

· 管理纵横 ·

基于 ESI 的化学与人工智能领域热点交叉主题识别与趋势研究

韩军伟* 汤国昌 赵世杰

西北工业大学 自动化学院, 西安 710072

[摘要] 探讨化学领域与人工智能的交叉现状,挖掘热点交叉研究主题,能够为基金项目资助决策提供有力依据,从而优化资源分配,促进化学领域前沿科学进展。本研究基于交叉领域的期刊文献和基金项目数据,从发文量、国家/机构合作、学科交叉三个维度揭示交叉态势,通过期刊文献关键词共现与聚类分析识别热门交叉研究主题,并进一步结合中国、美国在各主题下的基金项目资助情况研究交叉领域的发展趋势。研究发现化学领域与人工智能交叉的研究随时间发展显著增加,识别出泛函理论、虚拟筛选等七个热点交叉主题,揭示了人工智能在化学领域的应用潜力以及发展趋势,并给予基金资助应针对不同的主题定制不同资助策略的启示。

[关键词] 人工智能;化学领域;知识图谱;交叉主题识别;自然科学基金

自 20 世纪以来,得益于计算水平的显著提升和大数据的广泛应用,人工智能得到迅速的演进和发展,在各个领域都展现出了巨大的应用前景。2023 年,为贯彻落实国家《新一代人工智能发展规划》,科学与技术部会同国家自然科学基金委员会(National Natural Science Foundation of China, NSFC,以下简称“自然科学基金委”)启动“人工智能驱动的科学驱动”(AI for Science)专项部署工作,紧密结合化学、物理、数学等基础学科关键问题,围绕药物研发、新材料研发、基因研究等重点领域科研需求展开,布局“人工智能驱动的科学驱动”前沿科研体系^[1]。化学领域作为自然科学的重要分支,研究物质的性质、组成、结构、变化以及与能量之间相互关系,涵盖了从基础研究到实际应用的各个层面。化学领域与人工智能交叉结合推动着化学领域科研范式变革,为解决相关研究难题提供了新的契机^[2, 3]。化学领域研究所面对的十分复杂的物质体系和实验过程很难通过物理化学原理进行精准的分析判断,而人工智能可以挖掘化学实验中产生的海量实验数据的相关性,帮助化学家做出合理分析



韩军伟 教授,博士生导师,西北工业大学自动化学学院院长,信息融合技术陕西省科技创新团队负责人,无人机信息安全西安市重点实验室主任。研究方向为人工智能、模式识别、类脑计算、医学影像处理、遥感影像处理等。

预测,从而大大加速化学研发过程^[4, 5]。例如人工智能方法可以帮助预测化学反应产物、反应的产率以及反应物的活性等相关信息,从而帮助对反应条件进行优化^[6]。通过融合传统化学领域相互相对独立的专家知识,辅助设计出更合理、更有效的合成路线,从而聚焦于更复杂的合成任务^[7]。因此,探讨分析化学领域与人工智能的交叉态势具有重要价值,通过挖掘化学领域与人工智能的热点交叉主题,能够为基金资助决策提供有力依据,从而优化资源分配,更有效的促进化学领域前沿科学进展。

近年来,化学领域与人工智能交叉态势受到越来越多学者关注,如 Baum 等^[8]通过对过去二十年人工智能相关化学出版物的增长和分布进行统计分析,发现分析化学和生物化学融合人工智能的程度

收稿日期:2023-11-30;修回日期:2024-08-15

* 通信作者,Email: jhan@nwpu.edu.cn

本文受到国家自然科学基金项目(721A0001)的资助。

最高。刘小平等^[9]基于国际上人工智能在化学研究中的重要进展,分析了人工智能如何促进化学领域研究范式的变革。Neeru等^[10]概述了人工智能如何帮助学者完成从分子设计到分子性质预测等相关研究。但是在已有研究中,研究人员往往只聚焦于如反应预测、分子性质预测、结构解析等化学领域中的一个或某几个方面的研究,对人工智能的应用方式进行分析 and 阐述,缺乏全面性,限制了对交叉领域的全面了解。化学领域与人工智能交叉本质上是领域间知识的渗透与融合,基于知识产出能够更好地表征领域之间的知识流动,因此基于期刊文献进行交叉主题挖掘是了解化学领域与人工智能交叉现状与趋势的可行手段。然而,化学领域涵盖范围极广,研究主题广泛,相关术语使用多样化。常见的基于关键词检索文献的策略,在关键词全面性优先时容易引入大量不相关文献,造成筛选和分析困难。而基本科学指标数据库(Essential Science Indicators, ESI)提供了过去十年被引次数排名前50%的化学领域期刊目录,这些期刊通常具有较高的学术影响力和权威性且研究方向覆盖全面,能够在避免检索盲点的同时保证收集文献的质量和效率。因此,本研究将 ESI 化学领域期刊目录作为文献收集策略的主要依据。

常见的基于文献的主题识别方法为基于引文分析的识别、基于共词分析的识别以及基于文本挖掘的识别^[11]。基于引文的识别是指对文献间的相互引用进行统计并结合聚类等手段进行主题识别,黄福等^[12]对科学计量学领域文献进行分析发现文献共被引分析可以获得发展潜力巨大和成长性好的研究主题。基于共词分析的识别是指对文献集合中的关键词进行词频统计、共现分析和聚类分析等进行主题识别。王福等^[13]通过对竞争供应链领域的文献筛选高频关键词,并构建高频词共现网络,依据关键词的重要程度和各关键词联系的紧密情况将关键词聚类为三个研究主题。基于文本挖掘的识别主要指利用主题模型进行主题识别,如潜在狄利克雷分布(Latent Dirichlet Allocation, LDA)主题模型。裘惠麟等^[14]将期刊论文与专利文献作为多源数据源利用改进的 LDA 模型识别出机器学习热点研究主题。基于引文分析的主题识别由于引文关系形成时间长,因此得到的研究主题侧重于学科交叉的知识基础;基于文本挖掘的主题识别严重依赖于模型质量,主题内部的关键词关联性以及主题之间的区分性难以保证。而基于共词分析的主题识别由于是

