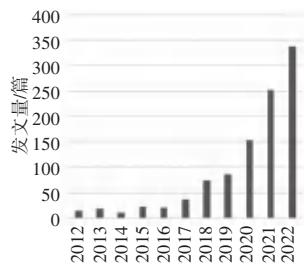


图 1 2012 年~2022 年多模态发文量统计

Fig. 1 Statistics of papers on multimodal from 2012 to 2022

2.2 关键词分析

关键词表达了一篇文献的核心要点,是文献研究内容特征的高度概括和总结,结合关键词检索可以将研究热点具体化^[7]。通常情况下,用中介中心性来衡量某一关键词在该领域中的重要性,中介中心性的计算公式^[8]如下:

$$C_B = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (1)$$

其中, $\sigma_{st}(v)$ 表示经过关键词 v 的 s 到 t 的最短路径条数, σ_{st} 表示关键词 s 到 t 的所有最短路径条数。中介中心性数值越高,表示该关键词的重要性也越大。

本文将时间切片设置为 1 年,通过数据筛选、去重等操作,对 1 027 篇文献期刊进行关键词的提取,表 1 所示为部分高频关键词。其中,频次是在所研究的文献数据中关键词出现的频率;年份为该关键词首次出现的年份;中心性超过 0.1 则代表该关键词较为主要。由表 1 可知,“深度学习”是这 1 027 篇文献中频次最高的关键词,频次为 1 091;而“多模态”为中心性最高的关键词,中心性为 0.9。结果表明,“深度学习”是现阶段多模态技术主要的研究方法;而关键词“多模态”处于关键地位,是连接不同领域的关键概念之一。

表1 部分高频关键词
Table 1 Some high-frequency keywords

序号	关键词	频次	年份	中心性
1	深度学习	1 091	2014	0.25
2	多模态	461	2012	0.90
3	特征融合	152	2012	0.05
4	人工智能	76	2016	0.08
5	特征提取	73	2014	0.38
6	机器学习	71	2012	0.10
7	神经网络	59	2017	0.02
8	迁移学习	56	2017	0.02
9	情感分析	56	2019	0.08
10	情感识别	55	2018	0.02
11	视觉问答	48	2017	0.06
12	图像融合	44	2013	0.33
13	目标检测	38	2019	0
14	语义分割	36	2018	0.05
15	图像分割	35	2018	0.01
16	信息融合	35	2013	0.10

2.2.1 共现分析

共现分析的目的是确定文献集中各主题之间的



图2 关键词共现分析图谱

Fig. 2 Keyword co-occurrence analysis map

2.2.2 聚类分析

关键词聚类是将关联紧密的关键词聚成一类,用来观察目前某个技术或学科领域形成了哪几个研究类团。关键词聚类分析通常涉及以下步骤:

1) 收集并准备数据集

使用对数似然比 (Log-Likelihood Ratio, LLR)

关系,其方法论基础是心理学的邻近联系法则和知识结构及映射原则^[9]。共现分析是指分析同一研究领域中的高频关键词共同出现的情况^[10],形成一个共词网络。

如图2所示,关键词共现图谱中的节点代表关键词,圆圈大小表示关键词的频次,圆圈越大则表示关键词出现的频率越高,说明该领域内对其关注度则越高^[11],由此可获取某个领域的研究热点。线条表示的是关键词之间的联系,连线粗细代表共现强度,可以通过线条看出哪些关键词之间是相互联系的。运行 CiteSpace 软件,得到多模态研究领域的关键词共现分析图谱,共有节点 281 个,连线 484 条。

由图2可知,圆圈最大即频次最大的节点是“深度学习”,这表明近年来多模态主要是基于深度学习进行的,深度学习已成为多模态技术的主要研究方法。连线最多的节点是“多模态”,这表明“多模态”是该领域的核心节点,与该节点相连的节点包括“图像融合”、“图像分类”、“人脸识别”、“多任务”、“医学影像”等多个关键词,说明这几个方面是现阶段多模态技术的主要研究方向,涉及到人工智能、医学、日常社交等多个领域。

