



论 文

基于PER-SAC的舰载机着舰复飞操纵策略研究

张志冰¹, 王家兴¹, 袁婷慧^{2*}

1. 航空工业沈阳飞机设计研究所, 沈阳 110035

2. 南京航空航天大学自动化学院, 南京 210016

* Email: sx2303134@nuaa.edu.cn

收稿日期: 2024-05-30; 接受日期: 2024-11-18; 网络版发表日期: 2025-05-21

摘要 为保证舰载机着舰复飞安全, 本文提出一种基于PER-SAC的舰载机着舰复飞操纵策略. 考虑到舰载机在航母不同距离处的复飞风险差异较大, 通过改进安全复飞准则优化了复飞边界. 将军用推力与升降舵操纵相结合, 提出了一种基于软行动者-评论家(soft actor-critic, SAC)强化学习算法的复飞操纵策略, 考虑复飞指标和软状态约束设计了一种引导型奖励函数, 提升了智能体的探索效率; 针对SAC算法采样效率低的问题, 引入优先经验重放(prioritized experience replay, PER)技术提出PER-SAC算法, 加快了SAC的收敛速度. 仿真结果表明, 本文设计的复飞策略能显著缩小复飞风险区, 提高复飞安全性.

关键词 舰载机, 复飞边界, 复飞操纵, 软行动者-评论家, 强化学习, 优先经验重放

1 引言

舰载机作为重要的海上军事力量, 在作战和侦察任务中发挥着关键作用^[1]. 在舰载机着舰过程, 舰尾流、风场等干扰可能会导致舰载机严重偏离理想下滑轨迹, 此时为避免撞舰, 需根据复飞边界进行决策, 并执行复飞操纵以驱使舰载机脱离危险区域^[2,3]. 因此, 确定舰载机复飞边界以及操纵策略是舰载机安全复飞的关键^[4].

目前, 国内外机构和学者已经对舰载机复飞问题进行了相关的研究. 文献[5]针对舰载机尾钩与甲板的安全间隙、飞行员反应时间等方面提出了安全复飞准则. 文献[6]对安全复飞准则进行了改进, 增加了复飞轨迹最低点的高度限制. 着舰风险评估与复飞决策技

术也一直是着舰复飞领域的研究热点. 文献[7]设计了基于改进型复飞边界准则的综合复飞决策系统, 判断舰载机是否进入复飞区, 复飞是否安全等. 文献[8]根据规划的切比雪夫复飞路径计算复飞边界条件并验证了方法的可靠性. 文献[9]和[10]提出了一种基于神经网络的舰载机复飞风险评估方法. 文献[11]提出了复飞决策关键位置(wave-off decision critical position, WODCP)的概念, 设计了基于WODCP的复飞风险估计策略. 文献[12]搭建了自动着舰仿真系统模型, 对着舰安全准则及复飞边界准则进行改进, 确定了着舰复飞决策步骤. 文献[13]研究了基于飞行安全区的复飞决策方法, 提出了基于复飞风险计算的复飞决策流程. 文献[14]定义了自动着舰安全窗口, 若舰载机超出安全范围, 则执行复飞操纵. 复飞操纵策略的合理性极

引用格式: 张志冰, 王家兴, 袁婷慧. 基于PER-SAC的舰载机着舰复飞操纵策略研究. 中国科学: 技术科学, 2025, 55: 1067–1076

Zhang Z B, Wang J X, Yuan T H. Wave-off manipulation strategy for carrier-based aircraft based on PER-SAC (in Chinese). Sci Sin Tech, 2025, 55: 1067–1076, doi: 10.1360/SST-2024-0133

大程度决定了复飞是否成功, 通过优化升降舵操纵可以提高复飞性能, 缩小危险区域. 文献[15]基于Radau伪谱法给出了舰载机着舰的最优复飞轨迹, 并提出一种基于偏差模型的动态面控制算法, 设计了速度误差子系统和高度误差子系统给出了推力操纵与升降舵操纵, 实现了轨迹的精确控制, 但是动态面控制对模型依赖程度较高, 算法比较复杂. 因此, 将复飞操纵问题转化为优化问题, 引入智能算法, 以提高自适应能力是解决复飞问题的另一种途径. 文献[16]和[17]将升降舵操纵定义为多目标优化问题, 并进行求解得到复飞升降舵操纵律. 文献[18]设计了一种基于非支配排序遗传算法的升降舵与军用推力相结合的复飞策略, 提高了舰载机复飞的安全性. 而强化学习算法相比于遗传算法可以更有效地处理连续状态下的控制问题, 更加适合处理舰载机着舰复飞问题. 文献[19]引入深度强化学习解决无人机自主飞行任务, 使用强化学习算法完成不同场景的舰载机自主控制任务, 实现了复杂环境下的操纵任务.

然而, 上述研究没有充分考虑舰载机安全复飞的全部约束条件, 且常规升降舵操纵优化方法无法得到安全复飞所需的最优升降舵操纵指令, 大大限制了舰载机的性能. 而基于神经网络的深度强化学习技术, 通过与环境进行不断交互, 可以生成特定情境下的全局最优策略, 已经在自动驾驶^[20]、飞行控制^[21]等领域取得了突出成果.

本文针对舰载机安全复飞问题展开了研究. 主要创新点有: (1)构建舰载机复飞系统, 根据着舰复飞真实场景对传统的安全复飞准则进行补充, 得到更为真实的复飞边界; (2)基于复飞指标构建马尔可夫决策过程模型, 采用SAC深度强化学习算法设计了军用推力结合升降舵操纵的复飞优化策略, 为安全边界约束下的复飞路径规划提供了快速的重新规划; (3)为了提升SAC算法的采样效率, 采用随机优先级抽样对常规采样方式进行改进, 使用重要性抽样修正了由状态分布改变导致的收敛偏差, 加快了算法的收敛速度.

