

DOI:10.11918/202510009

动态博弈下变后掠翼飞行器智能决策规避方法

张景辉,张秀云,刘达,宗群

(天津大学 电气自动化与信息工程学院,天津 300072)

摘要:为解决变后掠翼飞行器在动态拦截环境下的自主规避问题,本文提出一种智能变形决策算法,通过实时调节后掠角,将动态变形作为规避的核心手段。首先,针对后掠角可变的变体飞行器模型,基于最小二乘法拟合气动系数,并分析了气动参数对飞行器气动性能的影响,从而为智能变形决策提供依据。其次,考虑变后掠翼飞行器飞行速度、飞行区域边界等实际物理约束条件,构建面向突防任务的变体飞行器-双拦截器动态博弈场景,结合飞行器状态、拦截器状态及目标信息的状态空间,设计以规避效果、气动性能为优化目标的决策模型。仿真实验验证结果表明,本文算法能够在完成自主变形决策规避的同时,兼顾机动性和敏捷性,克服了传统变形策略依赖离线优化计算和根据预设任务切换,难以自适应应对高动态博弈环境的局限性。

关键词:变体飞行器;强化学习;自主变形决策;柔性动作-评价;规避

中图分类号: V249.1

文献标志码: A

文章编号: 0367-6234(2026)01-0035-12

Intelligent evasion maneuvering decision method for a variable-sweep wing aircraft under dynamic game conditions

ZHANG Jinghui, ZHANG Xiuyun, LIU Da, ZONG Qun

(School of Electrical and Information Engineering, Tianjin University, Tianjin 30072, China)

Abstract: To address the autonomous evasion problem for variable-sweep wing aircraft in dynamic intercept environments, this paper proposes an intelligent morphing decision algorithm. This algorithm leverages dynamic morphing, primarily through real-time adjustment of the sweep angle, as the core evasion strategy. Initially, aerodynamic coefficients for the variable-sweep-angle aircraft model are fitted using the least-squares method. The influence of these aerodynamic parameters on the aircraft's performance is then analyzed, providing the foundation for intelligent morphing decision-making. Subsequently, a dynamic game scenario is developed for a penetration mission involving the morphing aircraft and dual interceptors, incorporating practical physical constraints such as flight speed and operational area boundaries. A decision model is then designed, integrating a state space that includes aircraft status, interceptor status, and target information, with the optimization objectives of maximizing evasion effectiveness and aerodynamic performance. Finally, simulation results demonstrate that the proposed algorithm successfully achieves autonomous morphing-based evasion while maintaining high maneuverability and agility. This approach overcomes the limitations of traditional morphing strategies, which rely on offline optimization and predefined task switching, making it difficult to adapt to highly dynamic game environments.

Keywords: morphing aircraft; reinforcement learning; autonomous morphing decision-making; soft actor-critic (SAC); evasion

在现代航天技术的发展中,飞行器需要应对日益多样的任务需求和环境挑战^[1]。变体飞行器(morphing wing aircraft, MWA)是指通过改变飞行器外形,从而调整其气动参数,使飞行器获得更优的气动效率的一种飞行器^[2]。与传统飞行器相比,变体飞行器通常具有更强的机动性能和对环境的适应能力,其对提升飞行器总体性能具有巨大潜力^[3]。

因此,如何通过合理的变形策略提升变体飞行器应对复杂环境的能力成为当下研究热点。

根据飞行器的变形部位,变体飞行器可以分为外部结构变形和内部结构变形两大类^[4],外部结构变形主要通过改变飞行器的外部几何形态,如机翼、头锥和机身等变形部位,以提升其空气动力学性能,从而优化飞行器在不同飞行条件下的表现。变体飞

收稿日期: 2025-10-10; 录用日期: 2025-11-07; 网络首发日期: 2025-11-19

网络首发地址: <https://link.cnki.net/urlid/23.1235.T.20251118.1346.012>

基金项目: 国家自然科学基金(62373273, 62373268); 中国博士后科学基金(2024M762355)

作者简介: 张景辉(1999—),男,硕士研究生;宗群(1961—),男,教授,博士生导师

通信作者: 张秀云, zxy_11@tju.edu.cn

行器在面对复杂的飞行任务时,如何变形、何时变形成为解决问题的关键。变形决策算法的设计通常可分为两类:传统变形决策算法和智能变形决策算法。传统变形策略是根据特定的飞行任务和目的,基于预先得到的气动数据,预先设定最佳的气动方案和变形策略^[5]。然而传统决策方法需要已知飞行剖面,且只适用于特定的飞行任务和飞行条件,不能很好地适应灵活多变的飞行任务和复杂多样的环境^[6]。因此,变形技术正朝着智能化和自主化的方向发展。强化学习作为机器学习的一个分支,近年来发展迅速。在强化学习框架中,智能体通过与环境持续进行交互,并根据其在特定状态下采取的动作获得相应的奖励。通过这个过程,智能体能够迭代地更新其策略,提升其获取累积奖励的能力。强化学习是一种无监督的启发式算法,不需要建立精确的模型,具有良好的泛化能力^[7]。因此,强化学习可以作为变体飞行器变形决策的有效方法^[8]。针对智能变形决策相关问题,已有学者进行了大量的研究。Valasek 等^[9]在离线数据库基础上,利用机器学习方法学习变形动作,并结合自适应动态反演控制策略,实现了在参数不确定情况下的轨迹精确跟踪。Hou 等^[10]提出了一种基于深度神经网络(deep neural networks, DNNs)的智能变形决策方法,用于过程约束下高超音速助推滑翔变形飞行器的自主变形决策。Xu 等^[11]提出了一种基于深度 Q 网络(dueling-deep Q network, DDQN)的智能变形控制器,将飞行器的基本控制器和动力学模型作为环境,解决了控制周期短导致的收敛问题。Jiang 等^[12]提出了一种基于改进深度确定性策略(deep deterministic policy gradient, DDPG)算法的变形飞行器自主决策方法。通过结合长短期记忆网络(long short-term memory, LSTM)设计了 DDPGwTC (DDPG with a task classifier)算法,在保证不同任务环境下最优变形策略决策的同时,提高了算法收敛性。Wang 等^[13]提出了一种基于分层强化学习的变体飞行器轨迹规划方法,解决了威胁区交叉重叠全覆盖路径下,综合通行概率与变体飞行器外形优化的实时决策问题。Fu 等^[14]提出了一种基于强化学习的仿鸟变形飞行器变体决策与飞行联合控制框架,解决了在燃油消耗、机动性和敏捷性等多目标综合飞行任务下的最优控制与高效跟踪问题。Zhang 等^[15]提出了一种针对滑翔阶段变跨度变形飞行器的集成决策与控制框架,该框架运用了 DDPG 强化学习算法,并结合了在线气动不确定性估计,解决了变形机制与控制设计相互作用下的协同决策,以及在气动

不确定性环境下的综合性能优化问题。Cao 等^[16]提出了一种针对变形飞行器的一体化制导与控制方法,该方法运用了基于滑模控制(sliding-mode control, SMC)的鲁棒深度强化学习框架,通过离线训练生成端到端的控制指令,解决了在模型不确定性和外部干扰下的鲁棒控制问题。

上述智能变形决策方法虽然能够有效实现变体飞行器的自主变形决策,但现有相关研究主要集中在任务环境相对简单的滑翔段和制导段。在这些阶段,飞行器的变形策略通常基于离线优化计算,根据不同任务需求切换至预设的最优气动外形以提升飞行性能。然而,这类变形决策方法难以满足复杂规避任务下对飞行器机动性与灵活性的需求,尤其是在高对抗、红蓝双方动态博弈环境下的变体飞行器智能变形决策方法,现有的相关研究仍存在不足。Wang 等^[17]提出一种基于 DDPG 算法的智能变形决策方法,可以通过合适的变形决策指令在得到最优气动外形的同时获得更好的制导性能,但其研究局限于常规的制导问题,未考虑末段复杂的对抗场景。Zhang 等^[18]提出了一种基于动作-评价(actor-critic, AC)的变形策略,以最小飞行时间为目标,解决了由终端性能目标和长航程任务所导致的稀疏奖励无参考决策问题,但其仿真环境设定仅为非机动的平飞阶段任务。Yan 等^[19]提出了一种基于 Q 学习的最优翼型变形策略,使得变体飞行器能够在变体决策过程中根据飞行条件自主选择俯冲、爬升和巡航阶段的最优后掠角,同时可以保持变体过程中的稳定性,但该方法仍高度依赖离线计算的气动数据,对于动态的环境适应性有限。在复杂规避场景下,红蓝双方的动态博弈使得威胁环境快速演变,这对变体飞行器的自主变形决策能力提出了更高要求,红方飞行器需要根据蓝方的实时状态动态调整变形策略,从而优化气动性能,增强其机动规避能力。

