



基于深度强化学习的无人飞艇路径跟随控制

曹希雯^{1,2,3}, 周艳⁴, 温广辉^{5*}

1. 北京理工大学自动化学院, 北京 100081

2. 北京理工大学临近空间环境特性及效应全国重点实验室, 北京 100081

3. 北京理工大学长三角研究院, 嘉兴 314019

4. 东南大学数学学院, 南京 211189

5. 东南大学自动化学院, 南京 210096

* E-mail: ghwen@seu.edu.cn

收稿日期: 2025-06-30; 接受日期: 2025-09-11; 网络版发表日期: 2025-11-06

国家自然科学基金(批准号: 62325304, U22B2046)和江苏省应用数学科学研究中心课题(编号: BK20233002)资助项目

摘要 针对无人飞艇路径跟随问题, 提出一种融合视线法与深度强化学习的无模型控制方法. 首先, 基于视线法将路径跟随问题转化为航向控制问题; 其次, 依据飞艇运动学与动力学特性, 将航向控制问题描述为马尔可夫决策过程; 然后, 设计基于深度确定性策略梯度算法的无模型控制策略, 实现对期望航向的精准控制, 最终达成路径跟随目标. 仿真结果表明, 在三种典型期望路径下, 所提出的方法相较于传统PID控制能够更加迅速地收敛到预期控制目标, 并在跟踪精度与系统稳定性方面展现出优越性能, 验证了该方法的有效性与优越性.

关键词 无人飞艇, 路径跟随控制, 视线法, 深度强化学习, 深度确定性策略梯度

1 引言

临近空间是介于传统航空器最高飞行高度和传统航天器最低轨道高度之间的空天结合战略区域, 蕴含着巨大的民用和军事价值, 是当前国际竞争的前沿领域和空天一体化博弈的新兴战略要地^[1,2]. 在此背景下, 发展部署相关新域新质力量, 培育壮大临近空间关键技术, 对维护国防安全和服务人民生活与经济社会发展具有重要意义. 无人飞艇作为典型的低动态临近空间飞行器, 依靠轻质气体提供升力, 并通过操纵控制机构执行飞行任务. 相较于其他临近空间飞行器, 无人飞艇具有长航时与区域驻留的独特

优势, 在实时监测、通信中继及导航定位等领域具有显著应用价值. 路径跟随控制是充分发挥飞艇效能的关键技术之一, 其旨在设计控制策略使飞艇能够平稳趋近并精确跟随与时间无关的预设参考路径.

飞艇平均密度与大气密度相近, 且具有体积大、转动惯量大的特性, 运行过程易出现响应慢、动力低、风阻弱等问题. 同时, 飞艇动力学的强非线性与多参数强耦合特性, 显著增加了路径跟随控制的实现难度. 在飞艇路径跟随控制研究中, 部分文献以线性控制为基础, 基于飞艇动力学的线性化模型进行控制器设计以满足闭环系统性能指标, 此

引用格式: 曹希雯, 周艳, 温广辉. 基于深度强化学习的无人飞艇路径跟随控制. 中国科学: 技术科学, 2026, 56: 259–267

Cao X W, Zhou Y, Wen G H. Deep reinforcement learning-based path following control for autonomous airship (in Chinese). Sci Sin Tech, 2026, 56: 259–267, doi: 10.1360/SST-2025-0164

类控制方法因控制器结构简单、控制有效而被广泛应用^[3~7]。然而当系统模型复杂或环境动态变化时控制性能显著降低。相较于线性控制方法,多数非线性控制方法在处理强非线性和不确定性系统时具有较大优势,可在特定条件下实现高精度控制,并有望保证闭环系统具备期望的稳定性与强鲁棒性^[8~12]。值得注意的是,飞艇结构的复杂性使系统建模与参数辨识极具挑战性。

强化学习作为一种机器学习方法,通过试错与环境交互获得策略的改进,且无需精确建模^[13]。目前,强化学习在无人飞艇控制领域的应用尚处于探索阶段,但在其他无人系统的控制问题上积累了丰富的经验。针对无人机路径跟随控制问题,Wang等人^[14]采用带有自适应奖励机制的强化学习方法实现了高精度路径跟随,并通过分层融合策略协调关键子任务,确保无人机能在复杂环境下稳定、精确地完成路径跟随任务。针对水面无人艇路径跟随问题,Wang等人^[15]提出双评论家确定性策略梯度积分补偿算法,有效提升控制器跟踪精度与鲁棒性。Zhao等人^[16]提出了带随机制动的深度强化学习方法,有效解决了无人艇编队路径跟随问题。