2 舰载机动力学模型

由于本文考虑的是舰载机着舰复飞的操纵策略, 主要考虑舰载机在平静空气环境中的纵向控制输入响应, 本节采用线性化方案构建纵向飞行动力学的线性

模型:

$$\begin{cases} \dot{V} = \frac{T \cos \alpha - D}{m} + g \sin \gamma, \\ \dot{\alpha} = \frac{g}{V} \cos \gamma + q - \frac{T \sin \alpha + L}{mV}, \\ \dot{\theta} = q, \\ \dot{q} = \frac{M}{I_y}, \\ \dot{h} = V \sin \gamma, \\ \gamma = \theta + \alpha, \end{cases} \quad (1)$$

式中, D , L , T 分别表示舰载机阻力、升力和发动机推力; V , α , θ , γ , q , h 分别表示舰载机空速、迎角、俯仰角、轨迹角、俯仰角速率和高度; m 表示舰载机质量; g 表示重力加速度; M 表示俯仰力矩; I_y 表示绕机体系 Y 轴的转动惯量.

舰载机的空气动力和力矩可以表示为

$$\begin{aligned} T &= T_{\max} \delta_T, \\ M &= \frac{1}{2} \rho V^2 S c \left(C_M^0 + C_M^a \alpha + C_M^q q + C_M^{\delta_e} \delta_e \right), \\ D &= \frac{1}{2} \rho V^2 S \left(C_D^0 + C_D^a \alpha + C_D^{a^2} \alpha^2 \right), \\ L &= \frac{1}{2} \rho V^2 S \left(C_L^0 + C_L^a \alpha + C_L^{\delta_e} \delta_e \right), \end{aligned} \quad (2)$$

式中, $T_{\max}$ 是发动机最大推力, ρ 是空气密度, S 是参考机翼面积, c 是平均气动弦长. C_M^0 , C_M^a , C_M^q , $C_M^{\delta_e}$, C_D^0 , C_D^a , $C_D^{a^2}$, C_L^a , $C_L^{\delta_e}$ 是气动力学参数.

通过小扰动方法对式(1)的非线性模型进行线性化, 得到配平状态的纵向线性化方程组

$$\dot{\mathbf{x}} = A_{\text{lon}} \mathbf{x} + B_{\text{lon}} \mathbf{u}, \quad (3)$$

式中, 状态变量 $\mathbf{x} = [\Delta V, \Delta \alpha, \Delta \theta, \Delta q, \Delta h]^T$ 表示舰载机空速、迎角、俯仰角, 俯仰角速率和高度增量; 输入变量 $\mathbf{u} = [\Delta \delta_e, \Delta \delta_T]^T$ 表示升降舵和油门增量; A_{lon} , B_{lon} 表示纵向线性化状态方程的系统矩阵和控制矩阵.

3 着舰复飞边界设计

3.1 复飞轨迹计算

舰载机复飞轨迹求解结构如图1所示, x_0 , h_0 , V_0 和 γ_0 分别表示舰载机执行复飞操纵的初始位置、高度、空速和轨迹角, Δx , Δh 和 $\Delta \gamma$ 分别表示舰载机位置

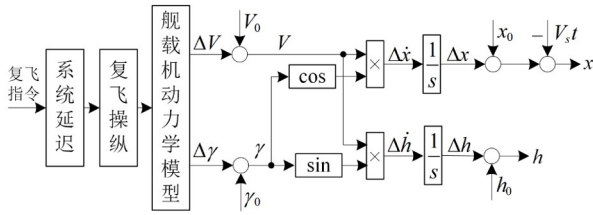


图 1 复飞轨迹求解结构

Figure 1 Structure for solving the go around trajectory.

增量、高度增量以及轨迹角增量, $\Delta \dot{x}$ 和 $\Delta \dot{h}$ 表示位置变化率增量和高度变化率增量, V_s 表示航母的前进速度. 在着舰过程中, 为保证舰载机能够及时响应复飞指令, 通常要求发动机工作点位于最大军用推力的 85%~90%.

确定轨迹时最重要的因素之一是初始和最终边界条件. 对于初始路径, 假设舰载机以一定的轨迹角和初始速度运动在离舰船甲板一定距离的位置. 结束条件是船舶甲板中心的零位置. 舰载机轨迹复飞开始于初始状态 x_0, h_0 .

舰载机复飞轨迹方程可描述为

$$\begin{cases} x(t) = x_0 + \int (V_0 + \Delta V) \times \cos(\gamma_0 + \Delta \gamma) dt - V_s t, \\ h(t) = h_0 + \int (h_0 + \Delta h) \times \sin(\gamma_0 + \Delta \gamma) dt. \end{cases} \quad (4)$$

3.2 复飞边界设计

在着舰过程中存在一组轨迹临界点, 一旦舰载机越过这些临界点, 则无论如何操纵都无法安全复飞. 以轨迹临界点为复飞起点的轨迹包线称为复飞边界, 复飞边界形成的区域称为复飞风险区^[22]. 复飞边界由安全复飞准则确定, 安全复飞准则是综合考虑舰载机复飞的安全高度、系统延迟等因素确定的, 具体准则如下:

- (1) 当舰载机飞越航母时, 飞机尾钩与舰尾之间至少有 3 米的高度间隙;
- (2) 飞行控制系统对复飞指令的理想响应时间为 0.2 秒;
- (3) 复飞操纵仅包含军用推力操纵.

然而上述准则未考虑舰载机在离航母不同距离时复飞风险的差异, 不足以确保复飞安全.

若舰载机在距离航母较近时执行复飞操纵, 在到达舰尾时, 舰载机下沉速率可能不为零, 此时高度会

继续下降, 尾钩或起落架可能会与甲板发生碰撞, 进而影响复飞安全; 若舰载机在距离航母较远时复飞, 舰载机的复飞轨迹最低点将接近海平面, 有坠海的风险. 在真实着舰下滑过程中, 将甲板下方 5 米水平线定义为安全基准线, 舰载机的复飞轨迹必须始终高于安全基准线.

因此, 在传统复飞准则的基础上进行补充:

- (1) 当舰载机飞至甲板上方且其下沉速率为零时, 尾钩与甲板至少有 3 米的高度间隙;
- (2) 复飞轨迹最低点应高于安全基准线.