以文献关键词所作的分析,结果直观且可靠性强。因此,本研究基于 CiteSpace(6. 2. R4 Advanced)^[15, 16]、VosViewer^[17]等工具对 2013—2022 年化学领域与人工智能交叉研究的文献进行统计、共现与聚类分析,揭示近十年化学领域与人工智能交叉态势,发掘交叉研究热点主题,并结合国内外自然科学基金数据深入了解两个领域的融合发展趋势,揭示研究重点和资源分配方向,为下一步的基金资助提供建议,从而优化资源分配,进一步促进跨学科融合与领域的创新和发展。

1 研究设计

本研究提出的化学领域与人工智能热点交叉主题识别与趋势研究框架,如图 1 所示,主要由期刊文献与基金数据获取及预处理、交叉态势分析、交叉主题识别、基金资助分析、结论与总结五部分构成。

1.1 数据获取

本研究期刊文献数据来源于 Web of Science (WoS)核心合集,其收录和覆盖了各个专业领域的文献资源。本研究首先基于人工智能领域关键词检索相关文献,检索策略如表 1 所示。接着基于 ESI 所划分的化学领域期刊目录对检索结果进行筛选,即可得到化学领域与人工智能交叉研究的文献数据。本研究基于该检索策略,经过去重、删除缺乏关键字段信息文献等预处理后,共得到化学领域与人工智能交叉的文献 12 895 篇。

本研究基金项目数据来源于中国自然科学基金委和美国国家科学基金会(National Science Foundation, U. S., NSF)。NSFC 和 NSF 分别为中国和美国支持和促进基础科学研究,推动科学技术的创新和发展的基金组织^[18],其中 NSFC 设置有化学科学部,NSF 设置有化学系。因此,分别在化学部门下依据人工智能关键词进行检索,即可得到化学领域与人工智能交叉研究的基金项目数据。本研究基于用于检索文献的人工智能领域关键词作为检索词,经过去重、筛选等预处理后,共得到化学领域与人工智能交叉的 NSFC 基金数据 254 项和 NSF 基金数据 191 项。

1.2 研究方法

本研究基于 VosViewer 工具进行国家/机构合作共现关系、学科共现关系的可视化,基于 CiteSpace 工具进行关键词共现关系可视化以及关键词聚类分析,其中共现关系是指两个或多个元素在同一文献中出现的情况。CiteSpace 在进行聚类

时依据聚类模块值(Q 值)和聚类平均轮廓值(S 值)两个指标反映网络结构和聚类的清晰度,当 $Q > 0.3$ 时聚类结构显著,当 $S > 0.7$ 聚类结果合理^[19]。

2 化学领域与人工智能交叉态势分析

2.1 文献发文量分析

将收集到的 12 895 篇文献按照年份进行划分,文献年度分布如图 2 所示,红色箭头为 2016—2022 年文献数量的指数拟合曲线。2016 年之前,文献数量较为平稳,仅有不足 500 篇。但 2016 年之后,文

献数量呈现显著的指数型增长趋势,化学领域与人工智能的交叉步入快速发展期,领域之间结合愈发频繁。

2.2 国家/机构合作共现分析

当文献的作者所属国家/机构并不唯一时,可看作不同国家/机构之间的合作。国家/机构合作共现图如图 3 所示。从国家合作的角度,中国(3 930 篇)、美国(2 802 篇)为发文量最高的国家,并且美国在网络中的中心性最高,说明美国在国际科研合作中扮演着桥梁的角色,促进了不同国家之间的科研

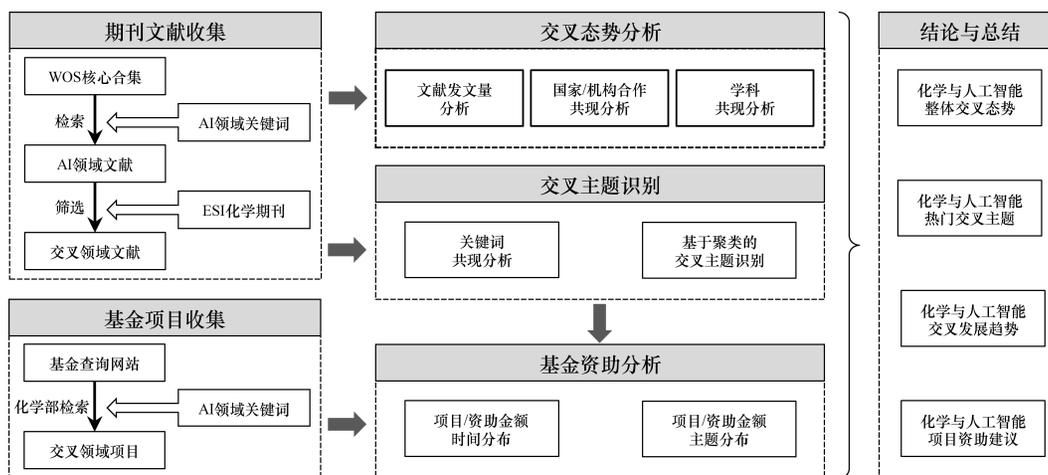


图 1 化学领域与人工智能热点交叉主题识别与趋势研究框架

表 1 文献检索策略

检索条件	检索策略
数据来源	Web of Science 核心数据集
文章类型	论文(Article)
出版年份	2013. 1. 1—2022. 12. 31
检索式	TS = (“Artificial Intelligence” OR “Deep learning” OR “Machine learning” OR “Transfer Learning” OR “Supervised Learning” OR “Unsupervised Learning” OR “Semisupervised Learning” OR “Reinforcement Learning” OR “Neural Network”)

