

DOI:10.19344/j.cnki.issn1671-5276.2025.04.045

基于改进机器学习的档案馆机器人自主避障控制研究

顾昕, 李雪, 刘俊鹏

(吉林蛟河抽水蓄能有限公司, 吉林 吉林 132599)

摘 要:针对目前档案馆机器人的避障路径较复杂、避障消耗时间较长等问题,研究结合机器学习中改进深度 Q 网络算法,搭建档案馆机器人自主避障算法。搭建档案馆机器人的空间运动模型和里程计算模型;对传统的深度 Q 网络算法进行优化,设计三重深度 Q 学习网络档案馆机器人避障算法。研究结果表明:所设计的算法具有较好的奖励曲线,当利用档案馆平面图搭建含有动态障碍点的实际动态障碍环境时,机器人自主避障寻优路径缩短到 2.5 m,避障耗时仅 8 s。

关键词:机器人;自主避障;路径规划;机器学习;档案馆

中图分类号:TP242.3 **文献标志码:**B **文章编号:**1671-5276(2025)04-0248-06

Research on Autonomous Obstacle Avoidance Control of Archive Robots Based on Improved Machine Learning

GU Xin, LI Xue, LIU Junpeng

(Jilin Jiaohe Pumped Storage Co., Ltd., Jilin 132599, China)

Abstract:This research by combining the improved deep Q-network algorithm in machine learning, builds an autonomous obstacle avoidance algorithm for archive robot aiming at the complicated obstacle avoidance path and long consumed time of current archive robot. The spatial motion model and mileage calculation model of the archive robot are constructed, the traditional deep Q-network algorithm is optimized and the triple deep Q-learning network archive robot obstacle avoidance algorithm is designed. The results show that the designed algorithm has a good reward curve, and its autonomy obstacle avoidance seeking the optimal path is shortened to 2.5 m, and the obstacle avoidance time consumed only 8 s upon in a real dynamic obstacle environment containing dynamic obstacle points with an archive floor plan.

Keywords:robot; autonomous obstacle avoidance; path planning; machine learning; archives

0 引言

随着人工智能技术的不断进步,机器人逐渐被广泛应用于各个领域以提高自动化水平^[1-2]。随着档案馆数字化转型的加速,档案馆机器人的应用越来越成为提升工作效率和准确性的关键。在此过程中,档案馆机器人的自主避障能力显得尤其重要。档案馆机器人的自主避障能力强弱关系到机器人是否能在复杂的档案环境中安全、高效地工作^[3]。在实际操作过程中,档案馆机器人如何高效、安全地完成资料的存取和搬运任务,避免在狭窄或复杂的档案库房环境中发生碰撞,是当前亟待解决的问题^[4-5]。近年来,关于档案馆机器人避障问题的研究已有若干进展,许多研究集中在传统的传感器融合技术和控制策略上,但对于如何结合深度学习技术优化机器人的避障性能,尤其是在动态环境中的适应性和决策效率,仍旧存在较大的研究空间^[6-7]。现有的避障方法,

如传统的深度 Q 网络算法,虽然在某些情况下表现良好,但在处理动态环境和实时适应性方面仍显示出局限性。针对这些挑战,本研究提出了一种基于改进深度 Q 网络的档案馆机器人自主避障控制算法,旨在通过该算法完成档案馆机器人在复杂工作环境中的避障任务,从而提高档案馆机器人的实际操作能力和工作效率。

1 基于改进 DQN 的档案馆机器人自主避障控制研究

1.1 面向档案馆机器人避障路径规划的机器学习框架构建

档案馆机器人是一种专为档案馆环境设计的自动化机器人,主要功能是帮助管理和操作档案资料。这类机器人通常能够自主导航,并执行诸如搬运档案、整理文件、检索资料等任务^[8]。传统的移动机器人往往采用单一的传感器进行数据的采集和传输。这类数据采集和传输方法不仅效率