结合上述安全复飞准则并进行大量模拟计算, 获取改进后的复飞边界, 如图 2 所示.

相比于传统复飞边界, 改进后的复飞边界范围扩大, 更符合真实的复飞情景.

4 基于 PER-SAC 的复飞操纵策略研究

4.1 基于 SAC 的升降舵操纵优化

舰载机复飞过程如图 3 所示, 为保证舰载机安全, 应基于复飞指标对复飞操纵策略进行优化, 本文所选用的复飞指标如下:

- (1) 舰载机从复飞初始位置到复飞轨迹的最低点时的水平距离 x ;
 - (2) 舰载机越过舰尾时, 飞机尾钩距舰尾的高度 h .
- 指标(1)是为了减少进场次数, 提高着舰效率; 指标(2)是保证舰载机越过舰尾时的高度最大, 降低撞舰风险.

传统方法仅采用军用推力执行复飞, 速度增大时迎角减小, 增加了复飞风险. 现代舰载机由于机身性能的提升, 在军用推力的基础上增加了升降舵操纵, 以减小高度损失. 常规优化方法忽略了升降舵偏转速率的限制, 且将复飞过程中升降舵的操纵变化量仅用多个参数简化表示, 大大限制了舰载机性能.

SAC 算法在 Actor-Critic 基础上结合最大熵原理, 在优化策略以获取更高累计收益的同时, 也会最大化策略的熵. 选择具有最大熵的策略能够保持策略的多样性, 以便智能体对外界环境进行充分探索, 从而避免局部最优^[23]. 考虑到复飞环境的复杂性和随机性, 基于 SAC 算法设计了军用推力结合升降舵的复飞操纵策略, 使用连续变化的曲线描述升降舵的偏转过程, 以获取最优复飞操纵策略.

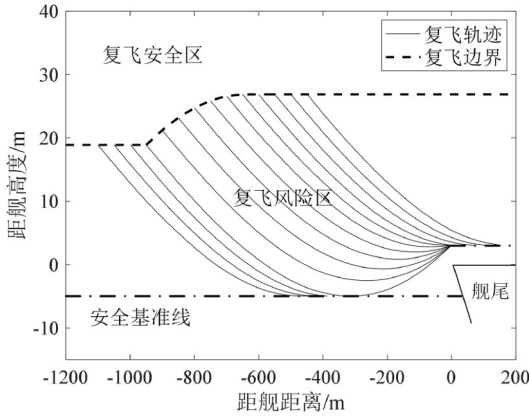


图2 改进复飞边界

Figure 2 Improved go around boundary.

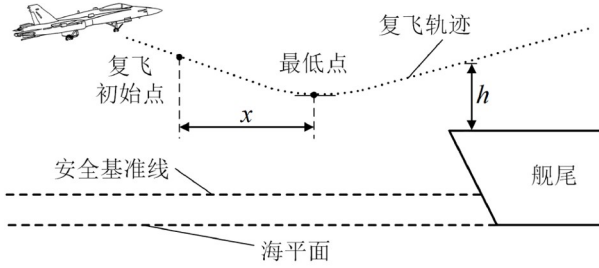


图3 复飞过程示意图

Figure 3 Schematic diagram of go around process.

4.2 马尔可夫决策过程建模

智能体与环境之间的交互被称为马尔可夫决策过程(Markov decision process, MDP) [24], MDP可以表示为 $M = (S, A, P, R, \gamma)$. 其中, S 为状态空间, A 为动作空间, $P(\cdot | s, a)$ 为状态转移概率分布函数, R 为状态转移的奖励回报, γ 为折扣因子. 在本文中, 舰载机状态的转移过程可以视为MDP:

$$P(s_{t+1}) = \mathbb{E}(S_{t+1} = s_{t+1} | S_t = s_t, A_t = a_t). \quad (5)$$

本文将SAC算法与优先经验回放技术相结合, 利用SAC算法实现复飞操纵完成的自主性与智能性, 并引入了一种优先体验回放方法来降低采样的随机性, 通过优先学习高价值的经验, 可以增强复飞操纵策略的性能. 基于PER-SAC的训练框架如图4所示.

本文的目的是基于复飞优化指标确定升降舵的最优操纵量, 因此在图4中, $s_t = \{x, h, \alpha, u, w, q\}$, 其中, u, w 分别是舰载机速度在机体轴坐标系下沿 X 轴和 Z

轴方向的分量. MDP动作空间即为升降舵相对于初始配平点的偏转角度, 即 $a_t = \Delta \delta_e$. 此外, 考虑到升降舵实际偏转限制, 因此对升降舵偏转角度进行限幅 $\delta_e \in [-24^\circ, 10.5^\circ]$ 和限速 $|\dot{\delta}_e| \leq 40^\circ/\text{s}$.

奖励函数反映了在 t 时刻的状态 s_t 下, 执行动作 a_t 可获得的回报, 本文基于复飞优化指标, 设计了如下奖励函数:

$$r_t = \omega_1 \dot{h} - \omega_2 \frac{1}{\alpha_{\max} - \alpha} - \omega_3 |\dot{\delta}_e|, \quad (6)$$

式中, ω_i 表示对应奖励项的系数; $\dot{h}$ 表示舰载机爬升速率; $\alpha_{\max} = 25^\circ$ 表示失速迎角. 第一项奖励是使得舰载机尽可能快速爬升, 第二项奖励是避免产生失速迎角, 第三项奖励是防止舵面偏转过快导致机身状态发生振荡. 本文把对迎角和舵面偏转速率的硬约束转换成软约束并结合到奖励中, 以获得最优的复飞性能.

