文章编号：1003－1251(2024)04－0032－06

基于 DQN 算法的农用无人车作业路径规划

庄金炜[1]，张晓菲[2]，尹琪东[1]，陈　克[1]

(1. 沈阳理工大学 汽车与交通学院，沈阳 110159；2. 中国人民解放军第六四零九工厂研究院，辽宁 抚顺 113000)

摘　要：传统农用无人车作业时常依据人工经验确定作业路线，面对复杂的作业环境时无法保证路径规划的高效性，且传统覆盖路径规划方法聚焦于覆盖率而忽略了车辆作业路线上的损耗。为此，提出一种以减少车辆在路线上的损耗为目标的最优全局覆盖路径规划方法。以深度 Q 网络(DQN)算法为基础，根据作业时车辆的真实轨迹创建奖励策略(RLP)，对车辆在路线上的损耗进行优化，减少车辆的转弯数、掉头数及重复作业面积，设计了 RLP-DQN 算法。仿真实验结果表明，对比遗传算法、A* 算法等传统路径规划方法，本文 RLP-DQN 算法综合性能较好，可在实现全覆盖路径规划的同时有效减少路线损耗。

关　键　词：农用无人车；路径规划；深度强化学习；DQN 算法

中图分类号：S232.7；TP242　**文献标志码**：A　**DOI**：10.3969/j.issn.1003－1251.2024.04.006

Route Planning of Agricultural Unmanned Vehicle Based on DQN Algorithm

ZHUANG Jinwei[1]，ZHANG Xiaofei[2]，YIN Qidong[1]，CHEN Ke[1]

(1. Shenyang Ligong University，Shenyang 110159，China；2. Chinese People's Liberation Army 6409 Factory Research Institute，Fushun 113000，China)

Abstract：In the traditional agricultural unmanned vehicle operation，the route is often determined according to manual experience，which cannot guarantee the efficiency of route planning in the face of complex operating environment. Moreover，the traditional coverage path planning method focuses on the coverage rate，but ignores the loss on the vehicle operation route. Therefore，an optimal global coverage route planning method aiming at reducing the loss of vehicles on the route is proposed. Based on the deep Q network(DQN) algorithm，the reward strategy(RLP) is created according to the real track of the vehicle during operation，and the loss of the vehicle on the route is optimized to reduce the number of turns，U-turns and repeated operation area of the vehicle. Simulation results show that compared with traditional path planning methods such as genetic algorithm and A* algorithm，the RLP-DQN algorithm in this paper has better comprehensive performance and can effectively reduce route loss while realizing full coverage path planning.

Key words：unmanned agricultural vehicles；path planning；deep reinforcement learning；DQN algorithm

随着智慧农业的快速发展，对农业生产效率的要求也逐步提高。农用无人车是我国农业生产的主要工具，车辆的全局覆盖路径规划对其自动化作业效率具有重要的影响，同时也影响全局农

收稿日期：2023－11－03

基金项目：辽宁省教育厅高等学校基本科研项目(JYTQN2023060)

作者简介：庄金炜(1999—)，男，硕士研究生；陈克(1965—)，通信作者，男，教授，博士，研究方向为汽车系统动力学与控制。

业生产任务的完成效率[1]。

传统农用无人车根据驾驶员经验选择作业路线,在复杂环境下难以确定最优路线,且传统路径规划方法通常采用往复式、牛耕式、内/外螺旋等方式,将整个工作区域分成多个小块,再逐个遍历[2]。这种方法是解决 NP-hard 问题的传统方式,应对规则边界且内部无障碍的作业环境可达到较好的效果,但忽略了车辆在作业过程中的非工作行为(如掉头、倒车)以及过渡阶段的路线损耗[3]。为提高作业效率、减少损耗,许多学者提出了新的路径规划算法。

陈凯等[4]考虑多种约束条件,提出了一种基于模拟退火法的混合规则路径规划算法,解决了传统模拟退火算法易陷入局部最优的问题,规划效果较好。谢金燕等[5]为提高果园割草机的工作效率、降低作业成本,提出了一种改进粒子群优化算法,与传统粒子群算法相比,作业路线长度减少了 7% 以上,但收敛速度有所下降。Wu 等[6]提出了一种基于遗传算法的覆盖路径规划算法,将传统遗传算法的染色体和单点突变扩展为染色体对和多点突变,与传统遗传算法相比,规划路径的重复作业面积减少了 38.54%,掉头次数减少了 35.00%。