视线法(line of sight, LOS)已广泛应用于路径跟随控制领域。其核心原理是模拟船舶驾驶中的视线导航行为,将位置误差转换为航向误差实现路径跟随。传统LOS法通常采用下一路经点作为LOS点^[17],但易受扰动导致横向跟踪误差显著增大。为提升鲁棒性,Fossen等人^[18]基于横向跟踪误差与固定前瞻距离计算LOS点位置,但当横向误差过大时,系统可能出现收敛困难。为此,Moreira等人^[19]提出收敛性更优的LOS圆方法,并在此基础上改进为可变LOS圆,以解决固定LOS圆与路径之间无交点问题。

受上述研究成果启发,针对无人飞艇路径跟随问题,结合LOS与深度强化学习算法,提出了无模型的路径跟随控制策略。本文贡献如下:针对飞艇路径跟随问题中位置与姿态的双重控制需求,引入LOS法实时计算期望航向角,将路径跟随问题转化为航向控制问题,简化控制结构;基于无人飞艇动力学特性,将航向控制问题描述为马尔可夫决策过程,采用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实现对航向精准控制,最终达成路径跟随控制目标。

2 预备知识与问题描述

2.1 运动学与动力学模型

无人飞艇作为一种低动态临近空间飞行器,其升力源于艇囊内填充的气体(氢气或氦气)产生的静浮力^[20]。区别于高空气球的被动漂移,飞艇能够通过发动机及螺旋桨推进系统与气动操纵面实现主动可控航行。如图1所示,地面坐标系记为{ERF},其中,原点 O_g 固定于地面, $O_g x_g$ 轴指向正北方向, $O_g y_g$ 轴指向正东方向, $O_g z_g$ 轴指向地心;体坐标系记为{BRF},其中,坐标系原点 O_b 为无人飞艇的体积中心, $O_b x_b$ 轴指向艇首, $O_b y_b$ 轴指向飞艇右舷方向,且垂直于 $x_b O_b z_b$ 平面, $O_b z_b$ 轴垂直向下。为简化问题,本研究仅考虑无人飞艇的水平运动。为此,做出如下假设:将飞艇视为刚体且忽略其弹性形变;假定飞艇体积中心与浮心(升力中心)重合,且浮力与重力满足平衡条件 $B_f = mg$ 。

根据以上假设,飞艇的三自由度运动学与动力学模型表述如下^[20,21]:

$$\begin{cases} \dot{\eta} = R(\psi)V, \\ M\dot{V} = C(V)V + \tau_a + \tau_s + \tau_p, \end{cases} \quad (1)$$

其中, $\eta = [x, y, \psi]^T \in \mathbf{R}^3$ 为地面坐标系下的位姿向量, (x, y) 为飞艇的位置, ψ 为航向角; $V = [u, v, r]^T \in \mathbf{R}^3$ 为体坐标系下的速度向量, u 为沿 x_b 轴的纵向速度, v 为沿 y_b 轴的横向速度, r 为绕 z_b 轴旋转的角速度;

$$R(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ 为地面坐标系与体坐标系间}$$

的旋转矩阵; $M = M_{RB} + M_A \in \mathbf{R}^{3 \times 3}$ 表示总质量矩阵, M_{RB}

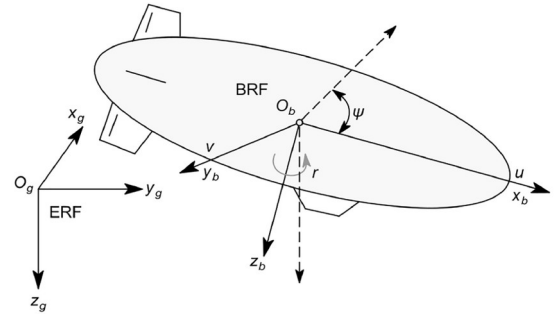


图1 无人飞艇和坐标系

Figure 1 The autonomous airship and coordinate system.

为惯性矩阵, M_A 为附加质量矩阵; $C(V)V \in \mathbf{R}^3$ 表示科里奥利力与离心力张量; $\tau_a \in \mathbf{R}^3$ 表示与空气相互作用产生的气动张量, 其值取决于相对速度; $\tau_s \in \mathbf{R}^3$ 表示由重力与浮力组成的静态张量; $\tau_p = [F_M, F_T, F_T O_y]^T \in \mathbf{R}^3$ 表示推进系统张量, F_M 和 F_T 分别代表主扭矩与尾部扭矩, O_y 表示尾部旋翼的位置在 $O_b x_b$ 轴上的投影。

2.2 LOS法

LOS法亦称为三点制导法, 其能将路径跟随问题转化为航向控制问题并实现优越的控制效果, 已成为路径跟随控制领域的常用技术^[22]。如图2所示, 设目标路径点为 $P_{k-1}(x_{k-1}, y_{k-1})$, $P_k(x_k, y_k)$ 和 $P_{k+1}(x_{k+1}, y_{k+1})$, 生成的路径描述为 $P_{k-1}P_kP_{k+1}$, 其中, P_k 表示第 k 个路径点, (x_k, y_k) 表示其位置。定义以 O_b 为圆心、 R_{LOS} 为半径的圆周, 且 R_{LOS} 满足

$$R_{\text{LOS}} = \begin{cases} 3L, & |d_e| \leq 3L, \\ |d_e| + L, & |d_e| > 3L, \end{cases} \quad (2)$$