4.3 SAC强化学习算法设计

标准强化学习算法的目的是学习一个最优策略 $\pi^*(a_t | s_t)$, 使得奖励的期望总和 $\sum_t \mathbb{E}_{(s_t, a_t) \sim \rho_\pi} [r(s_t, a_t)]$ 最大 [24], 其中, $\rho_\pi(s_t, a_t)$ 表示在策略 π 下轨迹出现 s_t 和 a_t 的概率. SAC算法在目标函数中引入了策略熵, 要求最优策略 $\pi^*(a_t | s_t)$ 不仅能最大限度地提高奖励回报, 还应具有最大熵, 引入策略熵后目标函数的定义如下:

$$J_\pi = \sum_t \mathbb{E}_{(s_t, a_t) \sim \rho_\pi} [r(s_t, a_t) + \beta \mathcal{H}(\pi(\cdot | s_t))], \quad (7)$$

式中, β 表示温度系数, 其决定了熵相对于奖励的重要程度; $\mathcal{H}(\pi(\cdot | s_t))$ 表示策略熵, 它描述了策略 $\pi(\cdot | s_t)$ 的不确定性, 熵越大则策略的选取越随机, 熵可通过下式计算:

$$\mathcal{H}(\pi(\cdot | s_t)) = \mathbb{E}_{a \sim \pi(\cdot | s_t)} [-\log \pi(a | s_t)]. \quad (8)$$

在该算法中, 目的是获取使得 J_π 最大化的最优策略, 即

$$\pi^*(a_t | s_t) = \operatorname{argmax}_{J_\pi}. \quad (9)$$

与标准强化学习算法类似, 在SAC算法中, 也采用软状态价值函数 $V_{\text{soft}}^\pi(s)$ 和软动作价值函数 $Q_{\text{soft}}^\pi(s, a)$ 来评价策略的好坏. 因此, 基于最大熵框架交替进行的策

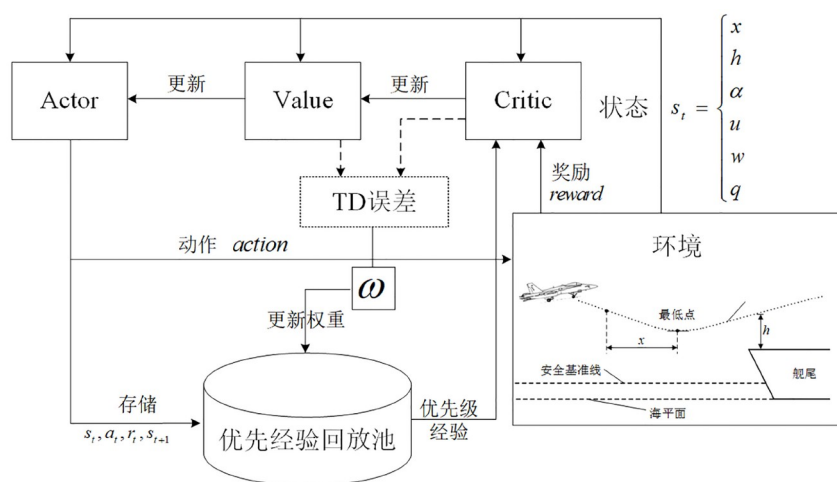


图 4 基于PER-SAC的训练框架

Figure 4 Training framework based on PER-SAC.

略评估和策略改进也被称为软策略迭代, 在软策略迭代的策略评估步骤中, $V_{\text{soft}}^{\pi}(s)$ 和 $Q_{\text{soft}}^{\pi}(s, a)$ 的更新如下:

$$Q_{\text{soft}}^{\pi}(s, a) = \mathbb{E}_{s' \sim p(s'|s, a)} [r(s, a) + \gamma V_{\text{soft}}^{\pi}(s')], \quad (10)$$

$$V_{\text{soft}}^{\pi}(s) = \mathbb{E}_{a \sim \pi}[Q_{\text{soft}}^{\pi}(s, a) - \beta \log \pi(a|s)]. \quad (11)$$

通过式(9)和式(10)的不断迭代, 最终可得到针对任意策略 π 下 $Q_{\pi}^{\pi}(s, a)$ 的收敛值.

在策略改进步骤中, 常规离线策略方法通过最大化动作价值函数 $Q^\pi(s, a)$ 来进行策略改进, 而SAC算法中, 策略不再采用高斯分布的形式, 而是采用基于能量的模型来表示, 即

$$\pi(a_t|s_t) \propto \exp(-\mathcal{E}(s_t, a_t)), \quad (12)$$

式中, $\mathcal{E}(s_t, a_t) = -\frac{1}{\beta} Q_{\text{soft}}(s_t, a_t)$ 表示能量函数, 可用神经网络来拟合.

由于大多数强化学习问题的解不唯一，当存在多个最优解或次优解时，常规的高斯分布只能将策略更新集中于更高的 Q 值附近，忽略了其他可行解，而基于能量模型的策略分布可以对所有可行解进行学习，具有更强的策略表达能力。

式(11)为SAC算法策略分布的理想形式, 但能量模型无法像常规的高斯分布一样直接对其采样, 所以在实际应用中, 仍然采用高斯分布来代替能量模型进行环境交互, 通过最小化KL散度使得该高斯分布尽可

能接近基于能量模型的策略分布. 因此, 针对每个状态, 根据下式进行策略改进:

$$\pi_{\text{new}} = \underset{\pi' \in \Pi}{\text{argmin}} \text{D}_{\text{KL}} \left(\pi'(\cdot | s_t) \left\| \frac{\exp \left(\frac{1}{\beta} Q^{\pi_{\text{old}}}(s_t, \cdot) \right)}{Z^{\pi_{\text{old}}}(s_t)} \right\| \right), \quad (13)$$

式中, Π 表示可选策略的集合, 本文为高斯分布集合; D_{KL} 表示KL散度; $Z^{\pi_{old}}(s_t)$ 表示配分函数, 用于归一化分布, 由于其相对于 π_{new} 是一个常数, 因此可直接忽略.

交替使用式(11)、式(12)的策略评估以及式(13)的策略改进，SAC算法最终可收敛至 Π 中的最优策略 $\pi^*(a_t|s_t)$.

