

基于改进Dueling-DQN的多无人机路径规划算法

付文浩, 葛礼勇, 汪文, 张淳

引用本文

付文浩, 葛礼勇, 汪文, 张淳. 基于改进Dueling-DQN的多无人机路径规划算法[J]. 计算机科学, 2025, 52(8): 326-334.

FU Wenhao, GE Liyong, WANG Wen, ZHANG Chun. [Multi-UAV Path Planning Algorithm Based on Improved Dueling-DQN](#) [J]. Computer Science, 2025, 52(8): 326-334.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于拥塞感知和缓存通信的多智能体路径规划](#)

Congestion-aware and Cached Communication for Multi-agent Pathfinding

计算机科学, 2025, 52(8): 317-325. <https://doi.org/10.11896/jsjcx.240900012>

[基于博弈论的混合粒子群的多无人机任务分配](#)

Multi-UAV Task Assignment Based on Hybrid Particle Swarms Algorithm with Game Theory

计算机科学, 2025, 52(7): 255-261. <https://doi.org/10.11896/jsjcx.240400079>

[基于改进蜣螂优化算法的无人机路径规划](#)

UAV Path Planning Based on Improved Dung Beetle Optimization Algorithm

计算机科学, 2025, 52(6A): 240900136-6. <https://doi.org/10.11896/jsjcx.240900136>

[移动机器人路径规划算法综述](#)

Review of Path Planning Algorithms for Mobile Robots

计算机科学, 2025, 52(6A): 240900074-10. <https://doi.org/10.11896/jsjcx.240900074>

[基于改进DDPG的多AGV路径规划算法](#)

Multi-AGV Path Planning Algorithm Based on Improved DDPG

计算机科学, 2025, 52(6): 306-315. <https://doi.org/10.11896/jsjcx.240500099>

基于改进 Dueling-DQN 的多无人机路径规划算法

付文浩 葛礼勇 汪文 张淳

南京邮电大学计算机学院 南京 210023

(zhc1088@njupt.edu.cn)

摘要 为了解决多无人机在三维未知障碍环境中对动态目标追击的路径规划问题,将人工势场法与深度强化学习算法结合,提出一种基于改进 dueling deep Q network(Dueling-DQN)的多无人机路径规划算法,用于解决多无人机合作捕捉动态目标的路径规划问题。首先,将人工势场法的思想融入到多无人机合作捕捉动态目标的训练奖励函数中,不仅解决了传统人工势场法复杂环境中表现不佳,易陷入局部最优的问题,同时解决了多无人机合作和无人机复杂环境避障问题。此外,为了使无人机之间能更好合作捕捉动态目标,设计了一种多无人机与动态目标的捕捉逃逸策略。仿真结果表明,与 Dueling-DQN 算法相比,提出的 APF-Dueling-DQN 算法有效降低了无人机航迹规划任务过程中发生碰撞的概率,缩短了捕捉动态目标所需规划路径长度。

关键词: 多无人机;Dueling-DQN;路径规划;避障;人工势场;捕捉逃逸

中图分类号 TP181

Multi-UAV Path Planning Algorithm Based on Improved Dueling-DQN

FU Wenhao, GE Liyong, WANG Wen and ZHANG Chun

College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Abstract To address the problem of path planning for multiple unmanned aerial vehicles(UAVs) in three-dimensional unknown obstacle environments when pursuing dynamic targets, this paper proposes a path planning algorithm based on an improved dueling deep Q network(Dueling-DQN) combined with the artificial potential field method and deep reinforcement learning algorithm. This is aimed at solving the problem of path planning for multiple UAVs cooperating to capture dynamic targets. Firstly, it incorporates the idea of the artificial potential field method into the training reward function for multiple UAVs cooperating to capture dynamic targets, which not only addresses the shortcomings of traditional artificial potential field methods in complex environments where they are prone to local optima, but also solves the problems of multi-UAV cooperation and UAV obstacle avoidance in complex environments. Additionally, to facilitate better cooperation among UAVs in capturing dynamic targets, a strategy for the capture and escape of dynamic targets by multiple UAVs is designed. Simulation results demonstrate that compared to Dueling-DQN algorithm, the proposed APF-Dueling-DQN algorithm effectively reduces the probability of collisions during UAV trajectory planning tasks and shortens the planned path length required to capture dynamic targets.

Keywords UAVs, Dueling-DQN, Path planning, Obstacle avoidance, Artificial potential field, Capture and escape

1 引言

无人机具有体积小、机动快、成本低、灵活性强的特点,因此在军事和民用领域都发挥了重要作用。在自然灾害、地震、洪水、大规模火灾等突发事件中,无人机可以第一时间到达现场^[1]。在军事对抗环境下,无人机以其良好的机动隐身性能在电子战中发挥重要作用^[2]。无人机路径规划问题是一个复杂的多约束组合优化问题。合理的航迹规划是无人机完成飞行和作战任务的关键技术^[3-4]。未知的环境威胁对无人机的实际飞行有很大的影响。有效的航迹规划算法可以

使无人机在飞行过程中避开地形障碍物和敌方威胁。因此,无人机必须具备快速独立决策的能力^[5]。

近年来,研究人员提出了许多关于智能体路径规划和避障的研究方法。Gu等^[6]提出了一种改进的快速扩展随机树算法来解决无人机的动态路径规划问题,通过引入目标启发信息和剪枝重构,使无人机在面对突发威胁时能快速调整航路,但是在复杂环境下可能会存在局部最优问题。Li等^[7]提出了基于合作共同演化编码粒子群优化算法来解决多无人机协同搜索动态目标的问题,建立马尔可夫过程用于模拟目标运动,并通过多个粒子群的信息共享提高了全局搜索性能,但

到稿日期:2024-06-17 返修日期:2024-09-22

基金项目:国家自然科学基金(61903198)

This work was supported by the National Natural Science Foundation of China(61903198).

通信作者:张淳(zhc1088@njupt.edu.cn)

是该算法在高动态环境中的适应性还需提升。Zhang 等^[8]提出了一种改进的蚁群优化算法来解决多无人机在山地环境下搜索多目标的问题,通过信息融合和信息扩散两种策略来同步无人机搜索不同区域获取的信息,并采用分层搜索策略提高搜索的效率,但是在大规模数据集的情况下,文中的聚类算法可能会需要大量的计算资源和存储空间。Tang 等^[9]提出了改进A*算法来解决无人机在路径规划时的效率低下问题,通过引入双向搜索和定向搜索机制在一定程度上提高搜索效率,同时引入路径平滑策略来减少冗余路径点,但是在高维空间中,A*算法的性能可能会受到维度灾难的影响,搜索的效率会大大降低。Xiang 等^[10]提出了一种改进的人工势场算法来解决无人机追踪动态目标的问题,首先采用基于聚类演化的目标关联算法将无人机划分成多个集群来匹配相应的目标,接着将人工势场算法和干扰敏感相结合来规划无人机追踪目标的轨迹,但是人工势场算法需要预先知道环境中的障碍物和目标位置,无法适应未知环境或者动态环境。Yildiz 等^[11]提出了一种混合算法来解决多无人机路径规划问题,结合了基于共识的虚拟领导跟踪群体算法和基于目标距离的快速探索随机树与粒子群优化算法,实现了在无人机群之间合作和协调飞行的基础上缩短规划的飞行路径,但是该算法可能会在无人机太靠近障碍物时陷入局部最小值,导致无人机无法有效避开障碍物。随着机器学习的发展,深度强化学习由于具有出色的泛化性和适配性被广泛应用于任务规划领域^[12]。Zhou 等^[13]提出基于导向强化Q学习的无人机路径规划方法,利用接收信号强度定义回报值,加快了学习算法的收敛速度,但是传统Q学习算法在训练过程中仍然存在收敛性差的问题。在无人机捕捉动态目标领域中,Kownacki^[14]提出了一种改进的人工势场算法来解决无人机跟踪动态目标的问题,该方法基于速度向量来控制滚转角、俯仰角和偏航角,结合人工势场法实现了无人机实时跟踪动态目标,但是该方法非常依赖跟踪