其中, L 为飞艇的长度, $d_e = \sqrt{x_e^2 + y_e^2}$ 为飞艇与目标路径的距离误差, x_e 与 y_e 分别为飞艇的纵向误差与横向误差。

选择圆周与路径交点中距离下一个目标路径点 P_{k+1} 更近的点 $P_{\text{LOS}}(x_{\text{LOS}}, y_{\text{LOS}})$, 称为视线引导点, 该点的位置满足

$$\begin{cases} (x_{\text{LOS}} - x)^2 + (y_{\text{LOS}} - y)^2 = R_{\text{LOS}}^2, \\ \frac{y_{\text{LOS}} - y_k}{x_{\text{LOS}} - x_k} = \frac{y_{k+1} - y_k}{x_{k+1} - x_k}, \end{cases} \quad (3)$$

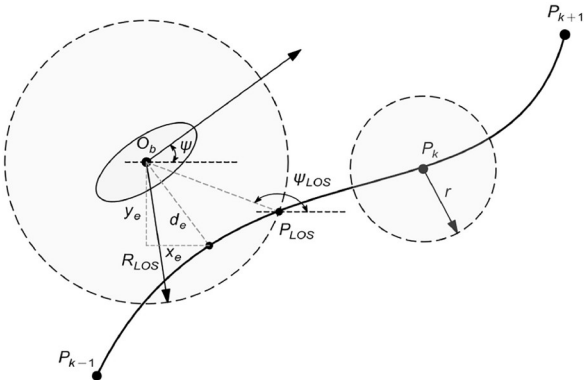


图2 LOS法示意图
Figure 2 Schematic of the LOS.

其中, (x, y) 为飞艇当前位置。当飞艇当前位置与路径终点的距离 $d < R_{\text{LOS}}$ 时, 路径终点视为视线引导点。

由视线引导点 $(x_{\text{LOS}}, y_{\text{LOS}})$ 及飞艇当前位置 (x, y) 可得期望航向角为

$$\psi_{\text{LOS}} = \arctan \frac{y_{\text{LOS}} - y}{x_{\text{LOS}} - x}. \quad (4)$$

LOS控制的目标是使得飞艇航向角 ψ 收敛至期望航向角 ψ_{LOS} 。

为避免飞艇与路径的距离误差 d_e 过大导致式(4)无解, 因此设定半径 R_{LOS} 满足式(3)。

2.3 问题描述

无人飞艇路径跟随控制的目标是驱使飞艇精准跟随期望路径 $\epsilon(\xi) = [x(\xi), y(\xi)]$, 即使得控制误差渐近收敛至零, 其中 ξ 代表路径参数。定义地面坐标系下的纵向跟踪误差 x_e 与横向跟踪误差 y_e :

$$\begin{cases} x_e = x - x_p(\xi), \\ y_e = y - y_p(\xi), \end{cases} \quad (5)$$

其中, (x, y) 表示无人飞艇的位置, $[x_p(\xi), y_p(\xi)]$ 为期望路径上与飞艇当前位置距离最近的点。定义航向角误差如下:

$$\psi_e(t) = \psi(t) - \psi_{\text{LOS}}(t). \quad (6)$$

采用LOS法将路径跟随控制问题转化为航向控制问题, 即使得

$$\lim_{t \rightarrow \infty} \psi_e = 0. \quad (7)$$

在LOS法下, 当航向角误差 ψ_e 趋近于0时, 路径距离误差 d_e 也趋近于0。该问题的核心在于如何通过协调控制主扭矩 F_M 与尾部扭矩 F_T , 使无人飞艇航向角 ψ 渐近收敛至期望值 ψ_{LOS} 。

2.4 强化学习

强化学习(reinforcement learning, RL)是一种广泛应用于序贯决策任务的交互式机器学习方法。其主要特点是试错与延迟奖励, 核心思想是通过智能体与环境的不断交互学习最优策略, 该交互过程通常描述为马尔可夫决策过程, 由五元组 $\langle S, A, P, \gamma, R \rangle$ 表示, 其中 S 为状态集合, A 为动作集合, P 为状态转移概率函数,

γ 为折扣因子, R 为奖励函数集合. 策略 π 定义状态到动作的映射关系. 在每次交互的时间步 t , 智能体根据当前策略 π 与状态 s_t , 选择动作 a_t , 继而依据奖励函数获得奖励 r_t , 根据状态转移概率函数 P 更新至状态 s_{t+1} . 总累积奖励, 即回报 G_t , 定义为 $G_t = \sum_{k=0}^N \gamma^k r_{t+k}$; 动作-价值函数 $Q^\pi(s, a) = E^\pi[G_t | s_t = s, a_t = a]$ 表示在策略 π 下从状态 s 开始, 执行动作 a 后的期望回报.