然而,前文的推导过程是基于离散动作空间和状态空间的,考虑到舰载机复飞过程中升降舵变化的连续性,需要将SAC算法推广至连续域,因此在本文的研究中,采用函数逼近器,如神经网络来近似表示软状态价值函数 $V_{\phi}(s_t)$ 、软动作价值函数 $Q_{\phi}(s_t, a_t)$ 及策略分布 $\pi_{\phi}(a_t|s_t)$,根据公式(9)可知,状态价值函数 $V_{\phi}(s_t)$ 可用 Q 函数 $Q_{\phi}(s_t, a_t)$ 表示,因此原则上不需要为状态价值函数设计单独的函数逼近器.在训练中, $Q_{\phi}(s_t, a_t)$ 函数的参数 ϕ 可以通过最小化式(14)所示的软贝尔曼残差

得到

$$J_Q(\varphi) = \mathbb{E}_{(s_t, a_t) \sim \mathcal{D}} \left[\frac{1}{2} \left(Q_\varphi(s_t, a_t) - \widehat{Q}(s_t, a_t) \right)^2 \right], \quad (14)$$

式中, $\widehat{Q}(s_t, a_t) = r(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1} \sim p} [V_\varphi(s_{t+1})]$, 值函数的参数 φ 通过更新 φ 进行隐式更新; $\mathcal{D}$ 表示重放缓存区; $V_\varphi(s_{t+1})$ 为目标网络, 目的是保持更新过程中软 Q 网络的稳定. SAC算法中 Q 网络的更新方式与DQN算法的梯度下降算法基本一致, 其中软 Q 网络损失函数的梯度表示为

$$\nabla_\varphi J_Q(\varphi) = \nabla_\varphi Q_\varphi(a_t, s_t) \left(Q_\varphi(s_t, a_t) - r(s_t, a_t) - \gamma V_\varphi(s_{t+1}) \right). \quad (15)$$

通过最小化式(13)中的KL散度以更新策略网络的参数, 策略网络 $\pi_\phi(a_t|s_t)$ 的损失函数如下:

$$J_\pi(\phi) = \mathbb{E}_{s_t \sim \mathcal{D}, a_t \sim \pi_\phi} [\beta \log \pi_\phi(a_t|s_t) - Q_\varphi(s_t, a_t)]. \quad (16)$$

由于策略 $\pi_\phi(a_t|s_t)$ 是一个分布, a_t 通过采样得到, 因此无法直接对式(16)直接求导, 因此使用重参数化技巧来采样, 此时有

$$a_t = f_\phi(\mu_t; s_t), \quad (17)$$

式中, μ_t 为服从正态分布的随机噪声向量; $f_\phi(\cdot)$ 表示关于参数 ϕ 的非线性函数. 将式(17)代入式(16), 可以得到策略网络的损失函数

$$J_\pi(\phi) = \mathbb{E}_{s_t \sim \mathcal{D}, \mu_t \sim \mathcal{N}} [\beta \log \pi_\phi(f_\phi(\mu_t; s_t)|s_t) - Q_\varphi(s_t, f_\phi(\mu_t; s_t))], \quad (18)$$

对变量 ϕ 求偏导, 得到式(19)的梯度

$$\begin{aligned} \nabla_\phi J_\pi(\phi) = & \nabla_\phi \beta \log(\pi_\phi(a_t|s_t)) \\ & + \left(\nabla_{a_t} \beta \log(\pi_\phi(a_t|s_t)) \right. \\ & \left. - \nabla_{a_t} Q(s_t, a_t) \right) \nabla_\phi f_\phi(\mu_t; s_t). \end{aligned} \quad (19)$$

在训练过程中, 分别朝式(15)和式(19)表示的梯度方向更新参数 φ 和 ϕ , 算法最终可收敛至最优解.

式(7)中提到, 温度系数 β 决定了熵相对于奖励的重要程度, 对SAC算法的性能具有重大影响. 因此在训练过程中, 应不断调节 β , 以适应不同场景下智能体的探索能力.

损失函数的推导参考文献[25], 最终可以得到关于 β 的损失函数

$$J_\pi(\beta) = \mathbb{E}_{a_t \sim \pi_t} [-\beta \log \pi_t(a_t|\pi_t) - \beta \overline{\mathcal{H}}], \quad (20)$$

式中, $\overline{\mathcal{H}}$ 为最小期望熵的阈值. β 朝着 $J_\pi(\beta)$ 的梯度方向更新, 以适应训练的不同阶段.

4.4 优先经验重放

常规SAC算法将历史探索经验存储在经验重放缓冲区中, 然后通过随机抽样获取样本. 然而, 这种采样方式未考虑样本的重要性, 导致采样效率低下. 先前研究表明, TD误差可以作为衡量经验样本重要性的指标, 但仅基于TD误差大小进行抽样会导致算法的过拟合^[26]. 为提高采样效率, 并保证算法的精度, 结合随机优先级抽样的经验重放技术, 提出PER-SAC算法. 随机优先级抽样将贪婪优先抽样与均匀随机抽样相结合, 每个经验样本的采样概率定义为

$$P(i) = \frac{p_i^\lambda}{\sum_{k=1}^K p_k^\lambda}, \quad (21)$$

式中, p_i 表示第 i 个经验样本的优先级, 由TD误差确定; λ 为优先级系数, 用于调节该经验的优先级, $\lambda = 0$ 时退化为随机抽样. 其中, $p_i > 0$ 可以表示为

$$p_i = |\delta_i| + \epsilon, \quad (22)$$

式中, $|\delta_i|$ 为TD误差; ϵ 是一个很小的正常数, 用于防止抽样概率为0.

由于优先经验重放改变了状态分布, 不可避免地会引入偏差, 为了纠正偏差, 引入经验 i 对应的重要性抽样权重

$$\omega_i = \left(\frac{1}{N} \cdot \frac{1}{P(i)} \right)^\sigma, \quad (23)$$

式中, N 为训练批次的样本数目; σ 为补偿系数. 此时式(14)的 Q 网络梯度可改写成 $\omega_i \nabla_\varphi J_Q(\varphi)$.

使用重要性抽样修正了由状态分布改变导致的收敛偏差, 并且保证了随机优先级抽样的采样效率.