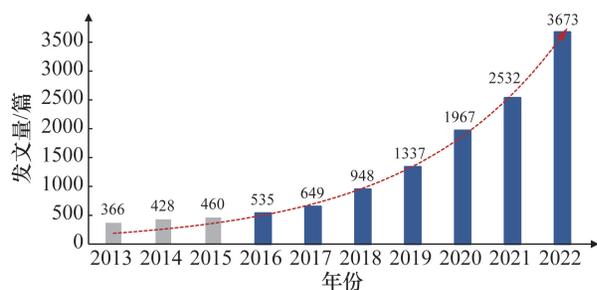


图 2 文献发文量年度分布

合作。从研究机构合作的角度,中国科学院(505 篇)、美国能源部(360 篇)是发文量最高的机构,且中心性也较高。这种广泛的国家/机构合作有助于促进跨国合作和知识共享,从而推动交叉领域的进步。

2.3 学科共现分析

文献与学科是多对多的关系,即一篇文献可以归类为多个学科,一个学科对应若干篇文献。基于 ESI 得到的化学领域与人工智能交叉的文献进行所属学科共现分析,可了解化学领域与人工智能交叉研究的学科构成。本研究所收集文献共涉及 60 个 WOS 分类下的学科,其中 57 个学科存在共现关系,学科共现图如图 4 所示。依据学科共现图,化学领域与人工智能的交叉研究形成了以交叉化学(工程化学)、物理化学、化学分析为中心的研究学科集群,其中交叉化学学科集群主要涉及与生物、环境、医疗等学科的交叉研究;物理化学学科集群重点研究量子领域的分子、原子性质;化学分析学科集群与光谱学、设备与仪器等学科交叉,主要涉及化学合成、成分分析等相关研究。

