

强化学习算法研究热点与发展趋势探析

佟 磊^{1,2}

¹河北软件职业技术学院软件工程系，河北 保定

²河北省智能互联装备与多模态大数据应用技术研发中心，河北 保定

收稿日期：2026年1月1日；录用日期：2026年1月28日；发布日期：2026年2月6日

摘要

为系统揭示2021年至2025年强化学习领域的研究态势与发展脉络，以中国知网收录的5099篇“强化学习”主题核心期刊论文为研究对象，采用文献计量法与可视化分析法，从文献时间分布、关键词共现与聚类、关键词突现等维度展开分析。研究发现：该领域文献呈快速扩张和稳定提质两阶段增长特征，前期聚焦基础算法改进与单一场景验证，后期转向细分场景深化与多技术融合；研究热点覆盖算法创新、多领域落地、技术协同与安全保障；前沿演进遵循技术融合到场景深化再到知识驱动的路径。研究可为科研人员把握领域热点、布局未来研究方向提供参考，为强化学习技术的产业化应用提供理论支撑。

关键词

强化学习，文献计量法，研究热点，发展趋势，技术融合

An Analysis of Research Hotspots and Development Trends in Reinforcement Learning Algorithms

Lei Tong^{1,2}

¹Department of Software Engineering, Hebei Software Institute, Baoding Hebei

²Hebei R&D Center for Intelligent Connected Equipment and Multimodal Big Data Application Technology, Baoding Hebei

Received: January 1, 2026; accepted: January 28, 2026; published: February 6, 2026

Abstract

To systematically reveal the research status and development context of the reinforcement learning field from 2021 to 2025, this study takes 5099 core journal papers themed “reinforcement learning”

文章引用：佟磊. 强化学习算法研究热点与发展趋势探析[J]. 计算机科学与应用, 2026, 16(2): 182-190.
DOI: [10.12677/csa.2026.162050](https://doi.org/10.12677/csa.2026.162050)

included in China National Knowledge Infrastructure (CNKI) as the research objects, and conducts analysis from the dimensions of literature time distribution, keyword co-occurrence and clustering, and keyword burst detection by adopting the bibliometric method and visual analysis method. The results show that the literature in this field has two-stage growth characteristics of rapid expansion and steady quality improvement: the early stage focuses on basic algorithm improvement and single-scenario verification, while the later stage shifts to the deepening of subdivided scenarios and multi-technology integration. The research hotspots cover algorithm innovation, multi-field application, technical collaboration and security assurance. The frontier evolution follows the path from technology integration to scenario deepening and then to knowledge-driven development. This study can provide references for researchers to grasp the hotspots of the field and layout future research directions, as well as offer theoretical support for the industrial application of reinforcement learning technology.

Keywords

Reinforcement Learning, Bibliometric Method, Research Hotspots, Development Trends, Technology Integration

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在人工智能技术从感知向决策深度跨越的进程中，强化学习(Reinforcement Learning, RL)凭借交互试错的核心机制，成为解决复杂动态决策问题的关键技术。强化学习的理论溯源可至 20 世纪 50 年代的“试错学习”理论，Sutton 与 Barto 在《Reinforcement Learning: An Introduction》中构建了完整的数学框架，明确了智能体、环境、奖励函数等核心要素，为领域发展奠定理论基础[1]。强化学习可应用于自动驾驶领域、机器人控制领域、推荐系统、市场交易和自动竞价等领域，其应用领域广阔，市场前景巨大[2]。

近年来，强化学习与深度学习、边缘计算、区块链等技术的融合创新，进一步拓展了其应用边界：深度强化学习解决了高维状态空间的决策难题[3]；基于强化学习的边缘计算实现了边缘计算任务的实时卸载及低能耗[4][5]；联邦强化学习兼顾了数据隐私与协同优化[6]。与此同时，2021 年后全球强化学习研究呈现爆发式增长，仅中国知网收录的核心期刊论文便从 2021 年的 587 篇增至 2025 年的 1324 篇，五年间增长 2.26 倍，文献内容涵盖算法改进、场景应用、安全优化等多个维度。面对海量且分散的研究成果，传统人工综述方法难以全面捕捉核心热点与前沿趋势，亟需借助可视化计量工具揭示领域知识结构。