综上所述,本文基于柔性动作-评价算法(soft actor-critic, SAC)^[20],提出了一种针对双拦截器拦截场景的智能变形决策算法。首先,针对变后掠翼的飞行器,建立三自由度质心运动模型,并分析了攻角、后掠角和马赫数等状态量对气动特性的影响。随后,通过结合所构建的动态博弈场景与 SAC 算法,提出了一种智能变形决策算法。最后,通过仿真实验和与其他强化学习方法的结果对比,验证了所提方法的有效性和优越性。

1 变体飞行器模型建立及气动性能分析

本文研究的飞行器采用刚体假设,且不考虑地

球的自转, 飞行器采用无动力滑翔模式, 忽略飞行过程中质量变化造成的影响^[21], 主要面向飞行器外部结构变形中的变后掠翼变形, 变形量为机翼后掠角 ξ 。即通过调整后掠角实现决策, 其结构示意图如图 1 所示。

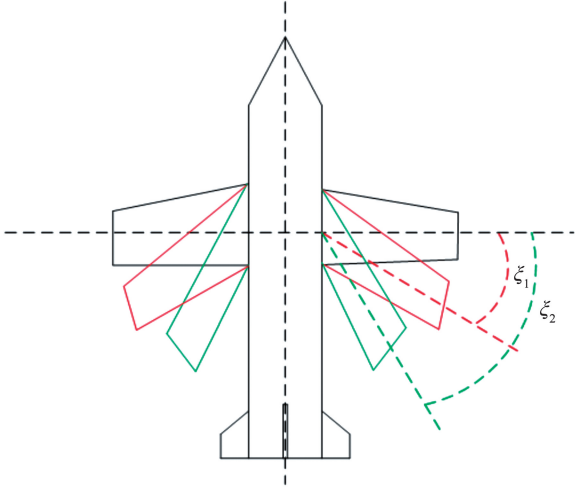


图 1 变体结构示意图

Fig. 1 Schematic diagram of morphing structure

1.1 变体飞行器质点运动动力学模型

质心平移运动方程将变体飞行器简化为一个可控质点, 通过分析该质点的运动状态, 获得其飞行路径。该模型能够描述质点运动的 3 个位置分量(地心距、经度、纬度)以及 3 个速度分量(速度、航迹角、航向角)。综上, 可构建如下变体飞行器三自由度运动模型^[22]:

$$\begin{cases} \dot{r} = v \sin \gamma \\ \dot{\lambda} = \frac{v \cos \gamma \sin \chi}{r \cos \phi} \\ \dot{\phi} = -\frac{v \cos \gamma \cos \chi}{r} \\ \dot{v} = -\frac{D}{m} - g \sin \gamma \\ \dot{\gamma} = \frac{1}{v} \left[\frac{L \cos \sigma}{m} - \left(g - \frac{v^2}{r} \right) \cos \gamma \right] \\ \dot{\chi} = -\frac{L \sin \sigma}{m v \cos \gamma} \end{cases} \quad (1)$$

式中: 飞行状态 $r, \lambda, \phi, v, \gamma, \chi, \sigma$ 分别表示地心距、经度、纬度、飞行速度、航向角、航迹角、倾侧角。 m 为飞行器质量, g 为引力加速度, L, D 分别为飞行器飞行过程中受到的升力、阻力。由地心距可以得到飞行高度 $h = r - R_e$, 其中 R_e 为地球半径。气动力 L, D 的计算方式如下:

$$L = q_d S_0 C_L \quad (2)$$

$$D = q_d S_0 C_D \quad (3)$$

式中: $q_d = 0.5 \rho v^2$ 表示动压, ρ 为大气密度, 可由标准大气公式计算得到; S_0 为飞行器气动参考面积; 基于最小二乘法拟合了升阻力系数 C_L, C_D , 其计算公式为:

$$C_L = c_{l_0} + c_{l_1} \beta + c_{l_2} Ma + c_{l_3} \alpha + c_{l_4} \xi + c_{l_5} Ma^{-1} + c_{l_6} Ma^2 + c_{l_7} \alpha^2 \quad (4)$$

$$C_D = c_{d_0} + c_{d_1} \beta + c_{d_2} Ma + c_{d_3} \alpha + c_{d_4} \xi + c_{d_5} Ma^{-1} + c_{d_6} Ma^2 + c_{d_7} \alpha^2 \quad (5)$$

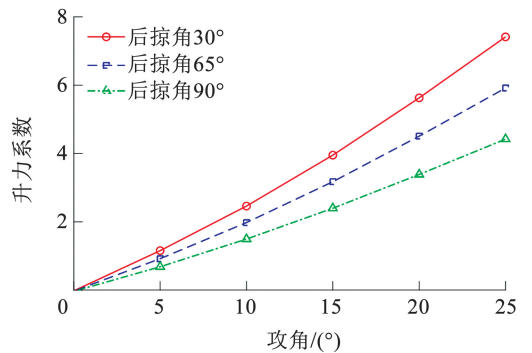
式中: $c_{l_i}, i \in \{1, 2, 3, 4, 5, 6, 7\}$ 为升力系数多项式的系数, $c_{d_j}, j \in \{1, 2, 3, 4, 5, 6, 7\}$ 为阻力系数多项式的系数, β 为侧滑角, Ma 为马赫数, α 为攻角, ξ 为后掠角。

本文的设计目标为: 考虑红方飞行器面对两个蓝方飞行器拦截的规避场景, 通过设计智能变形决策算法实现对攻角 α , 侧滑角 β 和后掠角 ξ 三个变量的智能决策, 从而改变飞行器的气动力和飞行状态, 最终在实现成功规避的同时获得更优的气动性能, 以保证后续飞行器的打击能力。

1.2 气动性能分析

相较于固定后掠角飞行器, 变体飞行器变形会对气动性能产生明显影响。根据已有的气动数据, 选取 $Ma = 8$ 时, 后掠角分别为 $30^\circ, 65^\circ, 90^\circ$ 的情况, 分析不同后掠角下气动系数随攻角的变化, 从而为飞行过程中的变体策略提供依据。

图 2 分别为升阻力系数随攻角变化曲线图及升阻比。由图 2(a)、(b) 可以看出, 升阻力系数始终为正值, 且与攻角呈现正相关关系, 与后掠角呈现负相关关系。在相同后掠角的情况下, 升阻力系数随攻角的增大而增大; 在相同攻角的情况下, 升阻力系数随后掠角增大而减小。由图 2(c) 可以看出, 升阻比值始终为正值。在相同后掠角的情况下, 升阻比随攻角的增大先增大后减小, 且存在一个使得升阻比最大的攻角。



(a) 升力系数随攻角变化

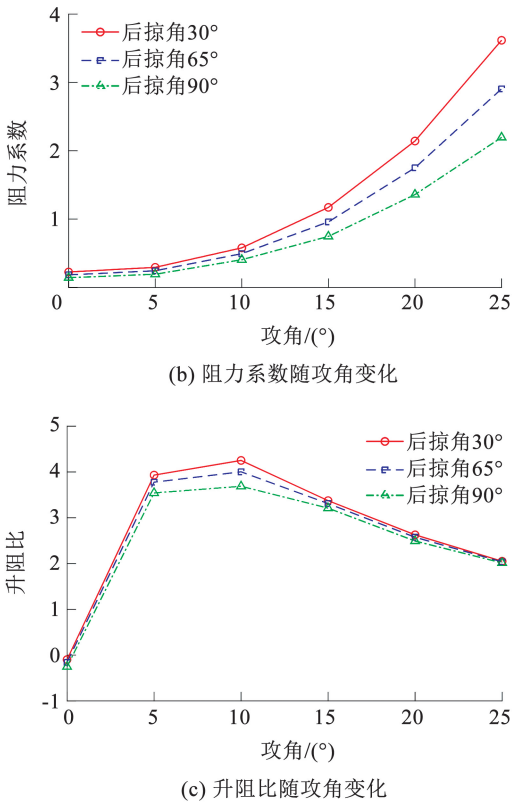


图 2 不同后掠角下气动系数随攻角变化曲线

Fig. 2 Variation of aerodynamic coefficient with angle of attack under different sweep angle

上述结果表明,后掠角的变化对飞行器纵向气动特性具有显著影响。智能变形决策则能够通过动态调整后掠角的变形幅度与时机,实现气动性能的在线优化,进而增强飞行器在多变环境中的任务执行能力。同时,通过合理选择攻角可以使飞行器在不同后掠角构型下均维持较高的升阻比。

2 智能变形决策算法设计

在建立变体飞行器模型和气动分析的基础上,针对突防任务下,面对两个蓝方飞行器拦截的场景,建立动态博弈模型,基于强化学习算法设计基于变后掠角的智能决策规避算法,实现变体飞行器的成功规避,并通过仿真实验,验证相较于其他强化学习方法,本文所提算法在完成自主规避的同时,可以获得更大的安全距离以及更好的气动性能。图 3 为规避场景示意图。