在实际生产中,作业环境的形状、环境内部的障碍物位置及作业范围的选择等都影响作业效率。为提高农用无人车的作业效率,本文基于强化学习的“试错”思想,量化车辆在路线上的损耗并依据损耗规划最优路径。首先,通过作业区域的俯视图建立栅格地图[7],根据地图信息将区域划分为待工作区、障碍区等;然后,基于深度强化学习的深度 Q 网络(deep Q network,DQN)算法框架,设计真实作业轨迹损耗的奖励策略,对车辆行走路线进行优化,减少掉头、转弯等动作,缩短整体行走距离和时间;最后,在实现全覆盖的基础上,进一步规划出损耗最小的工作路线,以实现更高效的农业生产。

1　深度强化学习 DQN 算法基础

强化学习是一种从环境状态对应到动作行为的学习过程,可概括理解为智能体与环境不断交互获得最大累计回报值的过程[8]。本文根据农用无人车的真实作业轨迹损耗设计全新奖励策略,结合强化学习中的深度强化学习 DQN 算法框架规划最优的作业路径。DQN 算法将深度学习思想和传统强化学习思想相结合,收敛速度较快、收敛效果较好,可更好地处理顺序决策类问题。

1.1　马尔可夫决策

马尔可夫决策过程(MDP)是强化学习中用于环境和智能体交互过程建模的一种数学框架[9],是强化学习问题的数学形式化描述。该过程由状态、动作、奖励、转移概率四个部分组成。状态表示系统在某一时刻的条件或属性,用于描述决策问题的各种可能情况;动作表示智能体可以采取的行动或策略,智能体根据当前的状态选择一个动作来影响系统的演变;奖励表示智能体根据当前状态和采取的动作所获得的即时回馈或衡量指标;转移概率表示在给定状态和采取的动作下,系统转移到下一个状态的概率。

将强化学习建模为马尔可夫决策过程,可以学习到最优策略,使得智能体在不同状态下选择最佳动作,以使长期累积奖励最大化。

1.2　Q-learning 与 DQN

Q-learning 的核心思想是利用经验数据来更新状态 - 动作值函数(Q 函数)[10]。Q 值大小与状态、执行动作后的即时奖励以及下一个状态的最大 Q 值有关,计算式为

$$Q(s_t,a_t)=r+\gamma\max Q(s_{t+1},a_{t+1}) \tag{1}$$

式中:$Q(s_t,a_t)$ 是在当前时刻(t)状态 s_t 下采取动作 a_t 时的 Q 值;r 是奖励值;γ 是奖励折扣系数;$Q(s_{t+1},a_{t+1})$ 表示在下一时刻($t+1$)状态 s_{t+1} 下采取动作 a_{t+1} 时的 Q 值。

使用下一时刻状态的最大 Q 值更新 Q 值表,表达式为

$$Q_{\text{new}}(s_t,a_t)=Q(s_t,a_t)+\alpha[r+\gamma\max Q(s_{t+1},a_{t+1})-Q(s_t,a_t)] \tag{2}$$

式中:$Q_{\text{new}}(s_t,a_t)$ 表示更新后的 Q 值;α 是算法的学习率。

通过不断与环境交互、观察结果、更新 Q 值,智能体逐渐学习到最优策略。Q-learning 具有较好的收敛性及泛化性,但当状态空间过大时,所需的存储和计算量会急剧增加,在处理连续动作空间和连续时间问题时存在困难。

