

基于深度强化学习的多无人机空战机动策略研究



雷毅飞¹,王露禾¹,贺泊茗¹,胡劲文¹,徐钊¹,吕明伟²,徐港²

1.西北工业大学,陕西 西安 710129

2.航空工业沈阳飞机设计研究所,辽宁 沈阳 110034

摘要:面对敌方空中力量的来袭,具有自主协同、灵活机动能力的无人机是参与空中作战的重要力量。面向多无人机协同高制胜率的对抗作战任务需求,并根据空战目标数量划分,重点对多无人机对单目标协同空战机动策略和多无人机对多目标协同空战机动策略展开研究。本文主要分析了空战过程中的关键战场要素,基于多机机动特性,建立无人机运动模型。根据无人机火控特点,分析无人机状态变化规律,建立无人机对敌攻击模型和动态对抗模型;针对多无人机对单目标自主协同空战问题,提出基于专家规则和强化学习相结合的多机自主机动策略。仿真结果表明,所提算法可以完成态势实时变化的多机对单目标空战任务。在作战双方数量相同的前提下,若敌方不具备智能机动行为,我方制胜率为 100%;即使双方采用相同的策略,如果我方数量多于敌方数量,我方仍有大的胜率,体现了协同策略的有效性。

关键词:空战策略;强化学习;自主机动;多机协同;态势评估

中图分类号:V32

文献标识码:A

DOI: 10.19452/j.issn1007-5453.2024.03.013

随着作战方式的进步,作战规模也在扩大,导致人力资源成本不断增加。减少人员伤亡是未来军事行动的关键目标之一,而“非接触”作战将成为实现该目标的主要手段。无人机(UAV)具有成本低、机动性高、隐蔽性强和不需要飞行员实时控制等优势。这些优势使得无人机广泛应用于各个领域,并越来越广泛地代替有人驾驶飞机用于执行侦察^[1]、监控、目标打击^[2-3]等军事任务,是“非接触”作战装备的典型代表^[4-5]。因为单架无人机执行任务和作战能力存在局限性,所以无人机的集群化和智能化作战成为近年来的研究热点。

人工智能(AI)技术在军事领域被广泛应用,在感知、理解、推理、判断等环节,敌我双方都将投入大量资源进行竞争。使用人机混合智能的优势可以使思考速度更快、预判更精确,并有效争夺作战的主动权。

世界上以美、俄为首的军事强国认为,未来战争将更智能化,并且国家间的军备竞赛将变成智能化军备竞赛。人工智能技术是一项具有颠覆性影响的技术,能够改变作战

规则^[6]。这些国家提前策划并推出了各种政策和研究项目,以将人工智能技术的最新发展用于军事领域。这些努力旨在为未来战争模式中的侦察、教育、训练、威胁评估、水雷战、情报分析和指挥控制等方面提供引领性规划,从而赢得智能化战争主动权^[7]。美国通过海军研究实验室(NRL)和国防预研局(DARPA)等机构致力于研究多无人机集群的最新技术,意在利用这项技术推进军事发展并确保国防实力领先于全球。俄罗斯正在大量投入资源,以确保与美国在人工智能相关的军事应用领域竞争的平衡,特别是在智能无人机技术方面。在此背景下,俄罗斯积极开展活动,以提高其军事人工智能水平,特别是结合无人机、导弹等武器系统,缩小差距并增强自身竞争力。我国无人机技术虽然起步较晚,但目前发展迅速,尤其是在智能多无人机技术方面取得了显著进展。沈林成团队是我国国内较早开展多无人机自主控制的研究团队,该团队认为多无人机自主控制主要有两种方法,分别为自上而下的分层递阶方法及自下至

收稿日期: 2023-06-09; 退修日期: 2023-11-23; 录用日期: 2023-12-26

基金项目: 国家自然科学基金(61803309); 航空科学基金(2019ZA053008, 20185553034); 陕西省重点研发计划项目(2020ZDLGY06-02); 中国博士后科学基金(2018M633574)

引用格式: Lei Yifei, Wang Luhe, He Bomng, et al. Research on multi-UAV air combat maneuver strategy based on deep reinforcement learning[J]. Aeronautical Science & Technology, 2024, 35(03): 111-118. 雷毅飞, 王露禾, 贺泊茗, 等. 基于深度强化学习的多无人机空战机动策略研究[J]. 航空科学技术, 2024, 35(03): 111-118.