目标的实时位置和速度信息。Li 等^[15]提出了一种增强的蚁群算法来解决战场环境下的多无人机路径规划问题,通过模糊逻辑记忆力机制获得低成本的飞行路线,同时采用分层扩展和矫正策略优化飞行路径,但是文章只针对离线规划的场景,无法实现动态环境下的路径规划。因此,在多无人机捕捉动态目标的路径规划任务中,选择一种收敛性好、成功率高的算法尤为重要。

在多无人机捕捉动态目标的任务过程中,会存在未知的三维障碍物,无人机不仅要绕开障碍物,还要在最短的时间内捕捉到动态目标。为了解决多无人机在未知三维空间中的自主实时飞行路径规划问题,提出一种基于改进 Dueling-DQN 的多无人机捕捉动态目标路径规划算法。根据飞行任务的要求,设计了无人机的状态空间、无人机的动作空间及其与环境交互的奖励规则。此外,将人工势场法与深度强化学习算法相结合,设计了一种多无人机与动态目标的捕捉逃逸策略。本文的主要贡献概述如下。

1)提出了一种人工势场法和 Dueling-DQN 算法相结合的多无人机捕捉动态目标路径规划算法,以解决人工势场法在复杂环境中表现不佳,易陷入局部最优的问题,并且能够提升深度强化学习训练过程中的奖励数值的稳定性和网络模型的收敛速度。

2)设计了一种多无人机与动态目标的捕捉逃逸策略,减小了多无人机执行任务过程中无人机之间发生碰撞的概率,缩短了无人机捕捉目标的路径长度。

2 系统模型

为了完成多无人机在三维空间中捕捉动态目标的实时飞行路径规划,避免无人机在航迹规划过程中陷入局部最优解,提高算法训练过程中奖励数值稳定性和加速网络模型的收敛,本文提出基于改进 Dueling-DQN 的多无人机捕捉动态目标路径规划算法。其系统模型如图1所示。

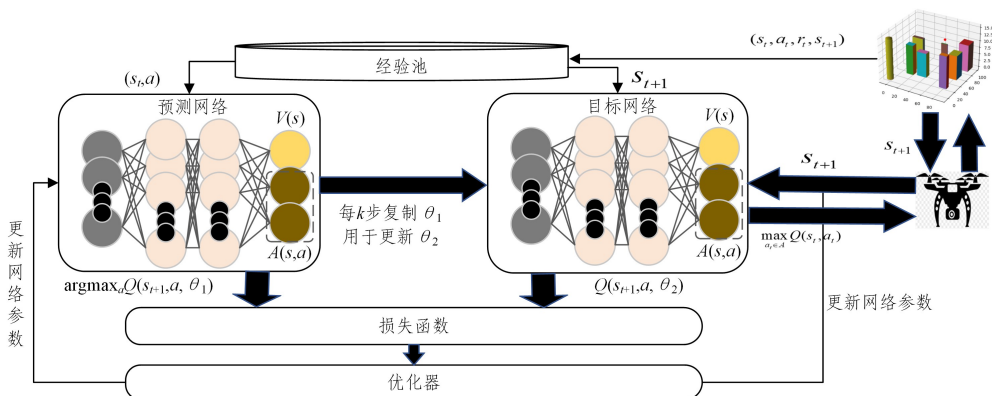


图1 系统模型

Fig. 1 System model

其中,Dueling-DQN 将Q值分解为状态的价值函数和每个动作的优势函数两部分。相较于传统DQN学习的目标为Q函数,此处将Q函数的神经网络拆分成两个子网络,分别学习价值函数和优势函数。优势函数网络会输出每个动作的价值,价值函数网络输出当前状态的价值,这样可以更好地估

计状态的价值和不同动作之间的优势^[16]。无人机路径规划过程中将面临多种情况下的动作选择,环境越复杂,特别是环境中障碍物多的情况,相应的整个飞行过程中产生的动作子集也更大,将不同动作的优势值作为无人机动作选择的一部分,使无人机在路径规划的过程中更加关注动作选择对当前

和长期回报的重要性,让整个训练过程更加稳定。在模型训练过程中,价值函数和优势函数两个子网络分别预测价值函数和优势函数的值,然后将两个结果相加就是 Q 函数的值。类似于传统 Q 网络中的 $\arg\max_a Q(s_{t+1}, a, \theta_1)$ 选择 Q 值最大的动作,将该动作带入目标网络,通过 $Q(s_{t+1}, a, \theta_2)$ 计算动作的 Q 值,计算损失函数,经过优化器优化,更新预测网络参数,并且,每 k 步复制预测网络参数 θ_1 用于更新一次目标网络参数 θ_2 。同时,由目标网络通过 $\max_{a_t \in A} Q(s_t, a_t)$ 选取 Q 值最大动作,由无人机在空间中执行该动作后将训练经验数据片段 (s_t, a_t, r_t, s_{t+1}) 传入经验池,模型分批从经验池中采样状态动作对 (s_t, a) 用于训练。为了减少相同障碍模型造成的训练关联性和无效训练,加快模型的训练速度,在训练过程环境中的障碍数量固定,但障碍位置随机。

2.1 无人机路径规划模型

为了更准确描述、客观评估和对比不同路径规划算法的性能指标,将复杂的多无人机路径规划问题公式化,本文使用多组合约束的数学优化来定义多无人机捕捉动态目标的路径规划模型。

