

环境风利用的浮空器区域驻留深度强化学习控制方法

柏方超, 杨希祥*, 邓小龙, 龙 远, 侯中喜
(国防科技大学 空天科学学院, 湖南 长沙 410073)

摘要:针对平流层浮空器在动态风场中的区域驻留问题,面向不同的控制通道,基于深度强化学习D3QN算法设计了环境风利用的浮空器区域驻留控制器,研究了不同的奖励函数对区域驻留控制器性能的影响。在以区域驻留时长为3 d、区域驻留半径为50 km的任务约束下,进行了区域驻留控制仿真。结果表明:与采用DDQN方法设计的区域驻留控制器相比,采用D3QN方法设计的控制器的性能显著提高,仅依靠高度调节控制轨迹的情况下,平均区域驻留半径可以达到25.26 km、驻留有效时间比为96.25%,在水平方向辅助动力推进的情况下,平均区域驻留半径可显著减小、驻留有效时间比可显著提高。同时,验证了基于深度强化学习设计的区域驻留控制器具有较强的鲁棒性,可通过不同的奖励函数设计控制器,满足不同的区域驻留任务需求。

关键词:平流层浮空器;区域驻留控制;D3QN;动力推进

中图分类号:V274 **文献标志码:**A **文章编号:**1001-2486(2025)02-078-11



论
文
拓
展

Station keeping control method based on deep reinforcement learning for aerostat using ambient wind

BAI Fangchao, YANG Xixiang*, DENG Xiaolong, LONG Yuan, HOU Zhongxi

(College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Aiming at the station keeping control problem of stratospheric aerostat in dynamic wind field, a station keeping controller designed based on deep reinforcement learning D3QN algorithm for different control channels of aerostat operated with ambient wind, studied the impact of different reward functions on the performance of regional resident controllers. Station keeping control simulation was carried out under the task constraint of a station keeping duration of three days and a station keeping radius of 50 km. Results show that: compared with the station keeping controller designed by DDQN method, the performance of the controller designed by D3QN method is significantly improved. When the control trajectory of aerostat is only adjusted by altitude, the average station keeping radius can reach 25.26 km, and the station keeping ratio is 96.25%. With the aid of horizontal propulsion, the average station keeping radius can be significantly reduced and the station keeping time ratio can be significantly increased. At the same time, the strong robustness of the station keeping controller based on deep reinforcement learning was verified, and the controller can be designed with different reward functions to meet the requirements of different station keeping tasks.

Keywords: stratospheric aerostat; station keeping control; D3QN; power propulsion

区域驻留控制是平流层浮空器飞行控制的关键技术之一,是实现和拓展其应用的基本要求^[1]。近年来,一种通过改变浮空器浮重平衡,调整浮空器驻空高度,从而利用准零风层风场环境控制轨迹的浮空器引起国内外学者的广泛关注^[2-3]。其相对于飞艇而言,系统简单、便于快速发放;相对于高空气球而言,利用准零风层的风场

特征^[4],可以通过较小能源消耗实现一定的轨迹控制^[2,5]。

通过利用风场环境实现区域驻留的浮空器控制方法可以概括为:传统基于中心点的控制方法与智能控制方法。基于中心点的控制方法不断计算当前位置与中心点的相对位置,当浮空器运动到区域驻留边界时,将其工作高度调节到指向中

收稿日期:2022-12-03

基金项目:国家自然科学基金资助项目(61903369,52272445);湖南省杰出青年基金资助项目(2023JJ10056)

第一作者:柏方超(1997—),男,河北邯郸人,博士研究生,E-mail:bail15932465997@163.com

*通信作者:杨希祥(1982—),男,河北衡水人,教授,博士,博士生导师,E-mail:nkyangxixiang@163.com

引用格式:柏方超,杨希祥,邓小龙,等.环境风利用的浮空器区域驻留深度强化学习控制方法[J].国防科技大学学报,2025,47(2):78-88.

Citation:BAI F C, YANG X X, DENG X L, et al. Station keeping control method based on deep reinforcement learning for aerostat using ambient wind[J]. Journal of National University of Defense Technology, 2025, 47(2): 78-88.

心点的风层^[6]。文献[7]提出将区域驻留控制分解为东西方向与南北方向的独立控制,在东西方向利用准零风层上下风速方向切变的特征,通过调节浮空器工作高度实现区域驻留,在南北方向通过螺旋桨抗风飞行实现区域驻留。文献[8]在高度方向调整浮空器驻空高度,在水平方向耦合动力推进系统,在东西方向、南北方向以及双通道方向设计区域驻留策略,研究表明,水平方向具有弱动力推进系统可以提高区域驻留控制效果。Du等^[6]将浮空器调节到期望高度,该高度上的风场方向接近于浮空器指向中心点的方向,在能源约束下,通过遗传算法对副气囊充放气控制策略进行了优化,仿真结果表明,浮空器在一天内可以实现区域驻留半径为30 km的目标^[9]。

高空风场短周期、小尺度范围内具有时变性和不确定性,传统控制方法并未考虑这些因素的影响^[10]。以强化学习为代表的智能控制方法在飞行器控制领域得到了广泛的研究^[11-14],在浮空器区域驻留控制领域具有巨大的应用潜力^[10, 13, 15]。

Loon项目^[16]对比分析了station seeker、tree-search controller与reinforcement learning算法设计的区域驻留控制器。结果表明,通过深度强化学习设计的控制器,在有限能源约束下控制效果更优^[10]。文献[17]通过调节空气囊空气质量,从而调整驻空高度,实现浮空器轨迹控制,在不确定风场条件下,设计了平流层浮空器区域驻留控制器,通过仿真分析验证了该方法具有较强的鲁棒性,分析了不同初始驻空点对区域驻留控制效果的影响。文献[18]采用双深度Q学习网络(double deep Q learning network, DDQN)方法设计了平流层浮空器区域驻留控制器,指出了在时变风场条件下,区域驻留控制器具有较强的鲁棒性,本文中所使用的方法较DDQN收敛速度更快,所设计的区域驻留控制器控制性能更优。

以环境风场利用的平流层浮空器为研究对象,建立了平流层浮空器动力学模型,基于决斗双深度Q学习网络(dueling double deep Q learning network, D3QN)方法设计了平流层浮空器区域驻留控制器,在时变风场以及不确定风场条件下开展了区域驻留控制仿真分析,研究了奖励函数以及动力推进对浮空器区域驻留控制效果的影响。

1 数学模型

1.1 平流层浮空器系统组成与工作原理

平流层浮空器系统如图1所示,其中球体分系统主要包括空气囊和氦气囊、设备吊舱与各部

分连接机构。囊体结构是球体分系统的主要部分,氦气囊用于填充氦气为浮空器提供浮力,空气囊用于调整浮空器的驻空飞行高度。吊舱设备用于搭载通信设备、安控设备、监测设备等。