3 基于DDPG算法的飞艇路径跟随控制

飞艇系统具有动力学模型复杂、参数高敏感等特性, 强化学习方法无需精确系统模型, 在飞艇控制中具有显著优势. 通过合理定义状态空间、动作空间以及奖励函数, 可将无人飞艇路径跟随控制问题描述为马尔可夫决策过程.

状态空间: x_e, y_e 分别表示无人飞艇位置与期望路径在 x_g 轴与 y_g 轴方向的误差, 距离误差被定义为 $d_e = \sqrt{x_e^2 + y_e^2}$; ψ_e 为实际航向角 ψ 与期望航向角 ψ_{LOS} 之间的误差. 因此, 状态定义为 $s = [d_e, \dot{d}_e, \psi_e, \dot{\psi}_e]$.

动作空间: 飞艇动力由主推进器产生的主扭矩 F_M 和尾部方向舵产生的尾部扭矩 F_T 控制. 因此, 动作定义为 $a = [F_M, F_T]$.

奖励函数: 设计分段奖励函数:

$$r = 1.5r_1 + r_2,$$

其中, r_1 与 r_2 分别为与 ψ_e 和 d_e 相关的奖励项, 具体定义为

$$r_1 = \begin{cases} 15, & 0 \leq |\psi_e| < \frac{\pi}{100}, \\ 10, & \frac{\pi}{100} \leq |\psi_e| < \frac{\pi}{20}, \\ -0.1|\psi_e|, & \frac{\pi}{20} \leq |\psi_e| < \frac{\pi}{10}, \\ -|\psi_e|, & \frac{\pi}{10} \leq |\psi_e| < \pi, \end{cases} \quad (8)$$

$$r_2 = \begin{cases} 15, & 0 \leq |d_e| < 0.01, \\ 10, & 0.01 \leq |d_e| < 0.1, \\ -0.1|d_e|, & 0.1 \leq |d_e| < 0.5, \\ -|d_e|, & |d_e| \geq 0.5. \end{cases} \quad (9)$$

该奖励函数的设计考虑了无人飞艇运动路径与期望路径的跟随误差, 误差越小, 奖励越高, 从而驱使算

法沿误差减小趋势收敛. 在奖励函数 r_1 中, 当航向误差 ψ_e 处于预设的极小范围 $[0, \pi/100]$ 时给予最高奖励, 此范围是可接受的最小航向角控制误差, 随着误差增大, 奖励值递减, 一旦误差超出最大阈值 $\pi/10$ 时, 将施加随误差值递增的惩罚, 以驱动飞艇实现精准航向控制; 奖励函数 r_2 的设计逻辑与 r_1 类似, 旨在促使飞艇的路径距离误差最终趋近于零. 此外, 将奖励函数 r_1 的权重系数设置为略大于 r_2 , 以引导飞艇优先最小化航向角误差 ψ_e 并快速跟踪期望路径.

传统强化学习算法通过值函数评估和策略更新实现环境交互学习, 但其处理能力常局限于低维状态与动作空间. 为应对实际控制任务中的高维空间挑战, 融合深度学习技术的深度强化学习应运而生. 在深度强化学习领域, 经典DQN算法虽能有效处理离散动作空间, 但无法直接适用于连续动作控制问题. 对此, Lillicrap等人^[23]于2016年提出DDPG算法. 该算法融合DQN的经验回放机制与确定性策略梯度理论, 具有对连续动作空间的处理能力, 有效弥补了DQN的缺陷^[24].

Actor-critic网络框架是DDPG算法的核心, 参数为 μ 的actor网络 π_μ 用于近似确定性策略, 在给定状态下输出确定性动作; 参数为 θ 的critic网络 Q_θ 通过逼近Q函数, 为策略优化提供梯度方向.