第一作者简介:顾昕(1978—),女,山东肥城人,副研究馆员,本科,研究方向为档案管理,1666181378@qq.com。

较低,并且容易忽略全局信息。为了更好地采集到档案馆内部的环境信息,使档案馆机器人能够根据指令顺利地完成任务,本研究首先改变了传感器数据的收集方式,分别搭建了档案馆机器人的空间运动模型和里程计算模型,并利用激光传感器和里程传感器同时收集位姿信息与里程信息。假设档案馆机器人自身坐标为 O_{xy} , 得到其在坐标系中的位姿计算如式(1)所示^[9-10]。

$$M_{O_{xy}} = [x_{t+1}, y_{t+1}, \theta_{t+1}] \quad (1)$$

式中: t 为时刻; θ_{t+1} 为档案馆机器人在 $t+1$ 时刻下的方向角,方向角定义为以 x 轴正半轴为正方向,机器人按照逆时针旋转所成夹角。 x_{t+1} 为 $t+1$ 时刻下档案馆机器人在 x 轴方向上的运动距离。 y_{t+1} 为 $t+1$ 时刻下档案馆机器人在 y 轴方向上的运动距离。机器人的线速度计算公式为

$$V = \frac{V_L + V_R}{2} \quad (2)$$

式中: V_L 和 V_R 分别为机器人左轮和右轮的轮速; V 为机器人线速度。左右轮的轮速计算公式为

$$\begin{cases} V_L = \omega d + V \\ V_R = V - \omega d \end{cases} \quad (3)$$

式中: ω 为角速度; d 为左右轮之间的距离。

轮式里程传感器是计算机器人移动位置的关键感应设备,该设备能够提供必要的即时位姿反馈信息,从而支持机器人完成避障路径规划任务。对于差速驱动的机器人来说,通常采用编码器计量的方式来计算其运动距离。编码器一般安装在各个动力轮的电机上,用于测量在一段时间内车轮的累积转角变化情况。利用里程传感器完成档案馆机器人移动距离的计算,所搭建的里程计算数学模型为

$$f(t) = \begin{bmatrix} \theta + \Delta\theta \\ x_{t-1} + r_k \sin(\Delta\theta) \end{bmatrix} \quad (4)$$

式中: $f(t)$ 为里程计算数学模型; r_k 为档案馆机器人的运动范围; x_{t-1} 为 $t-1$ 时刻下档案馆机器人在 x 轴方向上的运动距离。

环境条件为档案馆内部导航的机器人提供了路径设定的基本条件^[11]。假设将档案馆里错综复杂的环境作为机器人避障路径规划的背景,首先结合各类设备和软件搭建了机器人避障路径学习框架,然后再对避障算法进行优化,最终实现档案馆机器人自主避障控制研究。档案馆机器人避障路径规划学习框架结构如图1所示。

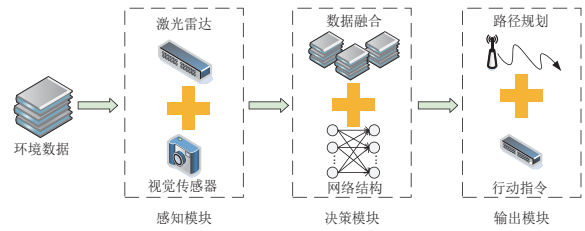


图1 档案馆机器人避障路径规划框架结构图

图1所示的框架可以分为3个主要模块,分别为感知模块、决策模块、输出模块^[12]。感知模块主要包括了收集周边环境数据的激光雷达和一个收集环境图像信息的视觉传感器。其中,激光雷达主要由激光传感器和里程计数器组成,激光传感器主要是收集机器人的自身定位信息和姿势信息,里程计数器主要是收集机器人与目标之间的距离信息。深度强化学习算法是决策模块的主要组成部分。在决策模块中,档案馆机器人通过不断尝试和收集错误来优化其决策过程,以实现避障和路径寻优的目的。当完成前面两个模块的任务后,需要将信息通过输出模块来输出。输出模块主要包括路径的规划和机器人的行动指令。箭头指向的曲线表示机器人经过学习后生成的路径,这条路径能够有效地绕开障碍物,引导机器人从起点移动到目标终点。