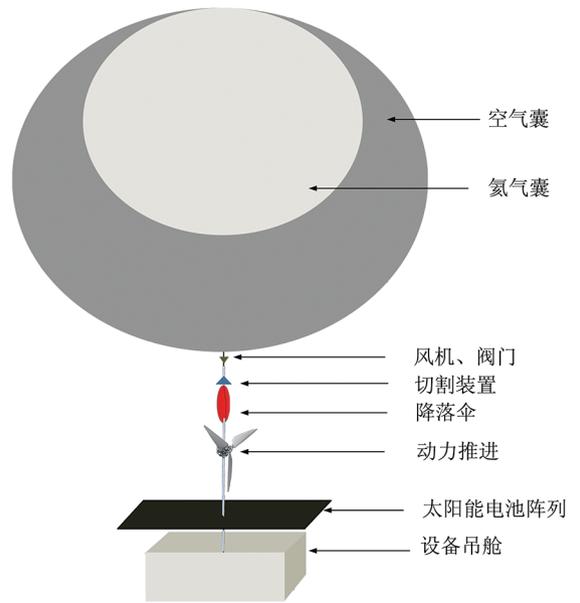


图1 环境风场利用的平流层浮空器结构

Fig. 1 Structure of stratospheric aerostat using environmental wind fields

高度调节分系统主要包括风机、阀门等。风机阀门采用一体化设计,风机向内部吸气必须先要将阀门打开。

能源分系统主要包括太阳能电池、储能电池、能源管理器等,主要作用是产能、储能与供能,用于浮空器跨昼夜长航时驻空飞行。

推进分系统主要包括电机、螺旋桨及转向机构,主要作用是水平方向的抗风飞行,为轨迹控制提供推力。

平流层浮空器工作原理如图2所示。平流层浮空器飞行在弱风层时,以较低的速度驻空,飞行速度接近于风速。当平流层浮空器需要向规划的方向运动时,根据合适的风速和风向,通过空气囊的吸气/排气调节高度,实现对风场能量的综合利用。当不存在所需方向的风场时,可以通过水平方向弱动力来抗风,以较小的能量消耗实现区域驻留。

1.2 动力学模型

忽略浮空器的热力学变化,将浮空器视为质点,将地球视为均质球体,假设浮空器驻空过程中能源充足,建立浮空器动力学模型如下。

1.2.1 纵向动力学

平流层浮空器纵向动力学方程为:

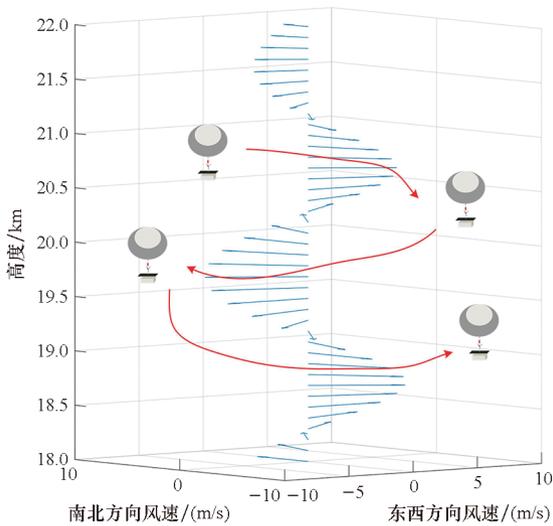


图 2 平流层浮空器工作原理

Fig. 2 Stratospheric aerostat working principle

$$\begin{cases} (m_{\text{tot}} + m_{\text{add}}) \frac{dv_h}{dt} + \dot{m}_{\text{tot}} v_h = B - m_{\text{tot}} g - D_h \\ \frac{dh}{dt} = v_h \end{cases} \quad (1)$$

其中, m_{tot} 为平流层浮空器的总质量, m_{add} 为附加质量, \dot{m}_{tot} 为总质量变化率, B 为总浮力, D_h 为平流层浮空器高度方向上的气动阻力, h 为高度, v_h 为垂直方向速度^[19]。

1.2.2 横侧向动力学

横侧向的动力学方程可表示为:

$$\begin{cases} (m_{\text{tot}} + m_{\text{add}}) \frac{dv_x}{dt} = \frac{1}{2} \rho_{\text{air}} (v_x + u)^2 C_d S + F_{\text{ps}} \\ \frac{dx}{dt} = v_x \\ (m_{\text{tot}} + m_{\text{add}}) \frac{dv_y}{dt} = \frac{1}{2} \rho_{\text{air}} (v_y + v)^2 C_d S + F_{\text{ps}} \\ \frac{dy}{dt} = v_y \end{cases} \quad (2)$$

其中, ρ_{air} 为驻空高度的大气密度, C_d 为浮空器阻力系数, S 为浮空器参考面积, u, v 分别为东西方向与南北方向的风速分量, v_x, v_y 分别为东西和南北水平方向上的速度, x, y 分别为东西和南北水平方向上的位移, F_{ps} 为推进系统推力。

平流层浮空器各部分质量、附加质量、气动阻力、阀门吸排气、水平方向动力系统推力建模方法可以参照文献[18]。

2 基于 D3QN 的区域驻留控制器设计

平流层浮空器在竖直方向的高度调节依靠阀门和风机的开关,为离散控制量;在水平方向动力系统推进系统的开关为离散控制量,推进系统的方向变

化为连续控制量,可以将其分解为离散动作空间。平流层浮空器在三维空间中的飞行轨迹为连续状态,为避免“维度灾难”问题,可以采用基于值函数的深度强化学习方法设计区域驻留控制器^[20]。

D3QN 深度强化学习算法结合了 DDQN 与决斗深度 Q 学习网络 (dueling deep Q learning network, Dueling DQN) 算法的优势,一方面能够有效减小动作值 Q 函数的过估计,另一方面,将 Q 值分解为状态值函数和优势值函数,可以先判断当前状态的好坏,然后根据动作空间判断增益进而选取动作。在这两种改进下, D3QN 可以加速找到收敛点^[20]。综上,采用 D3QN 深度强化学习方法设计平流层浮空器区域驻留控制器。

2.1 马尔可夫决策模型

平流层浮空器区域驻留控制问题可以描述为:时变风场条件下,求解浮空器动作序列(策略),使其在一定的空间范围内驻空时间越久越好。可以采用强化学习中的马尔可夫决策过程 (Markov decision process, MDP) 对区域驻留问题进行设计,相关元素设置如下。

2.1.1 环境状态空间

平流层浮空器具体状态参数如表 1 所示。

表 1 状态参数设置

Tab. 1 Status parameter setting

参数	取值范围	描述
高度 h	18 ~ 22 km	
位置坐标 $[x, y]$	-50 ~ 50 km	$\sqrt{x^2 + y^2} \leq 50$
空气质量 m	0 ~ 172 kg	
风速 S_{ws}		由风场条件确定
夹角 δ	0 ~ π	由风场与浮空器位置确定

其中,风场环境采用长沙附近的风场数据,每小时变化一次,浮空器当前位置的风场信息采用三次样条插值计算。

2.1.2 动作空间

根据高度控制系统,浮空器纵向的运动控制取决于阀门和风机的开关,可以分为 3 种动作:阀门开、风机开、阀门关。

根据推进系统设计,浮空器水平方向的运动控制取决于螺旋桨的开关与推进方向,可以分为 9 种动作:螺旋桨关、螺旋桨向北、螺旋桨向东北、螺旋桨向东、螺旋桨向东南、螺旋桨向南、螺旋桨向西南、螺旋桨向西、螺旋桨向西北。