首先,存在 N 架无人机可用,每架无人机都有一个初始位置 P_{init}^i 和目标位置 P_{target}^i ,无人机在飞行过程中生成的一系列点迹记为 $\{(P_1^i, T_1^i), (P_2^i, T_2^i), (P_3^i, T_3^i), \dots, (P_n^i, T_n^i)\}$,其中, T_n^i 表示无人机在 n 和 $n-1$ 时间段内产生的飞行路径长度。路径规划算法的目的就是为了让多无人机在更短的时间和更短的飞行距离内捕捉到动态目标。为了简化模型的复杂性,便于描述,假设无人机始终匀速飞行,且不存在环境因素对飞行产生影响,所以目标函数为所有无人机的总飞行距离,如式(1)所示。

$$\min \sum_{i=1}^N \sum_{j=1}^n T_j^i \quad (1)$$

本文中无人机从同一起点出发,为了保证多无人机合作追击动态目标的安全性,每个时刻同一位置只能有一架无人机访问,如式(2)所示。

$$P_i \neq P_{i'} \quad (2)$$

其中, $i' \in (0, i) \cup (i, N), t \in [2, n]$ 。

为了保证多无人机追击动态目标的合作通信,无人机之间应该存在合理的通信范围,即同一时刻无人机之间的距离不能大于最大通信距离阈值,如式(3)所示。

$$\|P_i, P_{i'}\|_2 \leq \mathcal{O}_c \quad (3)$$

其中, $i' \in (0, i) \cup (i, N), t \in [2, n]$; \mathcal{O}_c 为无人机之间的合作最大边界。

目标位置必须被访问,如式(4)所示。

$$P_1^i = P_{target}^i \cup P_2^i = P_{target}^i \cup P_3^i = P_{target}^i \cup \dots \cup P_n^i = P_{target}^i = 1 \quad (4)$$

其中, $i \in [1, N]$,式(4)表示必有一无人机在某一时刻的坐标与目标位置相同。

为了保证无人机飞行过程中的安全性,避免无人机与障碍物发生碰撞,无人机与障碍物之间存在最小距离限制。为了便于表示,将障碍物作为一个质点 $P_{obstacle}$,无人机与障碍物之间的距离如式(5)所示。

$$\|P_i, P_{obstacle}\|_2 > d_{min} \quad (5)$$

其中, d_{min} 为无人机与障碍物之间的安全距离阈值。

2.2 多无人机动力学模型

为了更好地完成多无人机对动态目标的捕捉路径规划任务,本文采用主机和僚机互相配合的方式。其中,主机负责对动态目标的捕捉,僚机协助,主机与僚机间实时通信,共享自己和目标实时位置信息。空间中无人机与待捕捉目标相对位置关系如图 2 所示。

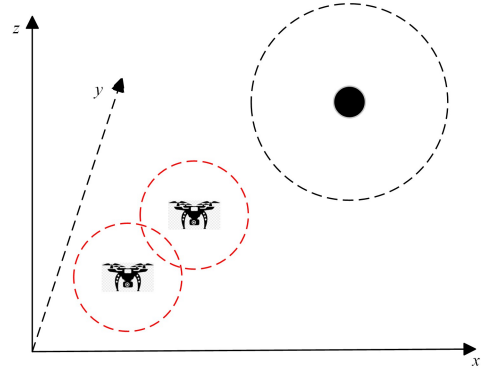


图 2 无人机与待捕捉目标相对位置关系示意图

Fig. 2 Schematic diagram of the relative position relationship between the UAV and the target to be captured

其中,僚机在主机侧面飞行,红色虚线圈是无人机的攻击范围,黑色点为待捕捉目标,黑色虚线圈是待捕捉目标的感知范围,待捕捉目标的感知范围大于无人机的攻击范围,便于待捕捉目标感知到无人机后逃逸。

假设无人机在某个水平面飞行,则无人机转弯角度受到过载约束。无人机飞行转弯示意图如图 3 所示。

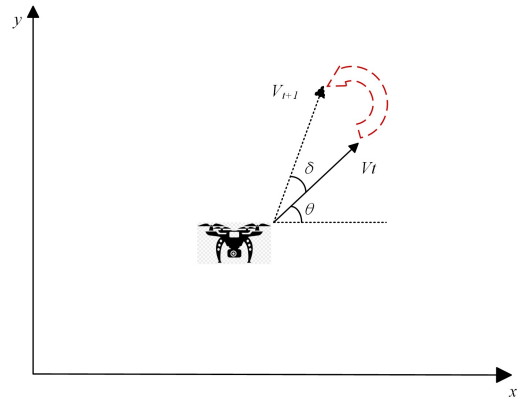


图 3 无人机飞行转弯示意图

Fig. 3 UAV flight turn diagram

其中, V_t 为无人机当前时间点速度矢量, θ 为 V_t 与水平面夹角,红色虚线箭头为无人机航向调整方向, V_{t+1} 为无人机航向调整后的速度矢量, δ 为无人机转弯角度。

转弯角的约束定义如式(6)所示。

$$\arccos\left(\frac{P_t^T P_{t+1}}{\|P_t\| \|P_{t+1}\|}\right) \leq \theta_{max} \quad (6)$$

其中, P_t 为无人机 t 时刻在空间中的位置, θ_{max} 为最大转弯角。

根据无人机动力学特性,将无人机在空间中某个位置的受力定义为:

$$m * \frac{d^2 \vec{P}_i}{dt^2} = F_{\text{total}}(\vec{P}_i, \vec{v}, t) \quad (7)$$

其中, m 表示无人机的质量, \vec{P}_i 表示无人机的位置矢量, \vec{v} 表示无人机的速度矢量, t 表示时间, F_{total} 表示作用在无人机上的合力。

3 基于改进 Dueling-DQN 的无人机路径规划算法

在传统的 DQN 算法中,只使用一个神经网络来拟合动作状态函数,这通常会导致高估 Q 值。因此,提出了 Dueling-DQN 方法来解决高估的问题。同时,将人工势场法的思想融入到多无人机合作捕捉动态目标的训练奖励函数中,解决了多无人机合作和无人机复杂环境避障问题。此外,还设计了一种多无人机与动态目标的捕捉逃逸运动策略。基于改进 Dueling-DQN 的无人机路径规划算法能够有效减少无人机航迹规划任务过程中发生碰撞的概率,缩短捕捉目标所需规划路径长度。

3.1 Dueling-DQN 算法

在 Dueling-DQN 算法中,首先构建两个结构相同但网络参数不同的竞争网络 $Q_1(s_t, a_t, \theta_1)$ 和 $Q_2(s_t, a_t, \theta_2)$ 。其中预测网络 Q_1 用于选取下一个状态的最大 Q 值动作,目标网络 Q_2 用于计算选取动作 Q 值^[16]。同时,将预测网络和目标网络分别拆分成状态值函数 V 和优势函数 A ,两个函数共享预测网络和目标网络的参数 θ_1 和 θ_2 ,在这个基础上,它们又有各自的参数,如式(8)中的 ρ_1 和 ρ_2 。Dueling-DQN 算法本质输出与 DQN 算法相同,Dueling-DQN 的 Q 值函数定义如式(8)所示。

$$Q(s, a, \theta, \rho_1, \rho_2) = V(s, \theta, \rho_1) + A(s, a, \theta, \rho_2) \quad (8)$$

其中, s 表示状态, a 表示动作, $V(s)$ 为状态值函数, $A(s, a)$ 为优势函数。 ρ_1 和 ρ_2 分别为状态值函数和优势函数的参数。

损失函数定义如式(9)所示。

$$L = \frac{1}{2} [r_1 + \gamma \max_{a \in A} Q(s_2, a_2, \theta_2) - Q(s_1, a_1, \theta_1)]^2 \quad (9)$$