DQN 算法利用神经网络代替传统 Q-learning 中的表格记录形式,两者之间的关系可表示为

$$Q(s,a)=f(s,a,w) \tag{3}$$

式中:f 表示以神经网络形式生成的非线性函数,用于近似代替 Q 函数;s 表示某时刻的状态;a 表示某时刻的动作;w 为 f 函数的参数。

DQN 算法比 Q-learning 算法增加了经验回放

机制和双神经网络训练机制,以提升算法效果。神经网络的作用是逼近 Q 值函数,训练过程旨在使神经网络的计算结果逐步逼近最优的 Q 值函数,从而得到最优策略。经验回放机制用于解决训练过程中样本的相关性和不稳定性问题[11-12]。智能体与环境交互产生的经验样本将被存放进经验缓冲区,在训练过程中不断从缓冲区随机抽取一批样本进行训练。DQN 算法引入目标价值神经网络,目标网络的 Q 值在一段时间内保持不变,降低了当前价值网络 Q 值与目标网络 Q 值的相关性。当前价值网络用于选择动作;目标网络用于计算目标值。经验回放机制有助于延迟更新目标网络,提高学习效果。

本文以 DQN 算法为基本框架,融合基于真实行为损耗设计的奖励策略设计路径规划算法。

2 基于真实行为损耗的奖励策略设计

在农用无人车实际作业过程中,存在前进、后退、转弯、掉头等行为,传统的路径规划方法忽略以上行为产生的损耗,而此类损耗是影响作业效率的重要因素。本文通过对农用无人车作业时的行为次序进行解析分类,设计基于真实行为损耗的奖励策略,以减少路线上非工作行为带来的损耗。

2.1 区域占据奖励策略

以栅格地图中单个栅格属性判断区域状态,若该栅格区域完成作业,智能体获得占据奖励,奖励值加 1。为使无人车优先完成全局覆盖任务,待作业区域数量为零时即视为完成路线覆盖任务,当无人车完成最后一块区域作业时将得到路线完成奖励,该奖励值远大于作业完成前的奖励值总和,以驱使智能体优先完成全局覆盖任务,避免其陷入局部最优。

2.2 行为奖励策略

以栅格地图为基础环境的路线规划中,智能体常见的运动方向有上、下、左、右、左上、右上、左下和右下八个方向[13],也会出现路线斜穿两个障碍物边界到达下一栅格的情况,如图 1 所示。本文研究对象为阿克曼转向的传统农用车,在栅格中的运动方向只有上、下、左、右四种,设置地图时尽量规避位于死区的待作业区域。

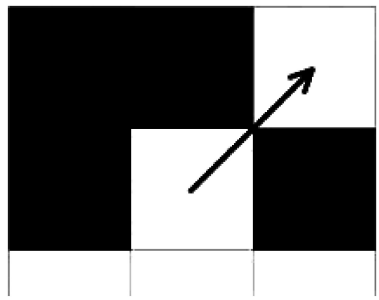

图 1 穿越障碍示意图

Fig. 1 Schematic diagram of crossing obstacles

农用车多数情况下携带拖拽式农具,无法完成差速原地转向,基于该情况,分析无人车在作业中可能出现的几种运动轨迹,确定行为奖励策略。在栅格地图中将车辆的上下移动设置成竖向运动行为类、左右移动设置成横向移动行为类,具体情况如图 2 所示。

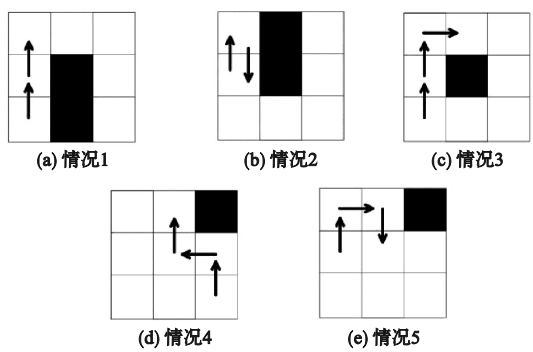

图 2 行为判定种类

Fig. 2 Types of behavior decision

图 2(a)和图 2(b)表示当前动作与上一动作属于同一行为类的两种情况。当前动作与上一动作完全相同时,当前动作判断为直行,奖励值减 1(情况 1);当前动作与上一动作不同,当前动作判断为倒车,奖励值减 8(情况 2)。

当前动作与上一动作不属于同一行为类时,分为以下两种情况。当前动作与第一个动作不属于同一行为类,当前动作判断为转弯,奖励值减 5,如图 2(c)所示。当前动作与第一个动作属于同一行为类时:如果当前动作与第一个动作相同,当前动作判断为转弯,奖励值减 5,如图 2(d)所示;如果不同,判断为掉头动作,奖励值减 10,如图 2(e)所示。