浮空器仅依靠高度调节实现区域驻留的动作空间有3个,水平方向增加东西方向与南北方向单通道控制时,浮空器动作空间组合各有9个,双通道控制时动作空间有27个。

2.1.3 状态转移概率

对于平流层浮空器的每一个动作和状态,当平流层浮空器执行完一个动作,可以确定下一步状态,状态转移概率 $P_{SS'}^a = 1$ 。

2.1.4 奖励函数

奖励函数反映强化学习控制的最终目的,耦合风向、浮空器位置信息设置奖励函数为:

$$R(s_t) = \max \{ N [w_{\text{wind}} \delta + w_{\text{dis}} (R_1 - S)] + b \} \quad (3)$$

式中, N 保证 w_{wind} 与 w_{dis} 奖励占比不变, w_{wind} 为风向权重系数, w_{dis} 为距离相权重系数, δ 为风向与当前位置方向的夹角, R_1 为区域驻留半径, S 为浮空器位置到区域驻留中心的距离^[18]。由于压缩机的开关须不频繁以达到节约能源的效果,增加了偏差 b , 以区分动作类型。

2.1.5 动作价值函数

基于动作价值函数的 Q 学习算法通过神经网络拟合状态-动作空间对应的值函数, 根据最大动作价值函数 $Q(s_t, a_t | \theta)$ 确定动作, θ 为神经网络参数, 训练过程中, 控制器与环境进行交互采样并保存数据, 将样本空间 $[s_t, a_t, r_t, s_{t+1}]$ 保存到经验池中, 样本空间中的元素分别为状态、动作、奖励以及下一步状态。

DDQN 解决动作价值函数 $Q(s_t, a_t | \theta)$ 问题的流程可以归纳为: 建立双 Q 网络, 分别为当前神经网络与目标神经网络, 通过梯度下降算法更新两个神经网络的参数 (θ 与 θ^-)。更新方程^[18] 如下:

$$y_t = r_t + \gamma Q(s_{t+1}, a_{t+1} | \theta^-) \quad (4)$$

$$a_{t+1} = \arg \max_{a_{t+1}} Q(s_{t+1}, a_{t+1} | \theta) \quad (5)$$

损失函数为:

$$L(\theta^-) = [y_t - Q(s_t, a_t | \theta)]^2 \quad (6)$$

在 DDQN 的基础上, 采用 Dueling DQN 算法结构, 将动作价值函数重新表示为式(7)^[19]。

$$Q(s_t, a_t) = V(s_t) + \left[A(s_t, a_t) - \frac{1}{|A|} \sum_{a' \in A} A(s_t, a') \right] \quad (7)$$

式中, $V(s_t)$ 为状态值函数, $A(s_t, a_t)$ 为优势函数。

2.2 基于 D3QN 区域驻留控制参数及评价指标

在神经网络训练时, 采用自适应矩估计 (adaptive moment estimation, Adam) 算法更新网络参数。经验池采用优先经验回放机制算法^[21]。

贪婪策略满足^[18]:

$$\varepsilon_{t+1} = \varepsilon_t (1 - \varepsilon_{\text{dec}} T) \quad (8)$$

式中: ε 初值为 0.98; ε_{dec} 为贪婪算法下降参数, 取值为 10^{-6} 。

神经网络由 1 个输入层、4 个隐藏层、1 个输出层组成。激活函数选择 ReLU 函数。其余相关参数如表 2 所示。

表 2 D3QN 算法参数

Tab. 2 D3QN algorithm parameters

参数	值
学习采样数 N_b	512
最大回合数 N_{max}	2 000
经验池大小 M	2×10^6
学习率	0.001
奖励偏差	-0.1

平流层浮空器区域驻留控制算法如算法 1 所示。

算法 1 平流层浮空器区域驻留控制算法

Alg. 1 Stratospheric aerostat station keeping control algorithm

输入: 初始化神经网络参数、最大回合数 N_{max} 、衰减因子 γ 、探索率 ε 、目标网络更新频率 P 、经验池 D
输出: 存储训练成型的网络, 平均奖励信息等参数

1. for episode = 1 to N_{max}
2. While not done
3. 初始化并获得当前环境状态 s_t
4. 设定仿真步长 $step$ 和最长仿真总时间 T_{sim}
5. 根据当前状态, 输入当前 Q 网络, 计算各个动作对应 Q 值, 通过贪婪算法选择动作 a_t
6. 执行动作 a_t , 得到新的状态 s_{t+1} 和奖励 r_t , 判断回合是否结束, 返回 done
7. 将 (s_t, a_t, r_t, s_{t+1}) 存入经验池 D
8. 计算当前 Q 网络的目标值
9. 根据随机梯度下降算法计算损失函数, 反向传播更新参数 θ
10. P 回合后, 将当前 Q 网络参数 θ 赋值给 θ^-
11. $s_{t+1} = s_t$
12. 存储训练成型的网络, 记录平均奖励信息等参数

图 3 为基于 D3QN 算法的平流层浮空器区域驻留控制流程。

为了量化分析控制器的区域驻留能力, 提出平均区域驻留半径 d_{avg} 和区域驻留有效时间比 t_{eff} 两个参数作为空间评价指标和时间评价指标^[18]。