CiteSpace 作为知识图谱分析的主流工具，可通过挖掘文献关键词共现、聚类、突现等潜在关联，以可视化形式呈现研究领域的发展脉络[7]。该工具已被广泛应用于人工智能领域的学术计量研究，王鹏等人基于 CiteSpace 梳理了我国人工智能高质量数据集的研究趋势和热点方向[8]；陈伊高等人借助其梳理人工智能技术与出版深度融合的趋势[9]。吴岩等采用 CiteSpace 的可视化分析，揭示配电网韧性研究领域的研究动向[10]。基于此，以 2021~2025 年中国知网核心期刊 5099 篇“强化学习”主题文献为对象，通过 CiteSpace 展开多维度可视化分析，旨在系统揭示该时期内强化学习算法的研究热点、前沿动态与机构合作特征，为科研人员选择研究方向、开展跨领域合作提供清晰的学术图景，同时为强化学习技术的产业化应用提供理论支撑。

2. 研究设计

2.1. 研究数据

研究所用数据基于中国知网数据库，通过中国知网数据库中的高级检索功能，研究主题设置为“强化学习”，检索范围为期刊论文。通过初步文献检索，在知网上能检索到该研究主题的最早期刊论文发表于1996年，检索得到14531篇相关文献。为了选择有代表性的期刊论文，将期刊来源类别限定为北大核心期刊、CSCD期刊和CSSCI期刊，能检索到的论文最早发表时间是1992年，共计7433篇文献，将文献研究时间范围设定为2021年1月1日至2025年11月15日，并对文献进行人工筛选、剔除不相关的会议、新闻公告等得到5099篇有效文献，最后选择“导出与分析”，以Refwork格式导出相关数据。

2.2. 研究方法

本研究综合采用文献计量法与可视化分析法，研究工具选用CiteSpace软件，该工具可通过挖掘文献潜在知识关联，以可视化图谱呈现知识结构与规律，适用于关键词共现、聚类及突现等维度分析。数据预处理阶段，启动CiteSpace后通过菜单栏“Data→Import/Export→CNKI”路径完成CNKI数据格式转换，并启用“Duplicate Detection”模块去除重复文献，确保数据有效性。参数设置方面，依据文献时间跨度将时间切片长度设为1年，节点类型选择“关键词”与“研究机构”，以TopN=50提取每切片被引频次前50的文献节点，采用“Pathfinder+Pruning sliced networks”方法优化网络结构，关键词突现分析中设置 γ 值为1、最小持续时长为1年。最终从三个维度展开分析：文献时间分布维度用于度量研究热度与阶段划分，关键词共现与聚类维度用于识别核心研究热点及专题结构，关键词突现维度用于定位研究前沿与演变趋势，形成系统的强化学习算法研究可视化分析框架。

3. 研究结果与分析

3.1. 文献时间分布分析

文献发文量的时间分布可直观反映领域研究热度与发展阶段。基于CiteSpace的“Time Series”模块分析，2021~2025年强化学习算法研究发文量呈“持续高速增长”特征，无明显回落，具体数据见图1。

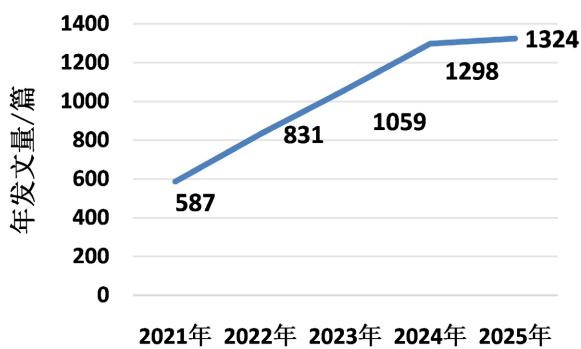


Figure 1. Publication volume in the field of reinforcement learning (2021~2025)