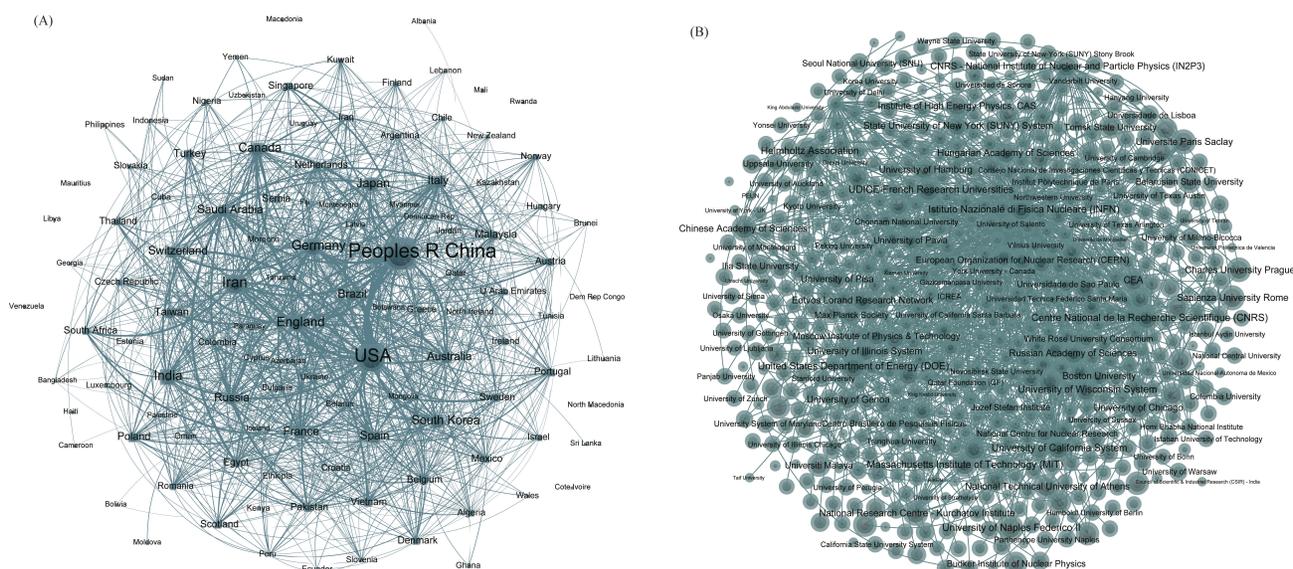


图3 国家/机构共现图:A. 国家共现图,B. 机构共现图

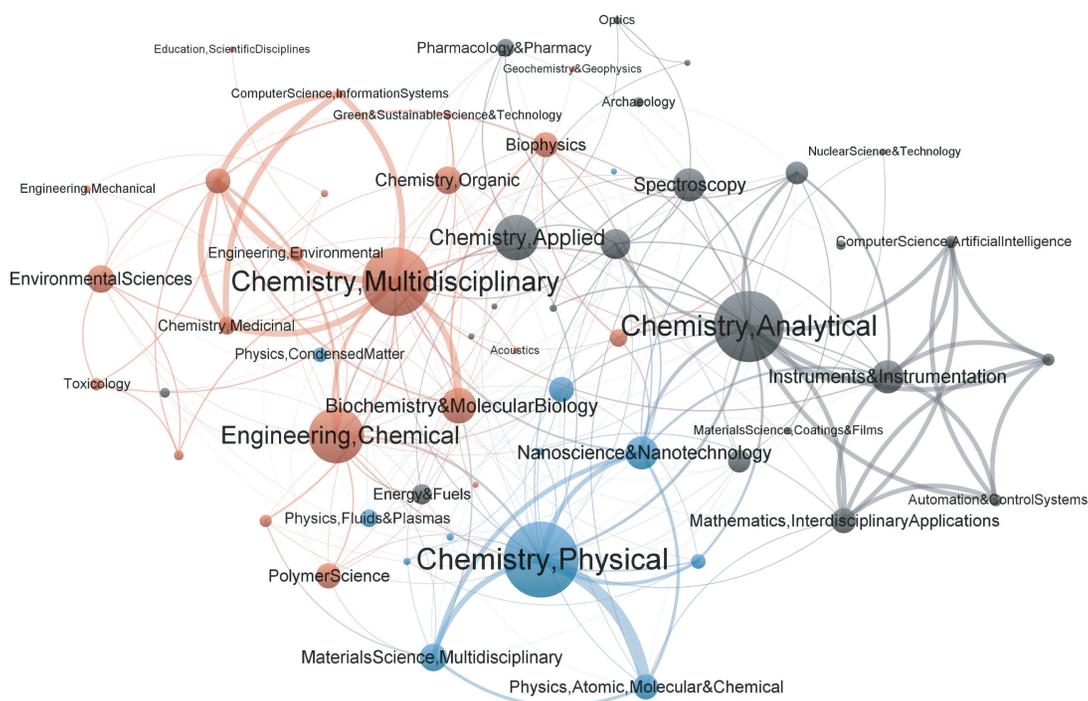


图4 化学领域学科共现图

3 化学领域与人工智能交叉热点主题挖掘

3.1 关键词共现分析

每篇论文的所有关键词之间构成了关键词共现关系,关键词共现图如图5所示(对单复数词、缩写词等进行了合并),高频关键词指代内容及含义如表2所示。综合来看,人工智能与化学领域存在紧密的交叉,一方面人工智能的技术在化学领域得到广泛应用,如卷积神经网络(Convolutional Neural

Networks, CNN)、循环神经网络(Recurrent Neural Network, RNN)等经典的神经网络在化学领域的应用以及迁移学习、强化学习等神经网络训练范式的选择;另一方面人工智能在化学领域的各个研究方向得到广泛使用,如在分子从头设计、合成路径、虚拟筛选、化合物性质预测等研究上取得了丰富的成果。

3.2 基于关键词聚类分析的热点交叉主题识别

在关键词共现图的基础上进行聚类,聚类的结

表 2 高频关键词指代内容及含义

关键词	词频	指代内容及含义
neural network	3 112	指代人工智能领域的神经网络模型,如 CNN、RNN、图神经网络等,人工智能技术被化学领域相关研究广泛采用。
machine learning	2 032	机器学习是人工智能的子领域,其专注于开发算法和模型,化学领域与人工智能进行交叉时更侧重于从数据中抽取模式和知识等工作。
prediction	1 382	指代结构预测、共形预测、熔体指数预测等,人工智能在化合物结构预测、分子生物活性预测、物理化学性质预测等方面存在普遍研究。
model	1 195	指代数学建模,基于人工智能构建特定任务的推理、预测等模型为人工智能在化学领域的典型应用。
design	838	指代分子设计、药物设计、从头设计等,人工智能在分子设计或化合物合成等方面有丰富的研究。
optimization	748	指代粒子群优化、全局优化、多目标优化等优化算法,人工智能在优化分子结构、化学反应条件等有丰富的研究
classification	671	指代化合物分类、质谱数据分类、性质分类等,人工智能可以辅助识别和理解化合物、光谱数据等
drug discovery	662	人工智能在药物发现的各个步骤中均有应用,包括寻找潜在分子、优化化合物结构和合成路线设计等
identification	651	指代人工智能技术用于区分相似化合物和特定性质分子的识别
system	541	指代基于人工智能构建化学领域特异于某项任务的系统

CiteSpace, v. 6.2.R7 (64-bit) Advanced
 January 29, 2024, 9:45:33 PM CST
 WoS: C:\Users\TGCH\Desktop\科技文献管理专项\Analyse\ESICHEMISTRY\data
 Timespan: 2013-2022 (Slice Length=1)
 Selection Criteria: g-index (k=15), LRF=3.0, L/N=10, LB=5, e=1.0
 Network: N=636, E=2233 (Density=0.0111)
 Nodes Labeled: 1.0%
 Pruning: Pathfinder
 Modularity Q=0.535
 Weighted Mean Silhouette S=0.7679
 Harmonic Mean(Q, S)=0.6306