算法^[12],将各个关键词根据相似性或关联性进行分组,形成不同的类别或“聚类”。各个聚类会有一个聚类标号^[13],聚类标号按照聚类中关键词的出现频率和相关度排列的。

2) 评估每个聚类,确定其中心关键词和研究主题。

由表 2 可知, 图像融合、特征提取以及神经导航是突现度较大的 3 个关键词, 三者代表了多模态技术的重要组成部分, 因此近几年受到了很广泛的关注, 并成为了近年来极具影响力的研究热点。

2.3 作者合作网络分析

在作者合作网络分析图谱中, 节点圆圈大小反

映了作者的出现频次, 被引频次最高的作者说明其在该领域内具有重要影响, 连线表示作者之间存在合作关系^[6]。CiteSpace 6. 2. R4 版本只显示合作较多的作者, 利用该软件对相关期刊文献的作者进行分析并绘制作者合作网络分析图谱如图 5 所示。



图 5 作者合作网络分析图谱

Fig. 5 Author collaboration network analysis map

从图 5 可见, 吴德权、姜晨、孙骏、包频、戴敏超、周志敏、冯素银以及张晓路几人合作较多。发文量较多的作者可以由软件中的作者发文量共现表获得, 中国多模态领域发文量前五的作者见表 3。

表 3 中国发文量前 5 名作者

Table 3 Top five authors in China

作者	发文量	首篇发表年份
王鑫	7	2019
徐常胜	5	2020
孙显	4	2022
杨超	4	2021
杨锋	4	2020

由表 3 可知, 王鑫、徐常胜是发文量最多的两位作者, 对多模态研究领域贡献较多。其中, 清华大学的王鑫^[17]在多模态数据与人工智能结合方面做了重要研究, 探讨了跨媒体关联分析与语义理解的技术, 以及多模态数据的统一表达与融合^[18]。这些研

究为多模态数据与人工智能的结合提供了关键支持。来自中国科学院自动化研究所的徐常胜^[19]提出了一种创新的多模态对抗式视频生成方法, 该方法在视频生成任务中具有较高的优越性; 另外, 其还构建了多模态医疗知识图谱^[20], 提升了问答系统的准确性。

但是, 由表 3 与图 5 可知, 发文量较多的作者之间的合作较少。

2.4 机构合作网络分析

分析研究机构的结构特征对于识别该领域有影响力的机构和理解机构间合作至关重要。研究机构的发文量表现了对多模态领域的研究程度, 机构合作网络分析图谱中节点大小反映了机构的出现频次, 节点越大则出现频次越高, 对多模态研究领域的贡献也就越大; 反之, 节点越小, 则表明在多模态领域发文量越少。同样的, CiteSpace 6. 2. R4 版本只显示合作较多的机构, 利用该软件对多模态研究领域的机构合作网络分析图谱如图 6 所示。

根据表2(突现词分析)的结果显示,“生理信号”和“自动驾驶”是近几年突现的关键词。这表明,近年来对人工智能领域自动驾驶的研究也成为当前的热门研究方向之一。来自电子科技大学的吴愚(2021)^[21]通过将激光雷达(LiDAR)点云数据和彩色(RGB)图像数据进行融合,提出了一种端到端的神经网络结构和数据填充算法的多模态融合策略,显著提升了自动驾驶领域中行人和车辆目标识别的精度和鲁棒性。其研究表明,将深度多模态目标检测和语义分割方法用于自动驾驶中的智能交通系统,能够实现更准确的障碍物检测以及场景理解。

(2)在医学领域中,多模态数据是指诊断报告、多种医疗成像设备如X光、计算机断层扫描、磁共振成像、超声成像和正电子发射断层扫描产生的医学影像等^[26]。从图2(关键词共现图谱)的分析中可以看到,多模态在影像组学、医学影像、脑电信号以及睡眠分期中应用颇广。

