

控制与决策

Control and Decision

一种时限可达域引导的航天器博弈决策学习方法

乔贝贝, 刘薛怡, 钱寒雨, 许静文, 肖冰

引用本文:

乔贝贝, 刘薛怡, 钱寒雨, 等. 一种时限可达域引导的航天器博弈决策学习方法[J]. *控制与决策*, 2025, 40(12): 3678–3688.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2025.0313>

您可能感兴趣的其他文章

Articles you may be interested in

[航天器输入受限的鲁棒自适应姿态跟踪控制](#)

Robust adaptive attitude tracking control of spacecraft with constrained inputs

控制与决策. 2021, 36(9): 2297–2304 <https://doi.org/10.13195/j.kzyjc.2020.0013>

[多无人机协同直播场景下自适应任务卸载决策](#)

Adaptive task offloading decision of multi-UAVs cooperation in live broadcasting scenario

控制与决策. 2021, 36(4): 974–982 <https://doi.org/10.13195/j.kzyjc.2019.1104>

[天临空协同对地观测任务规划模型与并行竞争模因算法](#)

Planning model and parallel competing memetic algorithm for space–near space–air based cooperative earth observation missions

控制与决策. 2021, 36(3): 523–533 <https://doi.org/10.13195/j.kzyjc.2020.0732>

[基于深度学习的四旋翼无人机地面效应补偿降落控制设计](#)

Robust landing controller design for quadrotor unmanned aerial vehicle ground effects compensation via deep learning

控制与决策. 2021, 36(11): 2637–2646 <https://doi.org/10.13195/j.kzyjc.2020.0184>

[分布式无人机的时变编队非线性控制设计](#)

Time-varying formation nonlinear control of distributed multiple UAVs

控制与决策. 2021, 36(10): 2490–2496 <https://doi.org/10.13195/j.kzyjc.2020.0136>

一种时限可达域引导的航天器博弈决策学习方法

乔贝贝¹, 刘薛怡¹, 钱寒雨¹, 许静文², 肖冰^{1†}

(1. 西北工业大学 自动化学院, 西安 710072; 2. 西北工业大学 航天学院, 西安 710072)

摘要: 针对在复杂空间博弈场景中, 脉冲推力驱动的航天器追逃博弈决策实时性受限和传统奖励函数较难适应远距离和高动态对抗学习环境的问题, 对航天器博弈对抗的智能机动决策和燃料优化展开研究. 首先, 建立轨道博弈动力学和机动约束模型; 然后, 提出一种具有时间约束的航天器单脉冲可达域求解方法, 并结合神经网络对轨道危险区进行量化拟合; 接着, 基于分布式系统架构设计层次强化学习框架, 采用近端策略优化 (PPO) 算法开展红蓝对抗学习训练; 最后, 对所提出机动策略进行验证. 仿真结果表明, 在二体动力学轨道博弈场景中, 危险区策略可使得平均燃料消耗降低 33.81%, 博弈策略相较于传统方法打靶率平均可提升 38.41%.

关键词: 航天器追逃; 轨道博弈; 机动决策; 可达域; 层次强化学习

中图分类号: V474.2+8 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.0313

引用格式: 乔贝贝, 刘薛怡, 钱寒雨, 等. 一种时限可达域引导的航天器博弈决策学习方法 [J]. 控制与决策, 2025, 40(12): 3678-3688.

A time-limited reachable domain-guided learning method for spacecraft game decision-making

QIAO Bei-bei¹, LIU Xue-yi¹, QIAN Han-yu¹, XU Jing-wen², XIAO Bing^{1†}

(1. School of Automation, Northwestern Polytechnical University, Xi'an 710072, China; 2. School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: To address the issues of real-time decision-making limitations in impulse-thrust-driven spacecraft pursuit-evasion games and the incapability of traditional reward functions to adapt to long-distance high-dynamic adversarial learning environments, this paper investigates intelligent maneuver decision-making and fuel optimization for spacecraft game confrontations. Firstly, the orbital game dynamics and maneuver constraint model are established. Secondly, a time-constrained single-impulse reachable domain solving method for spacecraft is proposed, and neural networks are integrated to perform quantitative fitting of orbital danger zones. Furthermore, a hierarchical reinforcement learning control framework is designed based on a distributed system architecture, and the proximal policy optimization (PPO) algorithm is employed to carry out red-blue adversarial learning training. Finally, the proposed maneuver strategies are validated. Simulation results demonstrate that in the two-body dynamics orbital game scenario, the danger zone strategy reduces average fuel consumption by 33.81%, and the game strategies improve the hit rate by an average of 38.41% compared with traditional strategies.

Keywords: spacecraft pursuit-evasion; orbital game; maneuver decision-making; reachable domain; hierarchical reinforcement learning control

0 引言

随着航天技术不断发展以及空间任务日趋复杂, 现代战争往往利用交会对接方式接近敌方航天器, 形成摧毁打击. 传统以距离为导向作为激励的模式难以适应高不确定性和强实时性的时敏空间任务^[1], 而在轨捕获技术^[2-3]以及空间交会技术^[4]是最优控制与动态博弈的深度结合, 可作为一种航天器追逃

博弈 (SPE Game) 问题对其进行分析. SPE Game 需要在仅知非合作目标有限状态的前提下, 实现有限时间内航天器的逼近或逃逸智能决策行为并最终完成博弈任务^[5]. 而卫星智能机动决策依赖大规模的地面训练和星上二次训练^[6], 且目前在时间约束下对非合作目标信息的获取手段较少. 因此, 亟需一种可高效提升训练航天器智能决策水平的训练赋能方法.

收稿日期: 2025-03-26; 录用日期: 2025-07-25.

基金项目: 国家自然科学基金项目 (62273277).

†通信作者. E-mail: xiaobing@nwpu.edu.cn.

针对 SPE Game 问题,目前多数文献采用了微分对策方法进行求解^[7-9],因其约束条件呈非线性、哈密顿-雅可比方程的耦合性,以及在 SPE Game 问题中非合作目标的状态读取受到限制,该方法通常很难求得解析解^[10].而深度强化学习作为另外一种解决 SPE Game 问题的方法,近些年得到了许多学者的重视^[11-12],该方法通过对环境进行建模,使得智能体在环境中不断试错,训练智能体在特定环境下的智能决策,完成最优博弈问题的求解.在深度强化学习 SPE Game 问题中,尽管现有的大部分深度强化学习方法能够有效完成博弈训练任务,但是,在连续状态空间中应用时仍然存在与表格强化学习^[13]类似的问题,即操作维数的增加会导致所需显式表示的操作数量呈指数级增长.鉴于深度强化学习在智能决策方面的学习能力和目前在连续状态空间的限制,一些学者利用可达域预测技术来辅助航天器进行机动^[14-16],提高其学习能力. Chernick 等^[17]基于可达集理论构建了相对运动轨道,将航天器燃料最优问题转化为几何路径规划问题,进一步降低了航天器在近距离相对运动中的碰撞风险;许展鹏^[18]在现有求解方法的基础上,提出采用奇诺多面体法计算航天器可达集,实现了对航天器近距离碰撞的预警和防护.但是,在 SPE Game 问题中,航天器机动持续时间远短于转移轨道周期,而现有的多数可达域求解方法往往是对整个轨道周期进行求解^[15-16],并不适用于 SPE Game.基于此,本文提出一种基于非合作目标时限可达域预测的层次强化学习博弈学习框架.

本文的主要内容如下:1) 优化现有单脉冲可达域预测方法^[15]的求解速度;提出一种基于时间约束下的单脉冲可达域求解方法,并将其作为机动策略引入 SPE Game,能够提高航天器的学习能力.2) 提出一种基于聚类算法、最小二乘法和神经网络的接近危险区^[16]神经网络拟合求解方法,在保证航天器追逃性能的同时能够降低燃料消耗.3) 将以上两种方法引入层级强化学习框架^[19],并结合粒子群优化(PSO)算法对奖励函数系数进行优化,在节省油耗的同时,能够大幅度提升追踪器和目标器的追逃性能.