2.1 动态博弈模型建立

强化学习的核心要素主要包括智能体与环境。在每次交互过程中,智能体首先观测当前环境状态,之后依据策略选择动作,在动作执行后,环境状态发生转移,并产生奖励信号以评估动作价值^[23]。该过程可表示为马尔可夫决策过程 (markov decision process, MDP),可以由一个五元组 $\langle S, A, P, R, \gamma \rangle$

构成,其中 S 为状态集, A 为动作集, R 为奖励函数, P 为状态转移概率函数, γ 则为折扣因子。

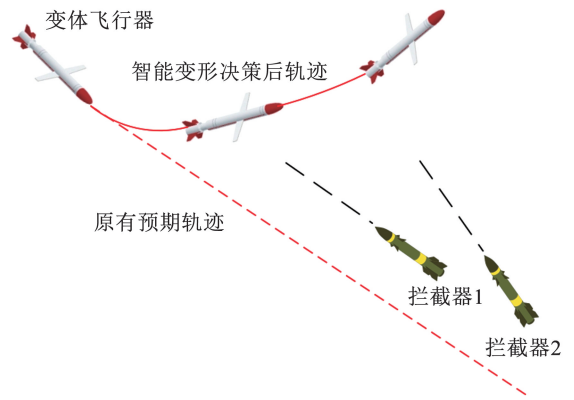


图 3 规避场景示意图

Fig. 3 Schematic of evasion scenarios

针对本文的变体飞行器规避蓝方飞行器的任务,根据高动态博弈环境下的信息以及双方状态信息,同时综合考虑变体飞行器的机动能力,建立面向变体飞行器规避任务博弈模型。假设场景包含一个红方变后掠翼飞行器和两个蓝方飞行器,得到博弈模型的马尔可夫决策模型的基本要素如下。

状态空间 S : 考虑飞行器规避决策需求,将变体飞行器状态信息 $O_T = (r_T, \lambda_T, \phi_T, v_T, \gamma_T, \chi_T)$ 、蓝方飞行器状态信息 $O_M = (r_M, \lambda_M, \phi_M, v_M, \gamma_M, \chi_M, \rho_M)$ 、以及目标点位置信息 $O_P = (r_P, \lambda_P, \phi_P)$ 作为变体飞行器规避的状态 S_Z , 即

$$S_Z = (r_T, \lambda_T, \phi_T, v_T, \gamma_T, \chi_T, r_{M_1}, \lambda_{M_1}, \phi_{M_1}, v_{M_1}, \gamma_{M_1}, \chi_{M_1}, \rho_{M_1}, r_{M_2}, \lambda_{M_2}, \phi_{M_2}, v_{M_2}, \gamma_{M_2}, \chi_{M_2}, \rho_{M_2}, r_P, \lambda_P, \phi_P) \in S \quad (6)$$

式中: $r_T, \lambda_T, \phi_T, v_T, \gamma_T, \chi_T$ 分别为变体飞行器的地心距, 经度, 纬度, 速度, 航迹角, 航向角; $r_M, \lambda_M, \phi_M, v_M, \gamma_M, \chi_M, \rho_M$ 分别为蓝方飞行器的地心距, 经度, 纬度, 速度, 航迹角, 航向角, 以及蓝方飞行器的威胁半径; r_P, λ_P, ϕ_P 分别为目标点的地心距, 经度, 纬度。

动作空间 A : 基于变体飞行器常用机动方式, 将攻角 α , 倾侧角 β , 后掠角 ξ 作为动作空间。其中, 后掠角的变化可以改变变体飞行器的展弦比, 进而影响其气动性能: 当后掠角减小时展弦比增大, 升力增加, 飞行器机动性提升; 当后掠角增大时展弦比减小, 阻力减小, 飞行器加速性能增强。通过智能切换后掠角状态, 变后掠翼飞行器可以兼顾机动性与高速性能, 满足突防任务需求。因此变后掠翼飞行器动作空间为

$$A = [\alpha, \beta, \xi] \quad (7)$$

状态转移概率 P : 表示变体飞行器在给定当前状态 s_i 下执行机动动作 $a_i \in A$, 达到下一时刻状态

s_{t+1} 的概率。将公式(1)飞行器的运动学方程作为动态博弈模型的状态转移函数。

折扣因子 $\gamma: \gamma \in [0, 1]$ 表示未来累积回报收益值相对于当前决策的重要程度。 γ 越大, 则表示策略考虑未来回报收益的比重越大; γ 越小, 则表示策略考虑未来回报收益的比重越小, 即更关注当前回报收益值。

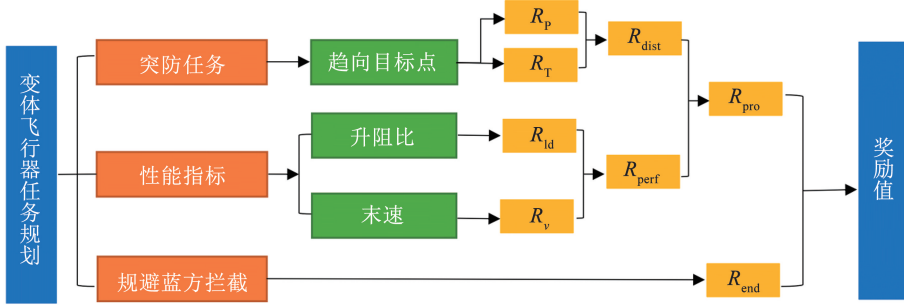


图 4 变体飞行器任务规划

Fig.4 Mission planning for morphing aircraft

根据任务规划确定单步决策立即回报收益值 R_t , 该值由过程奖励与终端奖励两部分组成为

$$R_t = R_{end} + R_{pro} \quad (8)$$

式中 R_{end} 为终端奖励: 当满足回合结束的要求时, 当前状态的立即回报收益值将根据回合是否成功获得一个较大的终端奖励或惩罚, 即

$$R_{end} = \begin{cases} -k_1, & \text{failed} \\ k_1, & \text{success} \end{cases} \quad (9)$$

式中 k_1 为常数系数。

过程奖励 R_{pro} 包含距离奖励 R_{dist} 和性能指标奖励 R_{perf} 。其中, 距离奖励 R_{dist} 计算公式为

$$R_{dist} = R_p + R_T \quad (10)$$

距离奖励 R_{dist} 由与目标点的距离奖励 R_p 、与蓝方飞行器的距离奖励 R_T 两部分组成, 如式(11)和式(12)所示:

$$R_p = -k_2 \times d_p^2 \quad (11)$$

$$R_T = \begin{cases} -k_3 \log \left(1 + \left(\left(\frac{1}{d_T} - \frac{1}{\rho_M} \right) \cdot 1000 \right)^2 \right), & d_T < \rho_M \\ 0, & \text{others} \end{cases} \quad (12)$$

式中: d_p 为与任务目标点的距离; d_T 为变体飞行器与蓝方飞行器的距离; ρ_M 为蓝方飞行器的威胁半径; k_2, k_3 为常数系数。从式(12)可以看出: 变体飞行器距离目标点越近, 奖励值越大; 变体飞行器与蓝方飞行器的距离越小, 奖励值越小。

性能指标奖励 R_{perf} 分为升阻比奖励 R_{ld} 和速度奖励 R_v 两个方面:

$$R_{ld} = k_4 \times \left(\frac{L}{D} \right) \quad (13)$$

立即回报收益值 R_t : 立即回报收益是变体飞行器由当前时刻状态在执行某一动作后转移至下一状态时从环境中接受到的反馈值。为了保障变体飞行器突防规避蓝方飞行器的能力, 考虑红蓝双方相对位置、目标点的距离和变体飞行器的性能等因素建立奖惩机制, 任务规划及奖励设置如图 4 所示。

$$R_v = k_5 \times \left(\frac{v_0}{v} \right) \quad (14)$$

式中: k_4, k_5 为常数系数, v_0 为飞行器的初速度。其中, 升阻比奖励可以保证飞行器在规避过程中保持更大的升阻比, 同时由于飞行器采用无动力滑翔方式, 速度奖励可以减少其在规避中的速度损耗。

2.2 约束条件

考虑实际作战场景中复杂限制因素, 在构造博弈模型时针对变体飞行器规避场景设计了如下约束条件:

速度约束: 在作战场景中, 红方变体飞行器和蓝方飞行器均不能超过自身最大速度, 即

$$\|v_T\| \leq V_{maxT}, \|v_{M1}\| \leq V_{maxM1}, \|v_{M2}\| \leq V_{maxM2} \quad (15)$$

式中: v_T 表示红方变体飞行器的速度, v_{M1}, v_{M2} 表示蓝方飞行器的速度; V_{maxT} 表示变体飞行器的最大速度, V_{maxM1}, V_{maxM2} 表示两个蓝方飞行器的最大速度。为了使突防过程更符合实际场景, 假设红方变体飞行器的最大速度大于蓝方飞行器的最大速度, 即:

$$\begin{cases} V_{maxT} > V_{maxM1} \\ V_{maxT} > V_{maxM2} \end{cases} \quad (16)$$

