



基于改进DDQN船舶自动靠泊路径规划方法

李康斌, 朱齐丹, 牟进友, 菅紫婷

引用本文:

李康斌, 朱齐丹, 牟进友, 等. 基于改进DDQN船舶自动靠泊路径规划方法[J]. 智能系统学报, 2025, 20(1): 73-80.

LI Kangbin, ZHU Qidan, MU Jinyou, et al. Automatic ship berthing path-planning method based on improved DDQN[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(1): 73-80.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202401005>

您可能感兴趣的其他文章

舰载机位姿实时视觉测量算法研究

Research on real-time vision measurement algorithm of shipborne aircraft pose

智能系统学报. 2021, 16(6): 1045-1055 <https://dx.doi.org/10.11992/tis.202103014>

一种基于深度学习目标检测的长时目标跟踪算法

A long-term object tracking algorithm based on deep learning and object detection

智能系统学报. 2021, 16(3): 433-441 <https://dx.doi.org/10.11992/tis.201910029>

生成对抗网络辅助学习的舰船目标精细识别

Fine-grained inshore ship recognition assisted by deep-learning generative adversarial networks

智能系统学报. 2020, 15(2): 296-301 <https://dx.doi.org/10.11992/tis.201901004>

弹性网络核极限学习机的多标记学习算法

Multi-label learning algorithm of an elastic net kernel extreme learning machine

智能系统学报. 2019, 14(4): 831-842 <https://dx.doi.org/10.11992/tis.201806005>

基于深度学习的椎间孔狭窄自动多分级研究

Deep learning based automatic multi-classification algorithm for intervertebral foraminal stenosis

智能系统学报. 2019, 14(4): 708-715 <https://dx.doi.org/10.11992/tis.201806015>

欠驱动船舶神经网络自适应路径跟踪控制

Neural network adaptive path tracking control for underactuated ships

智能系统学报. 2018, 13(2): 254-260 <https://dx.doi.org/10.11992/tis.201611011>

DOI: 10.11992/tis.202401005

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20241211.1826.002>

基于改进 DDQN 船舶自动靠泊路径规划方法

李康斌, 朱齐丹, 牟进友, 管紫婷

(哈尔滨工程大学智能科学与工程学院, 黑龙江哈尔滨 150001)

摘要: 船舶在自动靠泊过程中会受到风、浪、流和岸壁效应等因素的影响, 故需要精确的路径规划方法防止靠泊失败。针对全驱动船舶靠泊过程的基于双深度 Q 网络 (double deep Q network, DDQN) 算法, 设计了一种船舶自动靠泊路径规划方法。首先建立船舶三自由度模型, 然后将距离、航向、推力、时间和碰撞作为奖励或惩罚, 改进奖励函数。随后引入 DDQN 来学习动作奖励模型, 并使用学习结果来操纵船舶运动。通过追求更高的奖励值, 船舶可以自行找到最优的靠泊路径。实验结果表明, 在不同水流速度下, 船舶都可以在完成靠泊的同时减小时间和推力, 并且在相同水流速度下, DDQN 算法与 Q-learning、SARSA (state action reward state action)、深度 Q 网络 (deep Q network, DQN) 等算法相比, 靠泊过程推力分别减小了 241.940、234.614、80.202 N, 且时间仅为 252.485 s。

关键词: 自动靠泊; 路径规划; 深度强化学习; 双深度 Q 网络; 奖励函数; 水流速度; 状态探索; 推力; 时间; 独立重复实验

中图分类号: TP273 文献标志码: A 文章编号: 1673-4785(2025)01-0073-08

中文引用格式: 李康斌, 朱齐丹, 牟进友, 等. 基于改进 DDQN 船舶自动靠泊路径规划方法 [J]. 智能系统学报, 2025, 20(1): 73-80.

英文引用格式: LI Kangbin, ZHU Qidan, MU Jinyou, et al. Automatic ship berthing path-planning method based on improved DDQN[J]. CAAI transactions on intelligent systems, 2025, 20(1): 73-80.