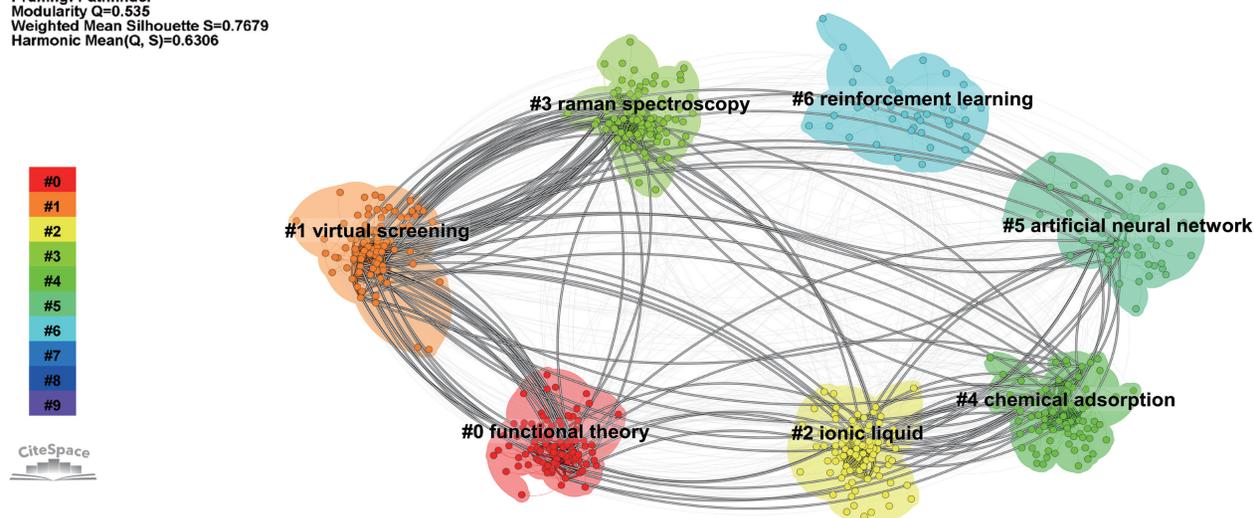


图 6 关键词聚类图

效率的缺点。而基于人工智能的方法可根据已知特性分子的性质、结构等信息学习到具有该特性的分子特征进而完成虚拟筛选。例如:2022 年发表在 *Nature Biomedical Engineering* 的文章中^[22],建立了一种基于人工智能融合多维分子信息的虚拟筛选算法,并结合细胞、线虫、小鼠多物种阿尔兹海默模型验证,成功快速筛选获得了具有该疾病治疗潜力的小分子化合物,为研究该疾病的治疗提供了一种高效的药物发现方案。

3.2.3 离子液体

离子液体是指在室温或接近室温下呈现液态的、完全由阴阳离子所组成的盐,也称为低温熔融盐。离子液体在化学反应溶剂、催化剂、电解质、提取介质等多个方面都扮演着重要角色。基于定量结构—性质/活性关系为预测如离子液体等分子系统性质的一种通用方法,该方法通过对已知化合物的性质/活性数据与包含分子结构信息的描述符进行建模,得到结构与性质之间的关联,进而实现对化合物未知性质的预测^[23]。例如:2023 年发表在

Nature Communications 的文章中^[24],开发了一种嵌入量子化学计算和图卷积神经网络的机器学习方法,得到了 49 个常温下为液态、常温下电导率大于 5 mS/cm 且电化学窗口大于 4 V 的离子液体。

3.2.4 拉曼光谱

拉曼光谱是指一种基于光和材料内化学键的相互作用而产生的一种散射光谱,其对分子键合以及分子结构非常敏感,因此每个分子都会产生其独特的光谱指纹。基于拉曼光谱可鉴别分子的结构特征或结构基团以及无机化合物的晶形结构等,能够在分子水平上研究电化学界面结构、吸附和反应等基础问题,并应用于电催化、腐蚀和电镀等领域。机器学习能够极大的改进传统的光谱分析方法,例如:2023 年发表在 *small* 的文章中^[25],利用拉曼光谱和机器学习算法实现了对来自五种细胞系的细胞外囊泡的化学特征识别,这些特征可以在液体活检样本中识别肿瘤细胞来源的细胞外囊泡,从而用于肿瘤的早期发现和预后。

3.2.5 化学吸附

吸附法被认为是处理含有如亚甲基蓝等污染物的废水的一种有效方法。在吸附实验优化中,传统的基于线性或多项式模型优化多个变量对响应变量影响的响应面法,对非线性关系的建模能力有限,且容易陷入局部最优解。因此利用机器学习的非线性关系建模和泛化能力来优化吸附参数具有显著的优势。另一方面,化学吸附可以在催化剂表面提供活性位点(活性位点是指对于特定的化学反应具有催化活性的位置),促进反应物质之间的相互作用,从而优化催化剂的设计。控制吸附过程可以提高催化剂的反应活性和选择性,例如于 2022 年发表在 *Nature Communications* 的文章中^[26],提出了基于吸附物化学环境的图卷积神经网络,该网络能够考虑到包含各种吸附物、结合位置、配位环境和基底形态在内的原子构型,实现系统的估计复杂多相表面催化反应的表面吸附结构,从而做到对在结构和化学复杂环境中异相催化影响因素的深入了解。

3.2.6 人工神经网络

人工神经网络是机器学习的一个分支,它是由生物神经系统启发而设计出的一种计算模型,用于模拟人脑神经元之间的信息传递和处理过程。人工神经网络是人工智能与化学领域研究交叉的一种基础计算模型,其优势在于能够从数据中学习模式和关联,适用于处理复杂、多维的化学信息,可以应用于反应预测、多因素下的模型预测(浓度、效率)等多

个方面。例如:2019 年发表于 *Chemical Science* 的文章中^[27],在给定反应物、试剂和溶剂种类的情况下,利用图卷积神经网络实现了对有机反应产物的预测。

3.2.7 强化学习

强化学习是机器学习的另一个分支,它依据在学习过程中获得的奖励或惩罚不断学习知识。在化学领域,强化学习可以应用于分子设计、反应路径预测等多个方面。例如:2018 年发表在 *Science Advances* 的文章中^[28],基于 ChEMBL 数据库中的 150 万个分子,提出了一种利用强化学习进行从头设计具有所需特性分子的方法。从头设计分子一直以来的缺陷为生成的化学分子无法合成,而该方法设计的分子被证明 95% 的分子结构有效。2020 年发表在 *Chemical Science* 的文章中^[29],提出了一种结合强化学习训练得到的价值网络、溶剂预测神经网络与蒙特卡洛树搜索变体的算法,该算法能够在固定搜索时间内找到有效的反应路径。在相同的搜索条件下,该方法显著提高了寻找有效反应路径的成功率。

4 基于热点交叉主题下的基金资助分析

本研究共收集到 NSFC 项目 254 项、NSF 项目 191 项,总资助金额分别为 15 225.3 万元和 8 386.5 万美元。将基金项目依据每个交叉主题下的研究方向进行划分,得到各交叉主题下的项目和资助金额分布。NSFC 项目及资助金额分布如图 7 所示。从时间维度观察,化学领域与人工智能相关项目数量以及资助金额未呈现出明显的发展规律。在各主题的项目分布方面,“泛函理论”主题在近十年持续有相关项目立项,该主题下项目数量为 62 项,占据总数的 24.41%。而其它主题在历年的立项状况表现不一,“虚拟筛选”“离子液体”“化学吸附”主题自 2015 年开始持续有项目立项;在各主题的资助金额方面,“泛函理论”主题处于领先地位,总资助金额为 3 648.60 万元,占据总数的 23.96%。其余主题除“虚拟筛选”主题占比 18.31% 以外,总资助金额占比均在 11.26% 左右,大约 1 714.50 万元。