在训练过程中, 智能体探索的状态转移经验 (s_t, a_t, r_t, s_{t+1}) 存储于回放缓冲区 D 中. 每次迭代时随机采样 N 个样本, 并最小化如下损失函数以更新参数 θ :

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N [y_i - Q_\theta(s_i, a_i)]^2, \quad (10)$$

其中, $y_i = r_i(s_i, a_i) + \gamma Q_{\theta'}[s_{i+1}, \pi_{\mu'}(s_{i+1})]$, $Q_{\theta'}$ 与 $\pi_{\mu'}$ 为相应的目标网络. Critic网络的参数按下式更新:

$$\theta \leftarrow \theta - \alpha_\theta \nabla_\theta J(\theta). \quad (11)$$

Actor网络的参数根据策略梯度上升进行更新:

$$\mu \leftarrow \mu + \alpha_\mu \nabla_\mu J(\mu), \quad (12)$$

$$\nabla_\mu J(\mu) \approx \frac{1}{N} \sum_{i=1}^N \nabla_\mu \pi_\mu(s_i) \nabla_a Q_\theta(s_i, a_i) \Big|_{a_i = \pi_\mu(s_i)}. \quad (13)$$

为提高训练稳定性, DDPG算法引入了目标网络机制. 其参数通过软更新方式缓慢同步:

$$\begin{aligned} \theta' &= \tau\theta + (1-\tau)\theta', \\ \mu' &= \tau\mu + (1-\tau)\mu', \end{aligned} \quad (14)$$

其中 $\tau \in (0, 1)$.

算法流程如算法1所示, 在每轮训练周期开始时, 首先初始化无人飞艇环境参数, 包括飞艇模型相关物理参数、期望路径及初始运动状态, 并设定周期步长. 随后重复执行以下步骤直至达到终止条件: (1) 在主网络中, 通过actor网络获取飞艇当前状态 s_t , 并根据当前策略选择动作 a_t ; (2) 执行当前动作并获得奖励 r_t , 根据飞艇系统模型得到下一时刻状态 s_{t+1} ; (3) 从经验回放池 D 中随机采样 N 个经验样本作为当前网络的训练数据; (4) 通过最小化损失函数更新critic网络参数, 利用策略梯度上升更新actor网络参数, 并对目标网络参数进行延迟更新.

在状态中引入距离误差 d_e 以提供距离信息, 增强飞艇的动态适应能力. 同时引入误差变化率 $\dot{\psi}_e$ 和 $\dot{d}_e$ 反映跟踪精度, 进一步提升决策效率.

通过定义连续状态空间、连续动作空间及分段奖励函数, 结合DDPG的无模型学习特性, 飞艇能够自主生成连续控制指令而无需预设复杂动力学模型. 同时, DDPG的探索机制有助于飞艇在路径跟随任务中做出高效决策.

算法1 基于DDPG的无人飞艇路径跟随算法

初始化: 初始化critic网络 Q_θ 与actor网络 π_μ 的权重参数 θ, μ , 目标网络 $Q_{\theta'}$ 与 $\pi_{\mu'}$ 的权重参数 $\theta' \leftarrow \theta, \mu' \leftarrow \mu$, 经验回放池 D , 以及随机高斯噪声 N

for episode = 1, M do:

获得初始状态 S_1

for t = 1, T do:

根据当前策略与噪声选取动作 $a_t = \mu(s_t) + N_t$

执行动作 a_t 并得到奖励 r_t , 根据飞艇系统模型获得后继状态 s_{t+1}

将样本 (s_t, a_t, r_t, s_{t+1}) 存储在 D 中

从 D 中随机采样一批数量为 N 的样本, 通过最小化如下损失函数更新critic网络参数 θ :

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N [y_i - Q_\theta(s_i, a_i)]^2$$

其中, $y_i = r_i + \gamma Q_{\theta'}[s_{i+1}, \pi_{\mu'}(s_{i+1})]$

通过策略梯度更新actor网络参数 μ :

$$\nabla_{\mu} J(\mu) \approx \frac{1}{N} \sum_{i=1}^N \nabla_{\mu} \pi_{\mu}(s_i) \nabla_a Q_{\theta}(s_i, a_i)$$

更新目标网络参数:

$$\theta' = \tau\theta + (1 - \tau)\theta'$$

$$\mu' = \tau\mu + (1 - \tau)\mu'$$

end for

end for

4 仿真验证

为深入评估基于DDPG的飞艇路径跟随控制策略的性能, 引入PID控制策略进行对比分析. 鉴于无人飞艇路径跟随任务的多样性, 设置三种期望路径: 直线路径、曲线路径与圆形路径. 飞艇模型参数设置如表1^[25,26], 强化学习网络参数设置如表2和3.

期望的直线路径、曲线路径与圆形路径分别表示如下:

$$\epsilon_1(\zeta) = [\zeta, 0],$$

$$\epsilon_2(\zeta) = [\zeta, 10\sin\frac{\pi}{50}\zeta],$$

$$\epsilon_3(\zeta) = [100\sin 0.02\zeta, 100\cos 0.02\zeta],$$

其中, $\zeta = kT$, $T = 0.01$ s为采样时间, k 为自然数. 无人飞艇初始位置设置为期望路径起点上方5 m处, 初始航向角指向水平方向^[27~29].