1.2 基于改进 DQN 的档案馆机器人自主避障路径规划算法设计

搭建好档案馆机器人避障路径规划框架后,需要考虑对其自主避障路径规划算法进行优化。传统的深度 Q 学习网络算法 (deep q-learning network, DQN) 可以使智能体在复杂环境中学习如何做出决策,从而在最大程度上优化未来的累计奖励,但也仍然存在训练样本多、学习效率低、训练灵活度差等问题^[13-14]。因此,在 DQN 的基础上,融合经验回放池、双重深度 Q 网络 (double deep q-network, double DQN) 和对决深度 Q 网络 (dueling deep q-network, dueling DQN), 设计了三重深度 Q 学习网络模型 (triple deep q-learning network, TDQN)。

Double DQN 网络相较于传统的 DQN 网络,在计算目标 Q 值时采用了不同的方法,以有效避免算法的过度估计问题。这种过估计现象主要源于移动机器人的 Q 网络参数需要在两个方向上更新,而网络对这些 Q 值结果过分依赖。为了解决这一问题,Double DQN 网络引入了一个结构与当前 Q 网络相同的目标 Q 网络来有效降低过度

估计的风险。在训练移动机器人的过程中,目标Q网络定期从当前网络获取参数更新。这种迭代方法使得当前网络训练的目标Q值由更新后的目标网络生成。此外,Double DQN网络通过将目标Q值的动作和值本身分开处理,有效地避免了单一网络可能导致的过度估计问题。

Dueling DQN改进了DQN的网络结构,将其Q网络拆分为两个部分,创建了一种竞争对抗的网络架构。其中一部分网络仅关注移动机器人的当前状态,不涉及动作选择,这部分被称为价值函数。另一部分既考虑状态又考虑动作选择,称为优势函数。这两部分的结果相结合,形成最终的价值函数。价值函数的表达式为

$$Q(S,a;\theta,\alpha,\beta)=V(S;\theta,\beta)+A(S,a;\theta,\alpha) \quad (5)$$

式中: S 为移动机器人的当前状态; a 为移动机器人的当前动作; θ 为移动机器人的公共网络参数; β 为移动机器人价值函数中的网络参数; α 为移动机器人优势函数中的网络参数; V 和 A 分别为价值函数和优势函数。Dueling DQN与DQN的不同之处在于前者将一个全连接层分解成 V 和 A ,然后将这两个分解出的函数重新组合,形成一个新的Q值输出。这种方法通过分解和重组增强了网络的功能。

TDQN算法通过引入DQN、Dueling DQN、Double DQN这3个并行的网络结构,其分别负责策略评估,值函数更新和目标策略优化,有效减少了过拟合的风险并提高了决策的稳定性。此外,算法中还引入了先进的探索机制和优化的奖励机制,使得机器人能够更快地适应环境变化,提高了在复杂动态环境中的避障能力。相较于传统的DQN算法来说,改进后的TDQN算法不仅能够增强模型的泛化能力,也能显著提升机器人在实际应用中的性能。TDQN结构如图2所示。

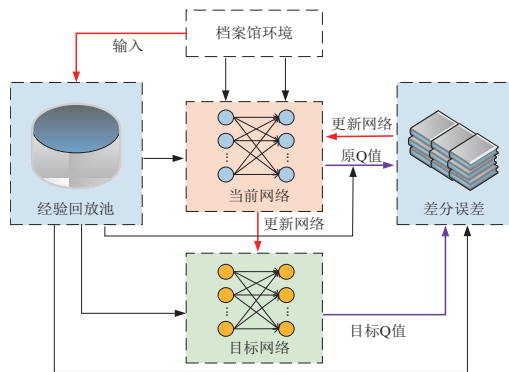


图2 TDQN结构图

在图2中,首先,将档案馆机器人与周围环境进行交互,然后在经验回放池中加入测得的时间差分误差并进行样本的优先排序。接着,利用经验采样概率的分布情况来选取样本值,并将该样本值作为TDQN算法的输入进行执行算法运算。通过不断地训练,能够使得当前深度网络中的Q值不断接近目标深度网络中的Q值,最终完成网络的训练。