3 RLP-DQN 算法训练过程

本文将基于真实行为损耗的奖励策略(real

loss policy,RLP)与深度强化学习 DQN 算法相结合,提出 RLP-DQN 路径规划算法,其训练流程如图 3 所示。

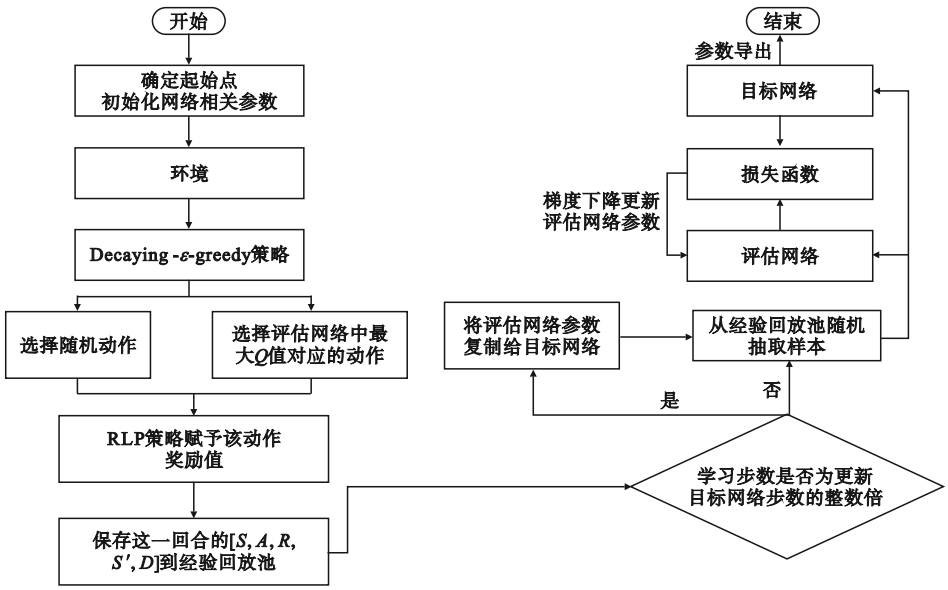

图 3　RLP-DQN 训练流程图

Fig. 3　Flowchart of RLP-DQN training

为降低训练数据间的相互关联性,训练过程中采用经验回放机制和随机抽样策略。对于智能体的动作选择,本文选用 Decaying-ε-greedy 策略,该策略与 DQN 算法的 ε-greedy 策略[14]区别在于,随着训练时间延长,Decaying-ε-greedy 策略算法中的 ε 值减小,可避免陷入局部最优。智能体将以 ε 的概率选择随机动作,以 $1-\varepsilon$ 的概率选择当前网络中估值最高的动作,表达式为

$$a_{t+1}=\begin{cases}\text{random}, & 0<p<\varepsilon\\ \text{argmax}(Q(a,s)), & \varepsilon\leqslant p<1\end{cases} \tag{4}$$

式中 p 为 0 ~ 1 的随机数,表示选择动作的概率。

首先进行参数初始化,包括学习率、最大迭代次数、记忆回放池容量以及奖励的折扣率等。

然后在设定的训练回合内,初始化评估神经网络及目标网络,定义两个神经网络的结构,包括输入层、隐藏层和输出层的节点数,两者具有相同的初始化权重参数。在经验回放池充满前,智能体在环境中完全依据 Decaying-ε-greedy 策略给出的动作进行移动,根据 RLP 奖励策略赋予该动作奖励值,同时这组动作会以[S,A,R,S',D]的形式存入经验回放池,其中 S 代表环境、A 代表动作、R 代表奖励、S'代表下一个环境、D 表示任务是否完成。经验回放池充满后,从中随机抽取一批经验样本,使用目标网络计算每个样本的目标 Q 值,进一步计算神经网络输出 Q 值和目标 Q 值之间的均方误差损失,更新神经网络的权重。