NSF 项目及资助金额分布如图 8 所示。与 NSFC 项目分布不同,NSF 项目数量及资助金额呈现出递增的趋势,且各主题间的项目数量和资助金额差异相对更为明显。而与 NSFC 项目分布相似的是,“泛函理论”主题在近十年也表现出持续立项,并且项目数量和资助金额占比最高,分别为 75 项 (39.27%) 和 3 106.43 万美元。

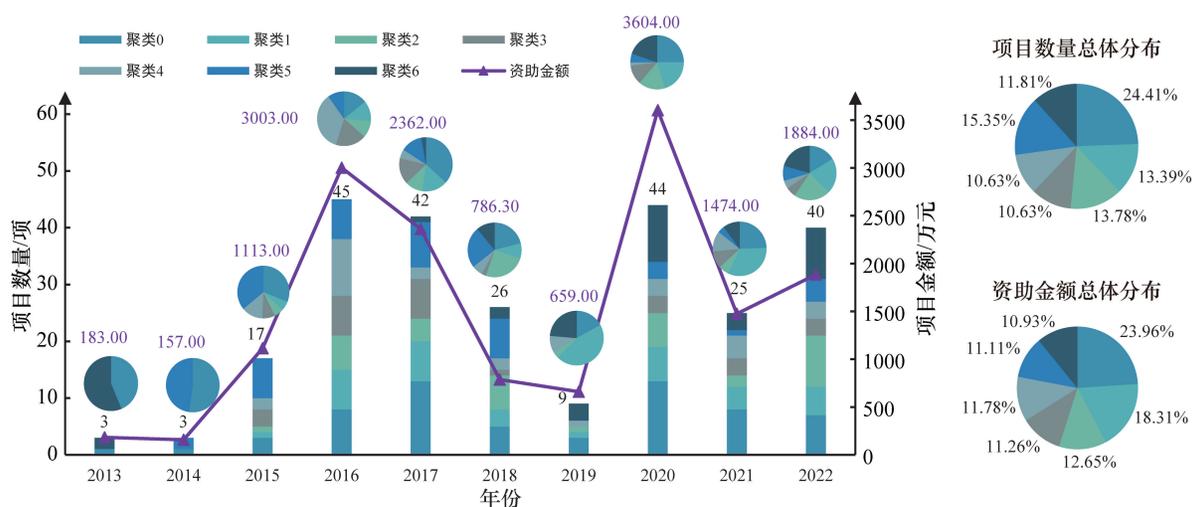


图7 NSFC项目及资助金额分布

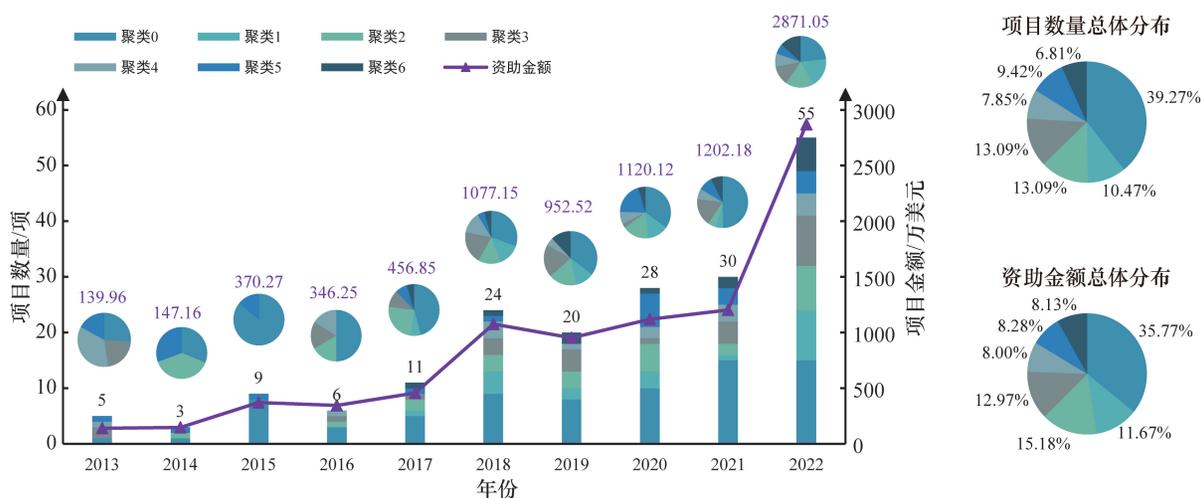


图8 NSF项目及资助金额分布

主题之间的项目数目和资助金额的差异一定程度上反映了研究主题的发展状况:一方面,其表现了研究主题的成熟度,即已有大量研究和知识积累的主题更容易获得资助用于进一步发展和深化该主题下的研究;另一方面,其反映了研究主题的复杂度,即不同的主题所需实验的规模、计算成本等存在差异,基金资助金额的需求因而也不同。在泛函理论主题下的研究中,电子结构的建模带来了严重的计算复杂性,而人工智能带来了高效率、高准确性和可扩展性的有效方法,在化学空间的探索方面发挥了重要作用。因此,基于项目分布和主题特点,泛函理论主题在化学与人工智能交叉领域的研究热度和关注程度最高,是交叉领域的核心研究主题。