其中, r_1 是执行动作 a_1 的奖励, γ 为折扣率, $\max_{a \in A} Q(s_2, a_2, \theta_1)$ 表示在状态 s_2 时挑选最大 Q 值动作 a_2 , $Q(s_1, a_1, \theta_1)$ 为预测网络计算的 Q 值。智能体处于状态 s_1 时,执行动作 a_1 进入状态 s_2 。

强化学习过程中的目标值是非固定的,会出现一些异常情况,导致产生的目标值与其他样本明显不同,这些异常情况可能是由于噪声、不稳定性或环境变化引起的。为了提高 Dueling-DQN 算法训练的数值稳定性和效果,使用 Hube 损失函数代替均方误差损失函数。

损失函数 L 的优化目标是预测网络参数 θ_1 ,通过梯度下降的方法,让 $Q(s_1, a_1, \theta_1)$ 不断逼近 $r_1 + \gamma \max_{a \in A} Q(s_2, a_2, \theta_2)$ 。每隔 k 步复制预测网络参数 θ_1 用于更新目标网络参数 θ_2 。

由于预测网络和目标网络分别被拆分为状态值函数 V 和优势函数 A ,损失函数 L 的优化目标是预测网络参数,因此预测网络拆分得到的状态值函数和优势函数的参数 ρ_1 和 ρ_2 的更新需要对损失函数 L 求偏导,如式(10)和式(11)所示。

$$\rho_1 = \rho_1 - \alpha \frac{\partial L}{\partial V_{\rho_1}} \quad (10)$$

$$\rho_2 = \rho_2 - \alpha \frac{\partial L}{\partial A_{\rho_2}} \quad (11)$$

其中, α 为学习率。

3.2 Dueling-DQN 算法网络结构

网络模型结构如图4所示。来自环境的图像,包括无人机的位置、速度和障碍物,被用作状态输入,并通过3个卷积层提取状态特征。网络中的隐藏层是完全连接层,采用 ReLU 函数激活功能。分别输出状态价值函数和优势函数,最后,再求和得到输出无人机针对不同动作的 Q 值。

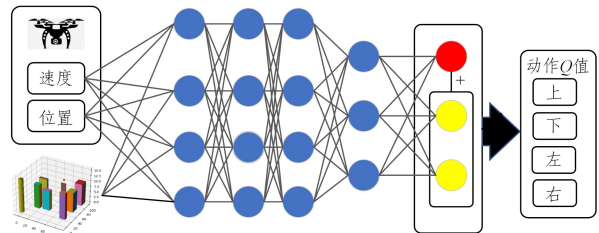


图4 Dueling-DQN 算法网络结构

Fig. 4 Dueling-DQN algorithm network structure

3.3 APF-Dueling-DQN 的无人机路径规划算法

人工势场法是将运动目标模拟成势场中的一个粒子,这个势场将路径规划过程中的目的终点的吸引力和障碍物的排斥力相结合来引导粒子在空间中的移动。粒子在势场中受到的引力场的定义如式(13)所示。

$$D_t = \sqrt{(x_o - x_t)^2 + (y_o - y_t)^2 + (z_o - z_t)^2} \quad (12)$$

$$U_{\text{att}} = \frac{1}{2} k D_t^2 \quad (13)$$

其中, D_t 表示无人机在时间 t 和待捕捉目标之间的距离, (x_o, y_o, z_o) 表示待捕捉目标的坐标, (x_t, y_t, z_t) 表示无人机在时间 t 的坐标, U_{att} 为待捕捉目标的引力场数值, k 为常数,表示引力系数。由式(13)可知,当无人机距离待捕捉目标距离越远时,引力场 U_{att} 值越大。

斥力势场用于帮助无人机避免与障碍物发生碰撞,环境中模拟的柱体为斥力源,排斥力的数值大小和无人机与障碍物的距离成反比。当无人机接近障碍物时,排斥力增大,无人机此时受到排斥力的影响避开障碍物。在势场中受到的斥力场如式(14)所示。

$$U_{\text{rep}} = \begin{cases} \frac{1}{2} \vartheta \left(\frac{1}{D_{u-o}} - \frac{1}{D_{\text{obs}}} \right)^2, & \text{if } D_{u-o} \leq D_{\text{obs}} \\ 0, & \text{if } D_{u-o} > D_{\text{obs}} \end{cases} \quad (14)$$

其中, U_{rep} 为斥力场数值; ϑ 为常数,表示斥力系数; D_{u-o} 为障碍物与无人机的距离; D_{obs} 为障碍物的影响范围。

吸引力和排斥力由力场取负梯度求得,如式(15)和式(16)所示。

$$F_{\text{att}} = -\nabla(U_{\text{att}}) = -k D_t \quad (15)$$

$$F_{\text{rep}} = -\nabla(U_{\text{rep}}) = \begin{cases} \vartheta \left(\frac{1}{D_{u-o}} - \frac{1}{D_{\text{obs}}} \right) \frac{1}{D_{u-o}^2} \left[\frac{\partial D_{u-o}}{\partial x}, \frac{\partial D_{u-o}}{\partial y}, \frac{\partial D_{u-o}}{\partial z} \right], & \text{if } D_{u-o} \leq D_{\text{obs}} \\ 0, & \text{if } D_{u-o} > D_{\text{obs}} \end{cases} \quad (16)$$

无人机在空间势场中的受力情况如图 5 所示。

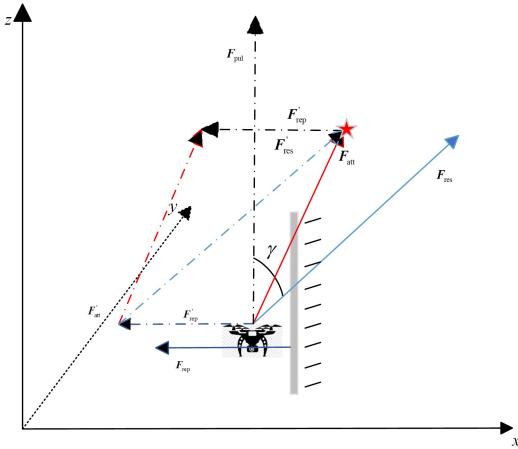


图 5 无人机受力示意图(电子版为彩图)

Fig. 5 UAV force diagram

在图 5 中,灰色体为障碍,红色五角星为待捕捉目标点,紫色实线 F_{rep} 表示无人机受到的由斥力源障碍物影响的排斥力矢量,红色实线 F_{att} 为待捕捉目标影响的吸引力矢量,黑色虚线 F_{pul} 表示无人机自身的动力矢量,蓝色实线 F_{res} 为无人机在空间中受到的势场合力矢量,紫色虚线 F'_{rep} 、红色虚线 F'_{att} 分别为虚拟排斥力矢量和虚拟吸引力矢量,由二者求得蓝色虚线 F'_{res} 虚拟势场合力矢量, γ 是无人机自身的动力矢量 F_{pul} 与势场合力 F_{res} 之间的夹角。