根据图4(关键词时间线图)的分析可以得到迁移学习在2015年应用在乳腺癌方面,2017年开始应用于脑肿瘤以及肿瘤等方面。利用多模态深度学习融合多种医学数据,提高了疾病诊断的准确性和早期预测的可靠性。

来自电子科技大学的陈航^[24]基于生成对抗网络,提出了一种面向医学影像分类的无源域数据领域自适应方法,提高了医学影像分类的准确率,其研究发现多模态的信息融合技术在医学领域的应用可以提升预测结果的精度。

(3)在日常社交中,多模态也得到了广泛的应用。图2(关键词共现图谱)显示,图像分类以及特征提取技术在家具风格、社交媒体以及动画技术中得到应用。利用多模态技术可以对多媒体视频、动画、图文等进行相应的分类^[18],继而根据人们的喜好进行大数据推荐同类信息。

此外,多模态技术也应用到了情感分析、行为识别、情绪识别以及脑电信号等方面。情感分析是类似于人类情感表达的一种方式。现阶段,各种深度学习方法已经成功地应用于学习高级特征表示以进行情感特征识别与分析。张晋敏等^[27]提出了一种基于BiGRU模型的多模态网络舆情情感分析框架,该框架结合文本和图像特征,通过线性融合技术进行特征融合,以实现更准确的情感分析。在人机交互中,智能机器人也可以通过语音识别、情感识别以及面部表情分析等多模态技术来了解用户当时状态的情绪,进而进行安慰或其他情感处理,给用户带来更好的体验。

3.2 发展趋势

多模态研究领域的未来发展趋势可由关键词突现图谱获取。通常,突现性是具有连续性的,所以突现持续到2022年的关键词便是未来多模态领域的前沿热点。表2突现词分析结果显示,“生理信号”与“自动驾驶”突现时间段分别为2021年~2022年、2020年~2022年,这表明“生理信号”与“自动驾驶”在未来几年可能依旧会广受关注,继续成为研究热点。随着人工智能以及自动化驾驶技术的发展,应用多模态信息融合将提高目标检测以及自动驾驶模型的性能^[28],实现智能化、自动化社会将成为未来的发展趋势。除此之外,对各种模态信息的有效融合也是多模态技术在各个领域发展的一大挑战。

4 结束语

综上所述,随着人工智能与自然语言处理技术的不断进步,多模态逐步应用在人们生活的方方面面。尤其在人工智能方面,多模态技术的应用给人们的生活带来了极大的便利,未来可以对生理信号、自动驾驶等领域加强发展。另外,多模态技术逐渐在医疗方面有所应用,但应用技术尚未成熟,未来还需要在模型大小与模型性能、模态融合方式、训练以及推理速度以及数据隐私问题等方面进行探索。各个机构与作者之间也应该加强交流合作,形成更加完善的多模态研究体系。

参考文献

- [1] 刘佳琳,李喆. 多模态学习的研究热点、动态前沿与趋势分析:基于CiteSpace软件的可视化分析[J]. 卫生职业教育,2022,40(19):125-129.
- [2] 张虎成,李雷孝,刘东江. 多模态数据融合研究综述[J]. 计算机科学与探索,2024,18(10):2501-2520.
- [3] SCHERER S, WORSLEY M, MORENCY L P. 1st international workshop on multimodal learning analytics[C]//Proceedings of the 14th ACM International Conference on Multimodal Interaction. New York:ACM, 2012: 609-610.
- [4] WANG W, LU C. Visualization analysis of big data research based onCitespace[J]. Soft Computing, 2020, 24(11): 8173-8186.
- [5] 赵思越. 体育旅游研究的文献计量学可视化分析[D]. 天津:天津体育学院,2023.
- [6] 许引,熊冬春,林明,等. “智能+”时代我国虚拟教研室的研究前沿与趋势探究:基于CiteSpace知识图谱的可视化分析[J]. 广州广播电视大学学报,2023,23(1):25-31.
- [7] 马海韵,袁园. 大数据驱动下公共治理变革研究综述:基于2013-2020年CSSCI及核心期刊来源的文献计量分析[J]. 北京交通大学学报(社会科学版),2022,21(2):143-154.
- [8] PUZIS R, ZILBERMAN P, ELOVICI Y, et al. Heuristics for speeding up betweenness centrality computation [C]//Proceedings of 2012 International Conference on Privacy, Security, Risk and