1 问题描述与建模

1.1 基于可达域预测的航天器追逃博弈问题

航天器可达域是指航天器在给定初始条件和控制约束下,对所有可能的控制序列进行积分求解,当积分结果所对应的状态向量满足特定的边界判定条件时,该状态向量所对应的空间位置即为可达域包

络点,在一段时间内所有可能到达的可达域包络点集合称为可达域包络体^[20].在目前的研究中,航天器的初始轨道通常被认定为是椭圆形或圆形,在航天器能够执行脉冲机动的前提下,限制脉冲大小,则根据可达域的定义,其初始轨道必定在可达范围内,因为初始轨道相当于脉冲为0时的转移轨道,而可达域的外包络和内包络可认为是航天器在最大脉冲以及最小脉冲情况下机动得到的状态集合面,如图1所示.

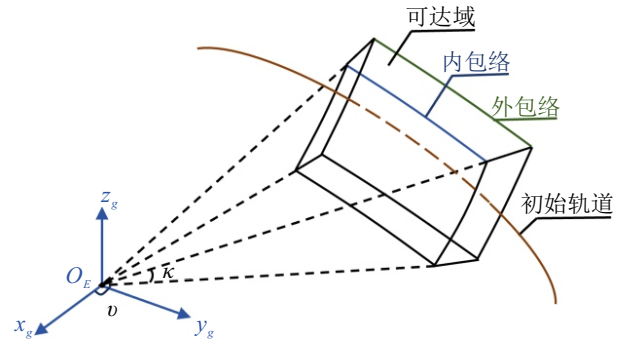


图1 非合作目标可达域示意图

本文的可达域均在近心点坐标系下表示^[21-22].由于求解包络面的连续点比较困难,现有的研究通常以包络面的离散点作为包络点的求解^[15],只要给定一个方位角 $[v, \kappa]$,便能够确定空间的一个方向,只要求得该方向上的可达矢径 $r_{\max}(v, \kappa)$ 和 $r_{\min}(v, \kappa)$,便能够确定可达域外包络以及内包络上的两个点.

在航天器追逃博弈问题中,追踪器和目标器可通过可达域预测方法来评估各自的策略有效性,从而对机动趋向进行指导,提高航天器追踪以及规避危险的效率^[14].本文考虑两个于地球静止轨道(GEO)附近绕地球运行的航天器,追踪器希望通过策略机动尽快地打击或逼近目标器,而目标器则希望在规定的约束内逃离追踪器的追踪或延长追踪时间,因此,目标函数可表示为

$$\max_p \min_e \left(\int_0^{t_f} dt \left| \min J(\mathbf{s}_p - \mathbf{s}_e, \mathbf{s}_p, \mathbf{s}_e, \text{fuel}) \right. \right). \quad (1)$$

其中: $\mathbf{s} = (x, y, z, v_x, v_y, v_z)$ 为航天器的状态; fuel 为剩余燃料; $J(\cdot)$ 为PE双方的机动策略代价函数,与航天器任务消耗正相关.

追踪器 P 需要最大化有效逼近的时间积累 $\int_0^{t_f} dt$, 即追求在时间 t_f 内快速缩短与目标器的距离,同时需要最小化任务消耗 $J(\cdot)$; 目标器 E 需要最小化追踪器的有效逼近积分,即通过增大距离或消耗追踪器燃料阻碍追踪器逼近,迫使追踪器在约束下无法有效逼近. P 、 E 双方通过代价函数优化自身的策略输出使其至最小,满足纳什均衡条件.因此,

在以此为前提的情况下,通过对目标函数进行优化求解即可得到 P 、 E 双方的最优策略机动。

1.2 坐标系定义

为描述航天器的相对运动以及空间可达域,本文建立以下坐标系:1)地心赤道惯性坐标系;2)近心点坐标系;3)轨道坐标系;4)CW (Clohessy-Wiltshire)参考坐标系.坐标系1)~坐标系3)的详细介绍可参见文献[15],各坐标系间的关系如下。

1)近心点坐标系与惯性坐标系可通过欧拉角序列3-1-3相互转换,旋转矩阵如下所示:

$$\mathbf{R}_g^J = \mathbf{R}_z(-\Omega) \cdot \mathbf{R}_x(-i) \cdot \mathbf{R}_z(-\omega). \quad (2)$$

2)轨道坐标系 S_0 与近心点坐标系 S_g 间的转移矩阵以及轨道坐标系 S_0 与 S_1 间的转移矩阵分别为

$$\mathbf{R}_0^g = \begin{bmatrix} \cos f & -\sin f & 0 \\ \sin f & \cos f & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{R}_0^1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i & \sin i \\ 0 & -\sin i & \cos i \end{bmatrix}. \quad (3)$$

3)CW参考坐标系:本文设定的追踪器与目标器均相对于星下点 O (50°E,0°N)运行于GEO轨道附近^[23],以该点 O 作为坐标原点建立CW参考坐标系.原点 O 与追踪器以及目标器的质心固联, x 轴与指向该星下点的地心矢径 r_1 相连, y 轴在GEO轨道平面内垂直于 x 轴,以目标器的运动方向为正, z 轴与 x 轴和 y 轴满足右手法则,如图2所示。

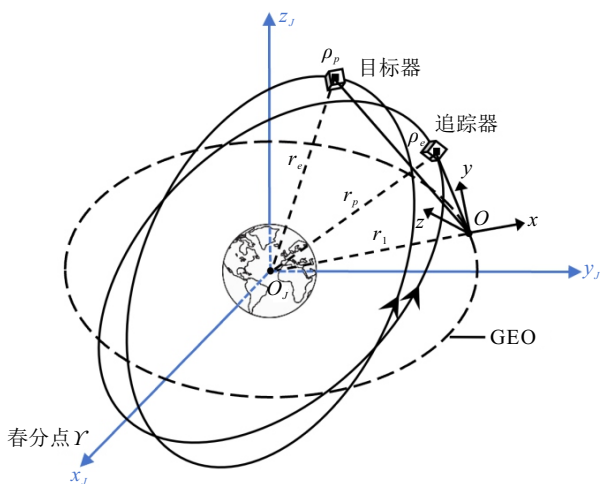


图2 CW参考坐标系

1.3 航天器博弈机动建模

在本文追逃(PE)博弈中,将目标器设于CW参考系坐标原点, P 、 E 双方均相对于原点以近圆轨道相对运行,且在整个运行过程中双方的相对距离相比于GEO轨道半径可以忽略^[6],因此,可通过CW方程近似描述双方的相对运动,如下所示:

$$\begin{cases} \ddot{x} - 2n\dot{y} - 3n^2x = a_x, \\ \ddot{y} + 2n\dot{x} = a_y, \\ \ddot{z} + n^2z = a_z. \end{cases} \quad (4)$$

其中: (x, y, z) 为航天器的相对位置信息; $n = \sqrt{\mu/r^3}$ 为GEO轨道平均角速度,与引力常数 μ 和轨道半径 r 相关; (a_x, a_y, a_z) 为航天器的三轴脉冲。

在进行轨道递推时,对于脉冲推力发动机而言,速度增量可认为是瞬间产生的^[6],因此,可通过对式(2)的相对运动方程进行积分将其以矩阵形式进行表达,如下所示:

$$\begin{bmatrix} \boldsymbol{\rho}(\tau) \\ \mathbf{v}(\tau) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}_{\rho\rho}(\tau) & \boldsymbol{\Phi}_{\rho v}(\tau) \\ \boldsymbol{\Phi}_{v\rho}(\tau) & \boldsymbol{\Phi}_{vv}(\tau) \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}(0) \\ \mathbf{v}(0) \end{bmatrix}. \quad (5)$$

其中: $\boldsymbol{\rho} = (x, y, z)$ 和 $\mathbf{v} = (\dot{x}, \dot{y}, \dot{z})$ 分别为航天器的相对位置以及相对速度; $\boldsymbol{\Phi}(\tau)$ 为相对运动的状态转移矩阵,具体表达式参见文献[24]。