图 1. 2021 年~2025 年强化学习领域发文量

结合数据可将该时期划分为两个发展阶段：

快速增长期(2021~2023年)：此阶段年均发文增长率达38.7%，发文量从2021年的587篇快速攀升至2023年的1059篇，两年间实现近翻倍增长。从文献类型来看，该阶段以基础算法改进类与单一场景应用验证类文献为主，前者聚焦深度强化学习算法的稳定性优化、传统Q-Learning等经典算法的参数调

整[11]，以及强化学习与深度学习融合的技术适配研究；后者则集中于无人机路径规划、智能电网初步调度等单一、低复杂度场景的技术落地验证，旨在验证强化学习技术在实际场景中的可行性[12][13]。深度强化学习核心算法逐渐成熟，技术门槛降低，为广大研究者提供了可复用的基础框架，吸引大量学者进入该领域开展入门级研究。而且，2021 年发布的《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》将人工智能列为新兴数字产业重点培育方向，明确提出促进数字技术与实体经济深度融合，政策导向与科研基金扶持共同激发了研究热情，而各行业对智能决策技术的初步探索需求，也促使学者优先开展技术可行性验证类研究，进而推动文献数量快速增长。

稳定增长期(2024~2025 年)：该阶段发文增长率放缓至 17.4%，发文量从 2024 年的 1298 篇温和增长至 2025 年的 1324 篇，增长节奏趋于平稳。文献类型呈现显著转型，以细分场景深度优化类与多技术融合类文献为主，前者聚焦机械臂轨迹跟踪精度提升、车联网动态资源分配效率优化等细分场景的技术深化研究，注重解决实际应用中的复杂问题[14][15]；后者则集中于强化学习与联邦学习、数字孪生等技术的融合创新，探索“强化学习 + X”的复合型技术方案，以突破单一技术的应用瓶颈[6][16]。经过前期研究积累，基础算法与单一场景验证的研究空间逐渐收窄，领域从数量扩张转向质量提升，学者更倾向于在细分领域深耕；同时，行业对强化学习技术的精度、效率、安全性提出更高要求，倒逼研究向深度优化与多技术融合方向发展，此外，边缘计算、数字孪生等支撑技术的成熟，也为强化学习的技术融合提供了条件，进一步推动文献类型向高复杂度、高实用性转变。

