文章编号:1674-8190(2022)05-047-12

自主空战连续决策方法

单圣哲[1,2],杨孟超[1],张伟伟[1],高传强[1]
(1.西北工业大学 航空学院,西安 710072)
(2.中国人民解放军93995部队,西安 710306)

摘 要: 未来空战正朝着无人化、自主化方向发展,自主空战决策方法是未来空战的重要支撑手段之一。传统空战决策方法由于维度限制,存在无法处理连续动作与远视决策的问题。基于Actor-Critic方法提出空战连续决策的统一方法架构,依据空战训练经验对状态空间、动作空间、奖励及训练科目进行合理设计,测试多种连续动作空间强化学习算法在高不确定性空战场景下的学习效果并进行可视化验证。结果表明:基于本文提出的方法架构,可以实现连续动作下的远视价值寻优,智能体可以在复杂空战态势下做出最优决策,对随机机动飞行目标有较高的击杀率,且空战机动轨迹具有较高的合理性。

关键词: 自主空战;强化学习;人工智能;深度神经网络
中图分类号: V212.1; TP18; E91 **文献标识码:** A
DOI: 10.16615/j.cnki.1674-8190.2022.05.05 **开放科学(资源服务)标识码(OSID):**

Continuous Decision-making Method for Autonomous Air Combat

SHAN Shengzhe[1,2], YANG Mengchao[1], ZHANG Weiwei[1], GAO Chuanqiang[1]
(1. School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China)
(2. 93995 Unit of the Chinese People's Liberation Army, Xi'an 710306, China)

Abstract: The future air warfare is developing in the unmanned and autonomous direction. The autonomous air warfare decision-making methods are one of the important support methods in future. Due to dimensional limitations, traditional air combat decision-making methods cannot handle continuous action and long-sighted decision-making problems. Based on the Actor-Critic method, a unified architecture for continuous decision-making in air combat is proposed in this paper. Combining air combat training experience, the state space, action space, reward and training subjects are rationally designed, and a variety of continuous action space reinforcement learning algorithms are tested in high uncertainty. The learning effect in the air combat scenario is visually verified. The results show that: based on the method architecture proposed in this paper, long-sighted value optimization under continuous actions can be realized, the agent can make optimal decisions in complex air combat situations, and has a high kill rate against random maneuvering flying targets. And the air combat maneuver trajectory is highly reasonable.

Key words: autonomous air combat; reinforcement learning; artificial intelligence; deep neural network

收稿日期: 2021-11-25; **修回日期:** 2022-01-24
基金项目: 国防科技重点实验室基金(6142219190302)
通信作者: 张伟伟, aeroelastic@nwpu.edu.cn
引用格式: 单圣哲, 杨孟超, 张伟伟, 等. 自主空战连续决策方法[J]. 航空工程进展, 2022, 13(5): 47-58.
SHAN Shengzhe, YANG Mengchao, ZHANG Weiwei, et al. Continuous decision-making method for autonomous air combat [J]. Advances in Aeronautical Science and Engineering, 2022, 13(5): 47-58. (in Chinese)

0 引 言

自主空战(AAC)是指战机依靠机载设备,感知战场态势,基于人工智能在战场中实时选择作战方案和战术动作的机制,其智能化程度决定了机制的优劣[1]。空战过程中,交战飞机需要在复杂的环境中通过连续高强度机动来力争态势,进而消灭敌人保全自己,故决策是自主空战中最为核心的部分[2]。

依据自主空战决策算法的核心内涵不同,可以将现有算法划分为基于数学求解、机器搜索以及数据驱动三大类[3]。

数学求解方法将空战决策视为博弈问题。该类方法通常基于博弈论对空战问题进行简化假设,使用微分对策方法[4]求解Nash均衡,根据假设类型不同可以将空战问题描述为追逃问题(Pursuit-Evasion Game,简称PE)[5-11]、双目标优化问题(Two Target Game)[12-15]和态势函数优化问题[16-18]等。微分对策方法本身属于解析求解方法,其结果具有清晰的数学形式和显式优越性,但由于微分对策方法在数学上具有局限性,尤其是在处理奇异曲面问题上的不完备性,限制了数学求解方法在复杂博弈问题中的应用。

机器搜索方法通常将空战中的可选方案离散化,通过试探输入得出每种机动方案的可能结果,并通过态势函数量化其结果,最终通过一定的搜索机制找出最有利的方案。根据搜索机制的不同该类方法可分为AML(Adaptive Maneuvering Logic)搜索[19-21]、博弈矩阵搜索[22-24]、启发算法搜索[25-27]等。由于空战机动中的控制量为连续变化量,存在选择方案无穷多的"维度爆炸"问题,针对该问题,文献[28-31]使用多种动作库来描述机动方案,但仍存在动作突变或灵活性较差问题。

数据驱动方法的优势在于可以摆脱常规方法对人类知识的高度依赖,主要有神经网络[32-33]、模糊矩阵[34]、强化学习[35-44]等。

强化学习方法,是基于马尔科夫决策过程(Markov Decision Process,简称MDP)中的价值迭代和策略迭代,让智能体与环境交互,利用环境的奖励反馈不断改进策略,以获得最大累计折扣奖励的方法。Liu P等[36]基于DQN(Deep Q-Learning)算法将深度强化学习用于空战决策,解决了连续状态输入的"维度灾难"问题,同时验证了奖励与行为的效用性,但DQN网络无法解决连续动作输出问题;张强等[37]基于强化学习中的Q-Network开展了超视距空战决策的研究,并基于Q-Network的输出量求解Nash均衡,选择敌我的空战动作,但动作的选取仍与真实空战有一定差距;B. Kurniawan等[38]将行动者—批评者(Actor-Critic,简称AC)架构引入空战决策研究,提高了训练效率,同时探索了奖励结构对学习速度的影响,但训练课目设置仍较为简单;Yang Q等[39-40]采用DDPG(Deep Deterministic Policy Gradient)算法解决了DQN算法无法实现空战的连续动作输出问题,提高了控制的精度和平滑性,同时也开展了DQN与Curriculum Learning相结合的研究,提高了决策模型在对抗中的获胜率。

综合考虑自主空战决策算法的研究进展与问题,为建立一种连续动作、强远视性、全动态的实时空战机动决策方法架构以适应于实际空战需求,本文基于AC强化学习架构,使用深度神经网络的非线性表现能力,构建由连续状态空间至连续动作空间的映射,采用线性方法将飞机各个状态机动包线动态归一化,与动作空间对齐,保证决策算法充分调用飞机机动潜力的同时避免动态失速;通过增大空战训练环境的不确定性,增强算法的全泛化能力,避免决策算法仅对特定起始条件有效;通过高不确定度的空战场景设置与训练,说明方法架构的合理性。

1 AC强化学习基本架构

强化学习问题分为连续时间问题和离散时间问题,通常将连续时间问题近似为离散时间问题,这样强化学习问题就可以统一表示为离散时间马尔科夫决策过程。本文基于连续时间的离散化与环境完全可知表述空战轨迹序列:

$$S_0, A_0, R_1, S_1, A_1, R_2, S_2, A_2, R_3, \cdots$$

基于马尔科夫性假设进一步引入MDP的演化概率为

$$p(s', r \mid s, a) = \Pr\left[S_{t+1}=s', R_{t+1}=r \mid S_t=s, A_t=a\right] \tag{1}$$