最后,当 Q 值收敛或达到终止条件时,从中选取最佳的策略,智能体将按照最新的策略进行移动。

4　仿真实验结果与分析

采用仿真实验验证 RLP-DQN 算法的效果。仿真实验环境为:Nvidia GTX 3060 显卡;Ubuntu 18.04,Python 3.6,Pytorch 1.1。

采用 Pytorch 框架搭建农用无人车覆盖仿真平台,作业环境构建效果如图 4 所示。地图尺寸为 15 × 18,黑色栅格表示作业环境中障碍物的位置,白色栅格表示待作业栅格。无人车每覆盖一个白色栅格,该栅格变为灰色,表示其已经被覆盖过,重复经过的栅格用深灰色表示。本次训练超参数设置如表 1 所示。

表 1　RLP-DQN 超参数设置

Table 1　RLP-DQN hyperparameters setting

参数	数值
折扣因子	0.9
学习率	0.005
抽取样本数量	128
经验回放池容量	1 000 000
训练回合数	80 000

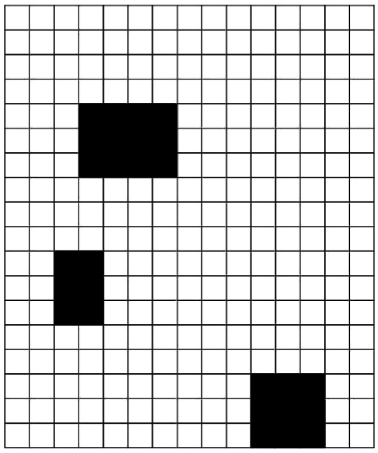

图 4 仿真栅格环境

Fig. 4 Simulated raster environment

采用 A* 算法[15]、遗传算法[16]、文献[17]中提出的聚类算法作为对比算法,本文 RLP-DQN 算法与对比算法在同一环境中规划的覆盖路径如图 5 ~ 8 所示。

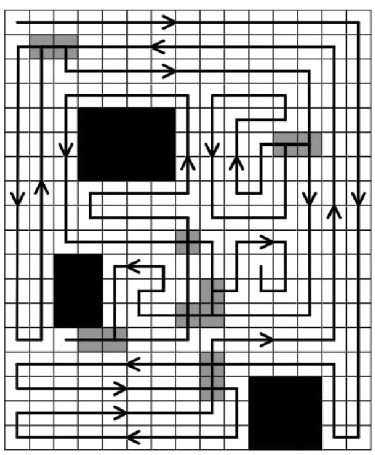

图 5 A* 算法规划路径

Fig. 5 The planning route of A* algorithm

四种算法规划的路径指标比较如表 2 所示。从路径的覆盖率来看,除了 A* 算法,其他算法均达到 100% 覆盖,可完成基本的覆盖要求。从二次覆盖数量来看,除了 RLP-DQN 算法,其他三种算法均出现了多次经过同一栅格的现象,在现实作业环境中,该现象意味着农用无人车重复经过了同一个区域,故采用单一算法进行农用无人车的作业路径规划效果不佳。从掉头次数和转弯次数来看,农用无人车由于自身体积较大并携带农具,掉头行为比转弯行为无论是时间还是能量消耗都更大。与A* 算法相比,RLP-DQN规划的路线掉头次数仅增加了 5.2%,但转弯次数减少了 66.7%;与遗传算法相比,RLP-DQN 规划的路线掉头次数