3.2. 关键词共现分析

关键词共现网络可揭示研究热点间的关联强度。基于 5099 篇文献，使用 CiteSpace 生成关键词共现图谱，见图 2。

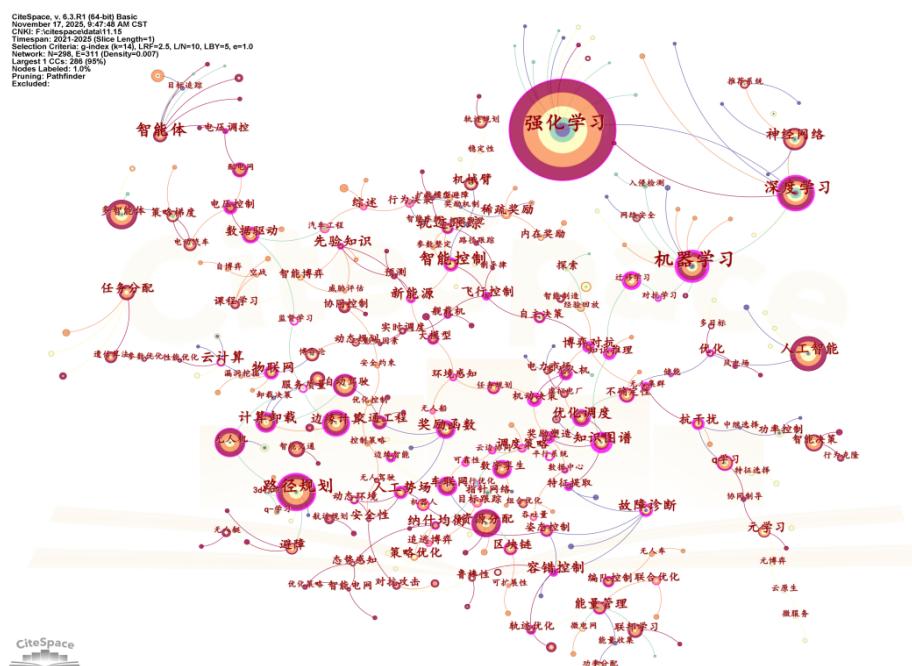


Figure 2. Keyword co-occurrence map
图 2. 关键词共现图谱

该图谱包含 298 个节点以及 311 条连线，网络密度为 0.007，表明关键词关联复杂且存在明确核心

枢纽。节点大小表示关键词频次，节点越大，代表该关键词在领域内关注度越高，连线粗细反映共现强度。

Table 1. Top 20 high-frequency keywords in reinforcement learning algorithm research (2021~2025)
表 1. 2021~2025 年强化学习算法研究高频关键词 Top 20

序号	关键词	词频	中心性
1	强化学习	1466	0.14
2	人工智能	199	0.05
3	路径规划	193	0.23
4	深度学习	178	0.27
5	多智能体	140	0.03
6	无人机	124	0.16
7	机器学习	118	0.40
8	资源分配	115	0.16
9	边缘计算	89	0.34
10	神经网络	84	0.06
11	自动驾驶	75	0.14
12	智能决策	51	0.04
13	计算卸载	50	0.27
14	车联网	44	0.34
15	智能交通	43	0.01
16	Q 学习	41	0.13
17	任务分配	41	0.04
18	数据驱动	38	0.30
19	联邦学习	38	0.03
20	奖励函数	35	0.25

根据 CiteSpace 输出的“Keyword Frequency”统计，词频前 20 位的关键词及中心性见表 1 所示。“强化学习”以 1466 次词频稳居核心，作为领域技术基石，与“深度学习”“机器学习”共同构成技术核心三角，其中“机器学习”以 0.40 的中心性成为网络最大枢纽，印证深度强化学习是主流技术范式；应用场景类关键词表现突出，“路径规划”(193 次)、“无人机”(124 次)、“资源分配”(115 次)跻身前列，反映强化学习在移动载体控制与资源优化领域的集中落地；“边缘计算”和“车联网”中心性均达 0.34，“数据驱动”中心性为 0.30，体现轻量化部署、场景适配与数据驱动优化的研究重点；“联邦学习”“Q 学习”“奖励函数”等关键词则揭示算法隐私保护、经典算法改进与核心组件优化的细分趋势，整体展现出技术深化、场景细化与多领域融合的发展格局。

3.3. 关键词聚类分析

关键词共现分析已成功识别出强化学习算法研究文献中的高频关键词，并将这些关键词界定为该领域的主要研究内容，但仅依靠单个高频关键词，难以明确强化学习算法的具体研究专题及其核心涵盖范围。聚类分析的核心价值在于直观呈现该领域的热点研究主题及关键词聚类特征，以已构建的关键词共

现知识图谱为基础,选取 K 标签来源(Keywords Source of Labels)作为聚类标签生成方式开展分析,经后续优化处理后,最终得到关键词共现聚类图,见图 3。