在面向移动机器人避障路径规划的TDQN网络中,收集的样本被存储在经验回放池里。此外,该网络假设每种状态在回放时被采集的概率都是一样的。在经验回放池中,采样概率的计算公式为^[15]

$$P(i)=\frac{p_i^\alpha}{\sum_k p_k^\alpha} \quad (6)$$

式中: $P(i)$ 为采样概率; p_i^α 为第 i 个样本下,利用优先程度为 α 时,样本的利用优先级。优先程度的取值范围在 $0\sim 1$ 之间。若样本的优先程度为 0 ,则需要对样本进行随机采样退化操作。此次研究将时间差分误差作为样本来表示优先级,得到优先级的比例优化公式为

$$p_i=|\delta_i|+\varepsilon_0 \quad (7)$$

式中: p_i 为比例优化下的优先级; δ_i 为时间差分误差; ε_0 为常数项。为了进一步降低样本分布不均所造成的误差值,引入采样权重进行优化^[16]。得到采样权重的计算公式为

$$w_i=\left(\frac{1}{N}\cdot\frac{1}{P(i)}\right)^\gamma \quad (8)$$

式中: N 为经验回放池里的样本总量; γ 为归一化调节常数; w_i 为采样权重。将 w_i 进行归一化处理,其处理公式为

$$w'_i=\frac{1}{\max(w_1,w_2,\cdots,w_N)}\cdot w_i \quad (9)$$

式中: w'_i 为归一化处理后的 w_i ; $\max(w_1,w_2,\cdots,w_N)$ 为所有权重中的最大值。经验回放机制的融入能够显著提升TDQN算法在机器人路径规划中的学习效率,通过打破样本间时间相关性,降低了训练过程的方差,从而提高了策略的稳定性^[17]。此外,该策略还有效地缓解了样本遗忘问题,保障了数据多样性,从而使得策略对于动态环境具有更强的适应性和鲁棒性。融合经验回放池的TDQN算法运行流程如图3所示。

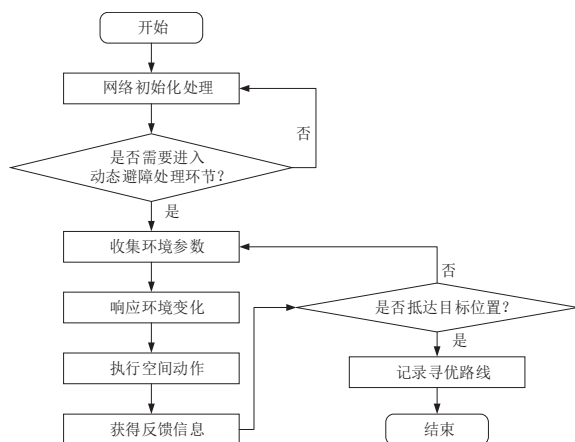


图3 TDQN 运行流程图

在图3中,需要先获取档案馆环境与机器人状态信息,随后进入初步路径规划阶段,以形成一个原始避障路径。该路径规划结果随即受到评估,以确定是否需要进入动态避障处理环节以响应潜在的复杂场景变化。在动态避障处理中,路径规划会被实时调整以适配环境中的变化,确保路径的可行性和安全性。这一处理过程完成后,系统将产生最终决策结果,并进行反馈以供执行。最终,当所有任务执行完毕后,系统会显示机器人已经抵达最终的目标位置,从而结束算法的运行。