至此,MDP可由四元组$\langle S, A, R, P_{S,R,S,A}\rangle$来表示,其中$P_{S,R,S,A}$是环境动力的四阶张量形式。

在AC架构下,agent由Actor和Critic两部分构成,分别负责动作生成与策略评估工作,其基本框架如图1所示。

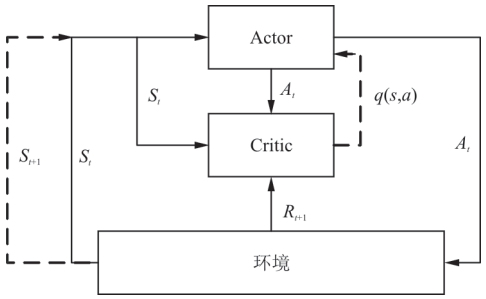


图 1　Actor-Critic 架构

Fig. 1　Framework of Actor-Critic

Actor 网络以当前时刻状态 s_t 为输入，输出当前时刻动作 a_t 的相关量；Critic 网络以当前状态—动作对 (s_t, a_t) 为输入，动作价值 $q(s_t, a_t)$ 为输出，其作用为评价当前状态采取的动作方案的价值。

$$q(s_t, a_t) = E\left[\sum_{\tau=0}^{+\infty}\gamma^{\tau} R_{t+\tau+1} \middle| S_t = s, A_t = a\right] \tag{2}$$

式中：$\gamma \in [0, 1]$ 为累加奖励的折扣因子，代表 agent 的远视程度。

agent 从环境观测得到当前状态样本 s_t，依据 Actor 网络输出的动作相关量选择动作样本 a_t，在环境中执行该动作，进而使环境演化至下一状态样本 s_{t+1}，并返回奖励样本 r_{t+1}。Critic 对 s_t 下行动方案 a_t 进行价值评估，并与奖励 r_{t+1} 进行对比更新自身网络参数，使其评价更接近真实值，本文采用单步时序差分的方式更新，其评估方式为

$$q(S_t, A_t) = R_{t+1} + \gamma q(S_{t+1}, A_{t+1}) \tag{3}$$

根据动作价值评估 $q(s_t, a_t)$，利用随机策略梯度（Policy Gradient，简称 PG）方法更新 Actor 网络参数，使其策略得到优化。

基于上述 Actor-Critic 架构产生了多种算法，其中比较有代表性的有邻近策略优化算法[45]（Proximal Policy Optimization，简称 PPO）、柔性行动者—批评者算法[46]（Soft Actor-Critic，简称 SAC）、深度确定性策略梯度算法[47]（Deep Deterministic Policy Gradient，简称 DDPG）以及双重延迟深度确定性策略梯度算法[48]（Twin Delay Deep Deterministic Policy Gradient，简称 TD3）等。

本文空战环境为连续动作环境，DDPG 算法的 Actor 网络输出为连续动作范围内的值，通过加入噪声 N 确定执行动作；TD3 算法在 DDPG 算法的基础上采用了两套价值函数，避免 DDPG 算法更新时陷入局部最优；SAC 算法的 Actor 网络输出为动作的均值与方差，以此均值和方差选取执行动作，并使用奖励工程在原奖励的基础上增加由动作分布确定的熵以鼓励智能体探索。

$$R = R + \alpha^{(\text{熵})} h\left[\pi(\cdot|S)\right] \tag{4}$$

2　状态空间及动作空间设计

AC 强化学习架构下，由于深度神经网络具有强大的非线性映射能力，可直接由连续状态空间向连续动作空间进行决策映射，通过数值方法实现空战环境连续决策。为在空战机动决策中使用 AC 架构，首先对空战的状态空间和动作空间进行设计。

2.1　空战状态空间设计

参量化空战状态空间，首先应建立空战态势的几何模型。敌我空战几何关系如图 2 所示，红色飞机为我方飞机，蓝色飞机为敌方飞机。$\boldsymbol{R}$ 为敌我距离矢量，其方向由我方指向敌方，同时也是我方的射击瞄准线方向；$\boldsymbol{V}_{\mathrm{r}}$ 为我方的速度矢量；$\boldsymbol{V}_{\mathrm{b}}$ 为敌方的速度矢量；两者参考系为地面坐标系 $Ox_{\mathrm{g}}y_{\mathrm{g}}z_{\mathrm{g}}$；$\varphi$ 为我方飞机的提前角，即我方飞机机体轴与射击瞄准线之间的夹角，其大小等于我方速度矢量 $\boldsymbol{V}_{\mathrm{r}}$ 与距离矢量 $\boldsymbol{R}$ 之间的夹角，该角度同时也是导弹瞄准时我方导弹的离轴角；q 为我方飞机的方位角，表示我方飞机相对敌机的方位，其大小等于敌方速度矢量 $\boldsymbol{V}_{\mathrm{b}}$ 与距离矢量 $\boldsymbol{R}$ 之间夹角。角度 φ 与 q 的计算公式如式(5)所示。

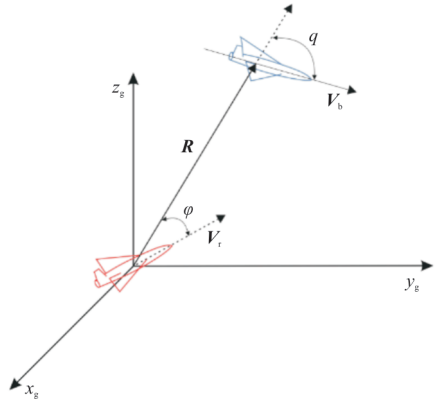


图 2　空战几何关系

Fig. 2　Geometry relationship of air combat

$$\begin{cases} \varphi = \arccos\dfrac{\boldsymbol{V}_{\mathrm{r}}\cdot\boldsymbol{R}}{\|\boldsymbol{V}_{\mathrm{r}}\|\cdot\|\boldsymbol{R}\|} & (\varphi \in [0, 180^\circ]) \\ q = \arccos\dfrac{\boldsymbol{V}_{\mathrm{b}}\cdot\boldsymbol{R}}{\|\boldsymbol{V}_{\mathrm{b}}\|\cdot\|\boldsymbol{R}\|} & (q \in [0, 180^\circ]) \end{cases} \tag{5}$$

由于环境需要满足马尔科夫性的要求，参量化空战状态空间需尽量包含空战的态势特征，而空战中位置、速度、加速度及角度是机动飞行的最主要特征，故将空战态势信息 $\boldsymbol{S}_t$ 描述为 $\boldsymbol{S}_t=\left[\boldsymbol{p}_{\mathrm{r}},\boldsymbol{V}_{\mathrm{r}},\boldsymbol{a}_{\mathrm{r}},\varphi_{\mathrm{r}},q_{\mathrm{r}},\boldsymbol{p}_{\mathrm{b}},\boldsymbol{V}_{\mathrm{b}},\boldsymbol{a}_{\mathrm{b}},\varphi_{\mathrm{b}},q_{\mathrm{b}}\right]$，其中 $\boldsymbol{p}_{\mathrm{r}}$ 和 $\boldsymbol{p}_{\mathrm{b}}$ 分别为我机和敌机的位置坐标；$\boldsymbol{V}_{\mathrm{r}}$ 和 $\boldsymbol{V}_{\mathrm{b}}$ 分别为我机和敌机的速度矢量；$\boldsymbol{a}_{\mathrm{r}}$ 和 $\boldsymbol{a}_{\mathrm{b}}$ 分别为我机和敌机的加速度矢量；φ_{r} 和 φ_{b} 分别为我机与敌机的提前角；q_{r} 和 q_{b} 分别为我机和敌机的方位角。