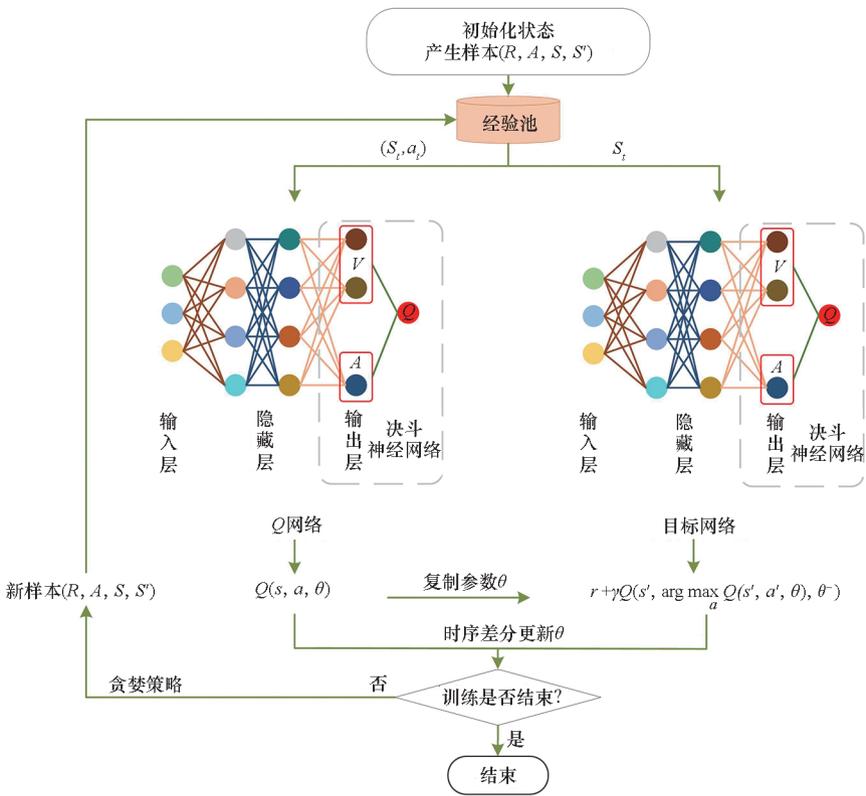


图 3 基于 D3QN 的平流层浮空器区域驻留控制流程

Fig. 3 Station keeping control flow based on D3QN for stratospheric aerostat

d_{agv} 为所有时间节点浮空器位置距区域驻留中心的平均值。

$$d_{agv} = \frac{1}{n} \sum_{i=1}^n \sqrt{x_i^2 + y_i^2} \quad (9)$$

t_{eff} 表示浮空器在 50 km 半径范围内驻留的时间占比。

$$t_{eff} = \frac{T_{all} - T_{50}}{T_{all}} \times 100\% \quad (10)$$

3 仿真结果与分析

3.1 无推进系统下的浮空器区域驻留控制

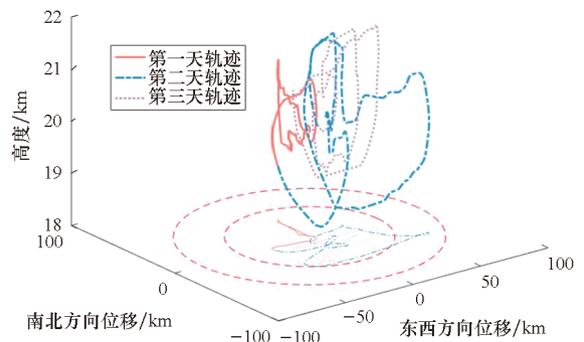
以驻空高度在 18 ~ 22 km, 总体积为 3 360 m³ 的平流层浮空器为研究对象。

采用强化学习区域驻留控制器得出的飞行仿真结果如图 4 所示(阀门开关状态取值为 1 代表排气、0 代表关闭、-1 代表吸气,下同),图 5 为相同风场条件下,传统基于中心点控制策略的水平轨迹。与基于中心点驻留控制策略、DDQN 设计的区域驻留控制器相比,通过 D3QN 设计的控制器,最大区域驻留半径 $d_{max} = 70.37$ km, 平均区域驻留半径 $d_{agv} = 25.26$ km, 区域驻留有效时间比 $t_{eff} = 96.25\%$, 浮空器 96.25% 的时间位于区域驻留任务区域内, 控制器性能明显提升。在调节

过程中, 相较中心点控制策略, 利用深度强化学习得到的控制器能够根据预测风场提前调节高度, 使得浮空器提前做出决策。图 4(c) 与图 4(d) 表示高度调节随时间的变化以及阀门的动态变化。本文采用的风场准零风层具有向上移动的趋势, 从图 4(c) 中可以看出, 浮空器工作高度向工作上限移动, 这与准零风层的移动趋势是相同的。

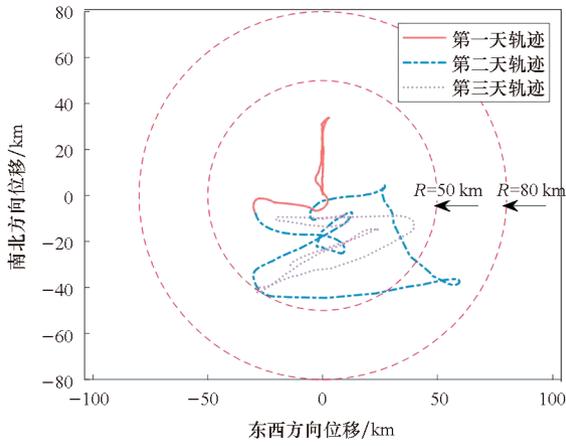
3.2 风场扰动影响分析

图 6 为在风场具有不确定性的情况下, 平流层飞艇在区域驻留控制器作用下的飞行仿真结果。东西和南北方向上的风速具有高斯分布形式的不确定性, 不确定性方差为 0.5。结果表明, 最大区域驻留半径为 65 km, 平均区域驻留半径为 23.5 km,

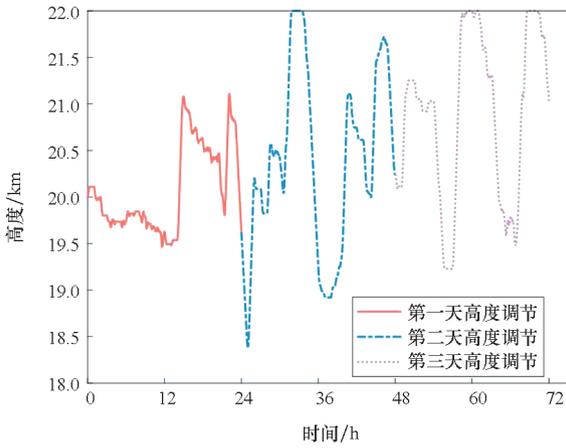


(a) 三维轨迹

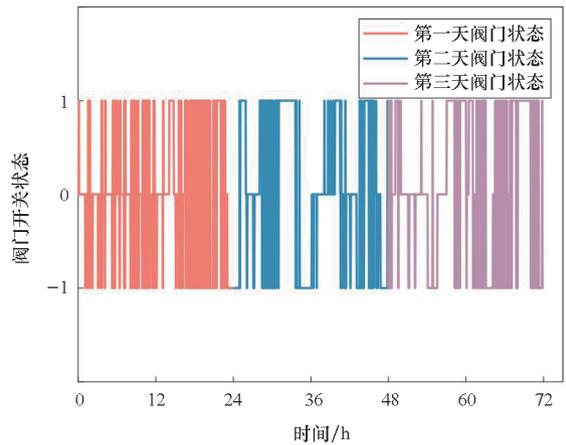
(a) Three-dimensional trajectory



(b) 水平轨迹
(b) Horizontal trajectory



(c) 高度随时间变化情况
(c) Change situation of altitude over time



(d) 阀门状态随时间变化情况
(d) Change situation of valve status over time

图 4 强化学习区域驻留控制器飞行仿真结果
Fig. 4 Flight simulation results of station keeping controller based on reinforcement learning