追踪器的目的是通过策略机动在规定的时间内逼近目标器或迫使其进入至接近危险区范围内,而目标器则希望在规定的时间内远离追踪器以及接近危险区.目标函数如式(1)所示。

此外,综合航天器在博弈任务中受到的各种制约因素,需要设定一些约束条件:1)航天器的决策频率 f_p, f_e 约束: $\Delta t > t_{\max}$;2)航天器的脉冲约束: $|\Delta \mathbf{v}|_{p_{\max}} < |\Delta \mathbf{v}|_{e_{\max}} < |\Delta \mathbf{v}|_{\max}$ 。

本文的目的是通过策略设计,使得航天器在满足上述约束的前提下,实现目标函数的最大优化。

2 时间约束下单脉冲可达域求解方法设计

目前,常见的航天器单脉冲可达域求解方法大多是在运行时间不受限制的假设下实现的,但是,在实际博弈任务中,受限于航天器材质、燃料的不足以及任务的时间要求,需要设计一种能够满足航天器博弈任务要求的时限可达域预测方法.假设追踪器的运行时间为 $t < t_{\max}$,其时限可达域预测集是无约束可达域预测集的一部分,如图3所示.其中:浅蓝色区域为无时间约束条件下追踪器所能抵达的空间位置集,阴影部分为受到时间约束的时限可达域.由于两个集合的包络重合部分距机动位置 f_0 较近,转移时间较短,该部分的可达位置相同.红线是时限可达域区别于蓝色区域的包络线($t = t_{\max}$).

2.1 无约束可达域

所提出无约束可达域求解方法以文献[21-22]提出的可达域求解框架为基础,验证了可达性的速度矢量轨迹为边界速度双曲线(TVH),并通过脉冲机动球与转移轨道面的二次曲线相交关系推导出可达性判断条件;经文献[15]进一步优化,通过引入变

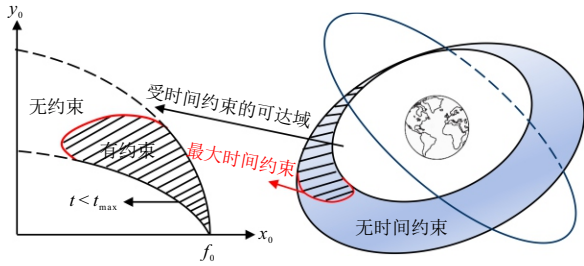


图3 时间约束下的单脉冲可达域示意图

步长二分法替代传统非线性迭代, 结合真近角遍历与局部寻优策略求解给定方位角的可达矢径极值, 如在机动位置自由场景中, 先筛选使得可达域包络矢径取极值的真近角区间, 再通过二分法提升精度. 相较于上述两种方法, 所提出方法的求解效率更高.

2.1.1 可达性判断条件

由轨道两点边值问题的结论可知^[20-21]: 对于所有能够到达给定目标位置矢量的转移轨道, 其初始速度矢量的集合构成了一条边界速度双曲线, 如图4所示. 由图4分析可知: 以轨道机动位置 r_0 为起点, 终端位置 r_f 为终点的两点边值问题, 若要使得 $[v, \kappa]$ 定义的 r_f 目标方向可达, 则必须保证机动后的速度始终能够落在边界速度双曲线 TVH 上, 且满足前提条件: 机动球与转移轨道的截面位于 TVH.

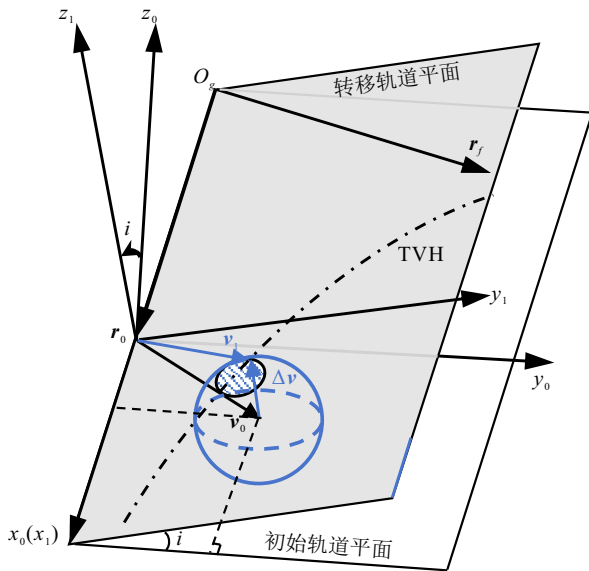


图4 转移轨道面与机动球位置关系示意图

由文献 [21] 可知, 速度脉冲 Δv 在截面内的分量 Δv_M 为

$$\Delta v_M = \sqrt{\Delta v^2 - (\mu/p_0)(1 + e_0 \cos f)^2 \sin^2 i}. \quad (6)$$

只需要满足根号下表达式非负即可满足轨道两点边值问题的条件, 即

$$0 \leq \tan^2 \kappa \leq \frac{\sin^2(v - f)}{[\mu/(p_0 \Delta v^2)](1 + e_0 \cos f)^2 - 1}. \quad (7)$$

其中: $[v, \kappa]$ 为描述机动后未来可达位置矢量 r_f 的空间指向, 详细说明参见图1; p_0, e_0, f 分别为初始轨道的半通径、偏心率、真近点角; Δv 为脉冲增量.

2.1.2 可达矢径求解方法

若给定满足式(7)的方位角 $[v, \kappa]$ 以及目标位置矢量 r_f , 则航天器可从 r_0 以机动速度 v_1 转移至 r_f , 基于二体轨道动力学模型^[24], 两点边值问题的终端矢量 r_f 可表示为

$$r_f = \frac{h^2}{\mu(1 - \cos \Delta\theta) + hv_t \cos \Delta\theta - hv_r \sin \Delta\theta}. \quad (8)$$

其中: h 为角动量的大小; v_r 和 v_t 分别为转移轨道面内速度矢量的径向分量以及切向分量; $\Delta\theta$ 为 r_0 与 r_f 的夹角, 且满足

$$\Delta\theta = \arccos(\cos(v - f) \cos \kappa). \quad (9)$$

目标位置矢量 r_f 是关于 f, κ, v, α 的方程^[20-21], 且在固定单脉冲机动下, 前3个变量给定, 只需求出 α 即可求出目标位置矢量的大小, 如下所示:

$$P(\alpha) = \left[2 \frac{\mu}{hv_t} (1 - \cos \Delta\theta) - \frac{v_r}{v_t} \sin \Delta\theta \right] \cos \alpha - \sin \Delta\theta \sin \alpha. \quad (10)$$

在已知函数 $P(\alpha)$ (除 α 外)所有变量的情况下, 通过牛顿迭代法求解 $P(\alpha) = 0$ 即可得到 $\{\alpha_1, \alpha_2\}$, 则航天器在给定方位角下的近心点坐标值为 $r_{f_{\max}} = r(\alpha_1)P(v, \kappa)$, $r_{f_{\min}} = r(\alpha_2)P(v, \kappa)$, $P(v, \kappa)$ 为该可达方向上的单位向量, 由下式求解:

$$P(v, \kappa) = [\sin v \cos \kappa, \cos v \cos \kappa, \sin \kappa]. \quad (11)$$

不同于文献 [15] 通过迭代相同的初始值求 $P(\alpha) = 0$ 的解, 本文将上一步的 α 解作为本次迭代的初始值, 因为在相邻方向内 $P(\alpha) = 0$ 的解相近, 可通过使用上一个相邻方向的 α 值作为本次 $P(\alpha) = 0$ 迭代的初始值, 当本次的 α 解求出后, 也被作为相邻方向 $P(\alpha) = 0$ 迭代的初始值, 此方法的求解速率以及稳定性更高. 综上, 总结为算法1.

算法1 无约束可达域求解算法.

step 1: 给定步长 Δv 和 $\Delta \kappa$ 确定方位角 $[v, \kappa]$.

step 2: 根据式(7)判断可达性, 若通过, 则进入下一步; 否则, 返回至 step 1.

step 3: 结合所提出方法求解式(10), 确定航天器近心点坐标值.

step 4: 重复 step 1 ~ step 3.