可定义无人机在训练过程中受势场力影响的奖励为:

$$r_{APF} = \frac{1}{\arctan \gamma} \tau \quad (17)$$

其中, r_{APF} 为无人机受势场力影响的奖励; $\arctan \gamma$ 表示夹角开口大小; τ 是常数,为奖励系数。

假设无人机在空间中某一位置以恒定速度向待捕捉目标飞行。如图 6 所示,受势场影响的力为 F_1 和 F_1' ,无人机自身动力为 F_2 和 F_2' ,合力为 F_3 和 F_3' ,夹角为 α 和 α' ,且定义 $|F_1| = |F_2| = |F_1'| = |F_2'|$,夹角 $\alpha < \alpha'$,则合力 $|F_3| > |F_3'|$ 。即受势场影响的力与飞机自身动力夹角越小,则无人机所受合力就越大,飞行所需功耗就越小, $\arctan \gamma$ 值越小,受势场力影响得到的奖励分数就越高。

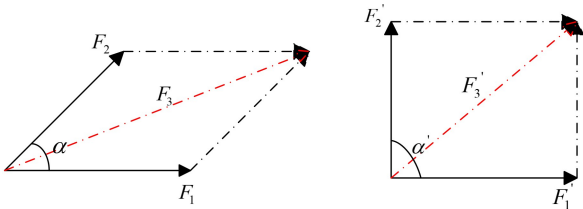


图 6 合力受夹角影响示意图

Fig. 6 Schematic diagram of resultant force affected by included angle

双无人机合作捕捉动态目标,目标在感知到无人机后进行逃逸。无人机采用的是分布式学习,通过强化学习的训练捕捉单个固定目标。主机与僚机之间为了更好的相互配合以捕捉动态目标,相互之间存在吸引力和排斥力,吸引力的作用是保持僚机始终在主机侧面配合主机执行捕捉动态目标任务;

排斥力的作用是避免主机与僚机之间发生碰撞,即主机与僚机之间的合作-安全边界。而待捕捉的动态目标相对于无人机也存在吸引力和排斥力。当无人机未进入动态目标的感知范围时,仅存在动态目标对无人机的吸引力;当无人机进入动态目标的感知范围时,对于动态目标自身而言,存在对无人机的排斥力,为了自身逃逸,对于无人机而言,仍然存在吸引力。

3.3.1 多无人机与动态目标的捕捉逃逸数学模型

假设在空间中存在 a 个无人机,忽略无人机的大小,将无人机视作空间中的一个点,且这 a 个无人机同时移动不存在时延的问题,无人机之间相互通信,共享自身位置与动态目标的实时位置。将无人机的位置定义为 $U_q \in R^n$,为了便于表示,假设动态目标共移动 b 次,将待捕捉的动态目标位置定义为 $T_j \in R^n$ 。则单个无人机的运动方程可以定义为:

$$\begin{aligned} \widetilde{U}_q = & \sum_{p=1, p \neq q}^a \sum_{j=1}^b f_1(U_q - U_p) f_2(U_q - T_j) \\ q = & 1, 2, \dots, a; j = 1, 2, \dots, b \end{aligned} \quad (18)$$

其中, $f_1(\cdot)$ 表示无人机之间的吸引力与排斥力函数, $f_2(\cdot)$ 表示无人机与待捕捉动态目标之间的吸引力与排斥力函数。根据无人机之间的合作-安全边界,可将无人机之间的吸引力和排斥力函数定义为:

$$f_1 = \mu_1 (e^{\frac{O_{c-s}}{\|U_q - U_p\|}} - e) \quad (19)$$

其中, μ_1 为增益系数, O_{c-s} 为合作-安全范围, $\|U_q - U_p\|$ 表示无人机之间的距离。当无人机之间的距离大于合作范围时, $\frac{O_{c-s}}{\|U_q - U_p\|} < 1$, 则 $e^{\frac{O_{c-s}}{\|U_q - U_p\|}} - e < 0$, 此时无人机之间为吸引力;

当无人机之间的距离小于安全范围时, $\frac{O_{c-s}}{\|U_q - U_p\|} > 1$, 则 $e^{\frac{O_{c-s}}{\|U_q - U_p\|}} - e > 0$, 无人机之间为排斥力。

将无人机与待捕捉动态目标之间的吸引力与排斥力函数定义为:

$$f_2 = \begin{cases} f_2', & \text{if } \|U_q - T_j\| > T_{per} \\ f_2'', & \text{if } \|U_q - T_j\| \leq T_{per} \end{cases} \quad (20)$$

$$f_2' = -\nabla \left(\frac{\mu_2}{2} \|U_q - T_j\|^2 \right) = -\mu_2 \|U_q - T_j\| \quad (21)$$

$$\begin{aligned} f_2'' = & f_2' + \left[-\nabla \left[\frac{\mu_3}{2} \left(\frac{1}{\|U_q - T_j\|} - \frac{1}{T_{per}} \right)^2 \right] \right] \\ = & -\mu_2 \|U_q - T_j\| + \mu_3 \left(\frac{1}{\|U_q - T_j\|} - \frac{1}{T_{per}} \right) \frac{1}{\|U_q - T_j\|^2} (\nabla \|U_q - T_j\|) \end{aligned} \quad (22)$$

其中, f_2' 为动态目标对无人机的吸引力; $\|U_q - T_j\|$ 表示无人机与动态目标之间的距离; T_{per} 为动态目标的感知范围,当无人机未出现在目标的感知范围内时,仅存在目标对无人机的吸引力; μ_2 为目标对无人机的吸引力增益系数, μ_3 为目标对无人机的排斥力增益系数; f_2'' 为动态目标对无人机产生的吸引力和排斥力的合力。定义无人机与待捕捉目标之间的吸引力是为了增强无人机与动态目标之间的关联性,无人机在做决策时考虑了与动态目标的距离,当无人机与动态目标的距离越远时,动态目标对无人机的吸引力越强;定义无人机与待捕捉目标之间的排斥力是为了让动态目标能在感知到无人机的存在时进行逃逸,且当无人机与动态目标距离越小时,动态

目标对无人机的斥力越大,鼓励动态目标逃逸,为了更好展示无人机是如何追击动态目标。

3.3.2 动态目标逃逸策略

在无人机捕捉动态目标过程中,动态目标在感知范围内发现无人机的存在,则立马进行逃逸摆脱无人机的捕捉。本文通过计算两个无人机与动态目标的相对位置形成的角度进行逃逸,逃逸方向映射二维水平面如图 7 所示。

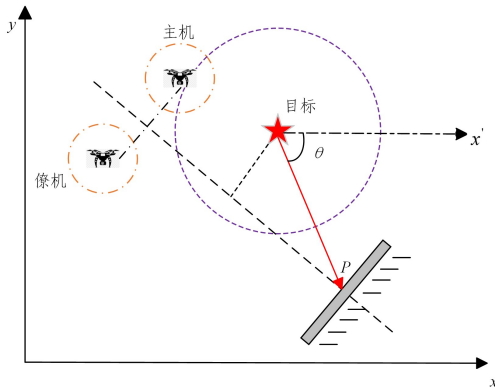


图 7 动态目标逃逸示意图(电子版为彩图)