区域驻留有效时间比为 83.75%，控制器具有较强的鲁棒性。与无扰动下的控制情况进行对比，最大区域驻留半径和平均区域驻留半径有所减小，这是

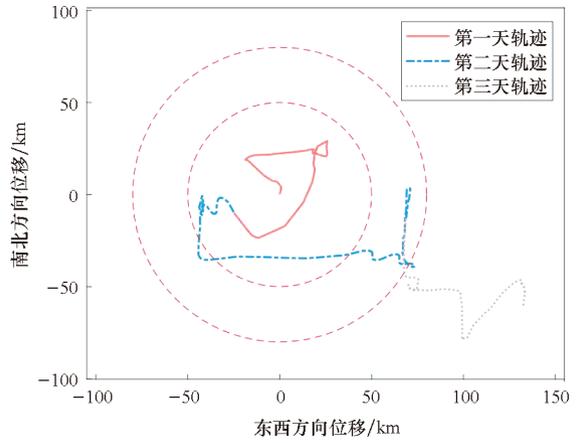


图 5 传统基于中心点控制策略的水平轨迹
Fig. 5 Horizontal trajectory based on traditional central point control strategy

由于设置的奖励函数中的距离项为 $w_{dis}(R_1 - S)$ ，在目标区域内停留的时间越久，奖励值越大，所以扰动情况下的区域驻留有效时间比减小了，总体指标接近于无风场扰动下的指标。

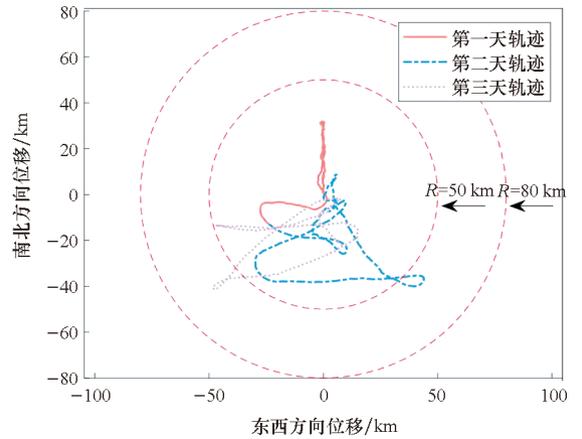


图 6 风场不确定性下飞行仿真结果
Fig. 6 Flight simulation results under uncertainty of wind disturbance

3.3 目标函数影响分析

图 7 为不同权重下的浮空器区域驻留控制器性能对比结果。由图 7 可以看出， $w_{wind}/w_{dis} = 0.5$ 时，浮空器 96.25% 的驻空时间在 50 km 驻留范围内，其余权重系数下，浮空器在 50 km 范围内停留时间较少。若浮空器的区域驻留范围要求是 20 km，权重系数比值 $w_{wind}/w_{dis} = 2$ 较合适。产生这种影响的原因是：虽然权重系数 w_{wind} 有利于浮空器找到合适的风层向区域驻留中心靠近，但是风速项的组合为 $w_{wind}\delta$ ，浮空器可能会在付出较小代价的情况下向远离区域驻留中心的方向移动，利用合适的风场向区域驻留中心靠近，反而会降低区域驻留控制器的预决策能力，区域驻留的控

制效果向中心点策略的效果靠近。对于特定任务下的区域驻留范围要求,可以选择不同权重系数比值下设计的区域驻留控制器。

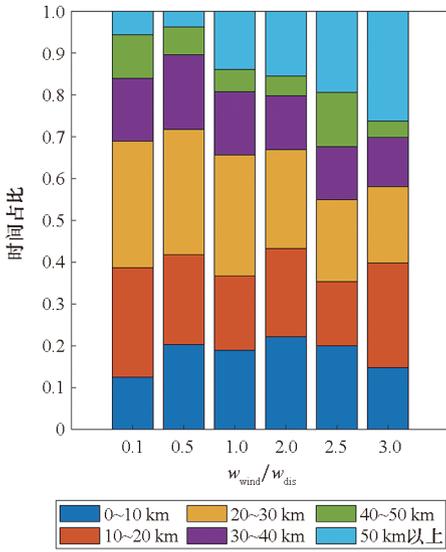


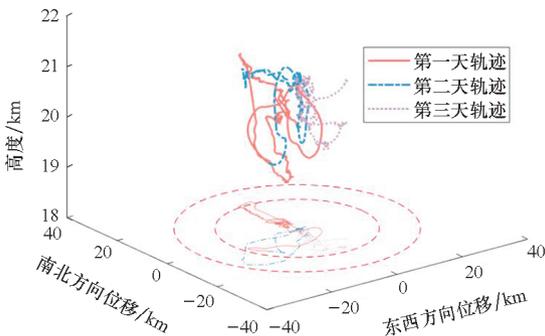
图 7 不同权重系数下的浮空器区域驻留控制器性能
Fig. 7 Performance of aerostat station keeping controller with different weight coefficients

3.4 推进系统作用下浮空器区域驻留控制

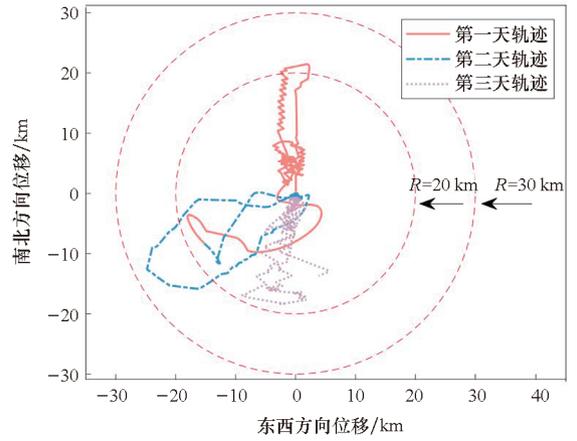
3.4.1 东西方向单通道控制区域驻留仿真

图 8 为东西单通道控制飞行仿真结果,由图 8(a)与图 8(b)可以看出,最大区域驻留半径 $d_{max} = 27.80$ km,平均区域驻留半径 $d_{avg} = 8.2$ km,区域驻留有效时间比 $t_{eff} = 100\%$ 。相比 3.2 节,在东西方向上增加推进系统,能够减小区域驻留半径,在东西方向上的驻留控制效果显著提高。

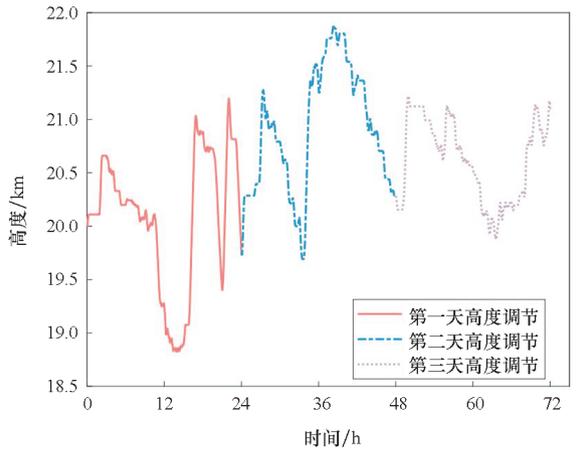
通过图 8(c)与图 4(c)的对比可以发现,东西方向增加动力推进后,浮空器高度调节的范围减少,对于东西方向风场的利用效率降低。同时,从图 8(d)与图 4(d)对比发现,阀门风机开关频率明显减小。