2.2 空战连续动作空间设计

2.2.1 动作维度选取及动态求解

动作维度应在保证可实现大多数机动动作的基础上尽量减少。参考飞行员在空战中的操纵习惯，多以驾驶杆和油门配合完成战术机动，故选取飞机法向过载 n_z、推力 T 与速度滚转角 ϕ_a 三个维度的连续量构成动作空间。

输入上述三个操纵量后，飞机动态仿真可在气流轴系下利用三自由度飞行动力学方程求解实现。

由于飞机机动范围较大，可能出现垂直向上或向下的姿态，采用欧拉角表征姿态会出现“万向锁”问题进而导致仿真求解中断。故采用四元数法[49]表征飞机姿态。

使用四元数，由地面坐标系到气流坐标系的旋转矩阵 $\boldsymbol{L}_{\mathrm{ag}}$ 可以表示为

$$\boldsymbol{L}_{\mathrm{ag}}=\begin{bmatrix} e_0^2+e_1^2-e_2^2-e_3^2 & 2\left(e_1e_2+e_0e_3\right) & 2\left(e_1e_3-e_0e_2\right) \\ 2\left(e_1e_2-e_0e_3\right) & e_0^2-e_1^2+e_2^2-e_3^2 & 2\left(e_2e_3+e_0e_1\right) \\ 2\left(e_1e_3+e_0e_2\right) & 2\left(e_2e_3-e_0e_1\right) & e_0^2-e_1^2-e_2^2+e_3^2 \end{bmatrix} \tag{6}$$

则飞机在地面惯性参考系下的速度矢量投影 $\boldsymbol{V}_{\mathrm{g}}$ 为

$$\boldsymbol{V}_{\mathrm{g}}=\boldsymbol{L}_{\mathrm{ag}}^{\mathrm{T}}\begin{bmatrix} V \\ 0 \\ 0 \end{bmatrix} \tag{7}$$

飞机所受外力可以简化为由重力 mg、气动升力 L、气动阻力 D 和发动机推力 T 构成，其中 L 和 D 可由操纵量法向过载 n_z 和飞行状态（飞行高度、飞行马赫数、飞行动压）确定。飞机所受合力在地面惯性参考系下的投影 F_{g} 为

$$F_{\mathrm{g}}=\boldsymbol{L}_{\mathrm{ag}}^{\mathrm{T}}\begin{bmatrix} T-D \\ 0 \\ -L \end{bmatrix}+\begin{bmatrix} 0 \\ 0 \\ mg \end{bmatrix} \tag{8}$$

飞机运动过程中，因所受外力与速度方向不共线而导致的速度轴的转动角速度 $\boldsymbol{\omega}_0$ 为

$$\boldsymbol{\omega}_0=\frac{\boldsymbol{V}_{\mathrm{g}}\times\boldsymbol{F}_{\mathrm{g}}}{mV^2} \tag{9}$$

则飞机气流坐标系相对惯性参考系的旋转角速度在自身坐标系的投影为

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}_a=\boldsymbol{L}_{\mathrm{ag}}\cdot\boldsymbol{\omega}_0+\begin{bmatrix} \dot{\phi}_a \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

则飞机的动力学方程可表示为

$$\begin{cases} \dot{V}=\dfrac{T-D}{m}+2g\left(e_1e_3-e_0e_2\right) \\ \dot{e}_0=-\dfrac{1}{2}\left(\omega_xe_1+\omega_ye_2+\omega_ze_3\right) \\ \dot{e}_1=\dfrac{1}{2}\left(\omega_xe_0-\omega_ye_3+\omega_ze_2\right) \\ \dot{e}_2=\dfrac{1}{2}\left(\omega_xe_3+\omega_ye_0-\omega_ze_1\right) \\ \dot{e}_3=\dfrac{1}{2}\left(-\omega_xe_2+\omega_ye_1+\omega_ze_0\right) \end{cases} \tag{11}$$

在实时求解飞机动态时，可以通过数值积分更新四元数和飞行速度标量 V，进而更新 $\boldsymbol{L}_{\mathrm{ag}}$ 矩阵，代表飞机的姿态。接下来可利用式(7)更新飞行速度矢量在地面坐标系下的投影 $\boldsymbol{V}_{\mathrm{g}}$，进而对时间进行数值积分，便可求解飞机的位置坐标。

2.2.2 机动包线动态归一化

基于上述连续动作空间，理论上可以实现除失速机动和非协调侧滑外的所有空战机动动作的仿真模拟，但在实际空战中，飞机操纵量的安全范围和极限使用范围是随飞行状态实时变化的。以法向过载为例，某型飞机法向过载的使用包线如图 3 所示，上方曲线代表不同高度最大可用正过载，下方曲线代表不同高度最大可用负过载。在实际飞行中，如果飞机使用过载超出过载包线范围可能发生失速偏离或结构受损等危险情况，同

样在油门和滚转角控制量上也存在着随飞行状态变化的限制条件。

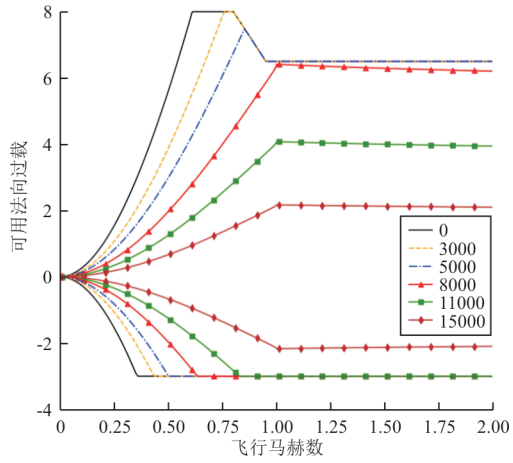


图 3　飞机可使用过载包线

Fig. 3　Envelope of load factor

由于存在限制条件,使得动作空间不再整齐,且量纲随飞行状态实时变化。本文在不破坏强化学习方法对环境马尔科夫性要求的前提下,采用动态线性归一化的方法,即实时将每个动作范围线性映射至[−1,1]区间内。以法向过载 n_z 为例,法向过载的动态归一化公式为

$$n_z^N=\frac{2n_z-\left(n_{z\max}+n_{z\min}\right)}{n_{z\max}-n_{z\min}} \tag{12}$$

式中:n_z^N 为归一化后的法向过载动作量;$n_{z\max}$ 和 $n_{z\min}$ 分别为飞机当时状态下的最大正、负过载可用值,两者都受飞行马赫数和飞行高度的影响。

采用类似方法也可将推力 T、速度滚转角 ϕ_a 动态归一化,进而得到动态归一化后的动作空间 $\boldsymbol{A}_{\mathrm{t}}=\left[n_z^N,T^N,\phi_a^N\right]$。采用该动作空间具有以下优点:

(1) 在不破坏环境马尔科夫性的同时,使所用动作范围保持在[−1,1]范围内,可以与深度神经网络的输出层激活函数 tanh 进行量纲对接,且数值范围处于激活函数的非饱和区域有利于加速训练;

(2) 可保证决策的输出动作均在包线以内,不会产生失速、结构超载等危险动作;

(3)完成常规机动动作的难度更低,如完成垂直动作时,若使用未归一化的动作空间需要根据动作阶段不断调整法向过载,而使用归一化后的动作空间只需保持法向过载为 0.8 即可使其保持在最优使用范围内。