2 档案馆机器人自主避障控制算法性能测试与避障效果分析

为了证明 TDQN 避障算法的有效性,本研究使用 Gazebo 仿真软件创建了一个详细的档案馆环境模拟。这个环境模拟了一个典型的档案馆布局,包括书架、走廊和其他常见障碍物。数据集记录了机器人在模拟环境中的位置、速度、方向以及

激光雷达和摄像头传感器的数据。为了生成 Gazebo 仿真数据集,研究控制机器人在仿真环境中按预定路径行驶,同时收集其传感器数据。这个仿真环境旨在模拟档案馆中的实际条件,包括复杂的走廊布局 and 不同类型的障碍物,以测试和比较不同避障算法的性能。档案馆机器人的网络参数设置如表1所示。

表1 参数设置表

机器人参数	数值
网络学习率	0.000 1
奖励系数	1.00
经验池大小	100 000
折扣因子衰减率	0.90
折扣因子初始值	1.00
折扣因子最终值	0.01

在仿真实验中,一共设置了两种不同的障碍物环境,分别为静态障碍物环境和动态障碍物环境。选择 DQN、动态窗口算法(dynamic window approach, DWA)、遗传算法优化下的粒子群算法(genetic algorithms - particle swarm optimization, GA-PSO)及 TOON 算法作为对比算法,首先得到了4种算法在两种避障环境下的奖励曲线如图4所示,图4(a)和图4(b)分别为静态障碍物环境和动态障碍物环境下各算法的奖励曲线,横坐标表示训练过程中机器人与环境交互次数,纵坐标表示对应交互次数下的奖励值(无单位)奖励曲线越快收敛且最终值越高表示算法避障学习效果越好。

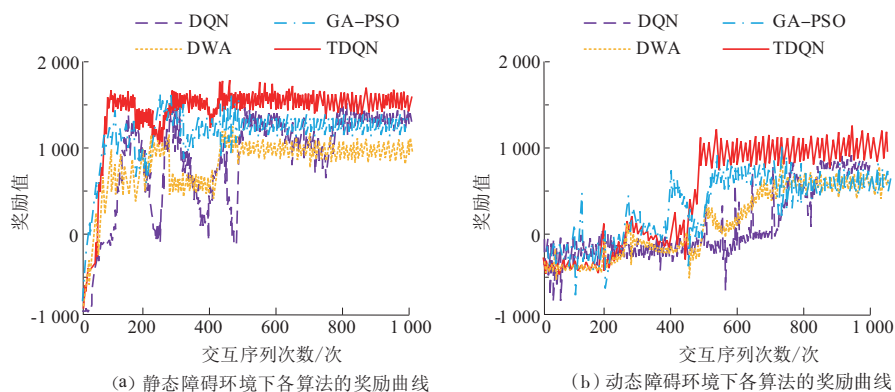


图4 不同避障环境下各算法的奖励曲线

由图 4(a)可知,随着交互序列次数的增加,4 种算法的奖励曲线都有不同程度的波动。其中,当交互序列次数分别为 785 次、512 次、486 次、410 次时,DQN、DWA、GA-PSO、TDQN 4 种避障算法能够各自达到稳定状态。由图 4(b)可知,随着交互序列次数的增加,4 种算法在动态障碍环境下的奖励曲线均会发生不同程度的波动。当交互序列次数分别为 926 次、738 次、698 次、455 次时,DQN、DWA、GA-PSO、TDQN 4 种避障算法均能够达到稳定状态。4 种算法在两种避障环境下的基准性能测试结果如表 2 所示。

表 2 4 种算法在两种避障环境下的基准性能测试结果

算法类型	静态		动态	
	准确率	响应时间/s	准确率	响应时间/s
DQN	0.83	1.05	0.78	1.25
DWA	0.87	0.59	0.82	0.71
GA-PSO	0.91	0.34	0.86	0.48
TDQN	0.98	0.03	0.96	0.05