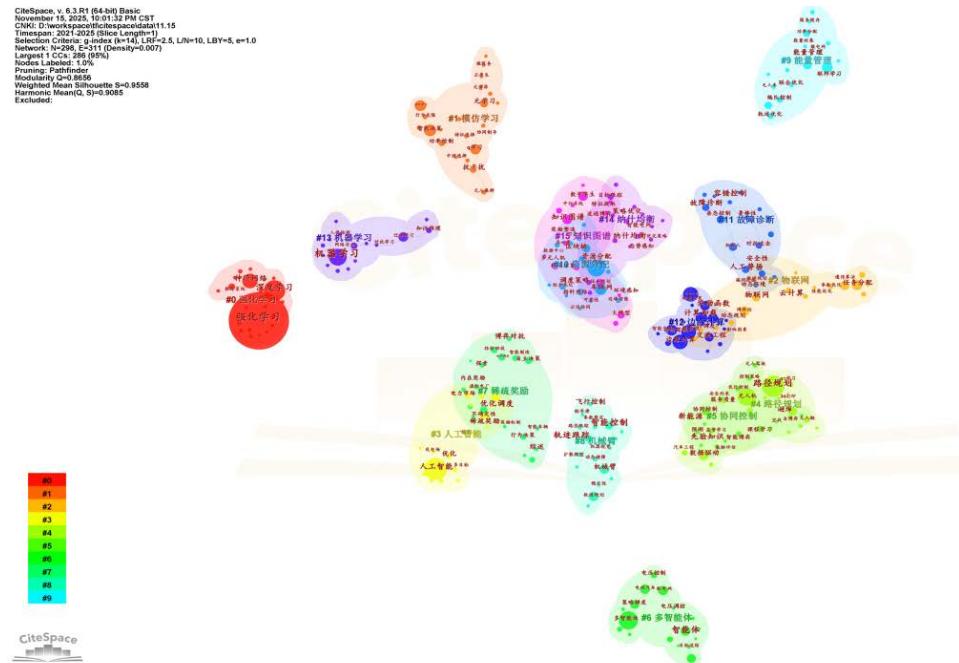


Figure 3. Keyword co-occurrence clustering map for reinforcement learning algorithm research
图 3. 强化学习算法研究关键词共现聚类图

通过对 5099 篇文献中关键词进行聚类最终得到 16 个有效聚类类别,这反映出强化学习算法的研究专题,同一聚类中词频值最大的关键词共计 16 个。这 16 个关键词聚类包括强化学习、模拟学习、物联网、人工智能、路径规划、协同控制、多智能体、稀疏奖励、机械臂、能量管理、资源分配、故障诊断、边缘计算、机器学习、纳什均衡、知识图谱等。在显著性方面,该聚类的模块度 Q 值为 0.8656 远大于 0.3,表明该聚类显著。与此同时,平均轮廓 S 值为 0.9085 大于 0.7,表明该聚类高效[17]。具体聚类专题及高频关键词见表 2 所示。

表中,“聚类编号”是系统自动生成的聚类标识,用于区分不同研究专题;“聚类大小”指聚类包含的成员数量,规模越大通常代表该主题越核心;“轮廓值”衡量聚类内部一致性与类间区分度,大于 0.7 则聚类质量较高;“聚类高频关键词”是聚类中权重最高的词汇,体现研究重点;“平均发表年份”反映主题首次出现或活跃的时间。通过对表 2 分析,2021~2025 年强化学习算法研究形成 16 个有效聚类,这些聚类可划分为四类:一是基础算法类(强化学习、模拟学习、稀疏奖励),聚焦强化学习与深度学习融合、经典算法改进等底层技术创新,为领域发展筑牢根基;二是场景应用类(人工智能、路径规划、多智能体、机械臂、能量管理、资源分配、纳什均衡),覆盖无人机、配电网、机械臂、车联网等多领域,体现技术向实体经济的深度渗透;三是支撑技术类(物联网、边缘计算、知识图谱),通过与边缘计算、数字孪生等技术融合,突破单一技术应用瓶颈;四是安全优化类(协同控制、故障诊断、机器学习),聚焦安全约束与故障应对,保障技术在关键领域的实用性。从平均发表年份看,从 2021 年向 2023 年逐步推进,清晰呈现出从基础算法探索到场景应用深化再到多技术协同创新的演进路径,整体形成“核心算法为基、场景应用为核、支撑技术为翼、安全优化为盾”的发展格局,彰显强化学习领域从技术验证向工业化、精细化应用的转型趋势。

Table 2. Clustering topics and their composition in reinforcement learning algorithm research (2021~2025)
表 2. 2021~2025 年强化学习算法研究聚类专题及构成