5 仿真结果

为验证所提出的复飞边界计算方法和操纵策略的

合理性和有效性, 本文以F/A-18为研究对象, 根据舰载机着舰复飞场景, 对所提方法进行仿真验证。

舰载机初始状态设为空速 $V_0 = 68.57 \text{ m/s}$, 迎角 $\alpha_0 = 7.54^\circ$, 下滑角 $\gamma_0 = -3.5^\circ$, 升降舵初始偏转角度 $\delta_{e0} = -4.54^\circ$. 发动机初始工作点设为最大军用推力的87%, 当接收到复飞指令, 经0.2 s系统延迟后立即增加至100%. 舰尾初始位置记为(0,0), 规定航母运行方向为 X 轴正方向, 垂直于海平面向上为 Y 轴正方向. PER-SAC算法超参数如表1所示。

为了评估该算法的性能, 本文将PER-SAC算法与常规SAC算法进行了比较, 训练过程两种算法的平均奖励回报如图5所示。

从图5可见, PER-SAC算法的奖励迅速增加, 并且在迭代次数340就开始收敛, 最终收敛至7938. 而常规SAC算法不仅收敛速度较慢, 而且平均奖励回报最终也只稳定在6305, 比PER-SAC算法更低, 可能是陷入了局部最优, 说明常规SAC算法收敛速度慢、采样效率低下, 而PER-SAC算法有效提高了优化效果和收敛速度。

训练完成后, 基于式(4)所示的复飞轨迹方程, 计算不同距舰高度和距舰距离下的舰载机复飞轨迹. 此外, 将本文提出的基于PER-SAC的升降舵操纵与仅采用军用推力操纵、基于待定参数法(method of undetermined parameters, MUP)的升降舵操纵以及基于常规SAC的升降舵操纵进行了比较. 图6(a)和(b)分别显示了舰载机初始位置为(-400,20)和(-200,10)的复飞轨迹。

从图6(a)和(b)可见, 在相同初始条件下, 采用基于PER-SAC的升降舵操纵能够更快拉起舰载机, 其次是

表 1 PER-SAC超参数

Table 1 PER-SAC hyper-parameters

超参数	取值
最大幕数	4000
重放缓存池容量	500000
训练批次大小	256
折扣因子	0.99
目标网络更新因子	0.001
学习率	0.0001
激活函数	ReLU
初始温度系数	0.2
优先级系数	0.6
补偿系数	0.45

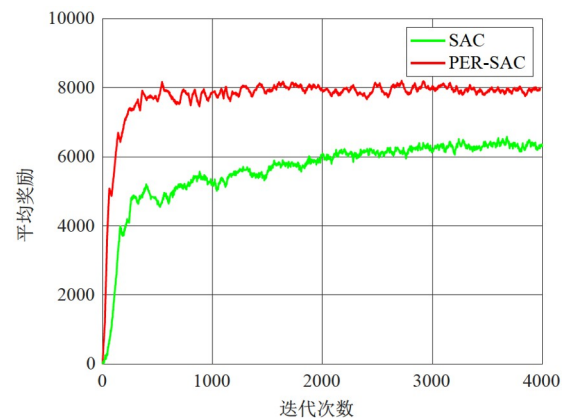


图 5 (网络版彩图)平均奖励变化曲线

Figure 5 (Color online) Average reward change curve.

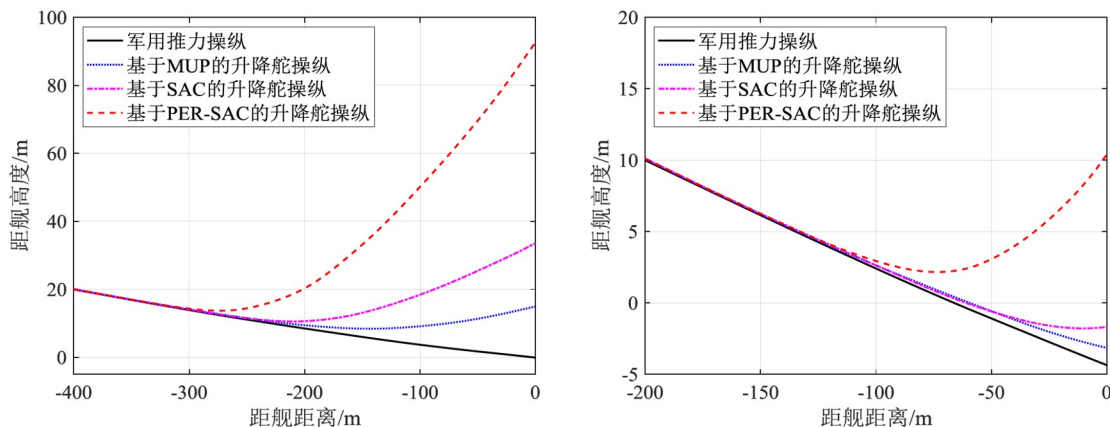


图 6 (网络版彩图)不同初始位置下复飞轨迹. (a)初始位置(-400, 20); (b)初始位置(-200, 10)

Figure 6 (Color online) Go around trajectory at different initial positions. (a) (-400, 20); (b) (-200, 10).

基于常规SAC的升降舵操纵, 而采用基于MUP的升降舵操纵和军用推力操纵的复飞性能较差. 此外, 从图6(b)中可见, 当舰载机的初始位置为 $(-200, 10)$ 时, 只能采用基于PER-SAC算法的军用推力结合升降舵的复飞操纵策略才能安全复飞, 说明该操纵策略大大提升了舰载机的复飞效果以及着舰的安全性.

采用本文所提出策略与MUP和SAC两种策略的舰载机升降舵变化和迎角响应对比曲线分别如图7(a)和(b)所示.

由图7(b)可见, 在复飞过程中迎角最大为 24.1° , 尚未超出失速迎角, 这得益于在奖励函数式(5)设计中引入了防止失速迎角的惩罚项.