上的分布式自组织方法。一种主要研究集中式规划算法解决无人机协同控制问题,另一种则研究个体对环境的动态反应和行为规则的方法来进行协调。对于分层递阶自主控制方法,该团队研究无人机自主控制的理论和方法,系统地覆盖了各个阶段,包括目标状态估计、协同任务分配、协同航迹规划和编队轨迹优化等,并研究了该技术在实际中的典型应用。符小卫及其团队^[8-9]使用威胁联网数学模型,开发出了多个多无人机控制方案,如无人机突防轨迹在线规划、无人机集群协同欺骗干扰控制等。黄长强等^[10-13]全面地研究了多无人机协同技术,遵循感知、评价、决策的过程,并重点研究了无人机集群的协同控制,其中包含多类型目标的搜索和决策、多目标任务分配、有人—无人战术协同,以及协同攻击技术等内容。此外,还有其他大量研究人员对多无人机自主控制方法开展了研究,如王玥总结了微型无人机(MAV)协同控制策略中常用的优化方法,并重点研究了与之有关的规划方法以及评价方法。吴坤等^[14]提出了一种基于混沌灰狼优化和几何规划的离线航路规划方法,用于协调多个无人机在饱和攻击任务中的航线规划。陈宇恒等^[15]提出基于改进贪心算法的无人机集群任务分配算法,优化飞行时间和距离,验证了其在任务场景上的有效性。中国电子科技集团有限公司(CETC)在2016年和2017年分别完成了67架和119架固定翼无人机集群飞行试验,打破了无人机集群飞行数量的纪录。试验成功演示了无人机集群智能起飞、集群自主飞行、感知与规避障碍等技术。

综上所述,基于多无人机作战环境复杂、决策变量维度高、任务目标多变、建模构造困难等原因,人工控制方法难以满足作战的实时性要求。针对以上问题,本文分析空战过程中的关键战场要素,基于多机机动特性,建立无人机运动模型。根据无人机火控特点,分析无人机状态变化规律,建立无人机对敌攻击模型和动态对抗模型;针对多无人机对单目标自主协同空战问题,提出基于专家规则和强化学习相结合的多机自主机动策略。

1 多无人机协同空战数学建模

1.1 无人机运动学模型建立

1.1.1 常用坐标系及其转换关系

为了设计无人机的控制律并对作战过程进行数值仿真和试验验证,需要建立运动方程来描述作战过程中无人机与敌机的相对运动。为此,下面给出本文用到的坐标系以及坐标系之间的转换关系。

(1)地面坐标系($O_e x_e y_e z_e$)

地面坐标系与地球表面固连,坐标系原点 O_e 通常选为某架无人机起飞时质心在地面上的投影点, $O_e x_e$ 轴向东, $O_e y_e$ 轴向北, $O_e z_e$ 按右手定则确定。

(2)惯性坐标系($O_o x_o y_o z_o$)

坐标原点 O_o 位于无人机质心上。 $O_o x_o$ 、 $O_o y_o$ 和 $O_o z_o$ 分别与 $O_e x_e$ 、 $O_e y_e$ 、 $O_e z_e$ 平行。

(3)机体坐标系($O_1 x_1 y_1 z_1$)

原点 O_1 取在无人机质心上, $O_1 x_1$ 轴与无人机的纵轴重合,指向无人机的头部为正; $O_1 y_1$ 轴位于机体纵向对称面内,垂直于 $O_1 x_1$,向上为正; $O_1 z_1$ 轴方向按右手定则确定。坐标系之间的转换可参见参考文献[16],这里不再赘述。

1.1.2 无人机运动建模

在研究多无人机的协同问题时,主要考虑的是无人机之间的行为,机间的相对位置和速度方向比较重要,所以将无人机视为质点是合理的。每架无人机都有相同的运动特征,所以无人机在 $O_e x_e y_e z_e$ 坐标系下的航路点运动模型如式(1)所示,其示意图如图1所示。

$$P_{U_i,t+1} = P_{U_i,t} + v_{U_i,t} \quad (1)$$

式中, $i \in \{1, 2, \dots, N\}$, N 表示我方无人机的数量, U_i 代表编号为 i 的无人机, t 是固定采样周期 T 下的离散时间指标, $P_{U_i,t}$ 和 $P_{U_i,t+1}$ 表示第 i 架无人机 U_i 在 t 时刻和 $t+1$ 时刻的位置, $v_{U_i,t}$ 代表无人机在 t 时刻的速度。同样地,第 j 个敌机 E_j 的位置和速度用 $P_{E_j,t}$ 和 $v_{E_j,t}$ 表示。在同一坐标系中,无人机的运动学模型可表示为

$$\begin{cases} \dot{x}_{i,t} = v_{i,x,t} \\ \dot{y}_{i,t} = v_{i,y,t} \\ \dot{z}_{i,t} = v_{i,z,t} \end{cases} \quad (2)$$

式中, $x_{i,t}$ 、 $y_{i,t}$ 、 $z_{i,t}$ 表示 t 时刻无人机 U_i 在参考坐标系中的三维坐标,有 $P_{i,t} = [x_{i,t}, y_{i,t}, z_{i,t}]$ 。 $\dot{x}$ 、 $\dot{y}$ 、 $\dot{z}$ 表示 $P_{i,t}$ 在三个坐标轴方向