2.3　空战连续机动动作验证

为验证动作空间设置的合理性和仿真方程的有效性,本文选取空中的高斤斗、斜斤斗和水平盘旋进行仿真验证。

以归一化动作为操纵量的斜斤斗及水平盘旋飞行轨迹分别如图 4～图 5 所示。

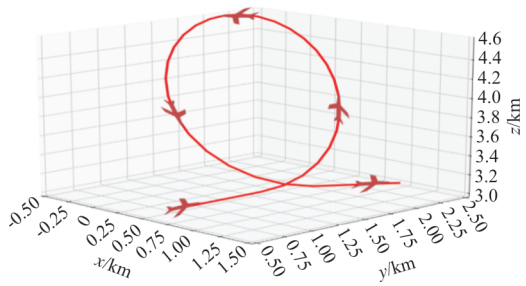


图 4　斜斤斗运动轨迹图

Fig. 4　Trajectory of inclined loop maneuver

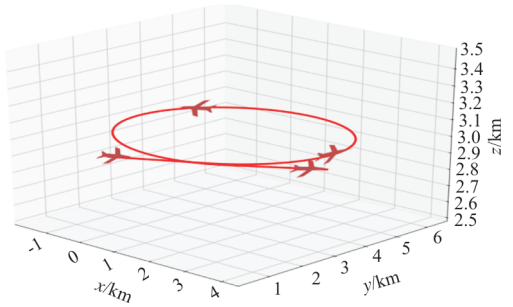


图 5　水平盘旋运动轨迹图

Fig. 5　Trajectory of sustained turn

从图 4 可以看出:飞机的机动轨迹平滑,在垂直位置未出现仿真中断现象。斜斤斗机动过程中三个欧拉角会急剧变化,尤其是接近垂直向上位置时,会出现欧拉角变化率无穷大的现象,表明本文采用的仿真方法具有良好的鲁棒性。

水平盘旋机动为空战水平机动的基础,实际中可通过增减推力 T^N 的大小进行加减速水平盘旋机动,也可通过改变滚转角 ϕ_a^N 和法向过载 n_z^N 的大小完成上升或下降转弯机动。

空战中所有常规机动动作均可以认为是水平动作和垂直动作的组合与变形。使用归一化的动作空间来参量化空战动作,经验证可以通过保持特定操纵量完成成套的垂直和水平机动动作。在实际空战中,机动动作多是成套动作的拆分和组合。通常飞行员动作切换的时间为秒量级,故本

文选取2 s为一个决策步长,决策步长之间的操作量变化受到飞机敏捷性制约。理论上使用本文动作空间设置方法,可以实现飞机全状态、全包线、任意空战动作的连续决策,且决策步长之间不会出现操纵量突变的现象。

3 空战环境与奖励设置

3.1 空战环境设置

AC强化学习架构下,agent需要不断与环境交互获得奖励,通过"试错"方式搜索最优策略。

航炮是应用最广泛的空对空武器,随着战斗机性能的不断提升,人们开始寻求更强的火力。即使航炮与瞄准设备在不断更新换代,航炮的空对空射击仍然对飞行员有较高的要求。实际需求与技术的进一步发展促使空对空导弹的出现。现代空战中,战斗机多会在装备航炮的同时携带导弹,由飞行员控制飞机做出机动以获得有利态势。纵观战斗机发展历程,虽然其作战能力在不断提升,但在近距空战中使用的战术机动并未体现出较大的差异性,仍以尾后攻击为主,因此空战环境中敌我双方飞机不同的性能参数设置并不会对强化学习结果产生很大的影响。同时考虑到在实际空战训练时,双方使用的飞机不会具有较大的性能差别,故对敌我飞机设置了相同的性能参数。在空战仿真环境中,机载武器考虑使用空空导弹与机炮的情况。同时为了与真实空战训练场景保持一致,本文主要针对从敌我飞机发现对方到一方构成武器攻击条件的空战机动过程进行训练寻优,不考虑武器发射后的后续规避过程。空战胜负的判据设定主要参考实际空战中BFM(Basic Fighter Maneuvering)课目的设置方法,即一方构成火力控制系统解算下的导弹发射条件[50-51]或机炮攻击条件,则认为该轮训练已分出胜负。参考空战训练中的相关课目,除上述情况外,若一方被迫撞地,也认为另一方对其完成"撞地击杀";若发生双机危险接近,则认为双机相撞,双方均"失败";若一方飞出边界,则认为该方任务失败。

3.2 空战奖励设置

参考空战实际训练中的课目设置方法,并考虑强化学习网络的收敛性,采用事件奖励为主,过程奖励为辅的奖励设置方法。

事件奖励只有在回合结束时才会给出,会受到折扣因子的影响而衰减。本文在奖励设计时选取较大量级的事件奖励值与合理的折扣因子,使智能体仍有足够的动力以结果为导向进行决策。而事件奖励在整个空战过程中是稀疏的,此时无明显的策略梯度来引导agent的演化方向。针对该现象,结合飞行员空战的先验知识合理设置过程奖励,在空战过程中实时给予智能体反馈以引导智能体探索最有可能获胜的方向。首先根据不同空战结果的重要程度,设置以结果为导向的事件奖励。空战中最"完美"的结果为使用导弹击杀敌机,若空战仿真结果为使用导弹击杀敌机,则agent获得奖励+2 000;若使用航炮击杀敌机,则agent获得奖励+1 000;若双方缠斗中,迫使敌方损失高度而最终撞地,则获得奖励+1 000;空战中应尽量避免与敌机相撞,若发生此结果,agent获得奖励-1 000。相对应,若被敌机导弹击杀,agent获得奖励-2 000;若被敌机航炮击杀,agent获得奖励-1 000;被迫损失高度撞地,agent获得奖励-1 000。

在仿真训练中,由于agent操纵飞机的自由度较大,飞机易出现进入小速度、低高度、超速、超出升限或者脱离初始空域等现象。在实际空战出现该类现象可能会影响飞行安全,故将该类现象统一称为"飞出边界"事件,并给予一定负奖励。发生该事件后将终止本轮空战仿真,重置空战环境,agent得到奖励-300,此处负奖励绝对值较小的原因是防止agent因避免"飞出边界"而限制飞机机动潜能。

综上,空战事件奖励的设置汇总如表1所示。

表1 事件奖励设置

Table 1 Reward setting of statements

结果类型	奖励值	结果类型	奖励值
导弹击杀	+2 000	被航炮击杀	-1 000
航炮击杀	+1 000	被撞地击杀	-1 000
撞地击杀	+1 000	双机相撞	-1 000
被导弹击杀	-2 000	飞出边界	-3 00

过程奖励的设置,主要以飞行手册中的"最佳机动点"为依据。实际双机机动对抗中,存在一个相对的位置区域,在该区域内飞机可以用最小的

机动过载来保持对敌机的持续跟踪，且容易达成导弹发射条件，该区域的中心即为“最佳机动点”。

使用某型导弹攻击某型飞机时，最佳机动点的坐标计算经验公式为

$$\boldsymbol{p}_{\text{best}}=\boldsymbol{p}_{\text{b}}-\frac{\boldsymbol{V}_{\text{b}}}{\|\boldsymbol{V}_{\text{b}}\|}\cdot\left[\alpha R_{\max}+(1-\alpha)R_{\min}\right] \tag{13}$$

式中：$\boldsymbol{p}_{\text{best}}$为最佳机动点在地面坐标系下的坐标；$\boldsymbol{V}_{\text{b}}$为敌方飞机速度矢量；$\alpha$为最佳攻击距离的比例系数，$\alpha\in[0,1]$。