5 结论

5.1 化学领域与人工智能的交叉研究愈发广泛

近十年化学领域与人工智能的交叉研究愈发丰富与热门。首先,文献发文数量呈现指数增长趋势,交叉研究产生了越来越多的研究成果。一方面,国家与机构之间的广泛合作以及科研项目与资金的投入支持,为交叉研究提供了强大的动力;另一方面,得益于如强化学习、深度学习等人工智能技术的进步以及大规模化学数据库的构建,促使人工智能在化学领域的应用更加强大与多样化。二者共同推动了化学领域与人工智能的协同作用,使得人工智能为解决一系列化学领域科学问题提供了有力的工具和机会。其次,研究成果的行业应用愈发多样。人工智能与化学领域的交叉研究不仅仅局限于研究和

发现阶段,还扩展到了药物发现、新型材料开发、化学工程与生产、环境监测与污染控制以及能源等多个领域,这种多样性反映出人工智能在提高效率、创新和可持续性等方面的巨大潜力,使各个行业都能受益于这一交叉领域的进展。

5.2 化学领域与人工智能的交叉融合潜力巨大

人工智能为化学领域提供了解决问题的全方位视角。本研究通过近十年相关文献识别出泛函理论、虚拟筛选、离子液体、拉曼光谱、化学吸附、人工神经网络、强化学习等七个热点交叉主题,涵盖了化学领域多方面的研究。从构建分子结构到优化反应路径,人工智能不仅能够建模分子结构与性质之间的联系确定分子设计方案,并完成合成路径的规划,而且对于合成产物能够进行分析表征与性质评估,并进一步进行优化和改进。这种多主题、多技术手段的交叉融合为化学研究提供了全新的思路和方法,展现出化学领域与人工智能交叉融合的巨大潜力。

5.3 人工智能属性由研究工具向研究范式转变

随着研究对象逐渐复杂化和高维度化,人工智能在化学领域的应用逐渐由辅助分析工具发展为研究范式。人工智能与化学领域交叉的动机取决于研究问题的“复杂性和计算性需求”以及“自动化和智能化需求”,多样化的大规模数据积累是交叉研究的基础。对于原子间复杂的多体相互作用的模拟,在传统方法受到计算资源的局限下,结合量子力学理论与机器学习的算法能够显著加速计算,是化学领域与人工智能交叉中长期处于重要地位的研究主题。而随着数据的积累以及实验自动化需求的不断增长,人工智能在化学领域开始转向基于大数据的研究,如性质预测、高通量筛选以及反应优化等研究逐渐转变为基于大规模数据挖掘潜在模式并建立预测模型来优化实验的效率与成本的研究范式,这种研究范式的变革将随着人工智能大模型发展进一步凸显。

5.4 不同主题的资金资助策略需不同

不同的交叉主题的科学背景与资源需求存在一定的差异,应针对不同的研究主题设置个性化的资助策略。不同研究主题的成熟度和复杂性不同,依据主题下的期刊文献、成果转化等数量和质量的差异应将主题所处的研究阶段划分为不同的层次。针对处于早期阶段的研究主题应侧重于高水平论文的产出,具有一定研究积累的研究主题应侧重于科研

成果转化。同时,不同研究主题的发展趋势是动态变化的,应及时针对不同发展阶段的主题调整资助策略,建立研究人员与科学界的持续反馈机制,基于需求和潜力等角度确保资源分配合理,确保项目资助策略与实际需求和趋势保持一致,以支持多样化和全面性的研究。最后,研究主题的发展受资助力度的显著影响,应鼓励跨学科合作,注重支持早期探索性研究,并且基金资助可考虑提供更多的支持来鼓励高风险、高回报的研究项目,激励研究人员在化学领域与人工智能交叉研究上进行更多的尝试,从而推动创新。

5.5 交叉研究的资助策略体系需要进一步完善

跨学科领域的交叉融合能够加速知识创新,应引入数据驱动的区域交叉热点识别机制,维持和保障热点主题的长期稳定资助。首先,基于知识图谱、机器学习等技术手段,通过定期监测交叉领域的动态发展来捕捉新兴前沿研究,再经专家评审、分析等实现前瞻性的资助战略布局。其次,针对热点或潜力交叉主题提供长期稳定的资助保障,突破短期资助带来的“重复性、移植性研究”科研瓶颈,推动深层次的领域融合和原创性成果创新。最后,以重点资助核心交叉主题为抓手,对比国外机构的资助强度、模式与趋势,优化基金资助分配,提高我国在世界范围内的科研竞争力,为科学研究的持续发展提供更多机遇。

6 结 语

本研究对基于 ESI 专业领域划分下的近十年化学领域期刊文献进行了发文量分析、学科共现分析、国家和机构合作共现分析、关键词分析。分析结果表明化学领域与人工智能交叉的研究随着时间的发展显著增加,形成了以化学跨学科、物理化学、化学分析为中心的研究学科集群,中国(中国科学院)、美国(美国能源部)在交叉研究领域扮演重要角色,识别出泛函理论、虚拟筛选、离子液体、拉曼光谱、化学吸附、人工神经网络、强化学习等七个热点交叉主题。之后,本研究还基于相关的 NSFC 与 NSF 基金项目进行了统计分析,从纵向发展角度对比了中美之间的交叉研究资助趋势差异,从横向对比角度揭示了泛函理论主题的核心地位以及不同主题间的资助强度差异。这些结果体现了化学领域与人工智能的交叉研究现况以及人工智能属性由研究工具向研究范式转变的发展趋势,突出了不同交叉研究