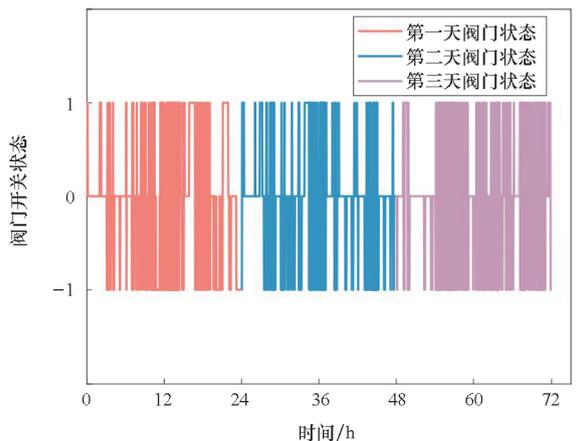
(a) 三维轨迹
(a) Three-dimensional trajectory



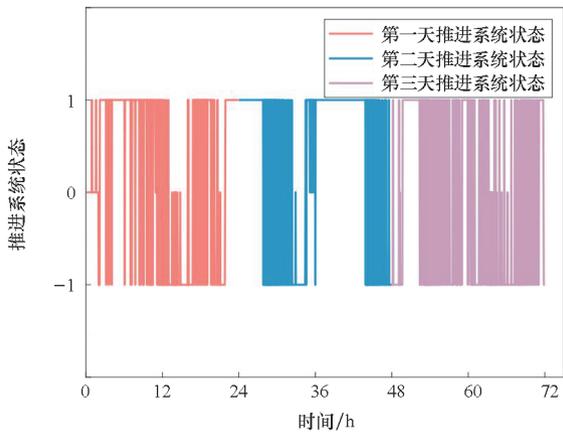
(b) 水平轨迹
(b) Horizontal trajectory



(c) 高度随时间变化情况
(c) Change situation of altitude over time

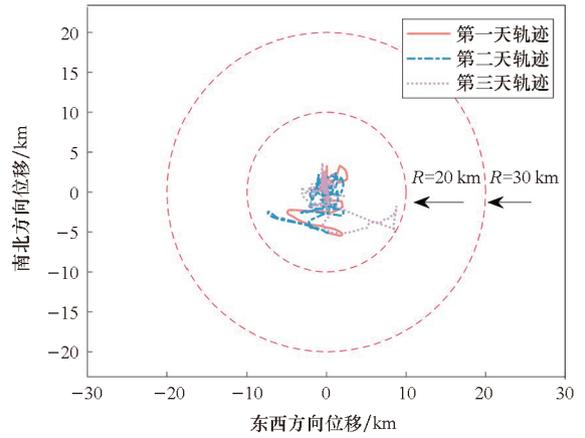


(d) 阀门状态随时间变化情况
(d) Change situation of valve status over time



(e) 推进系统状态随时间变化情况

(e) Change situation of propulsion system status over time



(b) 水平轨迹

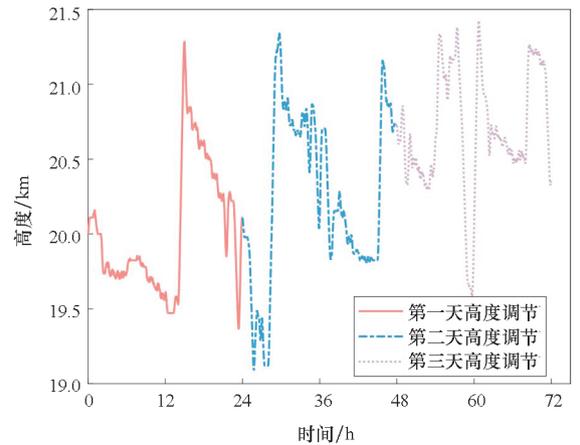
(b) Horizontal trajectory

图8 东西单通道控制飞行仿真结果

Fig. 8 Single-channel control flight simulation results in the east-west direction

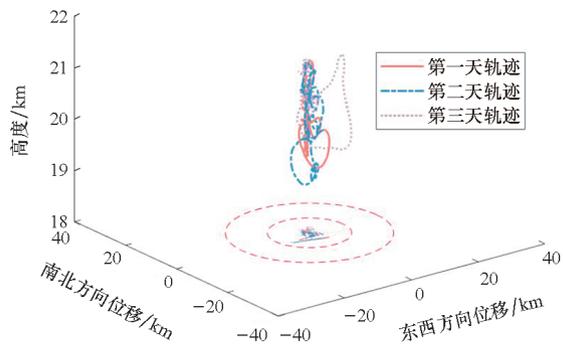
3.4.2 南北方向单通道控制区域驻留仿真

图9为南北单通道控制飞行仿真结果。从图9(a)与图9(b)可知,最大驻留半径 $d_{max} = 10 \text{ km}$, 平均区域驻留半径 $d_{avg} = 2 \text{ km}$, 区域驻留有效时间比 $t_{eff} = 100\%$ 。与无推进系统控制(图4(b))、东西方向单通道控制(图8(b))相比,区域驻留范围大幅减小,区域驻留控制效果明显提升。通过仿真结果可知,若要实现半径10 km范围内区域驻留控制的目标,在高度调节的同时,需在南北方向上用动力推进系统辅助控制轨迹。同时,该结果反映了准零风层风向变化特点:东西方向轨迹控制可以通过调节驻空高度,利用东西方向风向切变明显的特点来实现;南北方向的风向单一,难以通过高度调节控制轨迹,需用动力推进系统辅助控制。综合来看,在南北方向增加动力推进辅助控制轨迹的效果优于东西方向单通道的控制效果。