Fig. 7 Dynamic target escape diagram

其中,橙色虚线圈为无人机捕捉范围,红色五角星为待捕捉目标,灰色柱体为环境边界,红色箭头方向为动态目标逃逸方向,紫色虚线圈为动态目标感知范围, x' 表示 x 轴平行方向, θ 为目标逃逸方向与 x 轴夹角。通过虚线连接两个无人机,过此虚线作垂线与边界障碍相交于点 P ,坐标映射二维平面为 (x_{bor}, y_{bor}) ,待捕捉目标坐标为 (x_m, y_m) ,则动态目标映射水平面相较于 x 轴的逃逸方向可定义为:

$$\theta = \arctan \frac{y_{bor} - y_m}{x_{bor} - x_m} \quad (23)$$

待捕捉目标逃逸方向角选择 θ ,其是最快逃逸方向。

3.3.3 无人机捕捉动态目标策略

双无人机合作捕捉动态目标策略可以分为 3 个部分:一是无障碍影响,未进入动态目标感知范围时;二是有障碍影响,未进入动态目标感知范围时;三是待捕捉动态目标感知到无人机存在的情况。

首先是第一种,无障碍影响,未进入动态目标感知范围,双无人机捕捉动态目标策略映射二维水平面如图 8 所示。

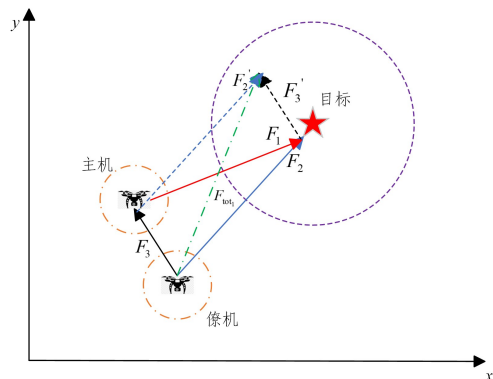


图 8 无障碍影响无人机捕捉动态目标示意图(电子版为彩图)

Fig. 8 Schematic diagram of accessibility affecting the capture of dynamic targets by UAVs

如图 8 所示,橙色虚线圈为无人机捕捉范围,紫色虚线圈为待捕捉动态目标感知范围。当无人机未进入动态目标感知范围内且不受障碍影响时,主机仅受动态目标吸引力 F_1 ,由红色箭头表示,所以主机的飞行航向始终朝向动态目标。而僚机不仅受到来自目标的吸引力 F_2 ,由蓝色箭头表示,还受到来自主机的力 F_3 ,由黑色箭头表示,通过构造 F_2, F_3 的虚拟力 F_2', F_3' 求得僚机此时受到的合力 F_{tot1} ,也是僚机的航向。

其次是第二种,有障碍影响,未进入动态目标感知范围,双无人机捕捉动态目标策略映射二维水平面如图 9 所示。

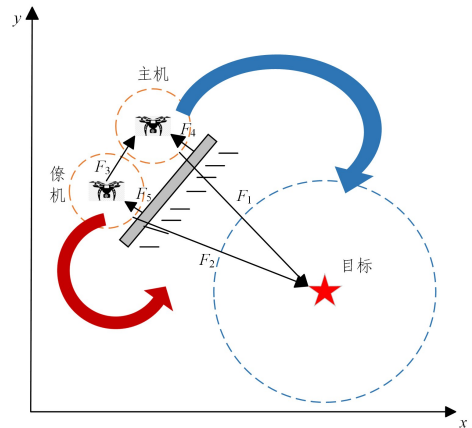


图 9 有障碍影响无人机捕捉动态目标示意图(电子版为彩图)

Fig. 9 There are obstacles affecting the UAV to capture dynamic targets schematic

在图 9 中,当有障碍影响,未进入动态目标的感知范围时, F_1 和 F_2 分别是动态目标对主机和僚机的吸引力, F_4 和 F_5 分别是障碍物对主机和僚机的排斥力, F_3 是主机和僚机之间的作用力。主机的航向为蓝色箭头方向,而僚机的航向则有两种,一种是跟随主机为蓝色箭头方向,另一种是选择红色箭头方向,其主要受无人机之间的合作-安全边界影响。从图 9 可以看出,僚机选择红色箭头方向作为航向明显其飞行代价更小,距离动态目标更近,但是可能会超出合作安全边界的范围,致使僚机往回飞靠近主机,造成更大的飞行代价,所述情况如图 10 所示。

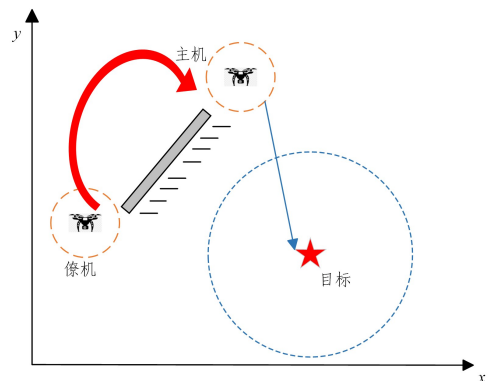


图 10 超出合作-安全边界示意图

Fig. 10 Beyond the cooperative security boundary diagram

图 10 中,蓝色箭头为主机航向,红色箭头为僚机航向调整方向,主机和僚机在开始时分别选择图 9 的两个航向,致使到达如图 10 所示位置时达到合作边界最大范围,僚机进行航

向调整飞往主机。

再者,当待捕捉动态目标感知到无人机存在的条件下,双无人机捕捉动态目标策略映射二维水平面如图 11 所示。

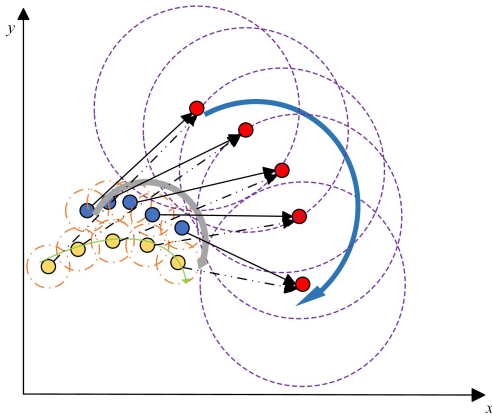


图 11 目标逃逸无人机路径实时重规划示意图(电子版为彩图)

Fig. 11 Schematic diagram of realtime replanning of target escape UAV path

动态目标匀速飞行,飞行速度为 3 m/s,目标感知范围为 4 m,无人机在追击动态目标过程中,由于式(1)约束函数的作用,主机随着目标的移动而实时调整飞行方向,获得最短的总飞行路程,而僚机由于合作-安全边界影响,配合主机执行追击任务。