攻角约束: 在作战场景中, 变体飞行器不能超过自身最大攻角, 即

$$\|\alpha_T\| \leq \alpha_{max} \quad (17)$$

式中: α_T 为变体飞行器的攻角大小, α_{max} 为变体飞行器的最大攻角。

边界约束: 为了防止规避过程时间过长, 限定飞行器在运动过程中不能越过指定地图范围, 即:

$$\begin{cases} r_{\min} \leq r_T \leq r_{\max} \\ \lambda_{\min} \leq \lambda_T \leq \lambda_{\max} \\ \phi_{\min} \leq \phi_T \leq \phi_{\max} \end{cases} \quad (18)$$

式中: r_T, λ_T, ϕ_T 分别为各飞行器的经度, 高度, 纬度; $r_{\max}, \lambda_{\max}, \phi_{\max}$ 和 $r_{\min}, \lambda_{\min}, \phi_{\min}$ 分别为地图范围的最大地心距, 最大经度, 最大纬度以及最小地心距, 最小经度, 最小纬度。

法向过载约束: 在机动过程中, 为保证飞行器安全飞行, 需要对变体飞行器的法向过载进行限制, 即

$$n_z = \frac{L}{mg} \leq n_{\max} \quad (19)$$

式中 n_{\max} 为飞行器的最大法向过载。

动压约束: 为了减少变体飞行器结构质量和执行机构承受的载荷, 同时需要对变体飞行器的进行约束^[24], 即

$$q = \frac{1}{2} \rho v^2 \leq q_{\max} \quad (20)$$

式中 q_{\max} 为飞行器的最大动压。

为实现变体飞行器的智能自主决策, 研究设计最大化博弈过程变体飞行器的总收益 (J), 即

$$J = \sum_{t=1}^{\infty} \gamma^{t-1} R_t \quad (21)$$

式中: R_t 为 t 时刻变体飞行器获得的立即回报收益值, 优化目标是使 J 最大。

综上所述, 本文针对变体飞行器规避拦截任务, 建立了完整的动态博弈模型, 通过定义马尔可夫决策过程的核心要素, 设计了如公式(8)所示的包含终端奖励和过程奖励的多层奖励函数, 其中式(9)~式(14)的奖励函数的设计综合考虑了距目标点距

离、红蓝双方距离以及升阻比和速度等性能指标。同时式(15)~式(20)严格约束了飞行速度、攻角范围、飞行区域边界、法向过载和动压等实际物理限制。最终以实现最大累积奖励为目标, 确保变体飞行器在满足各项约束条件的前提下, 通过智能变形决策实现最优的规避效果。

2.3 基于 SAC 算法的智能变形决策算法设计

为实现变体飞行器在复杂动态博弈环境下的自主、高效规避, 本文采用 SAC 算法设计智能变形决策方法。SAC 算法是一种基于 Actor-Critic 框架的最大熵强化学习算法。其核心优势在于强大的探索能力和学习的稳定性, 适合解决变形决策这种多维、连续的问题。

基于 SAC 的智能决策方法的整体框架如图 5 所示, 其算法核心由一个策略网络、两个动作价值网络及两个对应的目标动作价值网络构成。在训练过程中, 算法首先通过飞行器与动态博弈环境的交互来收集经验数据, 包含状态、动作、奖励等, 并将其存储于经验池中。而后, 动作价值网络从经验池中提取数据, 通过最小化价值损失函数 J_Q 来拟合当前策略下的动作价值函数 Q_{π} , 同时, 目标动作价值网络则通过软更新机制, 为动作价值网络自身的优化提供稳定且滞后的目标值基准, 防止过估计提高训练过程的稳定性。最后, 策略网络基于策略损失函数 J_{π} 进行参数更新, 通过最大化其在给定状态下输出动作的期望 Q_{π} 值, 生成飞行器的动作, 并结合熵正则化项鼓励探索, 不断优化其策略参数, 从而提升飞行器在动态博弈环境中的决策动作。

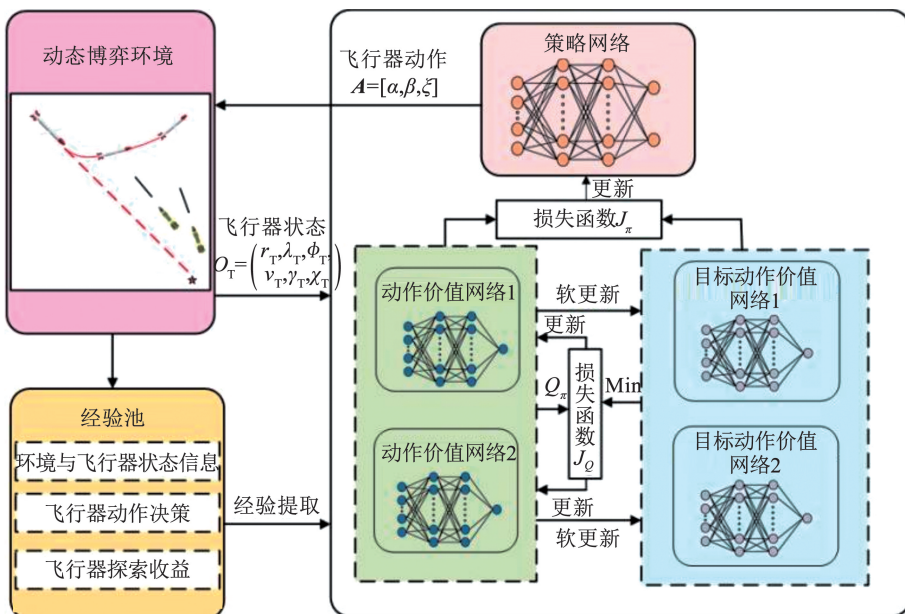


图 5 基于 SAC 算法的智能变形决策方法整体框架

Fig. 5 Framework of intelligent deformation decision method based on SAC algorithm

同时, SAC 算法在传统的累计奖励最大化目标的基础上, 引入策略熵的概念。在 SAC 训练过程中, 不仅希望最大化期望收益的总和, 还要最大化策略的熵^[25], 这使得变体飞行器能够充分探索变形策略, 避免因策略单一而陷入局部最优。其最优策略可以定义为

$$\pi^* = \operatorname{argmax}_{\pi} E_{\pi} [r(s_t, \mathbf{a}_t) + \delta H(\pi(\cdot | s_t))] \quad (22)$$

式中: δ 为温度系数, 其可以根据回报调节熵项, δ 越大, 意味着变体飞行器在训练初期策略的随机性越强。 $H(\pi(\cdot | s_t))$ 为策略 π 在状态 s_t 时的策略熵, 策略的熵值越高, 则策略分布的不确定性越高, 变体飞行器探索不同动作的可能性越大, 探索就越充分, 这对于飞行器在高动态博弈中寻找合适的规避策略至关重要。

为了定义 Q 函数之间的关系, 以求解最优的 Q 函数, 对于一个固定的策略 π , 此时的贝尔曼方程为

$$Q^{\pi}(s_t, \mathbf{a}_t) = r(s_t, \mathbf{a}_t) + \gamma E_{s_{t+1}} [V^{\pi}(s_{t+1})] \quad (23)$$

式中 γ 为折扣因子, 状态价值函数 V 为

$$V^{\pi}(s_t) = E_{\mathbf{a}_t \sim \pi} [Q^{\pi}(s_t, \mathbf{a}_t) - \delta \log \pi(\mathbf{a}_t | s_t)] = E_{\mathbf{a}_t \sim \pi} [Q^{\pi}(s_t, \mathbf{a}_t)] + H(\pi(\cdot | s_t)) \quad (24)$$

$$J_Q(\varphi) = E_{(s_t, \mathbf{a}_t) \sim m} \left[\frac{1}{2} (Q_{\varphi}(s_t, \mathbf{a}_t) - [r(s_t, \mathbf{a}_t) + \gamma E_{s_{t+1} \sim p} [V_{\varphi}^-(s_{t+1})]])^2 \right] \quad (26)$$

其中,

$$V_{\varphi}^-(s_{t+1}) = E_{\mathbf{a}_{t+1} \sim \pi_{\theta}} [Q_{\varphi}(s_{t+1}, \mathbf{a}_{t+1}) - \delta \log \pi_{\theta}(\mathbf{a}_{t+1} | s_{t+1})] \quad (27)$$