由表2可知基于PEC-SAC的升降舵操纵相比于SAC和MUP在允许的操纵范围具有更大的操纵量, 更快的操纵响应, 能将舰载机从复飞边界更快更好地拉

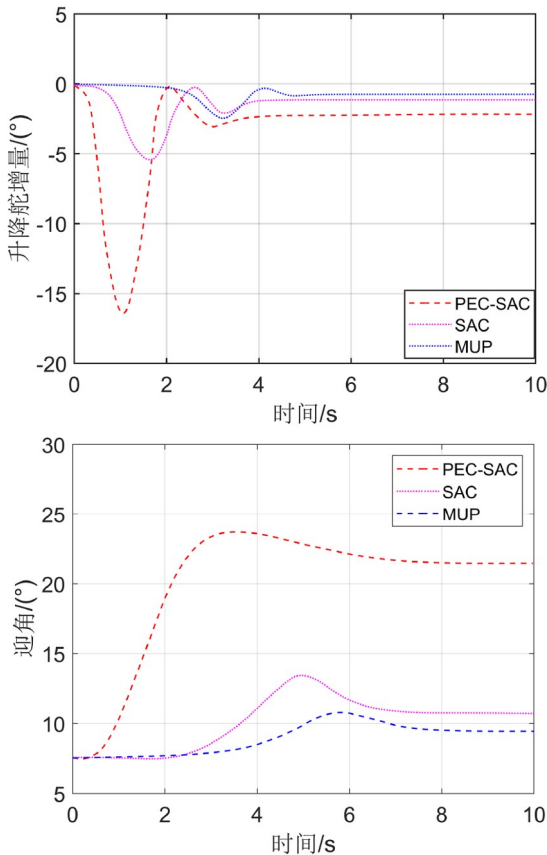


图7 (网络版彩图)升降舵变化对比曲线和迎角响应对比曲线。(a)升降舵变化曲线;(b)迎角响应曲线

Figure 7 (Color online) Comparison curves of elevator changes and angle of attack response. (a) Elevator response curves; (b) angle of attack response curves.

表2 PEC-SAC, SAC, MUP操纵性能对比

Table 2 Comparison of PEC-SAC, SAC, and MUP manipulation performance

方法	升降舵响应角度范围	迎角响应范围
PEC-SAC	$[-16.4^\circ, 0^\circ]$	$[7.1^\circ, 24.1^\circ]$
SAC	$[-5^\circ, 0^\circ]$	$[7.1^\circ, 13.7^\circ]$
MUP	$[-2.5^\circ, 0^\circ]$	$[7.1^\circ, 11.9^\circ]$

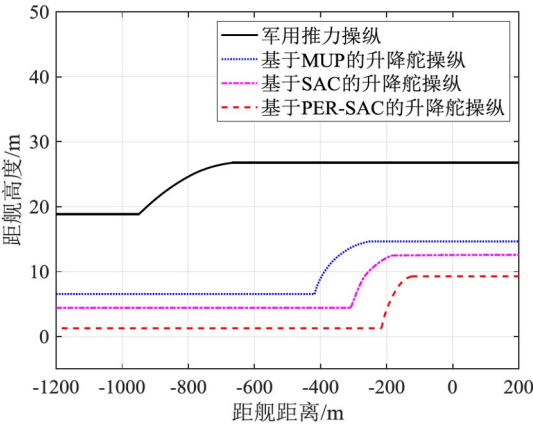


图8 (网络版彩图)不同操纵策略下的复飞边界

Figure 8 (Color online) The go around boundary under different manipulation strategies.

起, 拓宽了安全边界范围.

将不同距舰高度和距舰距离下得到的一系列复飞轨迹进行拟合, 最终得到的复飞边界如图8所示.

复飞边界区域越大, 说明舰载机复飞失败的可能性最大, 从图8中可以看出, 采用PER-SAC算法的升降舵操纵策略所获得的复飞边界区域最小, 显著提高了舰载机着舰复飞的安全性.

6 结论

本文研究了舰载机复飞边界和复飞操纵策略. 对常规安全复飞准则进行了改进, 通过大量仿真模拟得到改进的复飞边界. 为提高采样效率, 结合随机优先级抽样的经验重放技术, 提出了PER-SAC算法, 并基于PER-SAC算法设计了军用推力结合升降舵操纵的复飞操纵策略, 避免了传统升降舵操纵优化策略的缺陷. 仿真结果表明, 本文提出的复飞操纵策略能够在避免失速的前提下更快拉起舰载机, 可以显著减小复飞风险区域, 提高舰载机的复飞性能.