聚类编号	聚类名称	聚类大小	轮廓值	聚类高频关键词	平均发表年份
0	强化学习	23	1	强化学习、深度学习、神经网络、深度强化学习、推荐系统	2021
1	模拟学习	21	1	模仿学习、智能决策、Q 学习、行为克隆、抗干扰	2022
2	物联网	21	0.984	物联网、任务分配、舰载机、博弈论、动态规划	2022
3	人工智能	20	1	人工智能、电力市场、不确定性、强化学习、优化调度	2022
4	路径规划	18	0.978	路径规划、无人机、避障、强化学习、人工智能	2022
5	协同控制	18	0.907	协同控制、新能源、课程学习、安全约束、智能博弈	2023
6	多智能体	18	0.956	多智能体、配电网、电动汽车、智能体、电压控制	2022
7	稀疏奖励	18	0.919	稀疏奖励、自主决策、内在奖励、经验回放、智能制造	2022
8	机械臂	17	0.965	机械臂、轨迹规划、动态避障、智能控制、轨迹跟踪	2023
9	能量管理	17	1	能量管理、编队控制、联邦学习、联合优化、任务迁移	2022
10	资源分配	17	0.809	资源分配、车联网、区块链、网络切片、强化学习	2022
11	故障诊断	16	0.95	故障诊断、机器人、对抗攻击、姿态控制、航天器	2022
12	边缘计算	16	1	边缘计算、任务卸载、计算卸载、强化学习、智能交通	2022
13	机器学习	16	0.921	机器学习、网络安全、入侵检测、迁移学习、对抗学习	2022
14	纳什均衡	15	0.944	纳什均衡、策略优化、调度策略、智能电网、运行优化	2023
15	知识图谱	15	0.918	知识图谱、数字孪生、大模型、机动决策、具身智能	2022

3.4. 关键词突现分析

关键词突现可识别短时间内频次骤增的前沿议题。设置 $\gamma = 1$ 、最小持续时长 = 1 年，共得到 20 个突现词，核心特征见图 4。

根据该关键词突现分析图，2021 年作为突现关键词的集中爆发期，“深度学习”以 8.33 的最高强度贯穿全年并延续至 2022 年，成为深度强化学习技术创新的核心引擎，直接推动强化学习与深度学习在模型架构、训练机制上的深度耦合，为复杂环境下的决策任务提供了技术支撑。同期，“功率控制”“指针网络”“电力系统”等关键词聚焦能源领域，体现强化学习在智能电网调度、电力市场优化等场景的早期落地；“故障诊断”“目标检测”则指向工业检测、安防监控等领域的智能决策需求，是强化学习从实验室走向实际场景的关键探索。2022 年突现词“汽车工程”关联自动驾驶中的路径规划、车辆动态控制等核心任务，推动强化学习在交通领域的技术落地；“入侵检测”则聚焦网络安全场景的对抗学习研究，通过强化学习构建动态防御机制，反映领域对安全场景的关注。2023 年“运动控制”“知识迁移”成为新热点，“运动控制”服务于机器人、无人机的高精度轨迹优化与姿态调整，“知识迁移”则致力于提升算法在跨任务、跨环境下的泛化能力。2024~2025 年，“可解释性”针对强化学习黑箱性的行业痛点，是算法从实验室走向工业化应用的关键突破点，通过构建可解释性框架提升可信度；“网络切片”则契合网络的资源动态调度需求，为通信领域的智能决策提供技术支撑，推动强化学习在新型基础设施中的深度渗透。

Top 20 Keywords with the Strongest Citation Bursts



Figure 4. Keyword burst map for reinforcement learning algorithm research
图 4. 强化学习算法研究的关键词突现图

4. 结论与展望

本研究基于 2021~2025 年 CNKI 收录的 5099 篇“强化学习”主题核心期刊论文，通过 CiteSpace 可视化分析，得出核心结论：其一，文献增长呈快速扩张到稳定提质两阶段特征，五年发文量从 587 篇增至 1324 篇，前期聚焦基础算法改进与单一场景验证，后期转向细分场景深化与多技术融合；其二，研究热点形成四层架构知识体系，16 个显著聚类涵盖基础算法、场景应用、支撑技术与安全优化；其三，前沿演进遵循从技术融合到场景深化再到知识驱动路径，从物联网、边缘计算融合，到机械臂、协同控制场景落地，再到大模型、数字孪生深度耦合。