主题的基金资助策略需不同的重要意义,表现出化学领域与人工智能交叉在未来广阔的研究和应用前景。

参 考 文 献

- [1] 科技部: 加快推动国家新一代人工智能公共算力开放创新平台建设. 中国设备工程, 2023(8): 1.
- [2] 余惠敏. 人工智能加速基础研究变革. 经济日报, 2023-05-14(001).
- [3] 吉远辉, 朱家华, 穆立文, 等. 化工基础数据获取新范式: 机制+数据驱动. 中国科学基金, 2024, 38(04): 712—718.
- [4] 朱博阳, 吴睿龙, 于曦. 人工智能助力当代化学研究. 化学学报, 2020, 78(12): 1366—1382.
- [5] 王凯, 袁志宏, 王笑楠, 等. 化学品智能制造的科学基础. 中国科学基金, 2023, 37(05): 811—817.
- [6] 孔祥泰, 张润泽, 张玮, 等. 人工智能技术在化学反应预测中的应用. 中国现代应用药学, 2022, 39(21): 2856—2864.
- [7] 丁邵珍, 江小琴, 孟超, 等. AI及大数据技术在逆合成路线设计中的应用. 中国科学: 化学, 2023, 53(1): 66—78.
- [8] Baum ZJ, Yu X, Ayala PY, et al. Artificial intelligence in chemistry: current trends and future directions. *Journal of Chemical Information and Modeling*, 2021, 61(7): 3197—3212.
- [9] 刘小平, 刘耀虎, 郑企雨, 等. 人工智能化学: 变革研究范式, 加速物质发现. 化学通报, 2023, 86(6): 748—754.
- [10] Choudhary N, Bharti R, Sharma R. Role of artificial intelligence in chemistry. *Materials Today: Proceeding*, 2021: 1527—1533.
- [11] 李佳蕾, 安培浚, 肖仙桃. 学科交叉主题识别方法研究综述. 数据分析与知识发现, 2023, 7(4): 1—15.
- [12] 黄福, 侯海燕, 任佩丽, 等. 基于共被引与文献耦合的研究前沿探测方法遴选. 情报杂志, 2018, 37(12): 13—19, 35.
- [13] 王福, 李哲, 刘俊华, 等. 近十年来竞争供应链研究热点及其演化——基于关键词共现和社会网络分析. 供应链管理, 2022, 3(8): 5—19.
- [14] 裘惠麟, 邵波. 多源数据环境下科研热点识别方法研究. 图书情报工作, 2020, 64(5): 78—88.
- [15] Chen CM. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 2006, 57(3): 359—377.
- [16] 李杰, 陈超美. CiteSpace: 科技文本挖掘及可视化. 2版. 北京: 首都经济贸易大学出版社, 2017.
- [17] van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 2010, 84(2): 523—538.
- [18] 吴晶磊, 孟庆峰, 邸月宝, 等. 近十年NSF资助率和资助强度上升对我国科学基金资助工作的启示. 中国科学基金, 2024, 38(04): 687—695.
- [19] 陈悦, 陈超美, 刘则渊, 等. CiteSpace知识图谱的方法论功能. 科学学研究, 2015, 33(2): 242—253.
- [20] Kirkpatrick J, McMorro B, Turban DHP, et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, 2021, 374(6573): 1385—1389.
- [21] Zhang LF, Han JQ, Wang H, et al. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical Review Letters*, 2018, 120(14): 143001.
- [22] Xie CL, Zhuang XX, Niu ZM, et al. Amelioration of Alzheimer's disease pathology by mitophagy inducers identified *via* machine learning and a cross-species workflow. *Nature Biomedical Engineering*, 2022, 6(1): 76—93.
- [23] Koutsoukos S, Philippi F, Malaret F, et al. A review on machine learning algorithms for the ionic liquid chemical space. *Chemical Science*, 2021, 12(20): 6820—6843.
- [24] Li K, Wang J, Song Y, et al. Machine learning-guided discovery of ionic polymer electrolytes for lithium metal batteries. *Nature Communications*, 2023, 14(1): 2789.
- [25] Parlatan U, Ozen MO, Kecoglu I, et al. Label-free identification of exosomes using Raman spectroscopy and machine learning. *Small*, 2023, 19(9): e2205519.
- [26] Ghanekar PG, Deshpande S, Greeley J. Adsorbate chemical environment-based machine learning framework for heterogeneous catalysis. *Nature Communications*, 2022, 13(1): 5788.
- [27] Coley C, Jin WG, Rogers L, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chemical Science*, 2019, 10(2): 370—377.
- [28] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. *Science Advances*, 2018, 4(7): eaap7885.
- [29] Wang XX, Qian YJ, Gao HY, et al. Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning. *Chemical Science*, 2020, 11(40): 10959—10972.

Research on Identification and Trend of Chemistry and Artificial Intelligence Hot Cross Topics Based on ESI

Junwei Han* Guochang Tang Shijie Zhao

School of Automation, Northwestern Polytechnical University, Xi'an 710072

Abstract This study explores the current intersection between the fields of chemistry and artificial intelligence (AI) with the aim of identifying hot research topics. The goal is to provide strong evidence for decision-making in funding projects, thereby optimizing resource allocation and promoting cutting-edge scientific advancements in the field of chemistry. The research relies on cross-disciplinary journal literature and funding project data, revealing the interdisciplinary landscape through dimensions such as publication volume, international collaborations, and cross-disciplinary subjects. By employing co-occurrence and clustering analyses on keywords extracted from journal articles, the study identifies seven prominent cross-disciplinary research themes, including functional theory and virtual screening. Further analysis involves examining the funding situations in these topics for China and the United States. The findings highlight a significant growth in research at the intersection of chemistry and AI over time. The identified hot topics shed light on the potential applications of AI in chemistry and offer insights into future trends. The study suggests that funding strategies should be tailored to different themes, considering the distinct characteristics and requirements of each. This comprehensive approach not only enhances our understanding of the intersection between chemistry and AI, but also provides valuable guidance for strategic funding decisions to drive advancements in the field.

Keywords artificial intelligence; chemical field; knowledge mapping; cross topic identification; Natural Science Foundation

(责任编辑 陈鹤 张强)

* Corresponding Author, Email: jhan@nwpu.edu.cn