从表 2 可得,TDQN 在两种避障环境下均能拥有最高的避障准确率和最快的避障响应时间。

4 种算法在静态避障环境下的寻优路径如图 5 所示。

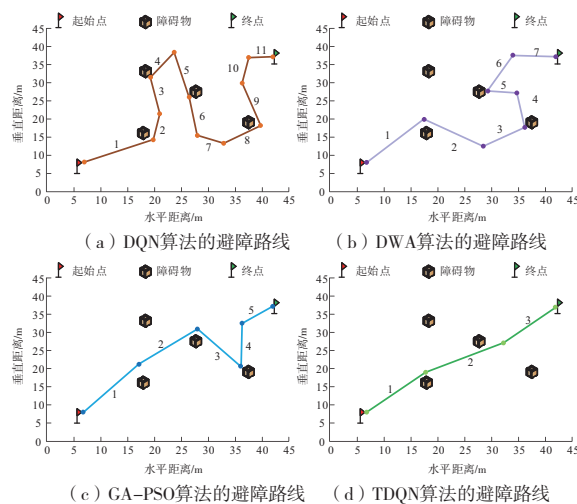


图 5 静态避障环境下各算法的寻优路线

由图 5(a)可知,DQN 避障算法需要经过 11 段路程才能够成功避开障碍物,抵达终点。由图 5(b)可知,DWA 避障算法需要经过 7 段路程来避开障碍物。由图 5(c)可知,GA-PSO 避障算法只需经过 5 段路程即可成功避开障碍物,抵达终点。由图 5(d)可知,TDQN 避障算法只需经过

3 段路程即可成功避开障碍物,抵达终点。综合来看,TDQN 避障算法下的路径寻优效果最好,其所行路径最短,所用时间最少。

为了进一步探索 GA-PSO 优化算法和本研究提出的 TDQN 优化算法在实际动态避障问题中的应用效果,利用某档案馆的平面图搭建了图 6 所示的动态避障环境,并对比了两种优化算法在实际避障问题中的路径长度和耗时情况。

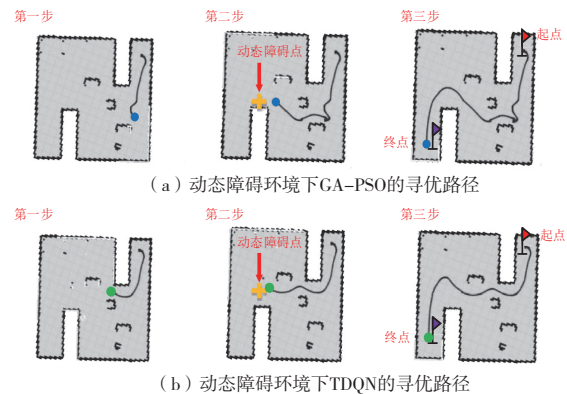


图 6 不同避障算法在实际障碍环境中的避障路径

结合图 6(a)和图 6(b)可知,相较于 GA-PSO,TDQN 在加入动态障碍点后仍能够很好地避开障碍物,从而为档案馆机器人规划出一条最短的避障路径。最终,GA-PSO 和 TDQN 的避障路径长度分别为 3.6 m 和 2.5 m,避障耗时分别为 25 s 和 8 s。

3 结语

为了优化档案馆机器人的避障效果,提高其工作效率,本研究对传统的深度 Q 学习网络算法进行了优化,并提出了 TDQN 自主避障算法。研究表明:TDQN 自主避障算法在档案馆机器人的避障任务中表现卓越。在静态避障环境中,DQN、DWA、GA-PSO、TDQN 4 种避障算法分别需要交互 785 次、512 次、486 次、410 次,才能够达到稳定的奖励曲线;在动态避障环境中,DQN、DWA、GA-PSO、TDQN 4 种避障算法分别需要交互 926 次、738 次、698 次、455 次,才能够达到稳定的奖励曲线。在静态环境中,TDQN 算法的准确率高达 0.98,响应时间低至 0.03 s,规划的路径段数最少,仅 3 段,显著少于其他算法。DQN、DWA、GA-PSO 则分别需要规划 11 段、7 段、5 段路径才能够完成避障任务。在动态避障环境的实际应用中,TDQN 算法的准确率高达 0.96,响应时间低至 0.05 s,所规划的路径更短,用时更少,仅需