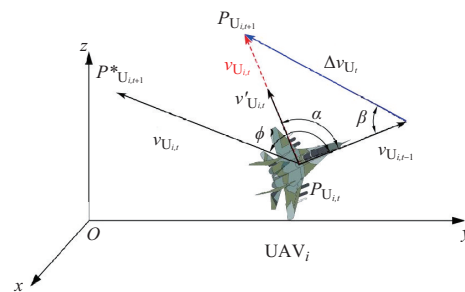


图1 无人机航路点模型

Fig.1 Waypoint model of UAV

的变化率。相应的动力学模型表示为

$$\begin{cases} \dot{v}_{i,x} = a_{i,x} \\ \dot{v}_{i,y} = a_{i,y} \\ \dot{v}_{i,z} = a_{i,z} \end{cases} \quad (3)$$

式中, $\dot{v}_{i,x}$, $\dot{v}_{i,y}$, $\dot{v}_{i,z}$ 分别为无人机三个坐标方向上的加速度。

1.2 无人机动态对抗模型

参考文献[17]中的攻击模型,无人机的机载传感器的探测范围是一个半径为 R_D 的球体,有效攻击范围是一个锥球 Ω ,其中锥球的锥顶与无人机的质心位置 $P_{U_{i,t}}$ 重合,轴方向为 $v_{U_{i,t}}$,角度为 φ_m ,被半径为 R_a 的球截断,如图2所示。其中 R_w 表示武器的最大攻击距离,在实际空战中,定义一个基于人工经验设定的最佳射击距离阈值 D_{attack} ,并且有 $D_{\text{attack}} \leq R_a$ 。

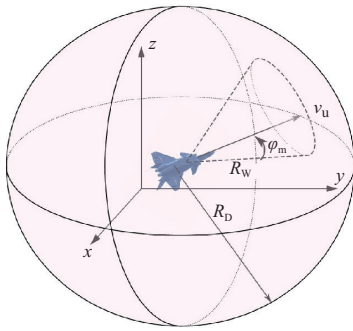


图2 无人机攻击模型

Fig.2 UAV attack model

当敌机 $P_{E_{j,t}}$ 出现在我方武器最佳攻击范围内时,即满足

$$\|P_{E_{j,t}} - P_{U_{i,t}}\| \leq R_a \quad (4)$$

$$\langle v_{U_{i,t}}, P_{E_{j,t}} - P_{U_{i,t}} \rangle \leq \varphi_m \quad (5)$$

无人机可以通过机载武器(如航火、航弹、航炮)对敌机发起进攻。其中, $\|P_{E_{j,t}} - P_{U_{i,t}}\|$ 表示敌我距离, $\langle v_{U_{i,t}}, P_{E_{j,t}} - P_{U_{i,t}} \rangle$ 代表敌机与火控范围轴心的夹角。同样,本文可以定义敌机的探测范围和锥形攻击范围。无人机 $U_{i,t}$ 对敌机 $E_{j,t}$ 的毁伤概率 c_{ij} 与机群的作战环境、作战态势以及机载火控的理想毁伤概率有关,可以用式(6)描述

$$c_{ij} = \beta_L \eta_{ij}(t) K_{ij} \quad (6)$$

式中, $0 < \beta_L < 1$ 为环境影响因子,表示作战环境对火控杀伤率的影响; η_{ij} 为我方无人机 i 对敌方无人机 j 的态势优势值,与无人机间的距离和角度有关,将在后文详细介绍; K_{ij} 为火控的理想杀伤概率。

2 多无人机对单目标协同空战机动策略研究

2.1 单机对单目标自主机动策略

2.1.1 状态空间与动作空间

根据1.2节的对抗模型,本文使用速度 v 、位置 P 、角度 φ 三个变量描述无人机状态。在空战过程中敌我双方的态势与相对状态有关,与绝对信息关系不大,因此本文将绝对信息转换为相对信息,使用 $v_{U,x} - v_{T,x}$, $v_{U,y} - v_{T,y}$, $v_{U,z} - v_{T,z}$, $x_U - x_T$, $y_U - y_T$, $z_U - z_T$ 和 $\varphi_U - \varphi_T$ 描述无人机的相对状态。为了提高网络学习的效率,本文将每个相对状态归一化至同一个范围

$$s = O(v, P, \varphi) \quad (7)$$

式中, O 为归一化映射函数,结果见表1。

表1 无人机自主攻击状态信息归一化

Table 1 Normalization of UAV autonomous attack state information

状态	定义
s_1	$\frac{v_{U,x} - v_{T,x}}{v_{\max} - v_{\min}} a_1 - b_1$
s_2	$\frac{v_{U,y} - v_{T,y}}{v_{\max} - v_{\min}} a_1 - b_1$
s_3	$\frac{v_{U,z} - v_{T,z}}{v_{\max} - v_{\min}} a_1 - b_1$
s_4	$\frac{x_U - x_T}{x_{\max} - x_{\min}} a_1 - b_1$
s_5	$\frac{y_U - y_T}{y_{\max} - y_{\min}} a_1 - b_1$
s_6	$\frac{z_U - z_T}{z_{\max} - z_{\min}} a_1 - b_1$
s_7	$\frac{\varphi_U - \varphi_T}{2\pi} a_1 - b_1$