式中 Q_{φ} 代表以 $\bar{\varphi}$ 为参数的目标动作价值网络。在 SAC 算法中, 运用了双层 Q 网络与目标动作价值网络的技术, 这种双层 Q 网络和目标动作价值网络的结构能够有效抑制 Q 值的过高估计, 提升训练的稳定性, 确保飞行器对机动动作的价值评估更加准确可靠。其中, Q 网络与目标动作价值参数 $\bar{\varphi}$ 、 $\bar{\theta}$, 是由一段时间内 Q 网络和目标动作价值网络的参数 φ 、 θ

$$J_{\pi}(\theta) = E_{s_t \sim m, \boldsymbol{\varepsilon}_t \sim N} [\delta \log \pi_{\theta}(f_{\theta}(\boldsymbol{\varepsilon}_t; s_t) | s_t) - Q_{\varphi}(s_t, f_{\theta}(\boldsymbol{\varepsilon}_t; s_t))] \quad (29)$$

式中 $\boldsymbol{\varepsilon}_t$ 为一个从标准正态分布 N 中采样得到的输入噪声向量, 目的仍是增强随机性, 以提高训练过程飞行策略的探索能力。为了进一步减少强化学习的超参数数量, 设定温度因子 δ 的损失函数为 $J(\delta)$, 使其实现自适应调整

$$J(\delta) = E_{s_t \sim m, \mathbf{a}_t \sim \pi} [-\delta \log \pi(\mathbf{a}_t | s_t) - \delta H_0] \quad (30)$$

最终, 当所有可训练策略网络与可训练价值网络均收敛后, 可视为飞行器可从环境中获取到最优的变形策略, 此时可将训练后的策略网络作为规避时的变形策略。

基于 SAC 算法的智能变形策略训练过程的伪代码如下:

于是, 对于每个飞行状态 s_t , 根据当前策略网络 π_{old} , 计算软更新后的策略分布 π_{new} 可以表示为

$$\pi_{\text{new}} = \operatorname{argmin}_{\pi'} D_{\text{KL}} \left(\pi'(\cdot | s_t), \frac{\exp\left(\frac{1}{\alpha} Q^{\pi_{\text{old}}}(s_t, \cdot)\right)}{Z^{\pi_{\text{old}}}(s_t, \cdot)} \right) \quad (25)$$

式中: π' 表示在策略空间中的可选策略, $\pi'(\cdot | s)$ 用于表示策略 π' 在给定状态 s_t 时的动作选择概率分布, $D_{\text{KL}}(\cdot)$ 用于计算两个概率分布之间的差异, $Q^{\pi_{\text{old}}}(s_t, \cdot)$ 表示在当前旧策略 π_{old} 的 Q 函数, $Z^{\pi_{\text{old}}}(s_t, \cdot)$ 是归一化常数, 确保分布的总和为 1。通过最小化 KL 散度, 使新策略接近于旧策略的动作选择概率分布, 优化策略以提高在给定状态下选择动作的效果, 并在策略迭代的过程中逐步改善策略的性能。

SAC 算法利用柔性策略迭代方法交替优化动作价值网络 $Q_{\varphi}(s, \mathbf{a})$ (Q 网络) 和策略网络 $\pi_{\theta}(s)$ 进行学习, φ 和 θ 分别是动作价值网络和策略网络的参数, 策略网络输出的是策略概率分布的均值和标准差, 动作价值网络输出的是根据当下策略采取的动作的价值。

Q 网络可以通过最小化贝尔曼残差对柔性 Q 函数进行逼近, 其损失函数定义为

的指数滑动平均获得, 这种方式也被称为软目标更新^[26], 即:

$$\begin{aligned} \bar{\varphi} &\leftarrow \tau \varphi + (1 - \tau) \bar{\varphi}, \\ \bar{\theta} &\leftarrow \tau \theta + (1 - \tau) \bar{\theta} \end{aligned} \quad (28)$$

式中 τ 为软更新系数, $\tau \ll 1$ 。

利用重采样技术 $\mathbf{a}_t = f_{\theta}(\boldsymbol{\varepsilon}_t; s_t)$ 从策略概率分布中采样, 可以得到策略网络的损失函数

SAC 算法

- 1: 初始化网络参数和记忆库 $\varphi_1, \varphi_2, \theta, \bar{\varphi}_1, \bar{\varphi}_2, \bar{\theta}, \delta, m \leftarrow \phi$
- 2: 对每一次迭代
- 3: 对每一个环境步数
- 4: 飞行器根据当前环境状态从策略中采样动作 $\mathbf{a}_t \sim \pi_{\theta}(\mathbf{a}_t | s_t)$
- 5: 接受环境转移的下一个状态 s_{t+1} 以及反馈 r_t
- 6: 将经验数据 $(s_t, \mathbf{a}_t, r_t, s_{t+1})$ 存储在经验池 m 中
- 7: End
- 8: 对每一个更新步数
- 9: 更新 Q 函数 $\varphi_i \leftarrow \varphi_i - \mu \nabla_{\varphi_i} J_Q(\varphi_i)$, for $i \in \{1, 2\}$
- 10: 更新策略函数 $\theta \leftarrow \theta - \mu \nabla_{\theta} J_{\pi}(\theta)$
- 11: 更新目标动作价值网络参数 $\bar{\varphi}_i \leftarrow \tau \varphi_i + (1 - \tau) \bar{\varphi}_i$, for $i \in \{1, 2\}$, $\bar{\theta} \leftarrow \tau \theta + (1 - \tau) \bar{\theta}$
- 12: End
- 13: End

3 仿真与分析

3.1 实验环境与参数设置

为验证本文提出所提出的智能变形决策算法的有效性,对基于 SAC 的智能变形决策算法的效果进行仿真校验。

首先,针对变体飞行器突防任务下规避蓝方飞行器的任务,设置飞行器和目标点的初始位置,飞行

过程需规避两个蓝方飞行器,根据末段突防任务的假设,在初始时刻蓝方飞行器开始发射,且蓝方两飞行器发射间隔为 3 s,这一设置旨在模拟真实的末段动态博弈环境,蓝方飞行器的制导方法采用比例导航法(proportional navigation, PN)^[27],对于规避任务的奖励函数构造,其中, $k_1 = 10$, $k_2 = 1 \times 10^{-6}$, $k_3 = 5$, $k_4 = 0.03$, $k_5 = 0.05$,仿真过程中 SAC 算法的超参数如表 1 所示。

表 1 SAC 的超参数及其数值

Tab. 1 SAC hyperparameters and values

折扣因子	批尺寸	开始学习步数	Actor 网络各隐层单元数目	Critic 网络各隐层单元数目	学习率	温度系数(初值)	软更新率	回合数	单步采样间隔/s
0.99	256	256	64	[256,256]	0.000 1	0.05	0.004	2 000	0.05

SAC 算法主要超参数对训练性能的影响如下。

折扣因子:表示未来累积回报收益值相对于当前决策的重要程度。本文的折扣因子设置为 0.99,以确保策略对未来回报的充分考虑和长期规划的优化,提高算法在动态规避任务中的收敛性和成功率,这有助于飞行器在面对动态拦截时,优先考虑长期规避成功而非短期局部机动,提升变体飞行器的整体突防能力。

批尺寸:该参数用于 Q 网络和策略网络更新。大的批尺寸提供更稳定的梯度,适合复杂环境;小的批尺寸加速迭代但增加噪声。本文的批尺寸设置为 256,以确保梯度估计的稳定性和噪声的减少,提高训练过程的效率和稳定性,使算法在处理高维状态空间时更稳定,使变形决策在复杂博弈环境中快速收敛,提高规避成功率。

开始学习步数:该参数指定算法从随机探索切换到策略优化的步数阈值。在初始阶段,智能体采用随机动作填充经验池,确保多样性数据积累;之后,使用策略网络生成动作并进行网络更新。这有

助于避免早期学习不稳定,提高算法在复杂规避环境中的收敛性和泛化能力。

软更新率 τ :该参数控制目标网络的更新速度,小 τ 确保平滑更新,减少 Q 值过估计,提高稳定性;大 τ 加速收敛,但可能引入振荡。本文的软更新率设置为 $\tau = 2^{-8}$,保障在训练过程中参数平滑更新和 Q 值的过估计,提高训练过程中的稳定性,减少价值估计的偏差,确保飞行器在高动态环境下做出可靠的变形决策,提升规避的稳定性和机动性能。

学习率 μ :该参数控制网络参数更新的步长。 μ 越大,则参数更新幅度越大,可能导致训练不稳定或振荡; μ 越小,则参数更新更平滑,但收敛速度较慢。本文的学习率设置为 0.000 1,既保证了参数更新的稳定性和渐进优化,避免过拟合,也使得飞行器在面对高动态拦截环境时,逐步细化变形决策策略,提升规避路径的平稳性和成功率,同时适应复杂博弈中的不确定性。