2.2 时限可达域

时限可达域的时间约束 t_{\max} 主要包括两个部分: 一是航天器在初始轨道上的自由漂移时间 t_1 , 二是

机动操作后所需的转移轨道时间 t_2 ,且二者满足 $t_1 + t_2 \leq t_{\max}$.所提出时限可达域求解方法应用于博弈任务,所求可达域是以机动点 f_0 为起点的可达包络范围,因此 $t_1 = 0, t_2 \leq t_{\max}$.在受时间约束 t_{\max} 的影响下,航天器可达域在 $t = t_{\max}$ 的可达包络并不沿着原本的趋势向外扩散,而是逐渐收敛得到确定的包络边界,如图3所示.由于此约束下的可达矢径受时间影响,应设计方法求解航天器在满足 $t_2 = t_{\max}$ 情况下所能到达的空间位置集.本节提出一种可用于轨道博弈分析的时限可达域求解方法.

2.2.1 可达性判断条件

受限于最大时间 t_{\max} ,方位角 $[v, \kappa]$ 不能依次迭代,通过式(7)可知, κ 的最大值可由下式求出:

$$D_{\kappa_{\max}} = \frac{\sin^2(v - f)}{[\mu/(p_0 \Delta v^2)](1 + e_0 \cos f)^2 - 1}. \quad (12)$$

由于在抵达方位角 v 前航天器必须机动, f 应小于 v ,设定牛顿迭代法的初值为 $f_0 = v - \pi/2$,可求出与最大方位角 κ 对应的真近点角 f_1 .通过式(7)可知,在已知 $D_{\kappa_{\max}}$ 的前提下, κ 的最大值可表示为

$$\kappa_{\max} = \arctan \sqrt{D_{\kappa_{\max}}}. \quad (13)$$

给定方位角 v ,由式(13)可求出 κ_{\max} ,按照步长 Δv 和 $\Delta \kappa$ 确定 $[v, \kappa]$, $\kappa \in [-\kappa_{\max}, \kappa_{\max}]$,航天器可达性判断条件与式(7)相同.

2.2.2 时限可达矢径求解方法

根据开普勒方程,在已知航天器于坐标系 S_1 下的初始位置、机动速度以及近心坐标系下终端位置的条件,可求出航天器的转移时间 Δt ,如下所示:

$$\begin{aligned} \Delta t &= (E - E_0) - e(\sin E - \sin E_0) = \\ &= (E - E_0) + \frac{\sigma_0}{\sqrt{a}}[1 - \cos(E - E_0)] - \\ &= \left(1 - \frac{r_0}{a}\right) \sin(E - E_0). \end{aligned} \quad (14)$$

其中: $\sigma_0 = (\mathbf{r}_0 \cdot \mathbf{v}_0)/\sqrt{\mu}$; E, E_0 分别为终端和初始点的偏近点角; e 和 a 分别为转移轨道离心率和半长轴;真近点角 $f/2$ 与偏近点角 $E/2$ 同象限,关系如下所示:

$$\tan\left(\frac{f}{2}\right) = \sqrt{\frac{1+e}{1-e}} \tan\left(\frac{E}{2}\right). \quad (15)$$

时限可达矢径求解公式与式(8)相同,但是 α 需要根据转移时间求取,由算法1求出给定方位角的可达位置 $\mathbf{r}_{f_{\min}}$ 和 $\mathbf{r}_{f_{\max}}$,由式(2)和(3)的旋转矩阵可将航天器初始位置、机动速度转换至惯性坐标系下,如下所示:

$$[\mathbf{r}_J, \mathbf{v}_J]^T = \mathbf{R}_0^g \cdot (\mathbf{R}_0^1)^{-1} \cdot (\mathbf{R}_g^J)^{-1} \cdot [\mathbf{r}_1, \mathbf{v}_1]^T. \quad (16)$$

由式(16)进一步求解转移轨道半长轴 a 、离心率 e 以及初始真近点角 f_0 ,终端点的真近点角为 $f = f_0 + \Delta\theta$, $\Delta\theta$ 由式(9)求出,经式(15)可求出真近点角对应的偏近点角.综上,代入式(14)可求出航天器的转移时间 Δt .

对于转移时间满足 $\Delta t > t_{\max}$ 的可达矢径, α 值按照步长 $\Delta\alpha$ 从 $(\alpha_{\min}, \alpha_{\max})$ 迭代重新求解矢径,直至满足 $\Delta t < t_{\max}$ 和 $\Delta v < \Delta v_{\max}$, $\alpha_{\min}, \alpha_{\max}$ 分别为可达矢径 $\mathbf{r}_{f_{\min}}$ 和 $\mathbf{r}_{f_{\max}}$ 对应的 α 解.

综上,总结时限可达域求解方法为算法2.

算法2 时限可达域求解算法.

step 1: 给定初始轨道参数、最大脉冲 Δv_{\max} 以及最大转移时间 t_{\max} .

step 2: 方位角 v 在 $[0, 2\pi]$ 范围内按照步长 Δv 取值,并由式(13)确定 κ_{\max} ,方位角 κ 在 $[-\kappa_{\max}, \kappa_{\max}]$ 范围内按照步长 $\Delta\kappa$ 取值,确定方位角 $[v, \kappa]$ 并由式(11)计算 $\mathbf{P}(v, \kappa)$.

step 3: 由式(7)判断给定方位角是否满足可达条件,若满足,则继续step 4;否则,返回至step 2,对方位角重新取值.

step 4: 由算法1求解给定方位角的可达矢径 $\mathbf{r}_{f_{\min}}$ 和 $\mathbf{r}_{f_{\max}}$,并由式(14)~(16)求解所对应的转移时间 t_1 和 t_2 ,判断转移时间是否均满足 $t \leq t_{\max}$,若满足,则跳转至step 6;否则,继续step 5.

step 5: 若有可达矢径的转移时间满足 $t \leq t_{\max}$,则记录其所对应的近心点系坐标 $r(\alpha) \cdot \mathbf{P}(v, \kappa)$;对不满足 $t \leq t_{\max}$ 的可达矢径,按照步长 $\Delta\alpha$ 从 $(\alpha_{\min}, \alpha_{\max})$ 迭代重新求解矢径,并判断其转移时间是否满足 $t \leq t_{\max}$,将 α 值反向代入算法1中求解速度脉冲是否满足 $\Delta v < \Delta v_{\max}$,若均满足,则记录其对应的近心点坐标并跳转至step 6;否则,重新迭代 α 值求解矢径.

step 6: 若近心点坐标 $\mathbf{r}_{f_{\min}}$ 或 $\mathbf{r}_{f_{\max}}$ 存在,则其为方位角 $[v, \kappa]$ 的可达坐标;否则,该方位角不满足时间可达性,返回至step 2,重新迭代方位角 $[v, \kappa]$.

step 7: 重复step 2~step 6,直至方位角完全遍历,汇总所记录的近心点坐标即为航天器时限可达域.