表1中, $v_{\max}$ 和 $v_{\min}$ 分别为无人机的最大和最小速度。 $x_{\max}$, $y_{\max}$ 和 $z_{\max}$ 表示无人机作战的最大边界; $x_{\min}$, $y_{\min}$ 和 $z_{\min}$ 表示无人机作战的最小边界。 a_1 和 b_1 是两个正数,满足 $a_1 = 2b_1$ 。因此,强化学习的状态空间可以被定义为 $s = [s_1, s_2, s_3, s_4, s_5, s_6, s_7]$ 。

动作空间是无人机的机动库,如图3所示。根据无人机的运动控制模型,无人机的动作输出 $a \in A$ 为无人机在空间三个维度上的加速度

$$a = [a_x, a_y, a_z] \quad (8)$$

因此,无人机的动作空间可表示为

$$A = [a_x, a_y, a_z] \quad (9)$$

根据DDPG的网络架构,将 a 与 s 为 Critic 网络的输入,输出 O_c 是对 a 动作结果的评价。

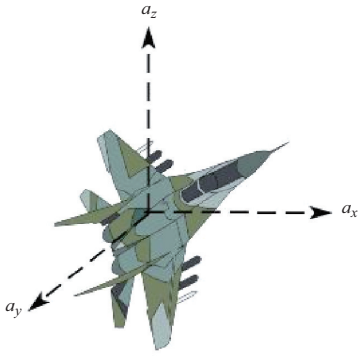


图3 无人机动作空间

Fig.3 Movement space of UAV

$$O_c = [Q_{a_x}, Q_{a_y}, Q_{a_z}] \quad (10)$$

2.1.2 奖励函数

奖励函数定义为

$$r_t = R_t + R_{\eta_t} \quad (11)$$

式中, r_t 由两部分组成, 第一部分 R_t 是对空战最终结果的评价。如果敌我距离满足 $D_{\min} < D_{U_i T_i} \leq D_{\text{attack}}$, 角度同时满足 $\varphi_{U_i} \leq \varphi_m$ 和 $\varphi_{T_i} \geq \pi - \varphi_m$, 那么认为无人机拥有最佳对敌射击条件, “我方胜利”条件成立。如果敌我距离满足 $D_{\min} < D_{U_i T_i} < D_{\text{attack}}$, 角度同时满足 $\varphi_{U_i} \leq \varphi_m$ 和 $\varphi_{T_i} \geq \pi - \varphi_m$, 那么认为“敌方胜利”条件成立。假设“我方胜利”的事件为 W , “敌方胜利”的事件为 L , 则奖励函数中的 R_t 一项记作

$$R_t = \begin{cases} C_1, & \text{If } W \\ -C_1, & \text{Else if } L \\ 0, & \text{else} \end{cases} \quad (12)$$

式中, C_1 为大于0的常数。 r_t 由式(11)可知是奖励函数, 其另一部分来自 η_t , 用作机动决策的即时评估, 记作 R_{η_t} 。并且有

$$R_{\eta_t} = \eta_t - \eta_{t-1} = w_1(\eta_{A_t} - \eta_{A_{t-1}}) + w_2(\eta_{D_t} - \eta_{D_{t-1}}) \quad (13)$$

式(13)表示无人机在空战过程中态势的变化情况。如果 $R_{\eta_t} > 0$, 说明 t 时刻的态势比 $t-1$ 时刻的优。无人机从 s_{t-1} 到 s_t 采取的机动策略 a_{t-1} 是合理的, 给予正值奖励。相反, 如果 $R_{\eta_t} < 0$, 给予负值惩罚。所以

$$r_t = R_t + R_{\eta_t} = R_t + w_1(\eta_{A_t} - \eta_{A_{t-1}}) + w_2(\eta_{D_t} - \eta_{D_{t-1}}) \quad (14)$$

2.2 引入专家规则的多机对单目标自主机动策略

在多机协同空战中, 无人机不仅要考虑敌方状态, 还要考虑友机的状态, 做到兼顾机间通信、机间防撞和目标打击。因此, 为进一步实现多机协同空战, 本文引入专家知

识, 设计了多种协同规则, 简称专家规则, 在保证多机正常飞行的前提下, 使算法适配多机对单目标对抗场景, 如图4所示。

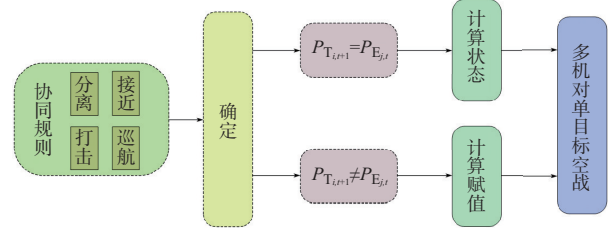


图4 基于专家知识的规则设计

Fig.4 Rule design based on expert knowledge

专家规则设计如下:

(1) 分离规则

使用 $P_{E_{i,t}}$ 表示 t 时刻敌方目标的位置, 如果两架无人机之间的距离小于预先定义的最小安全距离 $d_{\min}$, 则可能存在碰撞的潜在威胁。也就是说, 如果 $P_{U_{i,t}} - P_{U_{j,t}} < d_{\min}$, 则需要修改无人机下一时刻的航行目标点 $P_{T_{i,t+1}}$ 为

$$P_{T_{i,t+1}} = \lambda_1(P_{E_{i,t}} - P_{U_{i,t}}) + \lambda_2(P_{U_{i,t}} - P_{U_{j,t}}) \quad (15)$$

保证无人机在接近敌方目标的同时, 增加机间距离避免碰撞。

(2) 巡航规则

根据1.2节的无人机攻击模型, 当 $D_{U_{i,t} E_{j,t}} > R_D$, 无人机无法探测到任何一架敌机, 则无人机下一时刻的航行目标位置为

$$P_{T_{i,t}} = P_B \quad (16)$$

(3) 靠近/远离规则

若在 U_i 探测范围内存在 M 架敌机, $R_W < D_{U_{i,t} E_{j,t}} \leq R_D, j = 1, 2, \dots, M$, 则使用

$$P_{T_{i,t}} = \sum_{j=1}^M \omega_j (P_{E_{j,t}} - P_{U_{i,t}}) \quad (17)$$

确定下一时刻的飞行目标位置。其中

$$\omega_j = \frac{\kappa_{ij}}{\sum_{m=1}^M \kappa_{im}}, \sum_j \omega_j = 1 \quad (18)$$

(4) 攻击规则

若在 U_i 的攻击范围内存 M 架敌机, 即敌我距离和角度同时满足 $|D_{\min} < D_{U_{i,t} E_{j,t}} < D_{\max}, \varphi_{U_{i,t}} \leq \varphi_m, j = 1, 2, \dots, M$, 则无人机下一时刻的目标点定义为

$$P_{T_{i,t}} = P_{T_{j,t}}, \kappa_{ij} = \max(\kappa_{ik}), k = 1, 2, \dots, M \quad (19)$$

由式(19)可知,当敌机在无人机的攻击半径内,且满足角度攻击条件时,无人机对目标可以做出攻击,并且选择角度优势最大的敌机作为攻击目标。

3 仿真分析

3.1 参数设置

本文算法设计过程中所涉及的参数及数值见表2。在空战仿真过程中,决策周期 T 设置为1s,一集包含600个决策步数,如果满足任意一方无人机全部战损,则该轮训练结束。

表2 仿真试验中算法参数设置

Table 2 Algorithm parameter setting table in simulation experiment

参数	值	参数	值
$D_{\max}$	1000	$V_{\min}$	1
$D_{\min}$	200	$Z_{\max}$	12000
D_{attack}	500	$Z_{\min}$	1000
C_1	10	D_{th}	10000
w_2	20	$c_{\bar{p}}$	0.5
$y_{\max}$	300	$y_{\min}$	-200
b	1000	W_1	1
γ	200	W_2	12000
$V_{\max}$	500	a	1000
Φ_m	10	学习速度	10000
$x_{\max}$	20	$x_{\min}$	0.5
$z_{\max}$	300	$z_{\min}$	-200

3.2 结果分析

(1)神经网络训练结果

图5中从上到下的三种曲线分别展示的是 w_1 为分段函数,定值 $w_1=20$,正比于距离的连续函数三种情况下训练过程中平均奖励的变化趋势。奖励越大说明空战过程无人机的整体决策越优。该图的横轴表示训练的步数,纵轴表示的是平均奖励变化。当网络收敛后,基于分段函数的强化学习算法的奖励均值最高,说明学习效果最好。基于连续函数奖励反馈的强化学习算法的平均奖励为负值,说明无人机没有学到自主对抗策略。

(2)多机对单目标空战仿真验证

如图6(a)所示,我方无人机数量为4,多机自主机动到达终点,该过程中每架无人机与目标点的距离以及每架无人机与最邻近无人机直接的距离如图6(b)所示。无人机的位置和速度随机初始化,可以看出多机在飞向目标点的过程中也在相互靠近,因为有分离规则的约束,机间距离大于阈值5,