表 2 为仿真过程中飞行器的初始状态,变体飞行器的约束条件如表 3 所示。

表 2 飞行器的初始状态

Tab. 2 Initial state of the aircraft

变体飞行器初始速度/($\text{m}\cdot\text{s}^{-1}$)	变体飞行器初始位置	蓝方飞行器初始速度/($\text{m}\cdot\text{s}^{-1}$)	蓝方飞行器初始位置	蓝方飞行器威胁半径/m
2 500	($Re + 23\ 418\ \text{m}, 214.18^\circ, 63.29^\circ$)	2 300	($Re + 0\ \text{m}, 214.19^\circ, 63.88^\circ$)	500

表 3 变体飞行器的约束条件

Tab. 3 Constraints of the morphing aircraft

飞行器最大速度/($\text{m}\cdot\text{s}^{-1}$)	飞行器探索最大位置	飞行器探索最小位置	飞行器最大攻角/($^\circ$)	飞行器最大法向过载/g	飞行器最大动压/kPa
2 500	($Re + 40\ 000\ \text{m}, 216^\circ, 65^\circ$)	($Re + 10\ 000\ \text{m}, 214^\circ, 61^\circ$)	10	15	150

3.2 仿真结果分析

图 6 为训练阶段成功率曲线,可以看出,随着回合数的增加,训练成功率逐渐接近 1。同时,图 7 给出了每回合的步数曲线,随着回合数的增加,回合步数趋于稳定,这表明训练策略已实现收敛。

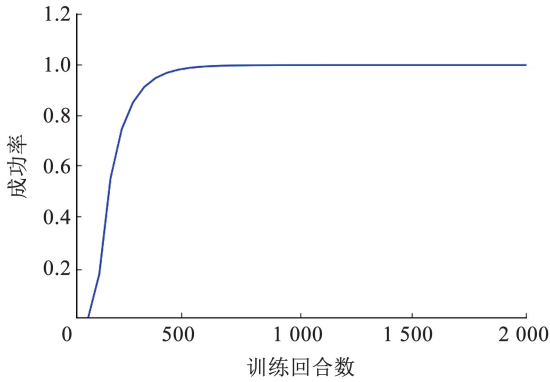


图 6 训练阶段成功率曲线

Fig. 6 Success rate curve of the training phase

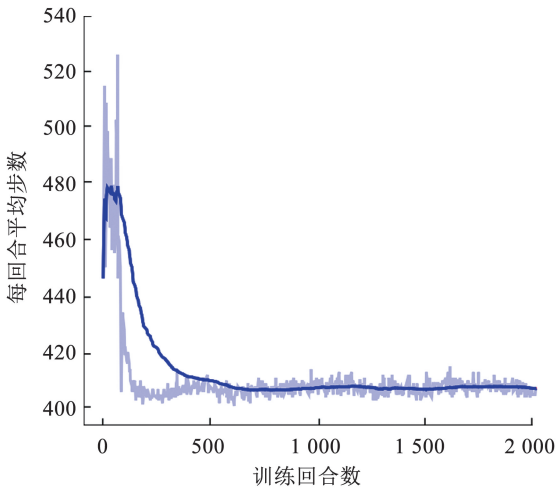


图 7 训练阶段每回合步数曲线

Fig. 7 Curve of steps per episode in the training phase

为了进行算法的训练效果对比,采用 TD3 (twin-delayed deep deterministic policy gradient) 算法^[28]对智能变形决策规避进行了训练,其中动态博弈场景的参数与 SAC 算法的完全相同,TD3 算法所选取的超参数信息也与本文所使用的 SAC 算法相同。二者在训练 2 000 回合后的奖励对比如图 8 所示,可以看出,TD3 算法在训练效果和收敛稳定性上均不及本文所采用的 SAC 算法。说明本文基于 SAC 的智能变形决策算法,相较于采用 TD3 算法具有更好的效果和稳定性。

图 9 为智能变形决策算法下的我方变体飞行器的法向过载变化曲线。法向过载是衡量飞行器机动性的关键指标。高过载能力使飞行器能够执行剧烈的规避动作和战术机动,而低过载则使其更专注于直线加速与高速巡航,法向过载的动态变化意味着

飞行器兼备机动性和加速能力。

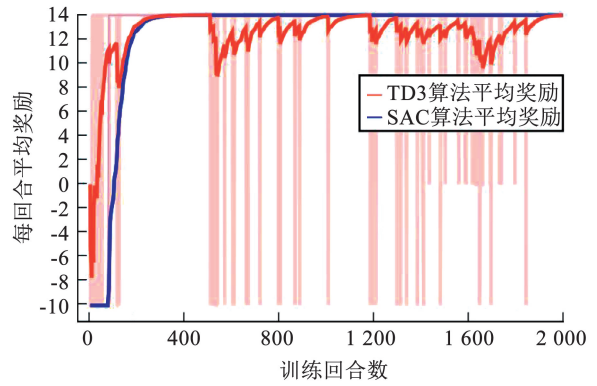


图 8 奖励曲线

Fig. 8 Reward curve

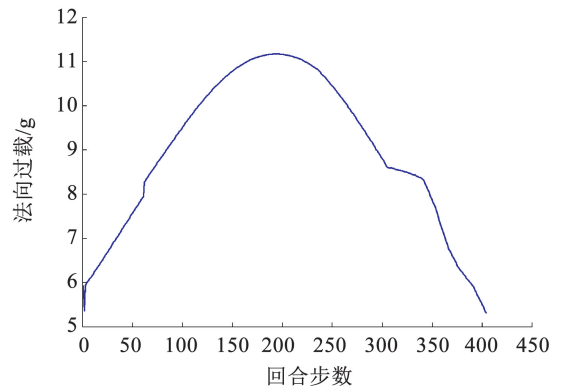


图 9 法向过载曲线

Fig. 9 Normal overload curve

从图 10 智能变形决策下后掠角变化中可以看出,在整个规避过程中,我方变体飞行器经历了两次明显的后掠角变化。结合图 10 和图 11 变体飞行器和两个蓝方飞行器之间的三维轨迹图可以看出,突防规避过程可以分为两个明显的阶段。在第 1 阶段,飞行器飞向目标点的过程中,突防初始阶段两个蓝方飞行器逐步接近我方飞行器。此时,基于智能变形决策,飞行器减小后掠角。同时,如图 9 所示,在攻角和后掠角的共同作用下法向过载持续增加,从而改变了原有飞向目标点的飞行状态,完成初步规避。

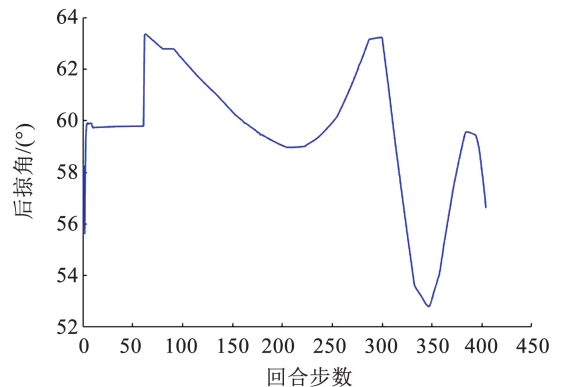


图 10 智能变形决策后掠角变化曲线

Fig. 10 Sweep angle variation curve for intelligent morphing decision

在第 2 阶段,蓝方飞行器完成机动转弯并开展追逐,我方飞行器随即进行第 2 次智能变形:后掠角再次减小,且减小的幅度更大。同时攻角增加,以保证向上机动飞行的状态。此阶段,飞行器优先保证飞行速度从而与蓝方飞行器拉开距离,而不再追求维持较高的法向过载。如图 9 所示,飞行器法向过载在第 2 阶段开始减小,仅在第 2 次智能变形期间出现一个短暂的延缓过程。结合图 11 和图 12 变体飞行器与两个拦截器之间距离变化,飞行器通过上述智能变形决策,在轨迹末端已成功摆脱两枚拦截器。此外,基于 2.2 小节的假设,突防过程中我方飞行器速度大于拦截器,因此突防成功。综上,基于智能变形方法的飞行器既能够获得较大的法向过载,增加飞行器的机动性,又可维持较高的飞行速度,提升其敏捷性,实现最终现高效突防。

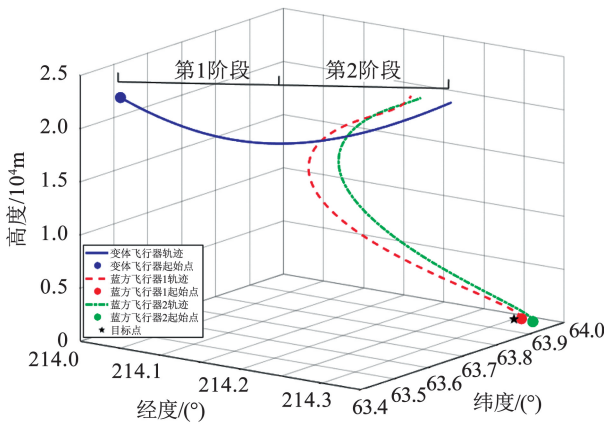


图 11 变体飞行器和两个蓝方飞行器的三维轨迹

Fig. 11 Three-dimensional trajectories of the morphing aircraft and two intercepting aircraft