未来研究可聚焦四大方向突破：技术融合层面，深化强化学习与大模型、数字孪生的耦合，借助大模型提升样本效率，通过数字孪生解决高风险场景实训难题；场景应用层面，深耕机械臂精度优化、车联网资源调度等细分领域，以定制化算法满足行业从可用到好用的需求；安全隐私层面，推动联邦强化学习、对抗鲁棒性优化常态化，构建全链路安全体系。综上，强化学习领域已形成完善知识体系，未来通过技术深耕、场景落地与生态共建，将实现从技术创新到产业价值创造的跨越，为人工智能决策层突破提供支撑。

基金项目

保定市科技计划资助，项目名称：基于深度学习技术的 PCB 表面缺陷检测系统研究，项目编号：2411ZG022。

参考文献

- [1] Sutton, R.S. and Barto, A.G. (1998) Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, 9, 1054. <https://doi.org/10.1109/tnn.1998.712192>
- [2] 申铠瑶. 基于强化学习的序列决策在两栖无人车跨域的应用研究[D]: [硕士学位论文]. 北京: 军事科学院, 2025.
- [3] 赵星宇, 丁世飞. 深度强化学习研究综述[J]. 计算机科学, 2018, 45(7): 1-6.

-
- [4] 陈刚, 王志坚, 徐胜超. 基于强化学习的移动边缘计算任务卸载方法[J]. 计算机测量与控制, 2023, 31(10): 306-311.
 - [5] 梁俊斌, 张海涵, 蒋婵, 等. 移动边缘计算中基于深度强化学习的任务卸载研究进展[J]. 计算机科学, 2021, 48(7): 8.
 - [6] 刘亮, 荆腾祥, 段洁, 等. 空天地一体化网络中基于联邦深度强化学习的边缘协作缓存策略[J]. 通信学报, 2025, 46(1): 93-107.
 - [7] 陈超美, 陈悦, 侯剑华, 等. CiteSpaceII: 科学文献中新趋势与新动态的识别与可视化[J]. 情报学报, 2009, 28(3): 401-421.
 - [8] 王鹏, 程思儒. 人工智能高质量数据集的发展趋势及热点——基于 CiteSpace 的知识图谱分析[J]. 技术经济与管理研究, 2025(4): 43-48.
 - [9] 陈伊高, 黄楚新. 人工智能技术与出版深度融合的焦点透视与趋势展望——基于 CiteSpace 的可视化分析[J]. 科技与出版, 2024(4): 112-120.
 - [10] 吴岩, 王光政. 基于 CiteSpace 的配电网韧性评估与提升研究综述与展望[J]. 中国电力, 2023, 56(12): 100-112+137.
 - [11] 毛国君, 顾世民. 改进的 Q-Learning 算法及其在路径规划中的应用[J]. 太原理工大学学报, 2021, 52(1): 91-97.
 - [12] 杨志鹏, 李波, 甘志刚, 等. 基于深度强化学习的四旋翼无人机航线跟随[J]. 指挥与控制学报, 2022, 8(4): 477-482.
 - [13] 邓柏荣, 陈俊斌, 丁巧宜, 等. 融合电网运行场景聚类的多任务深度强化学习优化调度[J]. 电网技术, 2023, 47(3): 978-990.
 - [14] 易佳豪, 王福杰, 胡锦涛, 等. 基于行为克隆的机械臂多智能体深度强化学习轨迹跟踪控制[J]. 计算机应用研究, 2025, 42(4): 1025-1033.
 - [15] 蔡玉, 官铮, 王增文, 等. 基于多智能体深度强化学习的车联网区分业务资源分配算法[J]. 计算机工程与科学, 2024, 46(10): 1757-1764.
 - [16] 钟子山, 唐建航, 金柯兵, 等. 基于深度强化学习的混合数字孪生边缘部署[J/OL]. 计算机工程: 1-12.
<https://doi.org/10.19678/j.issn.1000-3428.0252369>, 2026-01-23.
 - [17] Chen, Y., Chen, C.M., Liu, Z.Y., et al. (2015) The Methodology Function of Cite Space Mapping Knowledge Domains. *Studies in Science of Science*, 33, 242-253.