其中,蓝色点迹为主机飞行路径,黄色点迹为僚机飞行路径,红色点迹为动态目标逃逸路径,橙色虚线圈为无人机捕捉范围,紫色虚线圈为动态目标感知范围,黑色箭头实线和虚线分别为主机和僚机在当前位置的实时航向,皆为无人机飞行路径切线方向。因此,无论是主机还是僚机,它们的路径规划都随着动态目标的逃逸位置进行实时重规划。

4 实验与结果分析

本章对上述提出的基于改进 Dueling-DQN 的多无人机捕捉动态目标路径规划算法进行了仿真分析。为了验证所提出方法的有效性,将从训练过程中飞行路径规划的碰撞率、实际单个无人机捕捉单个固定目标飞行路径的长度和平滑度两个方面对所提出方法与 Dueling-DQN 算法的性能进行比较。

4.1 实验参数设置

4.1.1 实验设备

实验设备如表 1 所列。

表 1 实验设备

设备	配置
系统	Windows 11 64 bit
硬盘	500 GB
软件	Python 3.9
框架	Torch 1.13.0

4.1.2 无人机飞行参数约束

在仿真实验中,无人机的飞行路径规划空间设置为 $80 \times 100 \times 15$ 的立方体。假设无人机以均匀的速度飞行,不受自然因素的干扰。

为了保证仿真实验的有效性,无人机的飞行约束参数

设置如表 2 所列。

表 2 飞行约束参数

参数	值
飞行速度/(m/s)	5
俯仰角范围/ $^{\circ}$	$(-90, 90)$
方向角范围/ $^{\circ}$	$(0, 180)$
高度范围/m	$(0, 15)$
东西飞行范围/m	$(0, 80)$
南北飞行范围/m	$(0, 100)$
电池容量/mAh	5000
飞行能量消耗/(毫安时/步)	10
感知范围/m	3

4.1.3 路径规划算法训练参数

基于改进 Dueling-DQN 的多无人机捕捉动态目标路径规划算法参数如表 3 所列。

表 3 算法参数

参数	值
经验回放池大小	2×10^5
学习率	0.0003
折扣率	0.99
批量大小	128
训练周期	4000
网络参数更新间隔	100

4.1.4 路径规划算法奖励函数

为了使无人机之间能够互相配合,避开障碍,成功捕捉到动态目标,实现合理的路径规划,必须设置合理的奖励函数,奖励函数直接决定了模型的训练效果。

无人机之间的合作-安全边界约束以及待捕捉目标对于无人机的吸引力作用奖励如式(24)和式(25)所示:

$$r_{c-s} = \mu_1 \tau_{c-s} (e^{\frac{0}{\|v_{c-s} - v_p\|}} - e) \quad (24)$$

$$r_{att} = \tau_{att} (f'_{z_t} - f'_{z_{t-1}}) \quad (25)$$

其中, τ_{c-s} 为合作-安全边界奖励系数, τ_{att} 为吸引力奖励系数。在式(25)中,无人机与待捕捉目标间有吸引力作用致使无人机能够向目标靠近,而吸引力的大小与无人机和待捕捉目标间的距离有关,通过吸引力的大小变化判断无人机在相邻的时间段内是否产生至目标的有效位移并予以奖励。

为了鼓励待捕捉目标逃逸,设计目标逃逸奖励,当无人机进入待捕捉目标的侦测范围时,结合式(23)中待捕捉目标的最佳逃逸方向角,当动态目标实际逃逸方向和最佳逃逸角之间的夹角越小,动态目标就能越快逃离。动态目标逃逸奖励如式(26)所示。

$$r_{away} = \frac{\tau_{away}}{|\theta_t - \theta|} \quad (26)$$

其中, τ_{away} 是逃逸奖励增益系数, θ_t 为动态目标逃逸角。

其次需要考虑无人机是否成功捕捉动态目标、是否与环境障碍发生碰撞、是否超过无人机的最大步长,如式(27)所示。

$$r = \begin{cases} 200, & \text{if 无人机成功捕捉动态目标} \\ -200, & \text{if 当无人机发生碰撞} \\ -50, & \text{if 当无人机飞行步数超过最大步长} \end{cases} \quad (27)$$

无人机在训练过程中获得来自环境反馈的奖励包括无人机以及势场作用下无人机之间的合作-安全边界约束、待捕捉

目标对于无人机的吸引、无人机在势场环境下自身动力与势场作用合力夹角。所以无人机在训练过程中获得来自环境的奖励反馈如式(28)所示。

$$R=r+r_{\text{APF}}+r_{c-s}+r_{\text{att}}+r_{\text{away}} \quad (28)$$

其中, R 为总奖励。

4.2 实验结果比较与分析

如图 12 和图 13 所示,网络模型迭代训练 4000 步,通过无人机碰撞率对比,可见 Dueling-DQN 算法在训练进行 2000 步后才能有效降低无人机路径规划过程的碰撞概率。而 APF-Dueling-DQN 的多无人机捕捉动态目标路径规划算法在训练进行 500 步后可明显有效降低无人机路径规划过程中的碰撞概率。

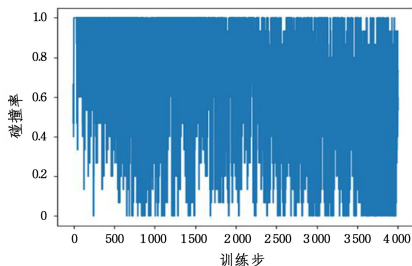


图 12 Dueling-DQN 碰撞概率

Fig. 12 Dueling-DQN collision probability

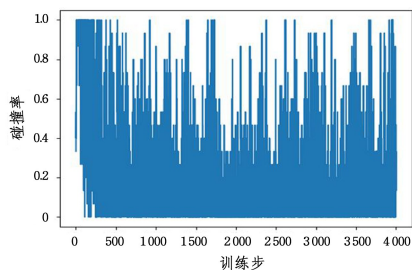


图 13 APF-Dueling-DQN 碰撞概率

Fig. 13 APF-Dueling-DQN collision probability

路径长度如表 4 所列。

表 4 路径长度
Table 4 Path length

路径	Dueling-DQN	APF-Dueling-DQN
a	98.5	87.4
b	119	107.5

使用单个无人机捕捉固定目标路径规划来对比 Dueling-DQN 算法和提出的 APF-Dueling-DQN 算法的性能,如图 14 所示。

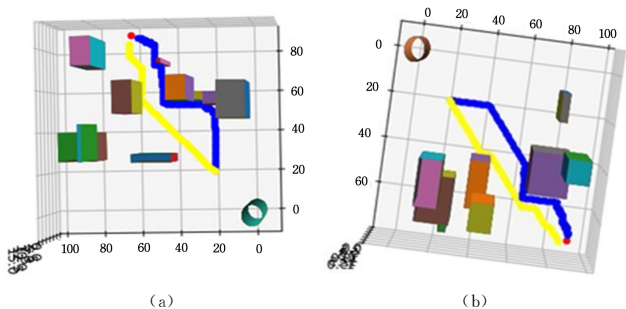


图 14 静态目标路径规划对比

Fig. 14 Comparison of static target path planning

其中蓝色路径为 Dueling-DQN 算法,黄色路径为提出的 APF-Dueling-DQN 算法,从图 14 中可以看出,改进后的算法的路径更加平滑;结合表 4 可知,由于奖励 r_{APF} 的作用,无人机与待捕捉目标的关联性更强,路径也相对更短。