过程奖励的设置思路为：当飞机位置与最佳机动点距离较远时，让奖励与该距离负相关，以引导 agent 以最快方式向最佳机动点接近；当距离较近时，为引导 agent 减小敌机视线率和导弹离轴角以构成导弹发射条件，此时的奖励要与敌机视线率和导弹离轴角的大小负相关，且距离越小奖励值越大，设置经验过程奖励公式为

$$\begin{cases} d=\|\boldsymbol{p}_{\text{r}}-\boldsymbol{p}_{\text{best}}\| \\ R_{t+1}=-\dfrac{d}{1\,000} & (d\geqslant 800) \\ R_{t+1}=-\dfrac{80\times(\varphi_{\text{r}}/\varphi_{\max}+l_{\text{os}}/l_{\text{os(max)}})}{d} & (d<800) \end{cases} \tag{14}$$

式中：d为我方飞机与最佳机动点的欧式距离；R_{t+1}为过程奖励值；φ_{r}为我机导弹离轴角；$\varphi_{\max}$为导弹最大离轴角；l_{os}为敌机视线率的大小；$l_{\text{os(max)}}$为导弹发射架的最大转动速率。

4　神经网络模型搭建

由于总体算法基于 Actor-Critic 框架，故需要建立两类神经网络模型。参考深度强化学习算法搭建神经网络，如文献[47]中使用 DDPG 算法在隐藏层较少时设置了 300 与 400 个节点，在不同难度的经典强化学习环境中获得了较好的收敛结果，如 4 维度状态空间的倒立摆环境，18 维度状态空间的机械臂环境等。而空战环境更为复杂且状态空间维度更大，同时考虑到通用性与封装性，本文 Actor 和 Critic 网络均采用相同规格的隐藏层，均设有两个隐藏层，每个隐藏层均有 512 个节点，在隐藏层后加入 ReLu 激活层，用来增强神经网络的非线性映射能力。

其中 Actor 网络以敌我飞机的总体态势为输入，由于S_t中的$\boldsymbol{p}_{\text{r}}$、$\boldsymbol{p}_{\text{b}}$、$\boldsymbol{V}_{\text{r}}$、$\boldsymbol{V}_{\text{b}}$、$\boldsymbol{a}_{\text{r}}$、$\boldsymbol{a}_{\text{b}}$均为三维矢量，对于敌我飞机均有方位角与离轴角，故 Actor 网络的输入维度为 22 维。Critic 网络的输入层维度为状态S_t维度与动作A_t维度的叠加，即 22+3=25。其隐藏层参数设置与 Actor 网络基本相同。Critic 的输出为状态—动作对的价值评估，输出维度为 1 维。

本文主要对比 DDPG、SAC、TD3 三种算法的效果，虽然都基于 AC 架构，但神经网络的构建仍有区别，SAC 将动作分布嵌入神经网络，而 DDPG 与 TD3 算法选择在外部添加噪声来实现连续动作的训练。

SAC 算法中 Actor 的输出不是确定性的动作，而是基于高斯分布的动作概率，故其输出为动作的均值与标准差，由于操纵量的设置共有 3 个维度，故均值与标准差均为 3 维输出。在均值与标准差输出层后，Actor 网络会基于高斯分布抽样选择出动作样本，并由 Tanh 激活层将动作归一化至[−1, 1]区间，与仿真环境进行量纲对齐，Actor 网络结构如图 6 所示。

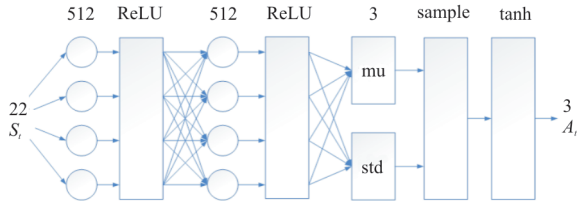


图 6　SAC 算法 Actor 网络结构

Fig. 6　Structure of actor neural network in SAC

为对比不同算法的效果，DDPG 算法与 TD3 的 Actor 网络架构与 SAC 基本相同，但输出为确定性动作，经 Tanh 激活层将动作归一化至[−1, 1]区间，在神经网络框架外引入高斯噪声，再经过范围限制输出动作，其 Actor 网络结构如图 7 所示。

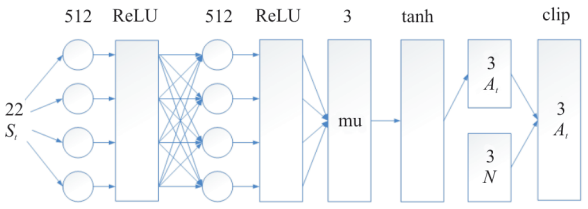


图 7　DDPG、TD3 算法 Actor 网络结构

Fig. 7　Structure of actor neural network in DDPG and TD3

三种算法 Critic 网络作用都为输出价值评估值，采用相同的 Critic 网络架构，如图 8 所示。

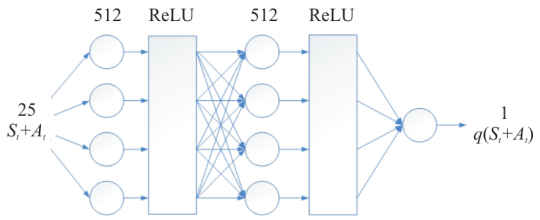


图 8 Critic 网络结构

Fig. 8 Structure of Critic neural network

5 结果与讨论

空战场景的设置，借鉴空战训练中的“热身练习”课目。空战训练中，在开始正式对抗课目之前，飞行员通常会进行 1～2 轮的热身练习，用以熟悉空战技术和适应空战节奏。本文研究以验证思路可行性为主，故借鉴空战中的热身训练课目设置空战场景。空战训练中常见的热身方式为，扮演敌方的飞机进行过载转弯，我方分别从劣势、均势、优势的起始态势下，对敌方飞机进行机动、跟踪、锁定、射击等操纵。

为提高决策算法的泛化能力，以更加适应实际空战中的复杂态势，设置高不确定度的空战场景以验证决策算法的有效性，具体设置如下。

敌方由蓝色飞机代表，轨迹为虚线，初始水平坐标为(0，0)，起始高度为 3 000 m，起始速度为 138～305 m/s，起始航向从 0～360°随机选取，0～60°随机选取滚转角做向左或向右等速水平盘旋机动；我方由红色飞机代表，轨迹为实线，起始水平坐标为(5 000，5 000)，起始高度从 2 500～3 500 m 随机选取，即我方初始高度优势随机设置，起始速度为 208 m/s，与敌机速度相比我方初始速度优势随机设置，起始航向从 0～360°随机选取，即我方初始角度优势随机设置。

在我方起始高度优势、速度优势和角度优势都具有高不确定度的空战场景下训练，各算法训练过程中学习曲线如图 9 所示，可以看出：各算法在本文设置环境下都有不错的收敛性，在 AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz 处理器及 NVIDIA GeForce RTX 3060 Laptop GPU 环境下训练，由于不同算法以及代码的差异性，训练时间有所差别，但 2～3 h 都可以达到收敛水平；DDPG 与 SAC 算法收敛性差别较小，TD3 由于使用了延迟更新与双网络学习稍显缓慢但最终也能趋于最优值。