2.3 轨道危险区

追踪器的无约束可达域是围绕其初始轨道的扁平形区域,将追踪器的可达域与目标器轨道的交集部分定义为轨道危险区^[16],如图5所示.由于追踪器的可达域有限,其与目标器轨道的位置关系可能呈现以下两种情况:1)不存在交叠区域;2)存在一段或两段交叠区域.在此基础上,危险区策略的设计需要优先判定危险区域的存在性,以确保后续策略的有效性和针对性.本文不深入探讨危险区的具体求解方法,但是经过验证,通过算法1计算得到的追踪器

可达域与目标器轨道在 xoy 平面内的交点数量, 与文献 [16] 所提出方法得出的数值解相吻合。

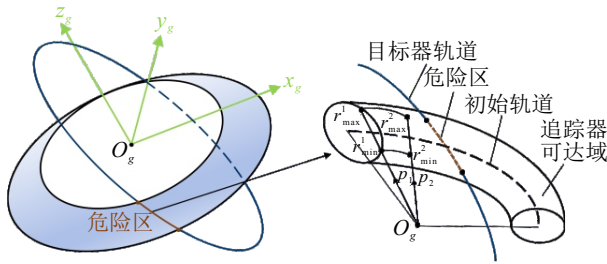


图5 接近危险区示意图

为提升模型训练效率和稳定性, 本文选用多层感知机 (MLP) 对平面无约束可达域的拟合曲线进行建模. 鉴于 MLP 具备优异的非线性拟合能力, 能够高效处理复杂曲线数据, 同时, 本文的网络输入参数分布较为规则, 无需引入复杂的深层网络结构, 采用简单的 MLP 即可准确拟合平面无约束可达域的包

络曲线. 最终, 通过拟合曲线与目标器轨道的交点, 可精准计算危险区域的数量。

首先, 给定不同的初始轨道参数并由算法 1 计算出追踪器对应的 xoy 平面无约束可达域, 得到初始数据集; 然后, 通过 DBSCAN (density-based spatial clustering of applications with noise) 聚类算法将初始数据集划分为内包络和外包络两组, 并使用最小二乘法拟合出内外包络曲线参数, 定义椭圆的参数方程为

$$\frac{(X \cos \theta + Y \sin \theta - x_0)^2}{a^2} + \frac{(Y \cos \theta - X \sin \theta - y_0)^2}{b^2} = 1; \quad (17)$$

最后, 构建输入和输出分别为 $[a, e, i, f, \text{fuel}]$ 和 $[X, Y, a, b, \theta]$ 的神经网络拟合平面可达域内外包络曲线, MLP 网络结构如图 6 所示。

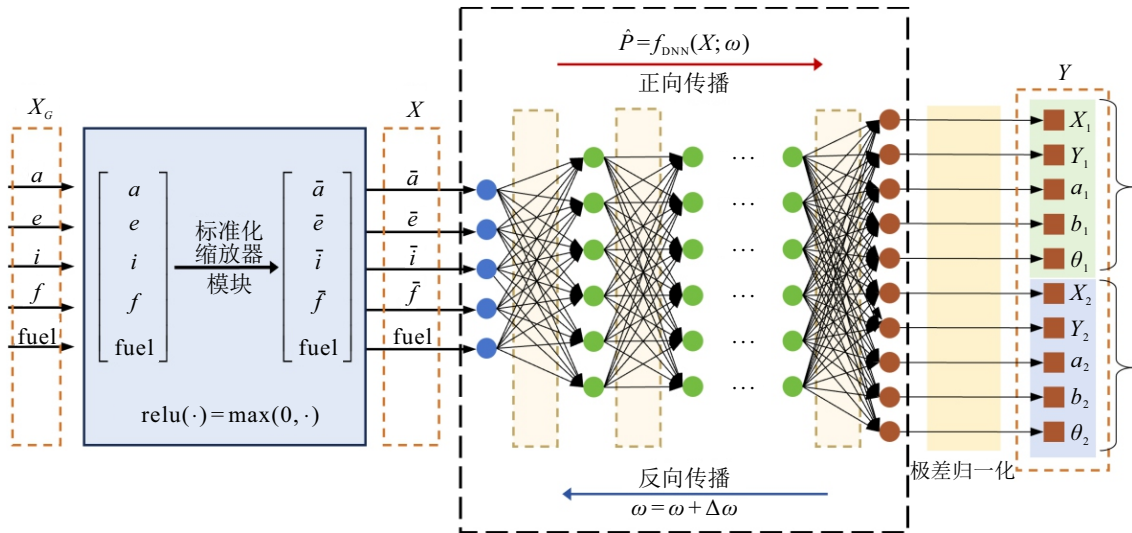


图6 危险区拟合神经网络示意图

在获得神经网络输出的内外包络拟合曲线参数后, 通过计算其与目标器轨道的交点即可确定轨道危险区域的具体数量。

3 基于层次强化学习智能决策框架

3.1 层次强化学习训练架构

本文层级强化学习框架基于任务 (goal) 进行划分^[19], 框架的目标是应对特定或人为定义的博弈场景, 具体而言, 追踪器需要逼近目标器至打击范围内, 获得获胜奖励值. 在此之前, 需要按照顺序执行以下几个步骤。

1) 任务 1: 在没有任何策略指导的情况下, 训练追踪器并收集追逃双方的轨道根数以及燃料信息. 完成后, 获得奖励 1。

2) 任务 2: 基于任务 1 获得的航天器状态计算

每一机动点处追踪器对应的无约束可达域, 并将其作为数据集来训练可达域拟合曲线网络. 完成后, 获得奖励 2。

3) 任务 3: 使用 PSO 优化算法寻找奖励系数的理论最优解. 完成后, 获得奖励 3。

4) 任务 4: 结合任务 2 和任务 3 的输出训练追踪器. 完成后, 获得奖励 4。

目标器的训练方式同理。

在本文层次强化学习训练中, 高层策略和底层策略以不同的时间跨度选择任务和 action. 高层策略基于 DQN (deep Q-network) 算法, 选择较长时间跨度的目标 (如任务周期); 底层策略则在较短时间跨度内选择动作 (如动作执行周期). 底层策略的训练方式差异化设计如下: 任务 1 和任务 4 采用 PPO

算法, 任务 2 基于神经网络近似, 任务 3 结合 PSO 算法进行策略参数优化. 高层策略通过内部奖励 (任务完成的阶段性回报/惩罚) 更新, 底层策略通过外部奖励 (动作即时环境反馈) 进行更新. 内部奖励是

当高层策略设定的任务被完成时, 高层网络收到的阶段性回报; 反之, 则收到惩罚. 外部奖励是底层策略在任务期间获得的回报或惩罚. 训练框架示意图如图 7 所示.

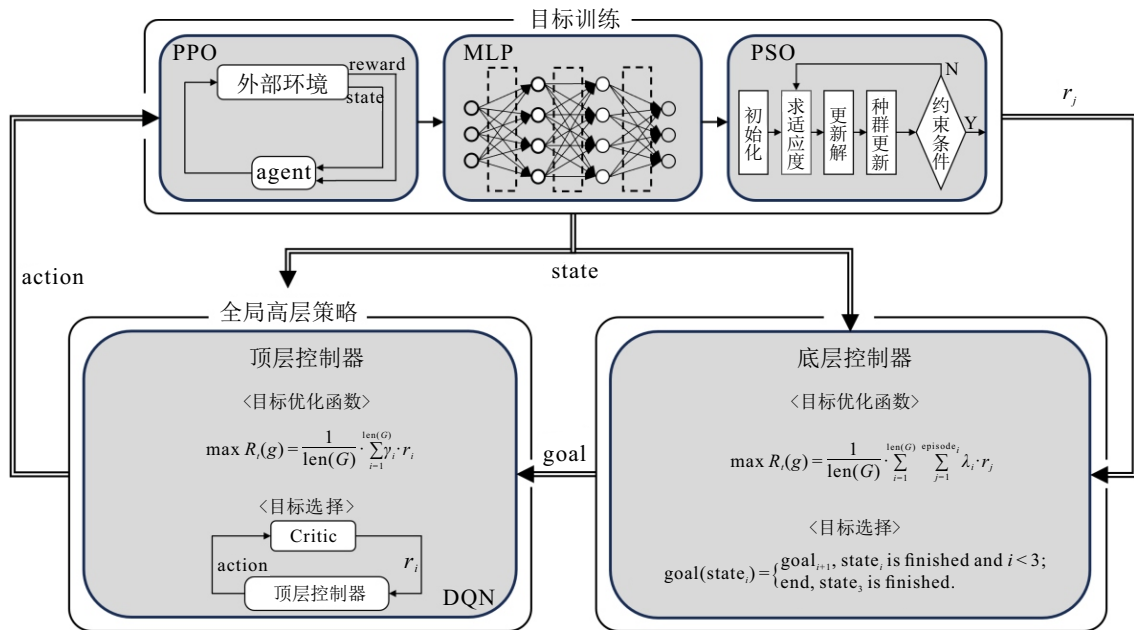


图7 层次强化学习框架

图 7 中: 底层控制器根据每个底层策略的输出 state 来选择目标 $goal \in G$, G 为所有底层策略的集合; 顶层控制器根据 state 和任务选择动作 action, 且在目标任务被实现或到截止状态前始终保持不变; Critic 网络负责评判目标任务是否被执行并考虑 state 的完成度来传递给顶层控制器一个阶段性奖励.