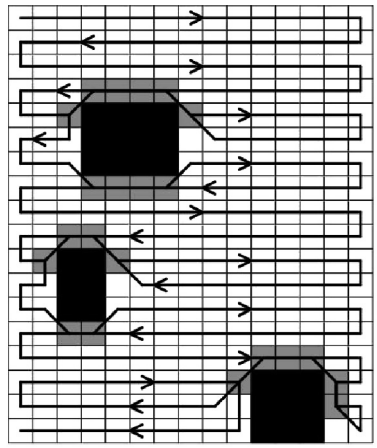

图 6 遗传算法规划路径

Fig. 6 The planning route of genetic algorithm

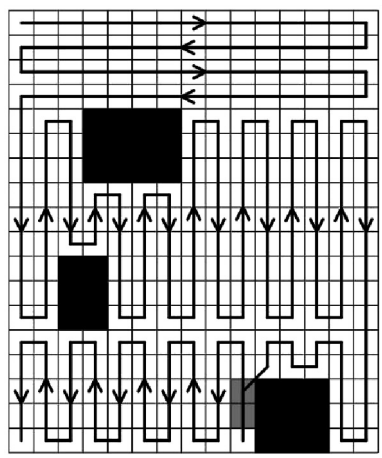

图 7 聚类算法规划路径

Fig. 7 The planning route of clustering algorithm

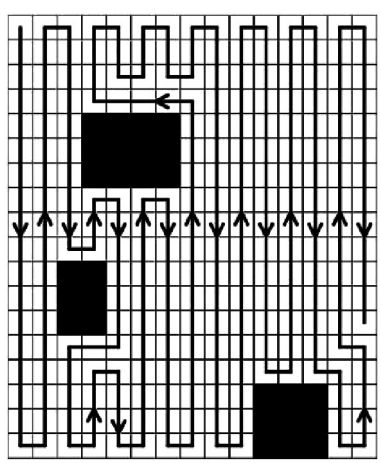

图 8 本文 RLP-DQN 算法规划路径

Fig. 8 The planning route of RLP-DQN algorithm

增加了 53.8%,转弯次数减少了 72.4%;与聚类算法相比,本文算法掉头次数减少了 37.5%,转弯次数增加了 166.7%(从数量上仅增加了 5 次)。从覆盖路线结束时农用无人车的位置来看,除遗

传算法位于场地中央,其余算法的规划路线终点皆位于地图环境的边界。可见,RLP-DQN 算法具有较好的综合性能,可在实现全覆盖路径规划的基础上减少非作业行为的损耗。

表 2　四种算法规划路径指标比较

Table 2　Comparison of planning routes of the four algorithms

算法	二次覆盖数量	掉头次数	转弯次数	覆盖率/%
A* 算法[15]	36	19	24	98.76
遗传算法[16]	12	13	29	100
聚类算法[17]	2	32	3	100
RLP-DQN	0	20	8	100

5　结论

本文针对不规则边界及内部存在障碍物的复杂作业环境下农用无人车作业路径规划问题,提出了一种基于 DQN 算法的 RLP-DQN 覆盖路径规划算法。构建栅格式地图,利用 Pytorch 库设计 RLP-DQN 算法框架,在路径规划过程中记录路径中的转弯次数和掉头次数,作为评估路径规划质量的关键指标。与其他算法对比结果表明,本文 RLP-DQN 算法规划路径无二次覆盖区域,且整体作业消耗等较其他算法有较大改善,在完成全覆盖路径规划的同时能够有效减少路线损耗。

参考文献(References):

[1] 何其全,张青,任志强,等. 智慧农业生产场景典型案例的启示与思考:以苏州市为例[J]. 农业科技管理,2023,42(5):40 – 44.
HE Q Q,ZHANG Q,REN Z Q,et al. Insights and reflections on typical cases of smart agricultural production scene,taking Suzhou city as an example[J]. Management of Agricultural Science and Technology,2023,42(5):40 – 44. (in Chinese)

[2] 邓红,孙栩. 基于鱼群算法的智能机器人全覆盖路径规划[J]. 计算机测量与控制,2023,31(7):222 – 227,297.
DENG H,SUN X. Full coverage path planning of intelligent robot based on fish swarm algorithm[J]. Computer Measurement & Control,2023,31(7):222 – 227,297. (in Chinese)

[3] 代勇,王铎,吴佳欣,等. 基于鲁棒 Tube-MPC 算法的无人车横向控制方法研究[J]. 沈阳理工大学学报,2023,42(3):28 – 34.
DAI Y,WANG D,WU J X,et al. Research on lateral control method of unmanned vehicle based on robust Tube-MPC algorithm[J]. Journal of Shenyang Ligong University,2023,42(3):28 – 34. (in Chinese)

[4] 陈凯,解印山,李彦明,等. 多约束情形下的农机全覆盖路径规划方法[J]. 农业机械学报,2022,53(5):17 – 26,43.
CHEN K,XIE Y S,LI Y M,et al. Full coverage path planning method of agricultural machinery under multiple constraints[J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(5):17 – 26,43. (in Chinese)