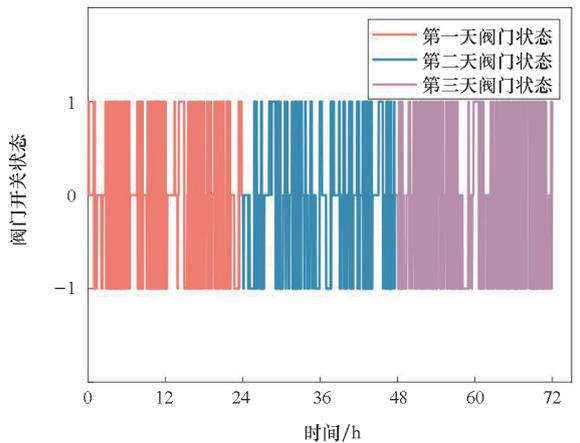
(c) 高度随时间变化情况

(c) Change situation of altitude over time



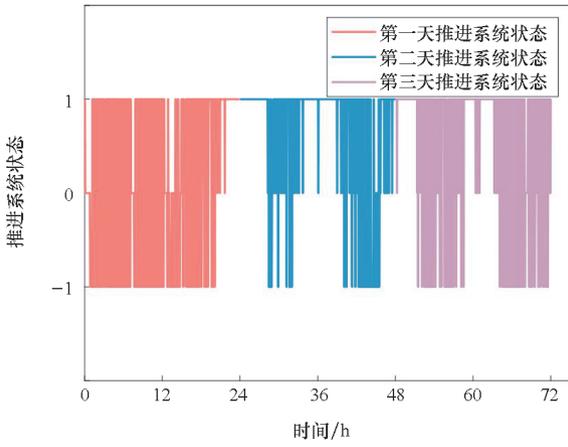
(a) 三维轨迹

(a) Three-dimensional trajectory



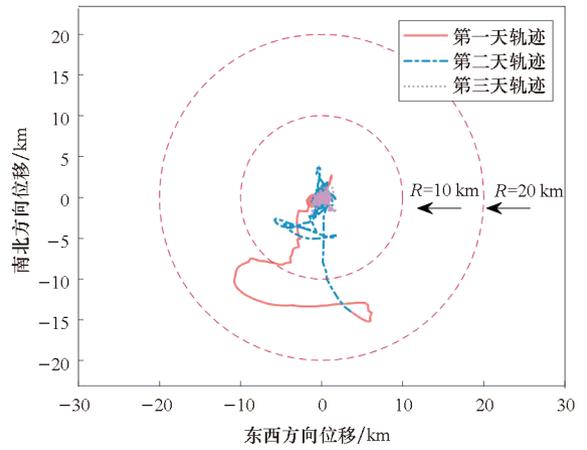
(d) 阀门状态随时间变化情况

(d) Change situation of valve status over time



(e) 推进系统状态随时间变化情况

(e) Change situation of propulsion system status over time



(b) 水平轨迹

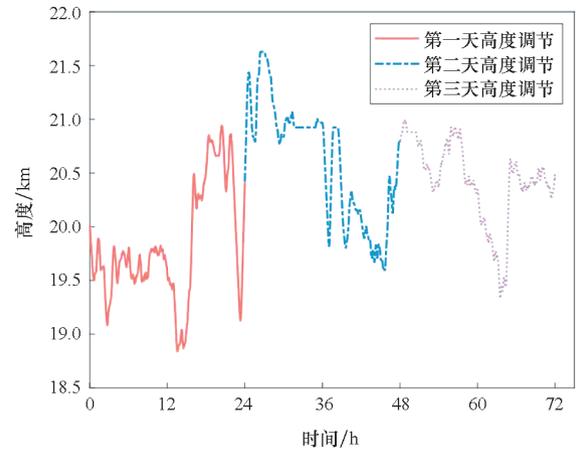
(b) Horizontal trajectory

图 9 南北单通道控制飞行仿真结果

Fig. 9 Single-channel control flight simulation results in the north-south direction

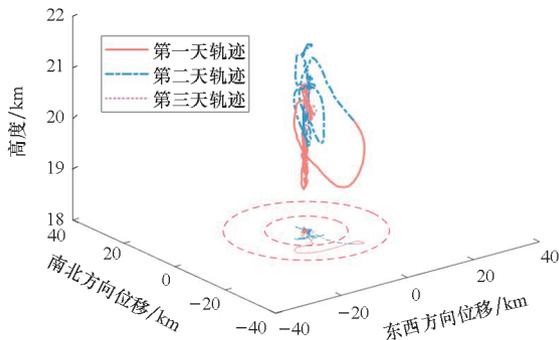
3.4.3 双通道控制区域驻留仿真

图 10 为双通道控制飞行仿真结果。从图 10(b)可以看出,最大区域驻留半径 $d_{\max} = 16.3 \text{ km}$,平均区域驻留半径 $d_{\text{avg}} = 1.7 \text{ km}$,区域驻留有效时间比 $t_{\text{eff}} = 100\%$ 。在双通道水平方向耦合动力推进下,可以进一步缩小区域驻留范围,平均区域驻留半径进一步缩小,但与南北方向单通道控制相差不大,并且通过高度随时间的变化情况可以发现,浮空器的高度调节范围较上述情况较小,且最大驻留半径有所增加。在仿真过程中并未考虑动力推进系统的能源消耗,需要在能源约束的情况下,进一步研究浮空器的区域驻留控制。表 3 为 D3QN 与 DDQN 设计的区域驻留控制器控制效果对比^[18],采用 D3QN 设计的区域驻留控制器控制效果显著优于 DDQN 设计的控制器。



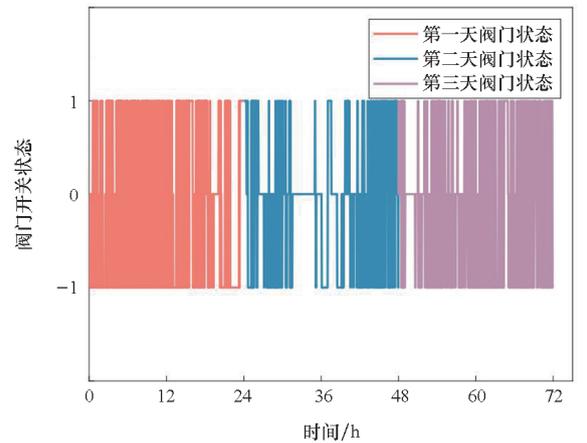
(c) 高度随时间变化情况

(c) Change situation of altitude over time



(a) 三维轨迹

(a) Three-dimensional trajectory



(d) 阀门状态随时间变化情况

(d) Change situation of valve status over time

图 10 双通道控制飞行仿真结果

Fig. 10 Dual-channel control flight simulation results

表3 D3QN 与 DDQN 设计的区域驻留控制器控制效果对比

Tab.3 Comparison of control effect for station keeping controllers designed based on D3QN and DDQN

算法	场景	评价指标		
		d_{max}/km	d_{avg}/km	$t_{eff}/\%$
D3QN	无推进系统	70.37	25.26	96.25
	东西方向推进	27.80	8.20	100
	南北方向推进	10.00	2.00	100
	双通道	16.30	1.70	100
DDQN	无推进系统	72.38	28.16	83
	东西方向推进	62.90	21.38	93
	南北方向推进	30.22	16.20	100
	双通道	17.63	8.84	100

3.5 控制器训练收敛特性分析

图 11 为各控制器的收敛情况。在 2 000 回合左右可以收敛,训练时间和处理过程随处理器频率的降低而增加。根据神经网络输出结果,区域驻留控制器选择动作时间在 5 ms 以内,符合浮空器实际工程中的控制需求。

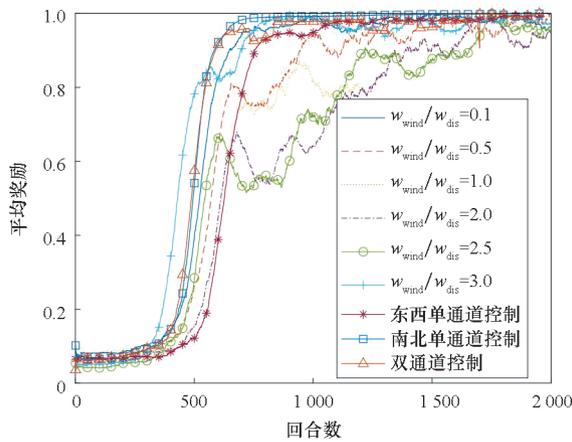


图 11 控制器收敛情况

Fig.11 Convergence situation of controller

4 结论

本文针对利用环境风场实现轨迹控制的浮空器,采用 D3QN 算法设计了区域驻留控制器。针对不同控制通道进行了 3 d 内的区域驻留仿真,验证了通过该方法获得的控制器具有环境适应性强、鲁棒性强的特点。