本文训练框架中的 Critic 网络与 AC (actor-critic) 算法中的 Critic 网络有所不同, 在 AC 算法中, critic 网络是一种用于评估 actor 网络输出值的模型; 而在本文中, Critic 网络仅仅是一种判断目标任务是否被执行的模型, 并据此判断为上层控制器提供相应的阶段性奖励.

顶层控制器引导策略的最优控制目标为

$$\max R_t(g) = \frac{1}{\text{num}(G)} \cdot \sum_{i=1}^{\text{len}(G)} \gamma_i \cdot r_i. \quad (18)$$

其中: $\text{num}(G)$ 为底层策略的数量, γ_i 为阶段奖励的衰减因子, r_i 为阶段性奖励.

底层控制器引导策略的最优控制目标为

$$\max R_t(g) = \frac{1}{\text{len}(G)} \cdot \sum_{i=1}^{\text{len}(G)} \sum_{j=1}^{\text{episode}_i} \lambda_i \cdot r_j. \quad (19)$$

其中: episode_i 为每个底层策略的训练轮数或迭代轮数, λ_i 为外部奖励的衰减因子, r_j 为智能体在任务期间获得的奖励.

3.2 底层策略

底层策略负责执行高层策略指定的具体子任务, 并反馈执行任务后的状态和奖励. 具体而言, 底层策略在接收到高层策略的高层动作指令后, 会根据当前环境状态选择并执行具体的底层动作, 以完成高层动作指令所定义的任务. 任务完成后, 底层策略会返回执行过程中获得的奖励信息和最终状态. 本文中, 将时限可达域预测方法和接近危险区拟合法作为两项奖励函数对追踪器的机动策略进行引导.

3.2.1 固定策略

在本文航天器的追逃博弈场景中, 通过 PPO 算法更新 AC 网络参数, 网络输入为追逃航天器的相对状态、各自的状态以及剩余燃料, 网络输出为各轴推力大小. 航天器状态由元组 $s = (x, y, z, \dot{x}, \dot{y}, \dot{z})$ 组成, 其中 (x, y, z) 为航天器相对 CW 参考点 $O(50^\circ E, 0^\circ N)$ 的相对坐标值.

奖励函数设计如下所示:

$$\begin{cases} R_p(t) = \delta_1 \cdot \rho(\hat{\mathbf{r}}_p, \hat{\mathbf{r}}_e) + \delta_2 \cdot \rho(\hat{\mathbf{r}}_e - \hat{\mathbf{r}}_p, \hat{\mathbf{v}}_p) + \\ \quad \delta_3 \cdot \rho(\hat{\mathbf{r}}_e - \hat{\mathbf{r}}_p, \hat{\mathbf{a}}_p) + \delta_4; \\ \delta_t = \begin{cases} 50, & \text{dis} < \text{dis_capture}; \\ 0, & \text{otherwise.} \end{cases} \end{cases} \quad (20)$$

其中: $\{\delta_i, i = 1, 2, 3, 4\}$ 为策略奖励系数; $\rho(\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2) = \hat{\mathbf{x}}_1 \cdot \hat{\mathbf{x}}_2 \in [-1, 1]$ 为单位向量 $\hat{\mathbf{x}}_1$ 与 $\hat{\mathbf{x}}_2$ 的相似度; 当博弈成功, 即相对距离进入打击范围时, 赋予追踪器

常值奖励 δ_t .

3.2.2 非合作目标信息预测策略

1) 危险区策略.

危险区表征在无机动脉冲提供的前提下,追踪器能够通过轨道自由飘逸逐渐逼近目标器.基于此,设计额外危险区奖励函数为

$$R_d(t) = \delta_4((1 - \cos(\pi \cdot n_d/2)) - 0.4)/2, \quad (21)$$

其中 n_d 为当前时刻危险区数量.且当追踪器与目标器间的相对距离进入约束范围 dis_r 内时,追踪器可通过无机动轨道飘逸来缩短二者间的距离.

2) 时限可达域预测策略.

通过算法2可求出目标器在给定时间下的可达域,基于已知结果,设计一种奖励策略,使得追踪器能够提前向可达域的平均位置点移动,设计如下所示:

$$R_r(t) = \delta_5 \cdot \rho(\hat{\mathbf{r}} - \hat{\mathbf{r}}_p, \hat{\mathbf{a}}_p), \quad (22)$$

其中 $\hat{\mathbf{r}} = \left(\sum_{i=1}^{Num} \hat{\mathbf{r}}_i \right) / Num$, Num为可达域中点集的个数, $\hat{\mathbf{r}}_i$ 为第 i 个位置的单位矢量.

同理,该策略也可用于确定追踪器的可达域,从而帮助目标器在未来一段时间内进行有效的危险规避机动.

4 仿真与分析

4.1 场景与参数设置

本文博弈场景设在GEO轨道附近,选择星下点 $O(50^\circ E, 0^\circ N)$ 作为CW参考点,其在惯性系下的位置和速度分别为 $(27\,098, 32\,306, 0)$ km和 $(-2\,350, 1\,970, 0)$ m/s.以参考点为球心建立半径为182 km的球体,追踪器的初始位置在球体表面随机设置,目标器与参考点重合,二者相对于参考点的速度均为0,两个航天器的运动状态由式(5)描述.

为验证所设计策略的有效性,设定两航天器的机动能力相同,单轴最大脉冲增量与燃料上限相同,分别为1.6 m/s和320 m/s;博弈场景打击距离 d_{cap} 和无控制约束范围 d_{range} 分别为50 km和100 km;仿真步长为100 s,每回合64步,博弈时间上限为6400 s.

本文实验设备如下:GPU为Nvidia RTX A6000,内存128 GB,针对本文架构进行训练.

4.2 算法有效性验证

1) 时限可达域验证.

为验证时限可达域的正确性,本文采取以下实验进行对比分析.航天器初始参数如表1所示.本文航天器机动时间为100 s,使用算法1求出无时间限

制下的航天器可达域并用算法2求出未来3000 s内的可达域,如图8所示.图8中:浅蓝色部分由文献[15]提供的方法求出,深蓝色部分由所提出时限可达域算法求出.观察发现二者的重合情况基本与图3以及文献[25]的结果相吻合,但是,本文对文献[15]的求解算法进行了优化,计算效率约提升了23.45%.

表1 航天器初始参数

参数/单位	值	参数/单位	值
a/m	1.279e7	$\Delta v/(m/s)$	300
i/rad	0.188	r_o/m	(1212, 6313, 1200)
e/rad	0.200	$v_o/(m/s)$	(10e6, 10e3, 0)
f/rad	1.392		

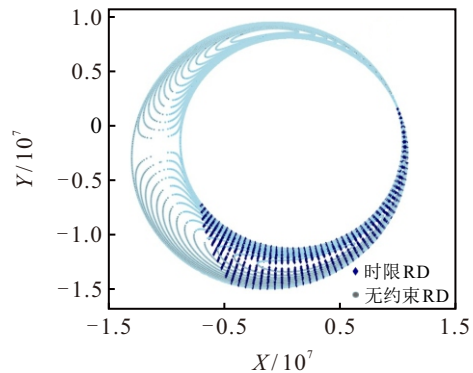


图8 算法1与算法2结果对比

使用蒙特卡洛打靶法进行验证,在机动点 f 处,对脉冲球均匀选取5000个样本点.将每个样本加到航天器初始速度并通过本文轨道外推模型外推5000 s,得到一条机动后5000 s的样本轨道.重复5000次,得到的轨道数据与算法2求出的时限可达域包络面进行对比,如图9所示.图9中:黑色部分为算法2求出的时限可达域包络面,蓝色部分为样本数据.