表1 无人飞艇相关参数

Table 1 Parameters of the autonomous airship

参数名称	数值
飞艇质量 m (kg)	7001
飞艇体积 V (m^3)	75000
转动惯量 I_z (kg m^2)	5.1×10^4
惯性因子 k_{11}	0.11
惯性因子 k_{22}	0.83
惯性因子 k_{66}	0.1
驻空高度 (km)	20
阻力系数 C_x	0.035
侧力系数 C_y	0.588
偏航力矩系数 C_n	0.062

表2 Critic网络参数

Table 2 Parameters of the critic network

参数	赋值
输入参数	状态向量 S , 动作向量 a
神经元个数	128
网络层数	5
激活函数	Relu
学习率	0.0001
优化器	Adam

表 3 Actor网络参数

Table 3 Parameters of the actor network

参数	赋值
输入参数	状态向量 S
神经元个数	128
网络层数	3
第1层激活函数	Relu
第2层激活函数	Tanh
学习率	0.0005
优化器	Adam
高斯噪声方差	0.2

仿真结果如图3~5所示。图3给出本文方法与PID方法在直线路径、曲线路径与圆形路径下的跟踪效果对比图。图4给出本文方法与PID方法在三种期望路径下的误差对比情况, 基于DDPG的控制算法的误差

曲线能迅速收敛并维持在极低水平且波动极小, 而PID控制器则存在明显的误差波动和较长的收敛时间。图5呈现DDPG算法的奖励值变化, 随着迭代次数增加, 奖励曲线逐渐上升并趋于稳定, 验证了算法具有良好的收敛特性。仿真结果表明, 在飞艇路径跟随控制任务中, 基于DDPG的控制算法相比传统PID控制方法展现出显著的性能优势, 在曲线、直线和圆形路径的跟踪轨迹上均表现出更高的精度、更小的振荡和更快的收敛速度。

表4给出两种方法下的平均路径跟随误差。表格数据显示, DDPG控制算法在所有测试路径类型下的平均路径跟随误差均低于PID, 尤其在圆形路径下对比最为显著, 这表明DDPG控制算法在处理飞艇非线性动力学和实现高精度、稳定路径跟随方面更有优越性, 且在三种不同路径下的良好表现也表明该算法具有更强的泛化能力。

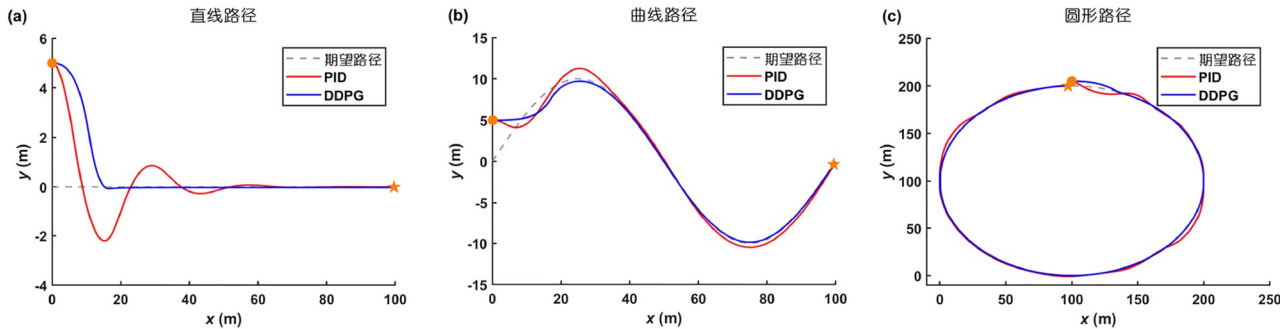


图 3 (网络版彩图)不同期望路径下的跟随效果
Figure 3 (Color online) Path-following effects under different reference paths.

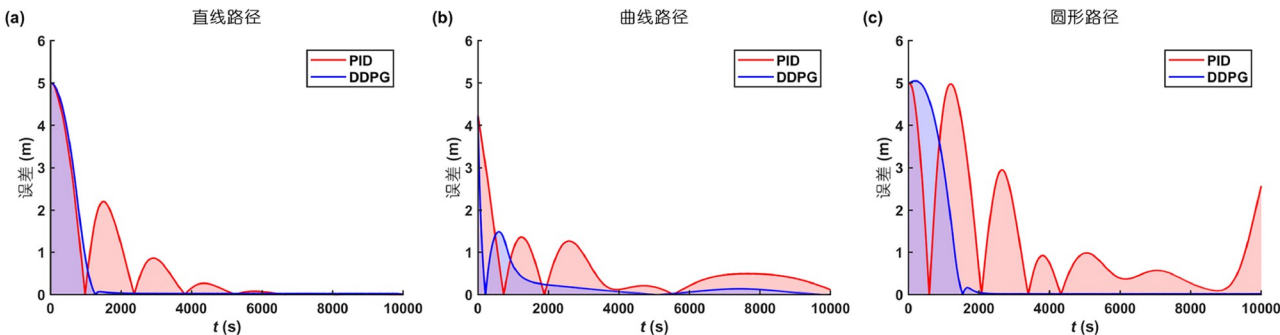


图 4 (网络版彩图)不同期望路径下路径跟随误差
Figure 4 (Color online) Path-following errors under different reference paths.