参考文献

- 1 Zhang Z, Peng Y, Liang T, et al. Research on lateral dynamics safety margins of carrier-based aircraft arresting. *Aeronaut J*, 2022, 126: 1593–1615
- 2 Duan H, Chen L, Zeng Z. Automatic landing for carrier-based aircraft under the conditions of deck motion and carrier airwake disturbances. *IEEE Trans Aerosp Electron Syst*, 2022, 58: 5276–5291
- 3 Yong Y, Yue G, Lin Z, et al. Research on impact of sea state on carrier-based aircraft landing performance with airwake coupled. In: Proceedings of the 33rd Chinese Control and Decision Conference (CCDC), 2021. 5900–5905
- 4 Zhen Z Y, Wang X H, Jiang J, et al. Research progress on automatic landing guidance and control of carrier-based aircraft. *Acta Aeronaut Astronaut Sin*, 2017, 38: 127–148 [甄子洋, 王新华, 江驹, 等. 舰载机自动着舰引导与控制研究进展. *航空学报*, 2017, 38: 127–148]
- 5 Johnstone R B. Development of the waveoff decision device. *J Aircraft*, 1969, 6: 367–371
- 6 Jiao X, Jiang J, Wang X H, et al. Research on comprehensive go-around decision for carrier-based aircraft. *Flight Dyn*, 2012, 30: 405–409 [焦鑫, 江驹, 王新华, 等. 舰载机综合复飞决策研究. *飞行力学*, 2012, 30: 405–409]
- 7 Jiao X. Research on carrier-based aircraft landing environment and go-around decision technology. Dissertation for Masteral Degree. Nanjing: Nanjing University of Aeronautics and Astronautics, 2013 [焦鑫. 舰载机着舰环境与复飞决策技术研究. 硕士学位论文. 南京: 南京航空航天大学, 2013]
- 8 Jaques, Emma. Receding horizon control for autonomous UAS shipboard landing with Tau guidance and polynomial trajectory planning. Dissertation for Masteral Degree. Pennsylvania: Pennsylvania State University, 2023
- 9 Jiang H, Li H, Su X. Landing risk assessment of carrier-based aircraft based on fuzzy multi-attribute group decision making. *Int J Comput Appl*, 2015, 8: 169–178
- 10 Li H, Wang Y, Su X D, et al. Research on go-around risk decision technology based on neural networks. *Comput Simul*, 2021, 38: 32–37 [李晖, 王雅, 苏晓东, 等. 基于神经网络的复飞风险决策技术研究. *计算机仿真*, 2021, 38: 32–37]
- 11 Li H, Su X, Zhang Y, et al. Research on wave-off risk estimation based on WODC, In: Proceedings of IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2019. 610–613
- 12 Li J T, Xing Z L, Ai J L. Decision method and fault analysis for go-around of carrier-based UAV landing. *Flight Dyn*, 2022, 40: 40–46 [李婧婷, 邢振林, 艾剑良. 舰载无人机着舰复飞决策方法及故障分析. *飞行力学*, 2022, 40: 40–46]
- 13 Li H, Zhao M, Li Y, et al. Research on wave-off decision making based on flight safe area. In: Proceedings of 2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2019. 313–317
- 14 Zhimin H, Guanxin H. Analysis of security window in automatic landing of the carrier-borne aircraft. *Res J Appl Sci Eng Technol*, 2014, 7: 2874–2879
- 15 Cui K K, Han W, Liu Y J, et al. Automatic go-around control algorithm for carrier-based aircraft based on DM-DSC. *J Beijing Univ Aeronaut Astronaut*, 2023, 49: 900–912 [崔凯凯, 韩维, 刘玉杰, 等. 基于DM-DSC的舰载机着舰自动复飞控制算法. *北京航空航天大学学报*, 2023, 49: 900–912]
- 16 Shen H L, Gong Z. Research on go-around decision and maneuvering for carrier-based aircraft. *Flight Dyn*, 2008, 26: 5–9 [沈宏良, 龚正. 舰载飞机复飞决策与操纵研究. *飞行力学*, 2008, 26: 5–9]
- 17 Du J, Jiang J. Simulation Study on the Elevator control performance of carrier-based aircraft during go-around and landing. *Comput Simul*, 2016, 33: 70–73 [杜洁, 江驹. 舰载机复飞着舰升降舵操纵性能仿真研究. *计算机仿真*, 2016, 33: 70–73]
- 18 Zhang L, Wang L, Xing Z, et al. Research on wave-off decision and elevator manipulation optimization of carrier-based UAV. *Adv Guid Navig Control*, 2023, 845: 5342–5354
- 19 Song M H. Research and implementation of autonomous flight control algorithm for UAV based on deep reinforcement learning. Dissertation for Masteral Degree. Beijing: Beijing Jiaotong University, 2022 [宋明惠. 基于深度强化学习的无人机自主飞行控制算法的研究与实现. 硕士学位论文. 北京: 北京交通大学, 2022]
- 20 Ye F, Cheng X, Chan C Y, et al. Automated lane change strategy using proximal policy optimization-based deep reinforcement learning. In: Proceedings of IEEE Intelligent Vehicles Symposium (IV), 2020. 1746–1752
- 21 Zhang S, Zhou P, He Y, et al. Experimental study on air combat maneuver decision based on deep reinforcement learning. *Acta Aeronaut*

- Astronaut Sin, 2023, 44: 122–135
- 22 Zhang S, Zhou P, He Y, et al. Air combat maneuver decision-making test based on deep reinforcement learning. *Acta Aeronaut Astronaut Sin*, 2023, 44: 128094 [章胜, 周攀, 何扬, 等. 基于深度强化学习的空战机动决策试验. *航空学报*, 2023, 44: 128094]
- 23 Yin H T, Li Y. Research on key technologies of carrier-based aircraft landing. In: *Proceedings of 6th International Conference on Robotics and Automation Engineering (ICRAE)*, 2021. 65–70
- 24 HAarnno T, Tang H, Abbeel P, et al. Reinforcement learning with deep energy-based policies. In: *Proceedings of the 34th International Conference on Machine Learning*, 2017. 1352–1361
- 25 Lilicrap T P, Hunt J J, Pritze A, et al. Continuous control with deep reinforcement learning. In: *Proceedings of 4th International Conference on Learning Representations (ICLR)*, 2016. 1–14
- 26 Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: *Proceedings of Machine Learning Research*, 2018. 1861–1870
- 27 Zhang L F, Feng Y H, Liang X X, et al. Sample strategy based on TD-error for offline reinforcement learning. *Chinese J Eng*, 2023, 45: 2118–2128

Wave-off manipulation strategy for carrier-based aircraft based on PER-SAC

ZHANG ZhiBing¹, WANG JiaXing¹ & YUAN TingHui²

¹ *AVIC Shenyang Aircraft Design and Research Institute, Shenyang 110035, China*

² *School of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*

To ensure the safety of landing and wave-off for carrier-based aircraft, a carrier-based aircraft wave-off manipulation strategy based on prioritized experience replay (PER) and soft actor-critic (PER-SAC) is proposed in this paper. Considering the large difference in wave-off risk of carrier-based aircraft at different distances from the carrier, the wave-off boundary is optimized by improving the safe wave-off criteria. By combining military thrust and elevator manipulation, a wave-off manipulation strategy based on a soft actor-critic (SAC) reinforcement learning algorithm is proposed, and considering the wave-off index and soft state constraint, a guided reward function is designed, which improves the exploration efficiency of the agent. Aiming to solve the problem of slow sampling efficiency of the SAC algorithm, combined with the PER technique, the PER-SAC algorithm is proposed to accelerate the convergence speed of SAC. The simulation results show that the wave-off strategy designed in this paper can significantly reduce the wave-off risk area and improve the wave-off safety.

carrier-based aircraft, wave-off boundary, wave-off manipulation, soft actor-critic (SAC), reinforcement learning, prioritized experience replay (PER)

doi: [10.1360/SST-2024-0133](https://doi.org/10.1360/SST-2024-0133)