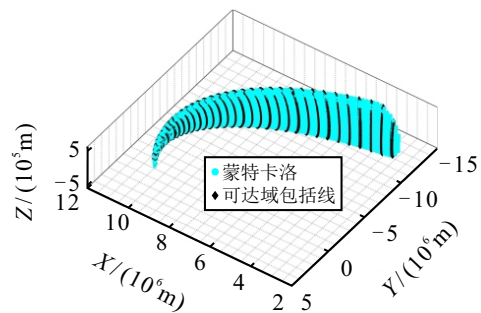


图9 蒙特卡洛打靶法验证

2) 危险区策略验证.

为验证所提出神经网络拟合法的效果,使用算法1获取可达域数据,输入已训练好的网络模型.然后,将模型的输出与原数据进行对比,图8为获取的

初始数据, 图 10 和图 11 分别为可达域包络点与拟合输出的对比. 其中: 淡灰色点为原数据的可达点, 蓝色椭圆由网络输出拟合参数绘制. 数值外推法和神经网络法的计算耗时分别为 14.643 ms 和 0.897 ms, 所提出方法降低了约 15.32 倍.

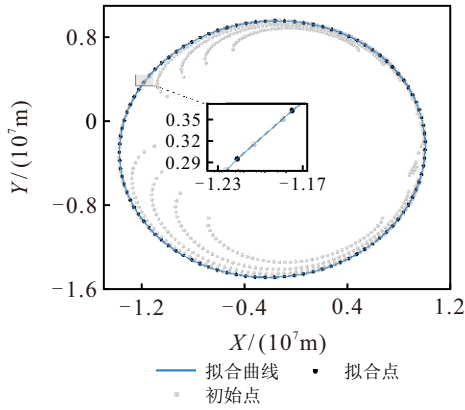


图10 平面可达域外包络拟合示意图

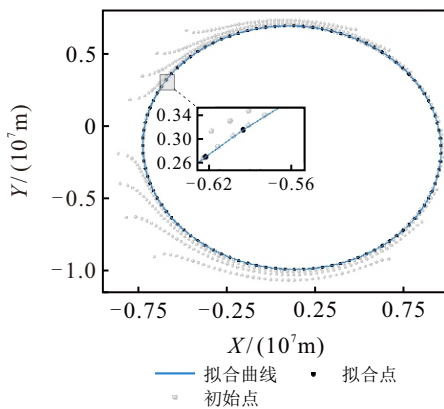


图11 平面可达域内包络拟合示意图

此外, 为验证危险区策略的有效性, 本文进行了两组最终距离相似的博弈测试, 一组应用危险区策略, 另一组作为对照不使用危险区策略, 燃料消耗情况如图 12 所示.

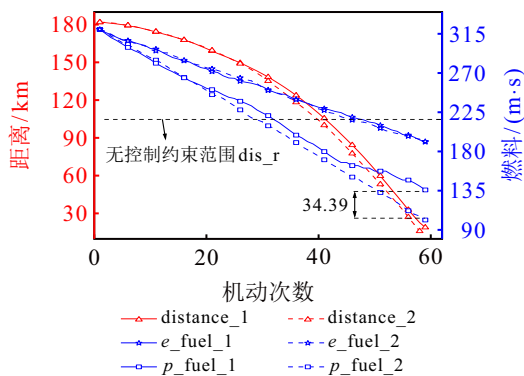


图12 危险区策略的燃料优化情况

图 12 分别展示了在不同策略下博弈双方的相对距离和剩余燃料的变化情况. 其中: distance 为相对距离, e_{fuel} 为目标器的剩余燃料, p_{fuel} 为追踪器的剩

余燃料, 实线数字 1 和虚线数字 2 分别表示有危险区策略和无危险区策略下的参数. 由图 12 可以观察到, 目标器的燃料消耗在两种策略下均呈现出相似的下降趋势. 然而, 当相对距离进入无控制约束范围 d_{range} 时, 采用危险区策略的追踪器在有危险区的情况下进入了自由飘逸阶段, 燃料消耗速度减缓, 而无危险区策略的追踪器燃料消耗仍然保持较快的下降速度. 最终, 采用有危险区策略的追踪器燃料相比于无危险区策略的追踪器燃料多了约 34.39 m/s, 油耗降低约 33.81%, 可以验证危险区策略的有效性, 追踪器能够更有效地控制燃料消耗, 在达到相对距离约束后的情况下保持较高的剩余燃料量.

3) 追逃验证.

为验证可达域策略的有效性, 分别针对追踪器和目标器进行验证. 追踪器和目标器分别使用可达域策略进行 5000 次打靶实验, 追踪器的目标是逼近目标器至 50 km 内, 目标器的目标是远离追踪器 130 km 以上. 两组实验结果如图 13 和图 14 所示, 使用可达域策略追踪器的打靶成功率为 84.11%, 而使用可达域策略的目标器的逃逸成功率为 92.45%.

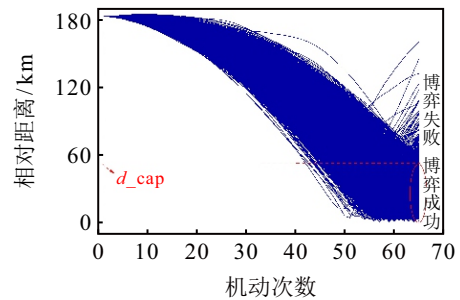


图13 可达域策略下追踪器打靶效果示意图

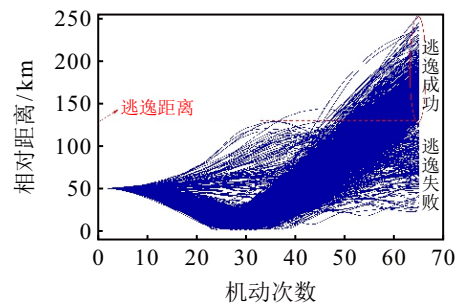


图14 可达域策略下目标器逃逸效果示意图

此外, 为验证所提出策略的有效性高于传统奖励函数 (以追逃双方距离为导向的奖励函数, 距离越近, 奖励越高; 反之, 则越低), 分别针对传统奖励函数和所提出 4 组策略组合方法进行 5000 次打靶实验, 所提出 4 种策略组合相比于传统设计的奖励函数的打靶率分别提升了 21.00%、22.25%、34.62% 和 42.2%, 打靶成功率如图 15 所示.

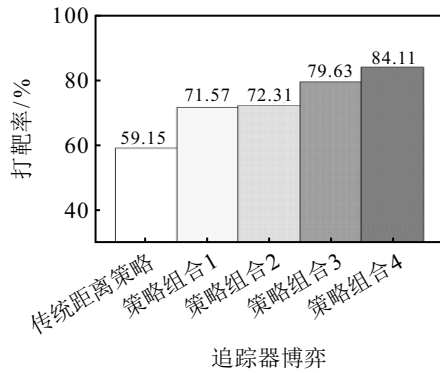


图15 不同策略下的追踪器打靶成功率

5 结论

本文针对机动脉冲和可达域作用下的轨道追逃博弈问题,建立了航天器相对动力学模型和相应的坐标系,在此基础上,研究了时限可达域的计算方法,并将危险区知识引入神经网络;结合航天器追逃博弈问题提出了可达域策略和危险区策略,并将其应用于层次强化学习智能决策框架中实现了航天器的高效率追逃博弈。

本研究是基于椭圆轨道和单脉冲机动的假设,并未考虑其他轨道场景,且强化学习策略依赖于单脉冲设计,对于多脉冲优化以及复杂扰动环境适应性不足。未来可拓展研究非椭圆轨道的可达域特性,探索异面多脉冲交会策略,结合物理约束改进神经网络泛化能力,并引入硬件约束与多目标优化指标,以提升模型在复杂航天任务中的工程适用性。

参考文献 (References)