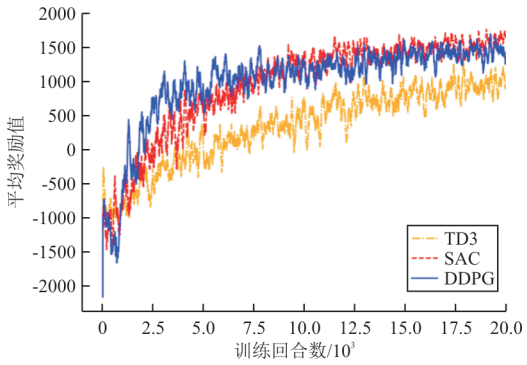


图 9 学习曲线对比

Fig. 9 Learning curves contrast

在此高不确定度空战场景中对不同算法训练完成模型进行测试，统计结果如表 2 所示，可以看出：agent 可以在多种算法下适应随机度的空战仿真环境，并且在大多数态势下做出最优决策，完成对敌击杀。

表 2 空战测试结果

Table 2 Results of air combat test

算法	导弹击杀/次	敌方飞出边界/次	我方飞出边界/次
SAC	83	5	6
TD3(less episodes)	68	0	23
TD3(more episodes)	96	0	0
DDPG	94	1	2

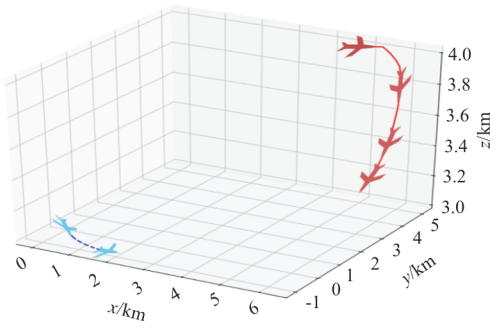
其中 TD3 算法由于延迟更新与双网络架构学习较慢，但在更多的更新回合下也能达到很好的效果；DDPG 与 SAC 算法整体差异不大，可以看出由于 SAC 算法选择将动作的不确定性嵌入神经网络，使得在测试时会有少许扰动让我方飞机飞出边界。

由于环境中机炮击杀范围比导弹击杀范围小，机炮击杀比导弹击杀要求更为严苛，agent 学习结果体现出使用导弹对敌击杀。

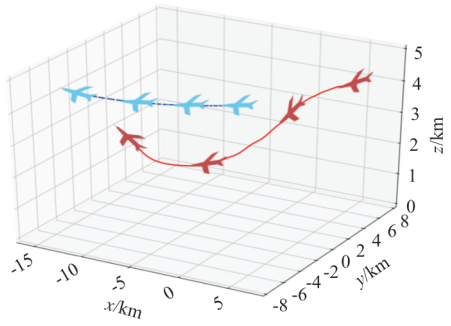
在初始条件为敌机随机位置、随机速度、随机滚转角做转弯机动时，agent 可以做出半滚倒转、低速 Yo-Yo 等实际飞行中飞行员常做的飞行动作。

敌机转弯空战动态如图 10 所示。图 10(a)中，我方飞机高度占优势，飞行方向与敌机基本平行，且敌机在我机后方，角度占劣势的前提下，agent 操纵我方飞机滚转 180°后保持较高过载，迅速下翻转，做了类似半滚倒转的机动动作，将机头指向敌机，完成导弹击杀。图 10(b)中，我方飞机高度与

敌机相近,飞行方向与敌机基本平行,且我机在敌机后方,角度占优势,但与敌机距离较远,无法构成导弹发射条件的前提下,agent操纵飞机向下俯冲增速,而后拉起机头指向敌机,通过类似低速Yo-Yo的机动动作,缩短双机距离,完成导弹击杀。



(a) 空战动态1



(b) 空战动态2

图10　敌机转弯空战动态

Fig. 10　Dynamic of air combat in enemy aircraft turn

为进一步说明空战环境与动作设置的合理性,除前文所述的敌方做随机水平盘旋机动外,还测试在不同场景下agent的表现,如双机迎面相遇、追击、爬升、防御等机动,如图11~图13所示。

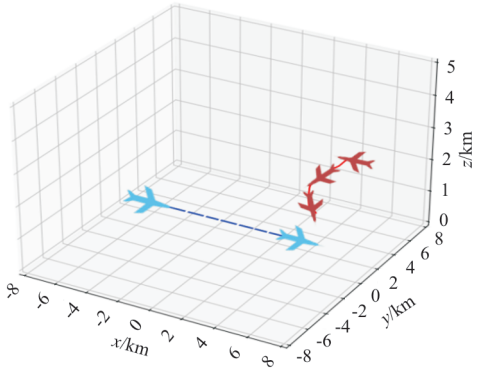


图11　迎面空战动态

Fig. 11　Dynamic of air combat in head-on

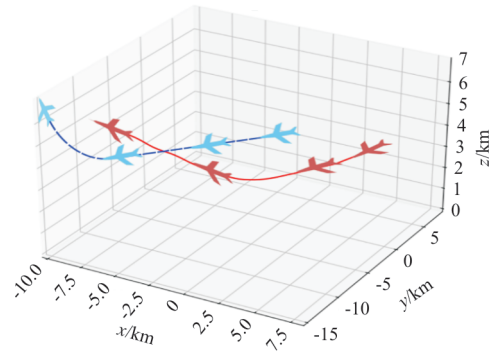


图12　爬升空战动态

Fig. 12　Dynamic of air combat in climb

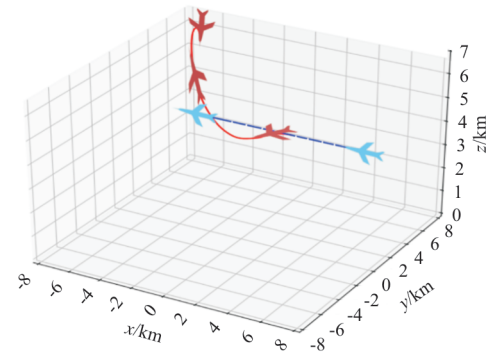


图13　攻防转换

Fig. 13　Attack and defense conversion

从图11可以看出:我机以大过载小转弯半径完成领先转弯将机头指向敌机完成击杀。

从图12可以看出:设置敌机在平飞一段距离后拉起,我机可以保持跟随敌机并用导弹完成击杀。

从图13可以看出:agent通过侧向拉起,直至敌机冲到我机前方后俯冲将机头指向敌机完成击杀。

6　结　论

(1) 基于AC强化学习架构,能够实现基于连续动作空间的空战机动决策,克服传统方法的"无限维度"问题,使空战基于连续动作的远视最优决策得以在较短时间内求解。

(2) 利用动态归一化方法,可以解决因飞机状态变化导致的动作空间不整齐问题,且可以降低执行成套机动动作的难度,有利于agent的训练学习。

(3) 通过高不确定度的空战仿真验证,训练完成的agent可以在复杂空战态势下对飞行目标保持

较高的击杀率;且agent在特定态势下可以做出实际空战中的常用机动动作,其机动方案具有较高的合理性。

本文通过高不确定性的空战场景设置与多种强化学习算法验证了在此环境下实现连续动作空战决策的合理性与可行性,但可以看到智能体所学习出的机动动作仍较为有限。为了与实际空战环境更为贴近,下一步的工作将主要针对多智能体自博弈方法展开研究,这将是未来智能空战的发展方向。

参 考 文 献

[1] TRSEK R B. The last manned fighter: replacing manned fighter with unmanned combat air vehicles[M]. Austin, TX, US: Bible Scholars, 2012.