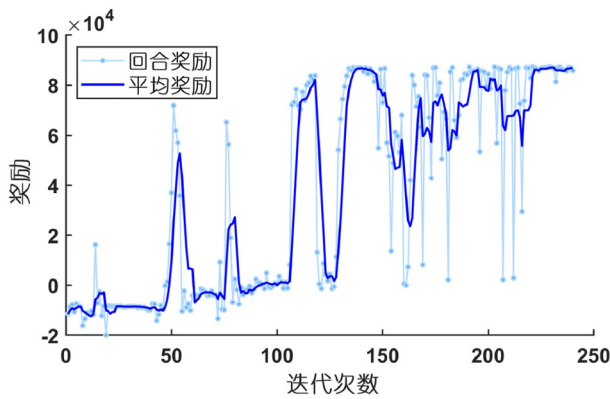


图 5 (网络版彩图) DDPG算法的奖励曲线

Figure 5 (Color online) The reward curve of the DDPG algorithm.

5 结论

针对无人飞艇路径跟随问题, 提出了基于DDPG

表 4 PID方法与DDPG方法下的平均路径跟随误差

Table 4 Average path-following errors under PID and DDPG methods

期望路径类型	PID	DDPG
直线	0.6078	0.3790
曲线	0.5850	0.2229
圆形	1.2830	0.5457

算法的控制框架。该框架融合了LOS法与DDPG算法, 利用LOS法将路径跟随问题转化为航向控制问题, 进而利用DDPG算法学习最优控制策略以实现期望航向角的精确控制, 最终实现路径跟随。与现有方法相比, 提出的控制方案无需依赖系统的精确数学模型, 显著降低了控制策略学习的复杂度, 有效提升了系统的路径跟随性能。未来工作拟进一步探索深度强化学习在复杂环境下飞艇集群协同控制中的应用。

参考文献

- Bellemare M G, Candido S, Castro P S, et al. Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature*, 2020, 588: 77–82
- Zhao D, Liu D X, Sun K W, et al. Research status, technical difficulties and development trend of stratospheric airship (in Chinese). *Acta Aeronautica Astron Sin*, 2016, 37: 45–46 [赵达, 刘东旭, 孙康文, 等. 平流层飞艇研制现状、技术难点及发展趋势. *航空学报*, 2016, 37: 45–46]
- Elfes A, Bueno S S, Ramos J J G, et al. Modelling, control and perception for an autonomous robotic airship. In: *Proceedings of the Modelling of Sensor-based Intelligent Robot Systems*. 2002. 216–244
- Miller C, Sullivan J, McDonald S. High altitude airship simulation control and low altitude flight demonstration. In: *Proceedings of the AIAA Infotech@Aerospace 2007 Conference and Exhibit*. 2007. 2766–2788
- Azinhira J R, de Paiva E C, Ramos J J G, et al. Mission path following for an autonomous unmanned airship. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2000. 1269–1275
- Hygounenc E, Soueres P. Lateral path following GPS-based control of a small-size unmanned blimp. In: *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2003. 540–545
- Zheng Z, Huo W. Planar path following control for stratospheric airship. *IET Control Theor & Appl*, 2013, 7: 185–201
- Beji L, Abichou A. Tracking control of trim trajectories of a blimp for ascent and descent flight manoeuvres. *Int J Control*, 2005, 78: 706–719
- Hygounenc E, Soueres P. Automatic airship control involving backstepping techniques. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2002. 1–6
- Lee S, Lee H. Back-stepping approach of trajectory tracking control for the mid-altitude unmanned airship. In: *Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit*. 2007. 6319–6325
- Azinhira J R, Moutinho A, de Paiva E C. A backstepping controller for path-tracking of an underactuated autonomous airship. *Intl J Robust Nonlinear*, 2009, 19: 418–441
- Zheng Z, Guan Z, Ma Y, et al. Constrained path-following control for an airship with uncertainties. *Eng Appl Artif Intelligence*, 2019, 85: 295–306
- Wen G H, Yang T, Zhou J L, et al. Reinforcement learning and adaptive/approximate dynamic programming: A survey from theory to