- [1] 袁利. 面向不确定环境的航天器智能自主控制技术[J]. 宇航学报, 2021, 42(7): 839-849.
(Yuan L. Spacecraft intelligent autonomous control technology toward uncertain environment[J]. *Journal of Astronautics*, 2021, 42(7): 839-849.)
- [2] Macias V, Becerra I, Murrieta-Cid R, et al. Image feedback based optimal control and the value of information in a differential game[J]. *Automatica*, 2018, 90: 271-285.
- [3] Shen H X, Casalino L. Revisit of the three-dimensional orbital pursuit-evasion game[J]. *Journal of Guidance, Control, and Dynamics*, 2018, 41(8): 1823-1831.
- [4] 常燕, 陈韵, 鲜勇, 等. 机动目标的空间交会微分对策制导方法[J]. 宇航学报, 2016, 37(7): 795-801.
(Chang Y, Chen Y, Xian Y, et al. Differential game guidance for space rendezvous of maneuvering target[J]. *Journal of Astronautics*, 2016, 37(7): 795-801.)
- [5] 夏元清, 谢超, 高寒, 等. 天空地一体化网络环境下多运动体系统跨域协同控制与智能决策[J]. 控制与决策, 2023, 38(5): 1176-1199.
(Xia Y C, Xie C, Gao H, et al. Cross-domain cooperative control and intelligent decision-making of multi-dynamic agents in space-air-ground integrated network environment: A review[J]. *Control and Decision*, 2023, 38(5): 1176-1199.)
- [6] 袁利, 耿远卓, 汤亮, 等. 航天器轨道追逃博弈多阶段强化学习训练方法[J]. 上海航天: 中英文, 2022, 39(4): 33-41.
(Yuan L, Geng Y Z, Tang L, et al. Multi-stage reinforcement learning method for orbital pursuit-evasion game of spacecrafts[J]. *Aerospace Shanghai: Chinese & English*, 2022, 39(4): 33-41.)
- [7] 郝志伟, 孙松涛, 张秋华, 等. 半直接配点法在航天器追逃问题求解中的应用[J]. 宇航学报, 2019, 40(6): 628-635.
(Hao Z W, Sun S T, Zhang Q H, et al. Application of semi-direct collocation method for solving pursuit-evasion problems of spacecraft[J]. *Journal of Astronautics*, 2019, 40(6): 628-635.)
- [8] 陈燕妮. 基于微分对策的有限时间自适应动态规划制导研究[D]. 南京: 南京航空航天大学, 2019: 12-25.
(Chen Y N. Research on differential games-based finite-time adaptive dynamic programming guidance law[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2019: 12-25.)
- [9] 张秋华, 孙松涛, 谌颖, 等. 时间固定的两航天器追逃策略及数值求解[J]. 宇航学报, 2014, 35(5): 537-544.
(Zhang Q H, Sun S T, Chen Y, et al. Strategy and numerical solution of pursuit-evasion with fixed duration for two spacecraft[J]. *Journal of Astronautics*, 2014, 35(5): 537-544.)
- [10] 韩楠, 罗建军, 柴源. 多颗微小卫星接管失效航天器姿态运动的微分博弈学习控制[J]. 中国科学: 信息科学, 2020, 50(4): 588-602.
(Han N, Luo J J, Chai Y. Differential game learning approach for multiple microsatellites takeover of the attitude movement of failed spacecraft[J]. *Scientia Sinica: Informationis*, 2020, 50(4): 588-602.)
- [11] Cheng Y, Sun Z J, Huang Y X, et al. Fuzzy categorical deep reinforcement learning of a defensive game for an unmanned surface vessel[J]. *International Journal of Fuzzy Systems*, 2019, 21(2): 592-606.
- [12] 陈珂昕, 刘晓路, 淳洁, 等. 考虑多类型任务的成像卫星群调度模型与算法[J]. 控制与决策, 2025, 40(6): 1913-1921.
(Chen K X, Liu X L, Chun J, et al. Model and algorithm for scheduling imaging satellite constellations based on multi-type tasks[J]. *Control and Decision*, 2025, 40(6): 1913-1921.)
- [13] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529-533.
- [14] Gong H R, Gong S P, Li J F. Pursuit-evasion game for satellites based on continuous thrust reachable domain[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2020, 56(6): 4626-4637.
- [15] 杜向南, 杨震. 航天器单脉冲机动可达域求解算法[J]. 力学学报, 2020, 52(6): 1621-1631.
(Du X N, Yang Z. An algorithm for solving spacecraft

- reachable domain with single-impulse maneuvering[J]. Chinese Journal of Theoretical and Applied Mechanics, 2020, 52(6): 1621-1631.)
- [16] 张赛, 杨震, 杜向南, 等. 基于轨道可达域的机动航天器接近威胁规避方法[J]. 航空学报, 2024, 45(4): 328778.
(Zhang S, Yang Z, Du X N, et al. Threat avoidance strategy of spacecraft maneuvering approach based on orbital reachable domain[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(4): 328778.)
- [17] Chernick M, D'Amico S. Closed-form optimal impulsive control of spacecraft formations using reachable set theory[J]. Journal of Guidance, Control, and Dynamics, 2020, 44(1): 25-44.
- [18] 许展鹏. 航天器近距离相对运动可达分析与最优控制方法研究[D]. 长沙: 国防科技大学, 2019: 54-61.
(Xu Z P. Spacecraft close-range relative motion: Reachability analysis and optimal control[D]. Changsha: National University of Defense Technology, 2019: 54-61.)
- [19] Kulkarni T D, Narasimhan K R, Saeedi A, et al. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation[C]. Advances in Neural Information Processing Systems. Barcelona, 2016: 3675-3683.
- [20] 温昶焯. 面向空间态势感知的可达范围理论和应用研究[D]. 北京: 北京航空航天大学, 2015: 22-45.
(Wen C X. Space situational awareness oriented research on theory and application of reachable domain[D]. Beijing: Beihang University, 2015: 22-45.)
- [21] Wen C X, Zhao Y S, Shi P. Precise determination of reachable domain for spacecraft with single impulse[J]. Journal of Guidance, Control, and Dynamics, 2014, 37(6): 1767-1779.
- [22] Wen C X, Zhao Y S, Shi P, et al. Orbital accessibility problem for spacecraft with a single impulse[J]. Journal of Guidance, Control, and Dynamics, 2014, 37(4): 1260-1271.
- [23] 耿远卓, 袁利, 黄煌, 等. 基于终端诱导强化学习的航天器轨道追逃博弈[J]. 自动化学报, 2023, 49(5): 974-984.
(Geng Y Z, Yuan L, Huang H, et al. Terminal-guidance based reinforcement-learning for orbital pursuit-evasion game of the spacecraft[J]. Acta Automatica Sinica, 2023, 49(5): 974-984.)
- [24] 张洪波. 航天器轨道力学理论与方法[M]. 北京: 国防工业出版社, 2015.
(Zhang H B. Theories and methods of spacecraft orbital mechanics[M]. Beijing: National Defense Industry Press, 2015.)
- [25] 杜向南. 航天器机动可达域表征与威胁规避[D]. 长沙: 国防科技大学, 2021: 20-35.
(Du X N. Spacecraft orbital reachable domain representation and risk avoidance with impulsive maneuver[D]. Changsha: National University of Defense Technology, 2021: 20-35.)

作者简介

乔贝贝 (2000-), 男, 硕士生, 主要研究方向为航天动力学与控制, E-mail: qiaobeibe623@mail.nwpu.edu.cn;

刘薛怡 (2002-), 女, 硕士生, 主要研究方向为航天动力学控制, E-mail: liuxueyi@mail.nwpu.edu.cn;

钱寒雨 (1997-), 男, 博士生, 主要研究方向为集群博弈规划与控制, E-mail: qianhanyu@mail.nwpu.edu.cn;

许静文 (1991-), 女, 助理研究员, 主要研究方向为航天测控与在轨智能管控技术, E-mail: xjw0324@nwpu.edu.cn;

肖冰 (1986-), 男, 教授, 博士, 主要研究方向为无人系统导航、制导与控制, E-mail: xiaobing@nwpu.edu.cn.