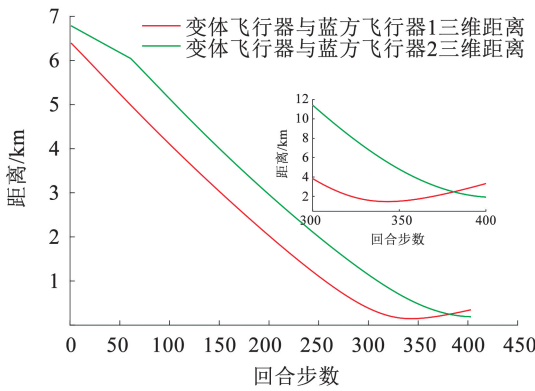
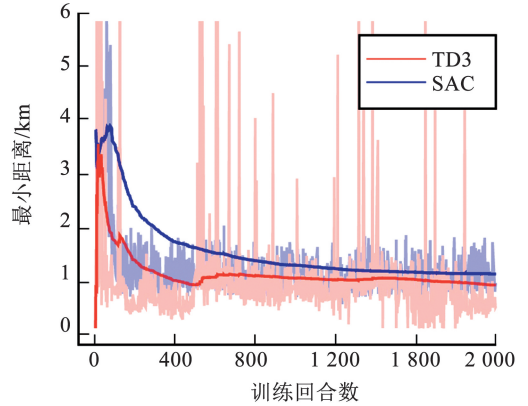


图 12 变体飞行器与两个拦截器之间距离的变化曲线

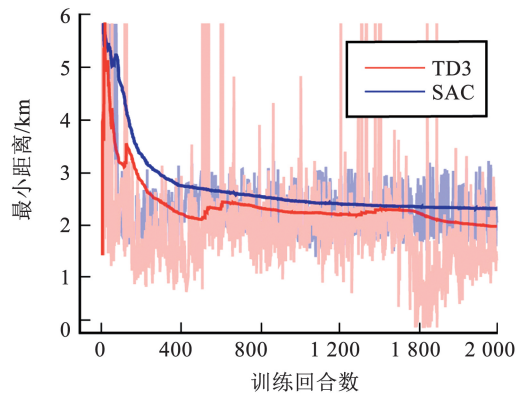
Fig. 12 Distance variation curve between morphing aircraft and two interceptors

图 13 为 SAC 算法和 TD3 算法下每回合中变体飞行器和两个蓝方飞行器之间的最小距离。图 13 结果显示,在训练过程中,基于 SAC 的智能变形决策算法始终保持更大的最小距离,优于 TD3 算法。

这表明 SAC 算法使变体飞行器与蓝方飞行器之间具有更大的安全距离,规避效果更佳。



(a) 每回合变体飞行器与拦截器1间最小距离



(b) 每回合变体飞行器与拦截器2间最小距离

图 13 两种算法下变体飞行器与拦截器之间每回合最小距离
Fig. 13 Minimum distance per episode between morphing aircraft and two interceptors during two algorithms

图 14 为变体飞行器在规避机动过程中基于 SAC 算法与 TD3 算法下变形决策的速度曲线对比。由于飞行器采用无动力滑翔模式,飞行器的速度始终处于减小的状态,可以看出,基于 SAC 的智能变形决策算法,相较于 TD3 算法下速度变化更为缓慢,这减小了飞行器在规避过程中的速度损耗,使飞行器具有更高的末端速度,意味着飞行器可以更有效地完成后续的打击任务。

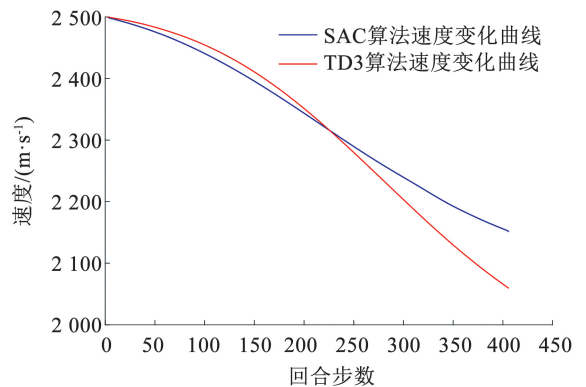


图 14 SAC 算法和 TD3 算法速度变化对比曲线
Fig. 14 Comparison of velocity variation curves for SAC and TD3 algorithms

图 15 展示了智能变形决策算法下攻角的变化曲线,图 16 比较了两种算法下变体飞行器的升阻比变化。结合图 15 以及 1.2 小节气动性能分析中的图 2(c),智能变形决策下的攻角,基本维持在可实现飞行器较大升阻比的范围内。通过设定的升阻比奖励和决策的攻角动作范围,飞行器在规避拦截过程中始终保持较高升阻比。尤其是在第 2 阶段,蓝方飞行器追逐拦截时,升阻比达到最大。从图 16 可以看出,相较于 TD3 算法,SAC 算法下的智能变形决策方法拥有更大的升阻比,这可显著提升飞行器的气动效率,使其能更灵活、高效的执行机动规避策略。

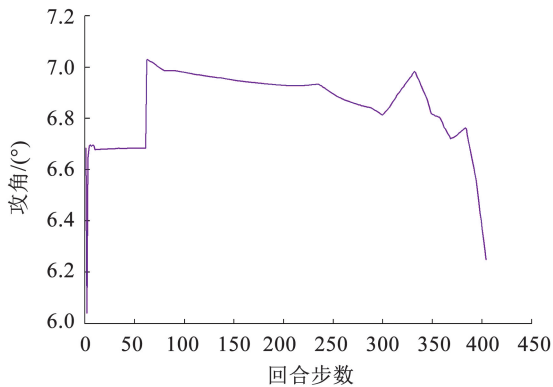


图 15 攻角变化曲线

Fig. 15 Attack angle change curve

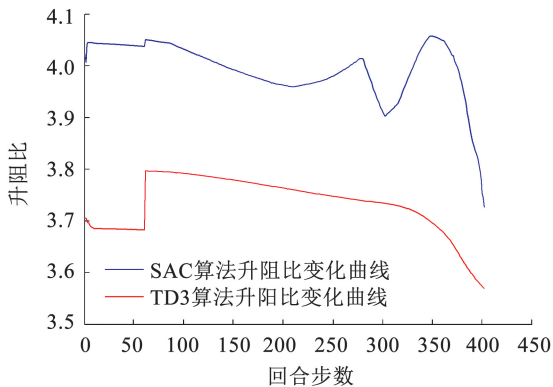


图 16 升阻比曲线

Fig. 16 Lift-to-drag ratio curve

4 结 论

针对变后掠角飞行器在动态博弈环境下的自主规避问题,本文设计了一种智能变形决策规避算法。通过将动态变后掠角作为规避的核心手段,实现飞行器在面对蓝方拦截时的智能变形决策机动规避。得到的主要结论如下。

1) 在建立变体飞行器三自由度质心运动模型的基础上,利用最小二乘法拟合了气动系数,并分析了气动参数对飞行器气动性能的影响,从而为智能

变形决策提供依据。

2) 考虑了变体飞行器飞行速度、攻角范围、飞行区域边界、法向过载和动压等实际物理约束条件,构建了面对突防任务的变体飞行器-双蓝方拦截器动态博弈模型,以综合考虑规避和气动性能为优化目标,提出了一种智能决策规避算法,在实现规避蓝方飞行器的同时,保证了最优的气动性能,提高了末端打击能力,克服了传统变形策略因依赖离线优化和预设任务,而难以自适应高动态环境的局限性。

3) 通过仿真实验,验证了本文所提出的智能变形决策规避算法的有效性。通过智能变形决策,变体飞行器在完成自主规避的同时,既能获得较大的升阻比和法向过载,以增加飞行器的机动性,又可维持较高的飞行速度,以提升其敏捷性和减少飞行器在规避过程的速度损耗。