如图 15 和图 16 所示,使用两个无人机捕捉动态目标,其中图 15 为俯视图,图 16 为正视图,蓝色和黄色点迹是无人机飞行路径,红色点迹是动态目标点迹。当无人机进入动态目标侦测范围时,目标开始逃逸,直到无人机捕捉到动态目标时触发终止条件。

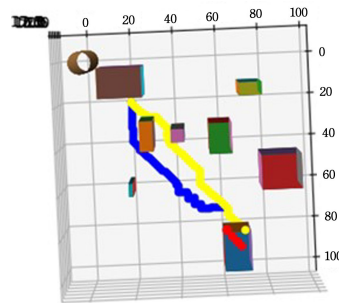


图 15 动态目标路径规划俯视图(电子版为彩图)

Fig. 15 Top view of dynamic target path planning

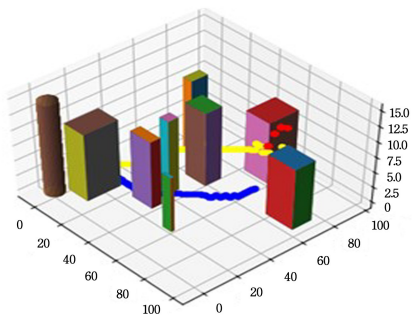


图 16 动态目标路径规划正视图(电子版为彩图)

Fig. 16 Dynamic target path planning front view

结束语 为了解决多无人机在三维未知环境下捕捉动态目标自主路径规划问题,提出了一种基于改进 Dueling-DQN 的多无人机捕捉动态目标路径规划算法。将人工势场法的思想融入多无人机合作捕捉动态目标的训练奖励函数中,不仅解决了传统人工势场法在复杂环境中表现不佳,易陷入局部最优的问题,而且解决了多无人机合作和无人机复杂环境避障问题。此外,设计了一种多无人机与动态目标的捕捉逃逸策略。与 Dueling-DQN 算法相比,APF-Dueling-DQN 算法有效降低了无人机航迹规划任务过程中发生碰撞的概率,缩短了捕捉动态目标所需规划路径长度。为了实现多无人机对动态目标的捕捉,无人机采用分布式控制方式,虽然一定程度上能够提高系统的鲁棒性和扩展性,但是分布式控制方式需要解决无人机协调问题,当执行任务的无人机数量增多后,可能会造成信息同步和决策一致性难题。尽管如此,所提网络模型仍然取得了显著的效果,在未来的研究中将探索更有效的无人机协调机制,并考虑将提出的基于改进 Dueling-DQN 的多无人机路径规划算法和狼群算法^[17]相结合,无人机个体之间通过模拟狼群的协作行为来提高整个无人机群的搜索效率,最后通过严格的实验设计来验证其有效性。

参 考 文 献

- [1] LI J Q, SUN T, HUANG X P, et al. A Memetic Path Planning Algorithm for Unmanned Air/Ground Vehicle Cooperative Detection Systems[J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(4): 2724-2737.
- [2] LI D C, YIN W P, WONG W E, et al. Quality-Oriented Hybrid Path Planning Based on A* and Q-Learning for Unmanned Aerial Vehicle[J]. IEEE Access, 2022, 10: 7664-7674.
- [3] CHEN J, YE F, LI Y B. Travelling salesman problem for UAV path planning with two parallel optimization algorithms[C]// Progress in Electromagnetics Research Symposium. Singapore: IEEE, 2017: 832-837.
- [4] XIONG C Y, XIN B, GUO M, et al. Multi-UAV 3D Path Planning in Simultaneous Attack[C]// IEEE International Conference on Control and Automation ICCA. Singapore: IEEE, 2020: 500-505.
- [5] THEILE M, BAYERLEIN H, NAI R, et al. UAV Path Planning using Global and Local Map Information with Deep Reinforcement Learning[C]// 2021 20th International Conference on Advanced Robotics (ICAR). New York: IEEE, 2021: 539-546.
- [6] GU Z L, LIU Y, SUN W B, et al. UAV Dynamic Route Planning Algorithm Based on RRT[J]. Computer Science, 2023, 50(S1): 65-69.
- [7] LI Y Y, CHEN W Y, FU B, et al. Research on Dynamic Target Search for Multi-UAV Based on Cooperative Coevolution Motion-Encoded Particle Swarm Optimization[J]. Applied Sciences-basel, 2024, 14(4): 1326.
- [8] ZHANG X Y, YUE W. Elite Dung Beetle Optimization Algorithm for Multi-UAV Cooperative Search in Mountainous Environments[J]. Journal of Bionic Engineering, 2024, 21(4): 1677-1694.
- [9] TANG J N, PENG Z X, LI M S, et al. Research on UAV path planning based on improved A-star algorithm [J]. Electronic Measurement, 2019, 46(8): 99-104.
- [10] XIANG L H, WANG F Y, XU W J, et al. Dynamic UAV Swarm Collaboration for Multi-Targets Tracking Under Malicious Jamming: Joint Power, Path and Target Association Optimization [J]. IEEE Transactions on Vehicular Technology, 2024, 73(4): 5410-5425.
- [11] YILDIZ B, ASLAN M F, DURDU A, et al. Consensus-based virtual leader tracking swarm algorithm with GDRRT* -PSO for path-planning of multiple-UAVs [J]. Swarm and Evolutionary Computation, 2024, 88: 101612.
- [12] ZHANG W, SONG K, RONG X. Coarse-to-fine UAV target tracking with deep reinforcement learning [J]. IEEE Transactions on Automation Science and Engineering, 2019, 16(4): 1522-1530.
- [13] ZHOU B, GUO Y, LI N, et al. UAV path planning based on guided reinforcement Q learning [J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(9): 506-513.
- [14] KOWNACKI C. Artificial Potential Field Based Trajectory Tracking for Quadcopter UAV Moving Targets [J]. Sensors, 2024, 24(4): 1343.
- [15] LI Y B, ZHANG Z T, SUN Q, et al. An improved ant colony algorithm for multiple unmanned aerial vehicles route planning [J]. Journal of the Franklin Institute, 2024, 361(15): 107060.
- [16] WANG X, GURSOY M C, ERPEK T, et al. Learning-based UAV path planning for data collection with integrated collision avoidance [J]. IEEE Internet of Things Journal, 2022, 9(17): 16663-16676.
- [17] ZHANG J, LU Y, CHE L, et al. Moving-distance-minimized PSO for mobile robot swarm [J]. IEEE Transactions on Cybernetics, 2021, 52(9): 9871-9881.



FU Wenhao, born in 2000, postgraduate. His main research interests include UAV task assignment and material delivery.



ZHANG Chun, born in 1985, Ph.D, associate professor. Her main research interests include swarm intelligence and Internet of Things.

(责任编辑:柯颖)