- applications in multi-agent systems (in Chinese). *Control Decis*, 2023, 38: 1200–1230 [温广辉, 杨涛, 周佳玲, 等. 强化学习与自适应动态规划: 从基础理论到多智能体系统中的应用进展综述. *控制与决策*, 2023, 38: 1200–1230]
- 14 Wang C, Wei Z, Jiang W, et al. Cooperative sensing enhanced UAV path-following and obstacle avoidance with variable formation. *IEEE Trans Veh Technol*, 2024, 73: 7501–7516
 - 15 Wang Y, Cao J, Sun J, et al. Path following control for unmanned surface vehicles: A reinforcement learning-based method with experimental validation. *IEEE Trans Neural Netw Learn Syst*, 2023, 35: 18237–18250
 - 16 Zhao Y, Ma Y, Hu S. USV formation and path-following control via deep reinforcement learning with random braking. *IEEE Trans Neural Netw Learn Syst*, 2021, 32: 5468–5478
 - 17 McGookin E W, Murray-Smith D J, Li Y, et al. Ship steering control system optimisation using genetic algorithms. *Control Eng Pract*, 1998, 8: 429–443
 - 18 Fossen T I, Breivik M, Skjetne R. Line-of-sight path following of underactuated marine craft. In: *Proceedings of the IFAC Conference on Maneuvering and Control of Marine Crafts*. 2003. 211–216
 - 19 Moreira L, Fossen T I, Guedes Soares C. Path following control system for a tanker ship model. *Ocean Eng*, 2007, 34: 2074–2085
 - 20 Sebbane Y B. *Lighter than Air Robots: Guidance and Control of Autonomous Airships*. Berlin: Springer, 2011. 15–20
 - 21 Zheng Z, Feroskhan M, Sun L. Adaptive fixed-time trajectory tracking control of a stratospheric airship. *ISA Trans*, 2018, 76: 134–144
 - 22 Gu N, Wang D, Peng Z, et al. Advances in line-of-sight guidance for path following of autonomous marine vehicles: An overview. *IEEE Trans Syst Man Cybern Syst*, 2022, 53: 12–28
 - 23 Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. In *Proceedings of the 4th International Conference on Learning Representations*. 2016. 1–14
 - 24 Wen G, Fu J, Dai P, et al. DTDE: A new cooperative multi-agent reinforcement learning framework. *Innovation*, 2021, 2: 100162
 - 25 Sun L, Zheng Z. Nonlinear adaptive trajectory tracking control for a stratospheric airship with parametric uncertainty. *Nonlinear Dyn*, 2015, 82: 1419–1430
 - 26 Yang X, Yang X, Deng X. Horizontal trajectory control of stratospheric airships in wind field using Q-learning algorithm. *Aerospace Sci Tech*, 2020, 106: 106100
 - 27 Yan Y, Yu S, Gao X, et al. Continuous and periodic event-triggered sliding-mode control for path following of underactuated surface vehicles. *IEEE Trans Cybern*, 2024, 54: 449–461
 - 28 Fang Z, Wang J, Wang Z, et al. Human-machine shared control for path following considering driver fatigue characteristics. *IEEE Trans Intell Transp Syst*, 2024, 25: 7250–7264
 - 29 Yu Z Q, Zhang Y M, Jiang B. Fractional-order intelligent fault-tolerant synchronization tracking control of UAV swarms (in Chinese). *Sci Sin Tech*, 2020, 50: 389–402 [余自权, 张友民, 姜斌. 群体无人机分数阶智能容错同步跟踪控制. *中国科学: 技术科学*, 2020, 50: 389–402]

Deep reinforcement learning-based path following control for autonomous airship

CAO XiWen^{1,2,3}, ZHOU Yan⁴ & WEN GuangHui⁵

¹ School of Automation, Beijing Institute of Technology, Beijing 100081, China

² State Key Laboratory of Environment Characteristics and Effects for Near-space, Beijing Institute of Technology, Beijing 100081, China

³ Yangtze Delta Region Academy, Beijing Institute of Technology, Jiaxing 314019, China

⁴ School of Mathematics, Southeast University, Nanjing 211189, China

⁵ School of Automation, Southeast University, Nanjing 210096, China

This paper proposes a model-free control approach by integrating the line-of-sight method with deep reinforcement learning to address the path-following problem for autonomous airships. Firstly, based on the line-of-sight method, the path-following control problem is transformed into a heading control problem. Then, considering the kinematic and dynamic characteristics of the airship, such a heading control problem is formulated as a Markov decision process. Building upon this, a model-free heading control strategy is designed using the deep deterministic policy gradient algorithm to achieve precise desired heading tracking, ultimately accomplishing the path-following objective. Simulation results demonstrate that under three typical expected paths, the proposed method achieves faster convergence to the expected control objective compared to the traditional PID control, while exhibiting superior tracking accuracy and system stability, thereby verifying its effectiveness and superiority.

autonomous airship, path following control, line-of-sight, deep reinforcement learning, deep deterministic policy gradient

doi: [10.1360/SST-2025-0164](https://doi.org/10.1360/SST-2025-0164)