[2] 国海峰, 侯满义, 张庆杰, 等. 基于统计学原理的无人作战飞机鲁棒机动决策[J]. 兵工学报, 2017, 38(1): 160-167.
GUO Haifeng, HOU Manyi, ZHANG Qingjie, et al. UCAV robust maneuver decision based on statistics principle[J]. Acta Armamentarii, 2017, 38(1): 160-167. (in Chinese)

[3] 董一群, 艾剑良. 自主空战技术中的机动决策: 进展与展望[J]. 航空学报, 2020, 41(s2): 4-12.
DONG Yiqun, AI Jianliang. Decision making in autonomous air combat: review and prospects[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(s2): 4-12. (in Chinese)

[4] 魏慎娜. 基于新型态势函数的空战微分对策问题研究[D]. 沈阳: 沈阳航空航天大学, 2018.
WEI Shenna. Research on air combat differential game based on differentiable situation function[D]. Shenyang: Shenyang Aerospace University, 2018. (in Chinese)

[5] MERZ A W. The homicidal chauffeur-a differential game[M]. San Francisco, California, USA: Stanford University, 1971.

[6] ARDEMA M D. Air-to-air combat analysis-review of differential-gaming approaches [EB/OL]. [2021-11-25]. https://www.researchgate.net/publication/4710199_Air-to-air_combat_analysis_-_Review_of_differential-gaming_approaches.

[7] LYNCH U H. Differential game barriers and their application in air-to-air combat[EB/OL]. [2021-11-25]. https://www.researchgate.net/publication/235144106_Differential_Game_Barriers_and_Their_Application_in_Air-to-Air_Combat.

[8] MORITZ K, POLIS R, WELL K H. Pursuit-evasion in medium-range air-combat scenarios[J]. Computers & Mathematics with Applications, Elsevier, 1987, 13(1/3): 167-180.

[9] SHINAR J. Evaluation of suboptimal pursuit-evasion game strategies for air combat analysis[C]// 11th Atmospheric Flight Mechanics Conference. Seattle, WA, USA: AIAA, 1984: 2126.

[10] ISAACS R. Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization[M]. North Chemsford: Courier Dover Publications, 1999.

[11] PACHTER M, YAVIN Y. A stochastic homicidal chauffeur pursuit-evasion differential game[J]. Journal of Optimization Theory and Applications, 1981, 34(3): 405-424.

[12] GETZ W M, PACHTER M. Two-target pursuit-evasion differential games in the plane[J]. Journal of Optimization Theory and Applications, 1981, 34(3): 383-403.

[13] YAVIN Y. A stochastic two-target pursuit-evasion differential game with three players moving in a plane[EB/OL]. [2021-11-25]. https://www.sciencedirect.com/science/article/pii/B9780080348629500165.

[14] SHINAR J, DAVIDOVITZ A. Unified approach for two-target game analysis[J]. IFAC Proceedings Volumes, 1987, 20(5): 65-71.

[15] GRIMM W, WELL K H. Modelling air combat as differential game recent approaches and future requirements[C]// Differential Games-Developments in Modelling and Computation. Berlin, Heidelberg: Springer, 1991: 1-13.

[16] 傅莉, 王晓光. 无人战机近距空战微分对策建模研究[J]. 兵工学报, 2012, 33(10): 1210-1216.
FU Li, WANG Xiaoguang. Research on close air combat modeling of differential games for unmanned combat air vehicles[J]. Acta Armamentarii, 2012, 33(10): 1210-1216. (in Chinese)

[17] PARK H, LEE B-Y, TAHK M J, et al. Differential game based air combat maneuver generation using scoring function matrix[J]. International Journal of Aeronautical and Space Sciences, The Korean Society for Aeronautical and Space Sciences, 2016, 17(2): 204-213.

[18] XU Guangyan, WEI Shenna, ZHANG Hongmei. Application of situation function in air combat differential games[C]// 2017 36th Chinese Control Conference (CCC). Dalian, China: IEEE, 2017: 9-14.

[19] BURGIN G, WILLIAMS W, SIDOR L. The adaptive maneuvering logic program in support of the pilot's associate program-a heuristic approach to missile evasion[C]// 24th Aerospace Sciences Meeting. Reno, NV, USA: AIAA, 1986: 423.

[20] MCMANUS J, GOODRICH K. Application of artificial intelligence (AI) programming techniques to tactical guidance for fighter aircraft[C]// Guidance, Navigation and Control Conference. Boston: AIAA, 1989: 3525.

[21] GOODRICH K, MCMANUS J. An integrated environment for tactical guidance research and evaluation[C]// Orbital Debris Conference: Technical Issues and Future Directions. Baltimore, MD, USA: AIAA, 1990: 1287.

[22] 赵明明. 不确定信息下多无人机空战动态博弈策略研究[D]. 沈阳:沈阳航空航天大学, 2014.
ZHAO Mingming. Study on multi-UAVs air combat dynamic game strategy based on uncertain information[D]. Shenyang: Shenyang Aerospace University, 2014. (in Chinese)

[23] 钱炜祺, 车竞, 何开锋. 基于矩阵博弈的空战决策方法[C]// 第二届中国指挥控制大会. 北京:中国指挥与控制学会, 2017: 409-413.
QIAN Weiqi, CHE Jing, HE Kaifeng. Air combat decision method based on game-matrix approach[C]// The Second China Command and Control Conference. Beijing: Chinese Institute of Command and Control, 2017: 409-413. (in Chinese)

[24] 车竞, 钱炜祺, 和争春. 基于矩阵博弈的两机攻防对抗空战仿真[J]. 飞行力学, 2015, 33(2): 173-177.
CHE Jing, QIAN Weiqi, HE Zhengchun. Attack-defense confrontation simulation of air combat based on game-matrix approach[J]. Flight Dynamics, 2015, 33(2): 173-177. (in Chinese)

[25] 邓可, 彭宣淇, 周德云. 基于矩阵对策与遗传算法的无人机空战决策[J]. 火力与指挥控制, 2019, 44(12): 61-66, 71.
DENG Ke, PENG Xuanqi, ZHOU Deyun. Study on air combat decision method of uav based on matrix game and genetic algorithm[J]. Fire Control & Command Control, 2019, 44(12): 61-66,71. (in Chinese)

[26] 高阳阳, 陈双艳, 余敏建, 等. 改进人工免疫算法的多机协同空战目标分配方法[J]. 西北工业大学学报, 2019, 37(2): 354-360.
GAO Yangyang, CHEN Shuangyan, YU Minjian, et al. Target allocation method of multi-aircraft cooperative air combat based on improved artificial immune algorithm[J]. Journal of Northwestern Polytechnical University, 2019, 37(2): 354-360. (in Chinese)

[27] 付跃文, 王元诚, 陈珍, 等. 基于多智能体粒子群的协同空战目标决策研究[J]. 系统仿真学报, 2018, 30(11): 4151-4157.
FU Yuewen, WANG Yuancheng, CHEN Zhen, et al. Target decision in collaborative air combats using multi-agent particle swarm optimization[J]. Journal of System Simulation, 2018, 30(11): 4151-4157. (in Chinese)

[28] AUSTIN F, CARBONE G, HINZ H, et al. Game theory for automated maneuvering during air-to-air combat[J]. Journal of Guidance, Control, and Dynamics, 1990, 13(6): 1143-1149.