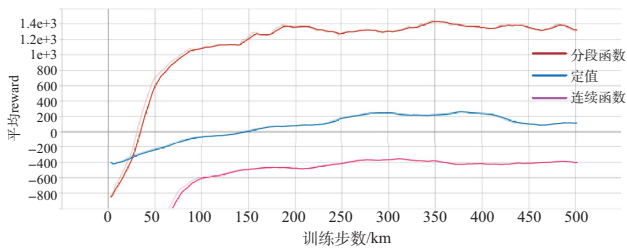
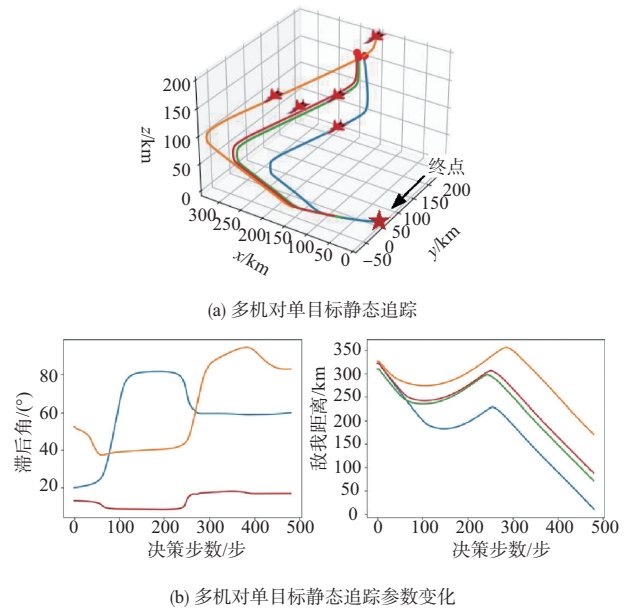


图5 神经网络训练过程中平均奖励变化趋势

Fig.5 Changing trend of average reward during neural network training



(a) 多机对单目标静态追踪

(b) 多机对单目标静态追踪参数变化

图6 多机对单目标静态追踪结果展示

Fig.6 Display of static tracking results of multi-UAVs on single target

机间防撞成立。因为没有进行时间约束,故第一架无人机到达终点后,最后一架无人机距终点200个距离单位。

同理,如图7所示,目标分别做直线运动和曲线运动,我方无人机随机初始化位置和速度方向。直线追踪场景中,我方每架无人机首先通过传感器获知友机的位置和速度信息,然后判断是否需要进行“分离”,可以看出无人机初始位置在符合巡航条件的情况下不断探测敌机位置并向其靠近,敌我距离不断缩小,机间最小距离如图8所示,由于有“分离”策略的约束,机间距离大于阈值5,保证无人机不会发生碰撞;每架无人机的滞后角如图8所示,可以看出无人机是从敌机后方对目标发动进攻的,最终无人机以 16° 的角度偏差将目标击毁。同理,分析曲线追踪的结果如图9所示,多机从敌机前侧进行拦截,滞后角为钝角,最终无人机以 175° 的角度