参 考 文 献

- [1] 张灿,王轶鹏,叶蕾. 国外近十年高超声速飞行器技术发展综述[J]. 战术导弹技术, 2020(6): 81
ZHANG Can, WANG Yipeng, YE Lei. A review of foreign hypersonic vehicle technology development in the last decade[J]. Tactical Missile Technology, 2020(6): 81. DOI:10.16358/j.issn.1009-1300.2020.1.545
- [2] 毛定坤,蔡光斌,冯志超,等. 变体飞行器传感器故障在线主动容错控制[J]. 哈尔滨工业大学学报, 2023, 55(8): 60
MAO Dingkun, CAI Guangbin, FENG Zhichao, et al. Online active fault-tolerant control for sensor faults of morphing aircraft[J]. Journal of Harbin Institute of Technology, 2023, 55(8): 60. DOI: 10.11918/202205056
- [3] 张家应,黄可,武冠振,等. 变体飞行器结构关键技术及研究进展[J]. Electronics Optics & Control, 2024, 31(1): 1
ZHANG Jiaying, HUANG Ke, WU Guanzhen, et al. Key technologies and research progress of morphing aircraft structure[J]. Electronics Optics & Control, 2024, 31(1): 1. DOI:10.3969/j.issn.1671-637X.2024.01.001
- [4] 彭宇宇. 高超声速飞行器气动变形方案设计与外形优化方法研究[D]. 长沙: 国防科技大学, 2019
PENG Wuyuyu. Research on aerodynamic deformation scheme design and shape optimization method of hypersonic vehicle[D]. Changsha: National University of Defense Technology, 2019
- [5] CHEN Xiaoyu, LI Chunna, GONG Chunlin, et al. A study of morphing aircraft on morphing rules along trajectory[J]. Chinese Journal of Aeronautics, 2021, 34(7): 232. DOI:10.1016/j.cja.2020.04.032
- [6] SHI Rongqi, WAN Weiyu. Analysis of flight dynamics for large-scale morphing aircraft[J]. Aircraft Engineering and Aerospace Technology: An International Journal, 2015, 87(1): 38. DOI:10.1108/aeat-01-2013-0004
- [7] 刘全,翟建伟,章宗长,等. 深度强化学习综述[J]. 计算机学报, 2018, 41(1): 1
LIU Quan, ZHAI Jianwei, ZHANG Zongzhang, et al. A survey of deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1. DOI:10.11897/SP.J.1016.2018.00001

- [8] 张旭辉, 韩卫波, 谭天宇. 智能变形飞行器决策规划与控制研究进展[J]. 宇航学报, 2025, 46(3): 395
ZHANG Xuhui, HAN Weibo, TAN Tianyu. Research progress on decision-making, planning and control for intelligent morphing aircraft[J]. Journal of Astronautics, 2025, 46(3): 395. DOI:10.3873/j.issn.10001328.2025.03.001
- [9] VALASEK J, TANDALE M, RONG J. A reinforcement learning-adaptive control architecture for morphing[J]. Journal of Aerospace Computing, Information, and Communication, 2005, 2(4): 174. DOI:10.2514/6.2004-6220
- [10] HOU Linfei, LIU Honglin, YANG Ting, et al. An intelligent autonomous morphing decision approach for hypersonic boost-glide vehicles based on DNNs[J]. Aerospace, 2023, 10(12): 1008. DOI:10.3390/aerospace10121008
- [11] XU Wenfeng, LI Yinghui, PEI Binbin, et al. Coordinated intelligent control of the flight control system and shape change of variable sweep morphing aircraft based on dueling-DQN[J]. Aerospace Science and Technology, 2022, 130: 107898. DOI:10.1016/j.ast.2022.107898
- [12] JIANG Weilai, ZHENG Chenghong, HOU Delong, et al. Autonomous shape decision making of morphing aircraft with improved reinforcement learning[J]. Aerospace, 2024, 11(1): 74. DOI:10.3390/aerospace11010074
- [13] 王翰桐, 禹春梅, 程晓明. 考虑威胁区通行概率的变体飞行器轨迹规划[J]. 航天控制, 2024, 42(2): 35
WANG Hantong, YU Chunmei, CHENG Xiaoming. Trajectory planning for morphing aircraft considering threat area passage probability[J]. Aerospace Control, 2024, 42(2): 35. DOI:10.16804/j.cnki.issn1006-3242.2024.02.005
- [14] FU Qien, SUN Changyin. Attention-based multi-objective control for morphing aircraft[J]. Biomimetics, 2025, 10(5): 280. DOI:10.3390/biomimetics10050280
- [15] ZHANG Baochao, GUO Jie, WANG Haoning, et al. Coordinated design of intelligent morphing decision and entry guidance for morphing hypersonic glide vehicles[J]. Journal of Aerospace Engineering, 2025, 38(6): 04025076. DOI:10.1061/jaece.aseng-6129
- [16] CAO Chengyu, LI Fanbiao, XIE Qihao, et al. Integrated guidance and control of morphing flight vehicle via sliding-mode-based robust reinforcement learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2025. DOI:10.1109/tsmc.2025.3540262
- [17] 王青, 刘国华, 屈东扬. 一种基于 DDPG 的变体飞行器智能变形决策方法[J]. 宇航学报, 2024, 45(10): 1560
WANG Qing, LIU Huahua, QU Dongyang. An intelligent morphing decision-making method for morphing aircraft based on DDPG[J]. Journal of Astronautics, 2024, 45(10): 1560. DOI:10.3873/j.issn.1000-1328.2024.10.005
- [18] ZHANG Baochao, GUO Jie, WANG Haoning, et al. Autonomous morphing strategy for a long-range aircraft using reinforcement learning[J]. Aerospace Science and Technology, 2024, 148: 109087. DOI:10.1016/j.ast.2024.109087
- [19] 闫斌斌, 李勇, 戴沛, 等. 基于增强学习的变体飞行器自适应变体策略与飞行控制方法研究[J]. 西北工业大学学报, 2019, 37(4): 656
YAN Binbin, LI Yong, DAI Pei, et al. Research on adaptive morphing strategy and flight control method for morphing aircraft based on reinforcement learning[J]. Journal of Northwestern Polytechnical University, 2019, 37(4): 656. DOI:10.3969/j.issn.1000-2758.2019.04.003
- [20] SONG Lifei, XU Chuanyi, HAO Le, et al. Research on PID parameter tuning and optimization based on SAC-auto for USV path following[J]. Journal of Marine Science and Engineering, 2022, 10(12): 1847. DOI:10.3390/jmse10121847
- [21] 何昊, 王鹏. 高速变形飞行器制导控制一体化设计方法[J]. 航空学报, 2024, 45(S1): 299
HE Hao, WANG Peng. Integrated design method of guidance and control for high-speed morphing vehicle[J]. Chinese Journal of Aeronautics, 2024, 45(S1): 299. DOI:10.7527/S10006893.2024.30692
- [22] 王婕. 弹性高超声速飞行器跟踪问题控制方法研究[D]. 天津: 天津大学, 2014
WANG Jie. Research on control methods for tracking problems of flexible hypersonic vehicles[D]. Tianjin: Tianjin University, 2014
- [23] 李永丰, 史静平, 章卫国, 等. 深度强化学习的无人作战飞机空战机动决策[J]. 哈尔滨工业大学学报, 2021, 53(12): 33
LI Yongfeng, SHI Jingping, ZHANG Weiguo, et al. Air combat maneuver decision-making for unmanned combat aerial vehicle based on deep reinforcement learning[J]. Journal of Harbin Institute of Technology, 2021, 53(12): 33. DOI:10.11918/202005108
- [24] 高长生, 陈尔康, 荆武兴. 高超声速飞行器机动规避轨迹优化[J]. 哈尔滨工业大学学报, 2017, 49(4): 16
GAO Changsheng, CHEN Erkang, JING Wuxing. Trajectory optimization for hypersonic vehicle evasive maneuvers[J]. Journal of Harbin Institute of Technology, 2017, 49(4): 16. DOI:10.11918/j.issn.0367-6234.201605015
- [25] BANERJEE C, Chen Zhiyong, NOMAN N. Improved soft actor-critic: Mixing prioritized off-policy samples with on-policy experiences[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 35(3): 3121. DOI:10.1109/tnnls.2022.3174051
- [26] 会鑫, 王宁, 王帅. 基于局部状态感知的无人艇深度强化学习路径规划[J/OL]. (2025-05-28) [2025-07-03]. <http://link.cnki.net/doi/10.19693/j.issn.13185.04390>
HUI Xin, WANG Ning, WANG Shuai. Path planning for unmanned surface vehicles based on local state-aware deep reinforcement learning[J/OL]. (2025-05-28) [2025-07-03]. <http://link.cnki.net/doi/10.19693/j.issn.13185.04390>
- [27] 闫天, 程昊宇, 高萌靖, 等. 基于预设性能的导弹拦截鲁棒智能制导律[J]. 宇航学报, 2024, 45(5): 753
YAN Tian, CHENG Haoyu, GAO Mengjing, et al. Prescribed performance-based robust intelligent guidance law for missile interception[J]. Journal of Astronautics, 2024, 45(5): 753. DOI:10.3873/j.issn.1000-1328.2024.05.010
- [28] 路坤锋, 贾晨辉, 黄旭, 等. 面向变构型飞行器的强化学习位置姿态一体化控制方法[J]. 宇航学报, 2024, 45(7): 1100
LU Kunfeng, JIA Chenhui, HUANG Xu, et al. Integrated position and attitude control method based on reinforcement learning for morphing aircraft[J]. Journal of Astronautics, 2024, 45(7): 1100. DOI:10.3873/j.issn.1000-1328.2024.07.011