2.5 m路径长度和8 s避障耗时即可完成避障任务,远远优于GA-PSO算法所需3.6 m路径长度和25 s避障耗时。综上,此次研究不仅证明了TDQN避障算法的高效性和实用性,也为档案馆机器人避障技术的进步提供了具有重要价值的结论。尽管本研究在档案馆机器人的自主避障控制领域取得了一定的进展,但也存在一些局限性和不足。例如算法是否同样适用于其他类型的避障任务和不同的机器人平台,后续还将引入更多不同类型的避障任务进行测试。

参考文献:

- [1] 王正家,吴顺尧,何涛,等. 自由空间中的机器人自主避障方法研究[J]. 传感器与微系统,2023,42(6): 54-57.
- [2] 郑渤龙,明岭峰,胡琦,等. 基于深度强化学习的网约车动态路径规划[J]. 计算机研究与发展,2022,59(2):329-341.
- [3] 吕刚,吕金壮,邓军,等. 部分环境信息未知下的机器人动态路径规划算法[J]. 自动化与仪器仪表,2022(3):186-190.
- [4] 廖列法,李浩瀚,李帅,等. 结合 Winner-Take-All 的足球机器人控制策略研究[J]. 计算机工程与应用,2021,57(7):136-143.
- [5] 李文彪. 基于深度强化学习的工业机器人避障路径规划方法[J]. 制造业自动化,2022,44(1):127-130.
- [6] SONG D L, GAN W H, YAO P, et al. Surface path tracking method of autonomous surface underwater vehicle based on deep reinforcement learning[J]. Neural Computing and Applications,2023,35(8):6225-6245.
- [7] 丁艳,毕杨. 基于蚁群算法的多移动机器人避障路径规划方法[J]. 自动化与仪表,2021,36(1):36-40.
- [8] 董永峰,杨琛,董瑶,等. 基于改进的DQN机器人路径规划[J]. 计算机工程与设计,2021,42(2):552-558.
- [9] MOHD ROMLAY M R, MOHD IBRAHIM A, TOHA S F, et al. Obstacle avoidance for a robotic navigation aid using fuzzy logic controller-optimal reciprocal collision avoidance [J]. Neural Computing and Applications, 2023,35(30):22405-22429.
- [10] 林俊强,王红军,邹湘军,等. 基于DPPO的移动采摘机器人避障路径规划及仿真[J]. 系统仿真学报,2023,35(8):1692-1704.
- [11] 张博,周红生. 基于RL-APF算法的巡检机器人避障研究[J]. 组合机床与自动化加工技术,2023(6):7-10.
- [12] 付雷,秦一杰,何顶新,等. 基于改进人工势场法的多机器人编队避障[J]. 控制工程,2022,29(3):388-396.
- [13] 李二超,王玉华. 改进人工势场法的移动机器人避障轨迹研究[J]. 计算机工程与应用,2022,58(6):296-304.
- [14] 申晓平. 智能移动机器人避障路线自动规划系统设计[J]. 机械制造与自动化,2023,52(5):28-31.
- [15] 陈梦清,陈洋,陈志环,等. 路网约束下异构机器人系统路径规划方法[J]. 自动化学报,2023,49(4):718-730.
- [16] ASHWIN S H, NAVEEN RAJ R. Deep reinforcement learning for autonomous vehicles: lane keep and overtaking scenarios with collision avoidance [J]. International Journal of Information Technology, 2023,15(7):3541-3553.
- [17] BASHNICK C, ULRICH S. Fast model predictive control for spacecraft rendezvous and docking with obstacle avoidance [J]. Journal of Guidance, Control, and Dynamics,2023,46(5):998-1007.

收稿日期:2024-01-02