[5] 谢金燕,刘丽星,杨欣,等. 改进粒子群优化算法的果园割草机作业路径规划[J]. 中国农业大学学报,2023,28(11):182 – 191.
XIE J Y,LIU L X,YANG X,et al. Orchard lawn mower operation path planning based on improved particle swarm optimization algorithm[J]. Journal of China Agricultural University,2023,28(11):182 – 191. (in Chinese)

[6] WU X Z,BAI J Q,HAO F Q,et al. Field complete coverage path planning based on improved genetic algorithm for transplanting robot[J]. Machines,2023,11(6):659.

[7] 岳伟韬,苏婧,谷志珉,等. 占据栅格地图的最佳栅格大小与地图精度[J]. 机器人,2020,42(2):199 – 206.
YUE W T,SU J,GU Z M,et al. Best grid size of the occupancy grid map and its accuracy[J]. Robot,2020,42(2):199 – 206. (in Chinese)

[8] 杨思明,单征,丁煜,等. 深度强化学习研究综述[J]. 计算机工程,2021,47(12):19 – 29.
YANG S M,SHAN Z,DING Y,et al. Survey of research on deep reinforcement learning[J]. Computer Engineering,2021,47(12):19 – 29. (in Chinese)

[9] 马昂,于艳华,杨胜利,等. 基于强化学习的知识图谱综述[J]. 计算机研究与发展,2022,59(8):1694 – 1722.
MA A,YU Y H,YANG S L,et al. Survey of knowledge graph based on reinforcement learning[J]. Journal of Computer Research and Development,2022,59(8):1694 – 1722. (in Chinese)

[10] JANG B,KIM M,HARERIMANA G,et al. Q-learning algorithms:a comprehensive classification and applications[J]. IEEE Access,2019,7:133653 – 133667.

[11] LADOSZ P,WENG L L,KIM M,et al. Exploration in deep reinforcement learning:a survey[J]. Information Fusion,2022,85:1 – 22.

[12] CARTA S,FERREIRA A,PODDA A S,et al. Multi-DQN:an ensemble of deep Q-learning agents for stock market forecasting[J]. Expert Systems with Applications,2021,164:113820.

[13] 周林娜,汪芸,张鑫,等. 矿区废弃地移动机器人全覆盖路径规划[J]. 工程科学学报,2020,42(9):1220 – 1228.
ZHOU L N,WANG Y,ZHANG X,et al. Complete coverage path planning of mobile robot on abandoned mine land[J]. Chinese Journal of Engineering,2020,42(9):1220 – 1228. (in Chinese)

[14] 刘建伟,高峰,罗雄麟. 基于值函数和策略梯度的深度强化学习综述[J]. 计算机学报,2019,42(6):1406 – 1438.
LIU J W,GAO F,LUO X L. Survey of deep reinforcement learning based on value function and policy gradient[J]. Chinese Journal of Computers,2019,42(6):1406 – 1438. (in Chinese)

[15] 陈镜宇,郭志军,尹亚昆. 基于混合算法的智能割草机全遍历路径规划及其系统设计[J]. 计算机科学,2021,48(S1):633 – 637.
CHEN J Y,GUO Z J,YIN Y K. Full traversal path planning and system design of intelligent lawn mower based on hybrid algorithm[J]. Computer Science,2021,48(S1):633 – 637. (in Chinese)

[16] TUNG W C,LIU J S. Genetic algorithm with modified operators for an integrated traveling salesman and coverage path planning problem[C]//Proceedings of the 16th International Conference on Applied Computing 2019. Madeira Islan,Portugal:IADIS Press,2019:205 – 216.

[17] 胡馨丹,杨盛毅,朱力,等. 基于栅格分区的覆盖路径规划方法[J]. 机械与电子,2022,40(5):13 – 16.
HU X D,YANG S Y,ZHU L,et al. Coverage path planning method based on grid region decomposition[J]. Machinery & Electronics,2022,40(5):13 – 16. (in Chinese)

(责任编辑:宋颖韬)