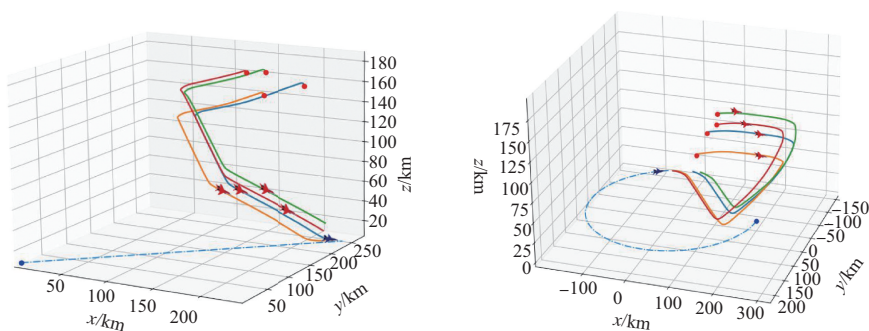


图7 多机对单目标简单对抗

Fig.7 Simple confrontation between multi-UAVs and single target

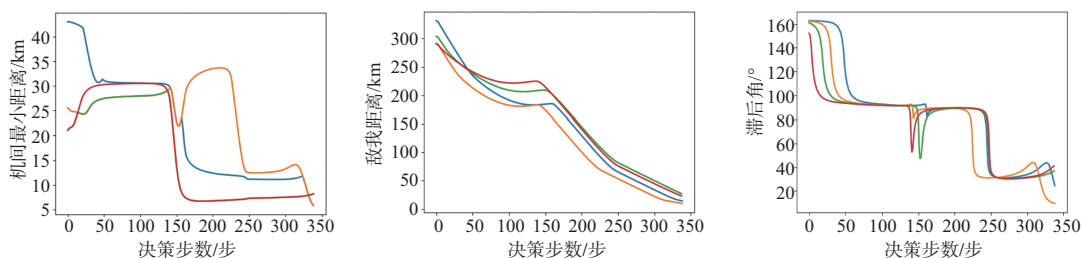


图8 多机对单目标直线追踪参数变化

Fig.8 Parameter changes of multi-UAVs linear tracking for single target

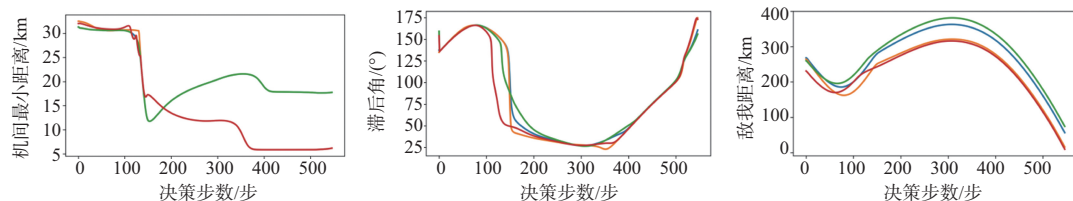


图9 多机对单目标曲线追踪参数变化

Fig. 9 Parameter change of multi-UAVs tracking single target curve

偏差将目标击毁,机间最小距离也均大于阈值。该组试验证明,专家规则的设计是可行的。但是,因为机间没有时间约束,所以无人机先后到达目标点的差距较大。

图10展示的是两架无人机对单目标自由空战的结果,我方胜利。作战双方都具备智能策略,但是我方无人机数量多于敌方。作战开始,双方距离较远,在无人机探测范围内无法感知对方的状态,双方根据巡航策略都向着对方基地靠近。然后在120s左右相遇,进行空中缠斗。空战过程中,我方 UAV1 在第 320s 被敌方击落,随后 UAV0 在第 330s 战胜敌方。图11给出了100次仿真试验的试验结果。

4 结束语

本文围绕多无人机对单目标协同机动策略和多无人机对多目标协同机动策略展开研究,主要结论如下:

(1) 针对多机空战对抗场景,建立无人机的运动控制模

型,使用加速度作为无人机的控制律。根据无人机空战任务需求,建立多机对单目标的空战任务模型。

(2) 针对多无人机空战对抗任务,提出基于专家规则和深度强化学习相结合的无人机机动策略,解决多无人机智能自主协同空战难题。设计深度强化学习算法,在奖励函数设计方面,创新性地使用基于比例因子自适应调整的奖励函数设计方法,解决了传统强化学习方法奖励稀疏且延迟带来的弊端。最后结合专家知识,设计多机协同作战规则,实现强化学习自主机动算法在大规模高动态变化场景下的应用。

AST

参考文献

- [1] Sun Long, Chang Pei, Jiang Huanhuan, et al. Research on intelligent target recognition technology for integrated reconnaissance/strike UAV[C]. Global Intelligence Industry

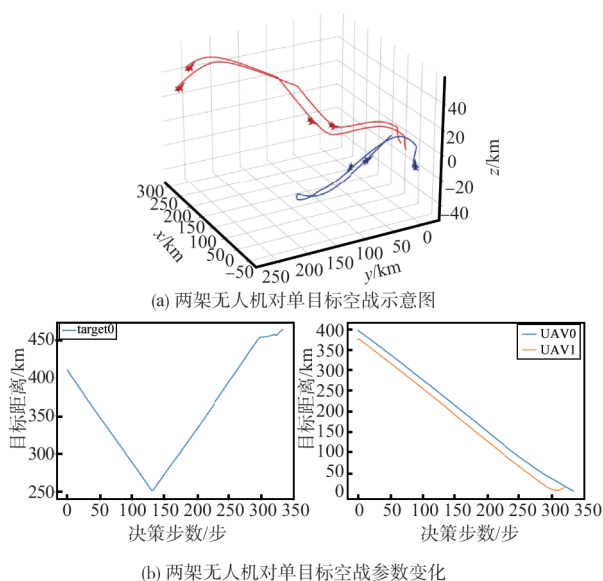


图10 两架无人机对单目标空战结果

Fig.10 Results of aerial combat between two UAVs against a single target

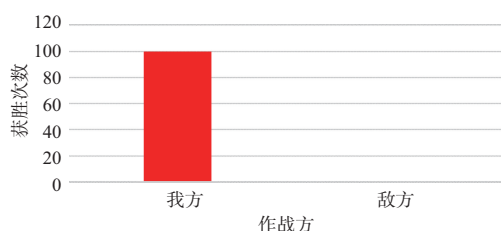


图11 两架无人机对单目标空战结果统计

Fig.11 Statistics of two UAVs' air combat results against a single target

Conference (GIIC 2018), 2018: 282-288.

- [2] Nolin P C. Unmanned aerial vehicles: Opportunities and challenges for the alliance[R]. NATO Parliamentary Assembly, 2012.
- [3] 罗德林, 张海洋, 谢荣增, 等. 基于多 agent 系统的大规模无人机集群对抗[J]. 控制理论与应用, 2015, 32(11): 1498-1504. Luo Delin, Zhang Haiyang, Xie Rongzeng, et al. Large-scale unmanned aerial vehicle swarm confrontation based on multi-agent system[J] Control Theory & Applications, 2015, 32(11): 1498-1504.(in Chinese)
- [4] Li Yue, Qiu Xiaohui, Liu Xiaodong, et al. Deep reinforcement learning and its application in autonomous fitting optimization for attack areas of UAVs[J]. Journal of Systems Engineering and Electronics, 2020, 31(4): 734-742.
- [5] Hu Dongyuan, Yang Renhong, Zuo Jialiang, et al. Application

of deep reinforcement learning in maneuver planning of beyond-visual-range air combat[J]. IEEE Access, 2021, 9: 32282-32297.