[29] 朱可钦, 董彦非. 空战机动动作库设计方式研究[J]. 航空计算技术, 2001(4): 50-52.
ZHU Keqin, DONG Yanfei. Study on the design of air combat maneuver library[J]. Aeronautical Computer Technique, 2001(4): 50-52. (in Chinese)

[30] 胡秉科. 歼击机一对一空战模拟系统及其应用[J]. 航空系统工程, 1992(5): 35-43.
HU Bingke. One to one air combat simulation system of fighter and its application[J]. Aerospace Systems Engineering, 1992(5): 35-43. (in Chinese)

[31] 朱宝鎏, 朱荣昌, 熊笑非. 作战飞机效能评估[M]. 北京:航空工业出版社, 2006.
ZHU Baoliu, ZHU Rongchang, XIONG Xiaofei. Combat aircraft effectiveness assessment[M]. Beijing: Aviation Industry Press, 2006. (in Chinese)

[32] MCMAHON D C. A neural network trained to select aircraft maneuvers during air combat: a comparison of network and rule based performance[C]// 1990 IJCNN International Joint Conference on Neural Networks. San Diego, CA, USA: IEEE, 1990: 107-112.

[33] SCHVANEVELDT R W, GOLDSMITH T E, BENSON A E, et al. Neural network models of air combat maneuvering[EB/OL]. [2021-11-25]. https://www.researchgate.net/publication/235068390_Neural_Network_Models_of_Air_Combat_Maneuvering.

[34] ERNEST N, CARROLL D, SCHUMACHER C, et al. Genetic fuzzy based artificial intelligence for unmanned combat aerial vehicle control in simulated air combat missions[J]. Journal of Defense Management, 2016, 6(1): 2167-2173.

[35] 左家亮, 杨任农, 张滢, 等. 基于启发式强化学习的空战机动智能决策[J]. 航空学报, 2017, 38(10): 217-230.
ZUO Jialiang, YANG Rennong, ZHANG Ying, et al. Intelligent decision-making in air combat maneuvering based on heuristic reinforcement learning[J]. Acta Aeronautica et Astronautica Sinica, 2017, 38(10): 217-230. (in Chinese)

[36] LIU P, MA Y. A deep reinforcement learning based intelligent decision method for UCAV air combat[C]// Modeling, Design and Simulation of Systems. Singapore: Springer, 2017: 274-286.

[37] 张强, 杨任农, 俞利新, 等. 基于Q-network强化学习的超视距空战机动决策[J]. 空军工程大学学报(自然科学版), 2018, 19(6): 8-14.
ZHANG Qiang, YANG Rennong, YU Lixin, et al. BVR air combat maneuvering decision by using Q-network reinforcement learning[J]. Journal of Air Force Engineering University (Natural Science Edition), 2018, 19(6): 8-14. (in Chinese)

[38] KURNIAWAN B, VAMPLEW P, PAPASIMEON M, et al. An empirical study of reward structures for actor-critic reinforcement learning in air combat manoeuvring simulation[C]// Advances in Artificial Intelligence. Cham: Springer International Publishing, 2019: 54-65.

[39] YANG Q, ZHU Y, ZHANG J, et al. UAV air combat autonomous maneuver decision based on DDPG algorithm[C]// 2019 IEEE 15th International Conference on Control

and Automation (ICCA). Edinburgh, UK: IEEE, 2019: 37-42.

[40] YANG Q, ZHANG J, SHI G, et al. Maneuver decision of UAV in short-range air combat based on deep reinforcement learning[J]. IEEE Access, 2019, 8: 363-378.

[41] PIAO H, SUN Z, MENG G, et al. Beyond-visual-range air combat tactics auto-generation by reinforcement learning [C]// 2020 International Joint Conference on Neural Networks (IJCNN). Glasgow, UK: IEEE, 2020: 1-8.

[42] 孙楚, 赵辉, 王渊,等. 基于强化学习的无人机自主机动决策方法[J]. 火力与指挥控制, 2019, 44(4): 144-151.
SUN Chu, ZHAO Hui, WANG Yuan, et al. UCAV autonomic maneuver decision-making method based on reinforcement learning[J]. Fire Control & Command Control, 2019, 44(4): 144-151. (in Chinese)

[43] 吴宜珈, 赖俊, 陈希亮, 等. 强化学习算法在超视距空战辅助决策上的应用研究[J]. 航空兵器, 2021, 28(2): 55-61.
WU Yijia, LAI Jun, CHEN Xiliang, et al. Research on the application of reinforcement learning algorithm indecision support of beyond-visual-range air combat[J]. Aero Weaponry, 2021, 28(2): 55-61. (in Chinese)

[44] 施伟, 冯旸赫, 程光权, 等. 基于深度强化学习的多机协同空战方法研究[J]. 自动化学报, 2021, 47(7): 1610-1623.
SHI Wei, FENG Yanghe, CHENG Guangquan, et al. Research on multi-aircraft cooperative air combat method based on deep reinforcement learning[J]. Acta Automatica Sinica, 2021, 47(7): 1610-1623. (in Chinese)

[45] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms (PPO) [EB/OL]. [2021-11-25]. https://www.researchgate.net/publication/318584439_Proximal_Policy_Optimization_Algorithms.

[46] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications [EB/OL]. [2021-11-25]. https://www.researchgate.net/publication/329705402_Soft_Actor-Critic_Algorithms_and_Applications.

[47] LILLICRAP T P, HUNT J J, PRITZEL A. Continuous control with deep reinforcement learning: US10776692B2 [P]. 2020-09-15.

[48] DANKWA S, ZHENG W. Twin-delayed DDPG: a deep reinforcement learning technique to model a continuous movement of an intelligent robot agent[C]// Proceedings of the 3rd International Conference on Vision, Image and Signal Processing. New York, NY, USA: Association for Computing Machinery, 2019: 1-5.

[49] 周须峰, 易华, 谢永强. 近距格斗空空导弹三自由度弹道仿真建模[J]. 四川兵工学报, 2014, 35(2): 9-12,26.
ZHOU Xufeng, YI Hua, XIE Yongqiang. Modeling on 3DOF trajectory simulation of short range dogfight air-to-air missile[J]. Journal of Sichuan Ordnance, 2014, 35(2): 9-12,26. (in Chinese)

[50] 胡朝晖, 李东文, 汪浩生. 通用空空导弹攻击区仿真研究[J]. 弹箭与制导学报, 2002(3):18-23.
HU Zhaohui, LI Dongwen, WANG Haosheng. Research on simulation attack area of general air-to-air missile[J]. Journal of Projectiles Rockets Missiles and Guidance, 2002(3): 18-23. (in Chinese)

[51] 顾佼佼, 刘卫华, 姜文志. 基于攻击区和杀伤概率的视距内空战态势评估[J]. 系统工程与电子技术, 2015, 37(6): 1306-1312.
GU Jiaojiao, LIU Weihua, JIANG Wenzhi. WVR air combat situation assessment model based on weapon engagement zone and kill probability[J]. Systems Engineering and Electronics, 2015, 37(6): 1306-1312. (in Chinese)

作者简介:

单圣哲(1991—),男,硕士研究生。主要研究方向:空战人工智能,飞行力学。

杨孟超(1998—),男,硕士研究生。主要研究方向:空战人工智能,飞行力学。

张伟伟(1979—),男,博士,教授。主要研究方向:气动弹性力学、流体力学与神经网络/大数据/人工智能等新兴领域的交叉方向,新概念飞行器空气动力学,理论与计算流体力学,流动主动/被动控制。

高传强(1988—),男,博士,教授。主要研究方向:空战人工智能,智能流动控制,气动弹性力学。

(编辑:马文静)