- Trust and 2012 International Conference on Social Computing, Piscataway, NJ: IEEE, 2012: 302-311.
- [9] 陈亚丽. 铸牢中华民族共同体意识研究的学科视角、热点议题与未来展望: 基于 CiteSpace 知识图谱的可视化分析[J]. 西南民族大学学报(人文社会科学版), 2024, 45(4): 219-229.
- [10] 荣康丽, 张虎彪. 国内垃圾分类治理研究现状、热点主题及发展趋势: 基于 CiteSpace 可视化分析[J]. 环境保护科学, 2024, 50(1): 1-8.
- [11] 卢小丽, 杨浩, 戴永恩, 等. 基于 CiteSpace 的国内患者安全教育研究现状的文献计量学分析[J]. 循证护理, 2024, 10(9): 1590-1595.
- [12] 谭坤彦, 杨孔雨. 基于神经网络的舆情情感分析研究热点与趋势: 基于 CiteSpace 的可视化分析[J]. 智能计算机与应用, 2022, 12(8): 33-42.
- [13] 黄明祥, 张健钦, 杨毅, 等. 1989—2022 年生态环境中人工智能应用的研究综述: 基于 Citespace 的知识图谱分析[J]. 环境保护科学, 2024, 50(2): 8-16.
- [14] NEWMAN M E J. Fast algorithm for detecting community structure in networks [J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2004, 69(6): 066133.
- [15] 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.
- [16] 王彦丁. 基于 CiteSpace 的我国基础教育知识图谱可视化分析[J]. 数字技术与应用, 2021, 39(7): 223-228.
- [17] 于俊清, 王鑫, 况琨, 等. 跨媒体智能关联分析与语义理解理论与技术研究进展[J]. 计算机辅助设计与图形学学报, 2023, 35(1): 1-22.
- [18] 朱文武, 王鑫, 田永鸿, 等. 多媒体智能: 当多媒体遇到人工智能[J]. 中国图象图形学报, 2022, 27(9): 2551-2573.
- [19] 于海涛, 杨小汕, 徐常胜. 基于多模态输入的对抗式视频生成方法[J]. 计算机研究与发展, 2020, 57(7): 1522-1530.
- [20] 张莹莹, 钱胜胜, 方全, 等. 基于多模态知识感知注意力机制的问答方法[J]. 计算机研究与发展, 2020, 57(5): 1037-1045.
- [21] 吴愚. 多模态融合感知技术研究与设计[D]. 成都: 电子科技大学, 2021.
- [22] 黄周. 基于视觉显著性的复杂场景目标检测方法研究[D]. 成都: 电子科技大学, 2022.
- [23] 杨锦涛. 基于多模态融合的情感分析关键技术研究[D]. 成都: 电子科技大学, 2021.
- [24] 陈航. 面向多模态医学影像分析的领域自适应方法研究[D]. 成都: 电子科技大学, 2022.
- [25] 杨帆, 王志社, 孙婧, 等. 红外与可见光图像交互自注意力融合方法[J]. 光子学报, 2024, 53(6): 221-232.
- [26] 丁国辉, 张琦, 房士超, 等. 多模态检索在医学领域的研究综述[J]. 计算机工程与应用, 2023, 59(1): 26-36.
- [27] 张晋敏, 李旭芳, 樊弟军. 基于 BiGRU 模型的多模态网络舆情情感分析[J]. 智能计算机与应用, 2024, 14(1): 191-193.
- [28] 代振钊. 面向自动驾驶的多模态融合感知技术研究[D]. 北京: 北方工业大学, 2024.