- [6] Otto R P. Small unmanned aircraft systems (SUAS) flight plan: 2016-2036. Bridging the gap between tactical and strategic[R]. Political Science, 2016.
- [7] Theraulaz G, Bonabeau E. A brief history of stigmergy [J]. Artificial Life, 1999, 5(2): 97-116.
- [8] 符小卫, 高晓光. 一种无人机路径规划算法研究[J]. 系统仿真学报, 2004, 16(1): 20-21. Fu Xiaowei, Gao Xiaoguang. A study on unmanned aerial vehicle path planning algorithm[J]. Journal of System Simulation, 2004, 16(1): 20-21.(in Chinese)
- [9] Xiao Qinkun, Gao Xiaoguang, Fu Xiaowei, et al. New local path replanning algorithm for unmanned combat air vehicle[C]. 2006 6th World Congress on Intelligent Control and Automation, 2006: 4033-4037.
- [10] Wang Yuan, He Lei, Huang Changqiang. Adaptive time-varying formation tracking control of unmanned aerial vehicles with quantized input[J]. ISA Transactions, 2019, 85: 76-83.
- [11] Zhang Hongpeng, Huang Changqiang. Maneuver decision-making of deep learning for UCAV thorough azimuth angles [J]. IEEE Access, 2020, 8: 12976-12987.
- [12] Han Bo, Huang Hanqiao, Lei Lei, et al. An improved IMM algorithm based on ST-SRCKF for maneuvering target tracking [J]. IEEE Access, 2019, 7: 57795-57804.
- [13] Huang hangqiang, Dong Kangsheng, Huang Hanqiao, et al. Autonomous air combat maneuver decision using bayesian inference and moving horizon optimization[J]. Journal of Systems Engineering and Electronics, 2018, 29(1): 86-97.
- [14] 吴坤, 池沛, 王英勋. 基于混沌灰狼优化的多无人机协同航路规划[J]. 航空科学技术, 2022, 33(10): 82-95. Wu Kun, Chi Pei, Wang Yingxun. Multi-UAV cooperative route planning based on chaotic grey wolf optimization[J]. Aeronautical Science & Technology, 2022, 33(10): 82-95. (in Chinese)
- [15] 陈宇恒, 陈进朝, 陈雪聪. 基于改进贪心算法的无人机集群协同任务分配[J]. 航空科学技术, 2022, 33(4): 13-18. Chen Yuheng, Chen Jinzhao, Chen Xuecong. UAV swarm cooperative task assignment based on improved greedy

- algorithm[J]. Aeronautical Science & Technology, 2022,33(4): 13-18.(in Chinese)
- [16] Andersen R A, Snyder L H, Li C S, et al. Coordinate transformations in the representation of spatial information[J]. Current Opinion in Neurobiology, 1993, 3(2): 171-176.
- [17] Fu Qiang, Fan Chengli, Song Yafei, et al. Alpha C2 an intelligent air defense commander independent of human decision-making[J]. IEEE Access, 2020, 8: 87504-87516.

Research on Multi-UAV Air Combat Maneuver Strategy Based on Deep Reinforcement Learning

Lei Yifei¹, Wang Luhe¹, He Boming¹, Hu Jinwen¹, Xu Zhao¹, Lyu Mingwei², Xu Gang²

1. Northwestern Polytechnical University, Xi'an 710129, China

2. AVIC Shenyang Aircraft Design and Research Institute, Shenyang 110034, China

Abstract: In face of the incoming attack of enemy air power, UAVs with autonomous coordination and flexible maneuvering capability are an important force to participate in air combat. Facing the demand of confrontation combat mission with high winning rate of multi-UAV coordination, and based on the number of air combat targets, we focus on the research of multi-UAV to single-target coordinated air combat maneuver strategy and multi-UAV to multi-target coordinated air combat maneuver strategy. This paper mainly analyzes the key battlefield elements in the process of air combat, and establishes the UAV motion model based on the characteristics of multi-machine maneuver. According to the fire control characteristics of UAV, analyze the change rule of UAV state, establish UAV attack model and dynamic confrontation model against the enemy; for the problem of multi-UAV to single-target autonomous coordinated aerial combat, put forward multi-autonomous maneuver strategy based on the combination of expert rules and reinforcement learning. The simulation results show that the proposed algorithm can accomplish the task of multi-aircraft aerial combat against single target with real-time change of situation. Under the premise of the same number of combatants, if the enemy does not have intelligent maneuvering behavior, our victory rate is 100%. Even if both sides use the same strategy, if our number is more than the enemy, we still have a large victory rate. This demonstrates the effectiveness of the coordinated strategy.

Key Words: air combat strategy; reinforcement learning; autonomous mobility; multiple machine collaboration; situation assessment

Received: 2023-06-09; **Revised:** 2023-11-23; **Accepted:** 2023-12-26

Foundation item: National Natural Science Foundation of China(61803309); Aeronautical Science Foundation of China (2019ZA053008, 20185553034); Key Research and Development Program Projects of Shaanxi Province(2020ZDLGY06-02); China Post-doctoral Science Foundation(2018M633574)