1) 与基于中心点的区域驻留控制策略、DDQN 算法设计的区域驻留控制器相比,基于 D3QN 算法设计的控制器能明显提高浮空器区域驻留能力,且在扰动情况下,控制器具有较强的鲁棒性;分析了奖励函数对区域驻留控制器性能

的影响,结果表明,可通过选择不同的奖励函数,实现不同的区域驻留控制效果。

2) 在南北方向单通道增加动力推进,能够有效减小区域驻留半径,控制效果优于在东西方向单通道增加动力推进,和双通道增加动力推进控制效果接近。

参考文献 (References)

[1] 《平流层飞艇技术》编写组. 平流层飞艇技术[M]. 北京: 科学出版社, 2019.
Stratospheric Airship Technology Writing Group. Stratospheric airship technology[M]. Beijing: Science Press, 2019. (in Chinese)

[2] 邓小龙, 杨希祥, 麻震宇, 等. 基于风场环境利用的平流层浮空器区域驻留关键问题研究进展[J]. 航空学报, 2019, 40(8): 022941.
DENG X L, YANG X X, MA Z Y, et al. Review of key technologies for station-keeping of stratospheric aerostats based on wind field utilization[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(8): 022941. (in Chinese)

[3] 罗昔柳, 刘俊涛, 张海艳, 等. 快速部署浮空器总体技术研究[J]. 宇航总体技术, 2019, 3(5): 17-22.
LUO X L, LIU J T, ZHANG H Y, et al. General technology study of rapid deployment stratospheric aerostat [J]. Astronautical Systems Engineering Technology, 2019, 3(5): 17-22. (in Chinese)

[4] 肖存英, 胡雄, 龚建村, 等. 中国上空平流层准零风层的特征分析[J]. 空间科学学报, 2008, 28(3): 230-235.
XIAO C Y, HU X, GONG J C, et al. Analysis of the characteristics of the stratospheric quasi-zero wind layer over China[J]. Chinese Journal of Space Science, 2008, 28(3): 230-235. (in Chinese)

[5] JIANG Y, LYU M Y, QU Z P, et al. Performance evaluation for scientific balloon station-keeping strategies considering energy management strategy[J]. Renewable Energy, 2020, 156: 290-302.

[6] DU H F, LYU M Y, LI J, et al. Station-keeping performance analysis for high altitude balloon with altitude control system[J]. Aerospace Science and Technology, 2019, 92:

- 644 - 652.
- [7] 王益平, 周飞, 徐明. 临近空间浮空器区域驻留控制策略研究[J]. 中国空间科学技术, 2018, 38(1): 63 - 69.
WANG Y P, ZHOU F, XU M. Research on control strategy of territory-hovering aerostat in near space[J]. Chinese Space Science and Technology, 2018, 38(1): 63 - 69. (in Chinese)
- [8] 邓小龙, 丛伟轩, 李魁, 等. 风场综合利用的新型平流层浮空器轨迹设计[J]. 宇航学报, 2019, 40(7): 748 - 757.
DENG X L, CONG W X, LI K, et al. Trajectory design of a novel stratospheric aerostat based on comprehensive utilization of wind fields[J]. Journal of Astronautics, 2019, 40(7): 748 - 757. (in Chinese)
- [9] DU H F, LYU M Y, ZHANG L C, et al. Energy management strategy design and station-keeping strategy optimization for high altitude balloon with altitude control system [J]. Aerospace Science and Technology, 2019, 93: 105342.
- [10] BELLEMARE M G, CANDIDO S, CASTRO P S, et al. Autonomous navigation of stratospheric balloons using reinforcement learning[J]. Nature, 2020, 588(7836): 77 - 82.
- [11] 张顶立. 基于深度强化学习的城市场景无人机避撞决策研究[D]. 德阳: 中国民用航空飞行学院, 2022.
ZHANG D L. Research on autonomous collision avoidance decision-making of UAV in urban airspace based on deep reinforcement learning [D]. Deyang: Civil Aviation Flight University of China, 2022. (in Chinese)
- [12] DOMINGO M C. Power allocation and energy cooperation for UAV-enabled mmWave networks: a multi-agent deep reinforcement learning approach[J]. Sensors, 2021, 22(1): 270.
- [13] YANG X W, YANG X X, DENG X L. Horizontal trajectory control of stratospheric airships in wind field using Q -learning algorithm [J]. Aerospace Science and Technology, 2020, 106: 106100.
- [14] TEETHI T I, 卢虎, 闵欢, 等. 基于改进强化学习的无人机规避决策控制算法[J]. 探测与控制学报, 2022, 44(3): 68 - 73.
TEETHI T I, LU H, MIN H, et al. An improved reinforcement learning method for drone avoidance decision control[J]. Journal of Detection & Control, 2022, 44(3): 68 - 73. (in Chinese)
- [15] LIU Y T, PRICE E, GOLDSCHMID P, et al. Autonomous blimp control using deep reinforcement learning [EB/OL]. (2021 - 09 - 27) [2022 - 06 - 15]. <https://arxiv.org/pdf/2109.10719v2>.
- [16] 李智斌, 黄宛宁, 张钊, 等. 2020 年临近空间科技热点回眸[J]. 科技导报, 2021, 39(1): 54 - 68.
LI Z B, HUANG W N, ZHANG Z, et al. Summary of the hot spots of near space science and technology in 2020 [J]. Science & Technology Review, 2021, 39(1): 54 - 68. (in Chinese)
- [17] XU Z Y, LIU Y, DU H F, et al. Station-keeping for high-altitude balloon with reinforcement learning[J]. Advances in Space Research, 2022, 70(3): 733 - 751.
- [18] 柏方超, 杨希祥, 邓小龙, 等. 基于深度强化学习的风场中浮空器驻留控制[J]. 北京航空航天大学学报, 2024, 50(7): 2354 - 2366.
BAI F C, YANG X X, DENG X L, et al. Station keeping control for aerostat in wind fields based on deep reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(7): 2354 - 2366. (in Chinese)
- [19] 张小达, 张鹏, 李小龙. 《标准大气与参考大气模型应用指南》介绍[J]. 航天标准化, 2010(3): 8 - 11.
ZHANG X D, ZHANG P, LI X L. Introduction to *Guidelines for the application of standard atmosphere and reference atmosphere models* [J]. Aerospace Standardization, 2010(3): 8 - 11. (in Chinese)
- [20] IQBAL A, THAM M L, CHANG Y C. Energy-and spectral-efficient optimization in cloud RAN based on dueling double deep Q -network [C]//Proceedings of the IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS), 2021: 311 - 316.
- [21] SCHAUL T, QUAN J, ANTONOGLIOU I, et al. Prioritized experience replay[EB/OL]. (2016 - 02 - 25) [2022 - 07 - 01]. <https://arxiv.org/pdf/1511.05952>.