Automatic ship berthing path-planning method based on improved DDQN

LI Kangbin, ZHU Qidan, MU Jinyou, JIAN Ziting

(School of Intelligent Science and Engineering, Harbin Engineering University, Harbin 150001, China)

Abstract: During the automatic docking process, ships are influenced by factors such as wind, waves, currents, and quay wall effects, necessitating precise path-planning methods to prevent docking failures. For fully actuated ships, a ship's automatic docking path-planning method is designed based on the double deep Q network (DDQN) algorithm. Firstly, a three-degree-of-freedom model for the ship is established, and then the reward function is improved by incorporating distance, heading, thrust, time, and collisions as rewards or penalties. DDQN is then introduced to learn the action-reward model and use the learning results to manipulate ship movements. By pursuing higher reward values, the ship can autonomously find the optimal docking path. Experimental results show that under different water flow velocities, ships can reduce both time and thrust while completing docking. Moreover, at the same water flow velocity, compared with Q-learning, SARSA, and deep Q Network (DQN), the DDQN algorithm reduces thrust by 241.940N, 234.614N, and 80.202N respectively during the docking process, with the time being only 252.485 seconds.

Keywords: automatic berthing; path planning; deep reinforcement learning; double deep Q network; reward function; current velocity; state exploration; thrust; time; independent repeated experiments

船舶自动靠泊问题涉及船舶在浅水区的低速运动, 风、浪、流和岸壁效应的干扰比较大, 对船

舶的操纵和控制比较困难^[1-3]。靠泊路径规划问题是船舶自动靠泊的关键问题之一^[4], 是在一定的限制下, 规划一条满足设定指标的从起点到目标点的最优路径。因此对于船舶靠泊路径规划不仅要规划靠泊路径, 还要考虑路径点对应的船舶

收稿日期: 2024-01-03. 网络出版日期: 2024-12-12.

基金项目: 国家自然科学基金项目 (52171299).

通信作者: 朱齐丹. E-mail: zhuqidan@hrbeu.edu.cn.

©《智能系统学报》编辑部版权所有

状态。此外,船舶的大质量、大体积导致其惯性大、加速度小,对控制响应有一定的延迟。因此应用常用的机器人路径规划算法来规划船舶靠泊路径是困难的^[5]。

在船舶路径规划中,优先考虑直线路径、最小化转弯和避免复杂环境^[6]。传统的船舶路径规划方法以 A* 算法为主导方法,需要完整的环境信息作为先验知识,但在未知的海洋环境中很难获得周围的环境信息,因此只适用于简单和静态的环境^[7-9],此外传统路径规划方法主要考虑路径的距离、平滑度和可达性,而不考虑智能体的动态特性,因此,这些规划方法得到的路线往往不能满足船舶运动学 and 安全性要求^[10]。

目前,人工智能技术尤其是强化学习在船舶路径规划中表现出一定程度的优越性^[11-12]。强化学习主要由智能体、环境、动作、状态和奖励 5 个部分组成,强调智能体从环境到行为映射的学习,不需要任何人类知识或预先设定的规则,并通过最大化价值函数,寻求最正确或最佳的行动决策^[13-16]。Chen 等^[10]采用 Q-learning 算法进行船舶路径规划,经过多轮训练,可以让船舶自主找到合适的路径或导航策略。Li 等^[17]利用人工势场算法对深度 Q 网络(deep Q network, DQN)算法进行改进,实现了具有海上自动驾驶能力的无人船自主避碰路径规划。Zhang 等^[18]为实现港口环境下船舶路径规划,提出了一种基于分层深度强化学习(deep reinforcement learning, DRL)的自主导航决策模型,实验结果表明,改进的 DRL 算法可以有效地提高船舶路径规划的安全性和避障能力。Zhou 等^[19]提出了一种基于 DQN 的无人船协同路径规划方法,该方法定义了 13 个动作,让船舶在其中选择当前状态下的最优动作。Guo 等^[20]将深度确定性策略梯度(deep deterministic policy gradient, DDPG)与人工势场相结合进行船舶路径规划,但主要关注局部避碰。

目前强化学习在船舶路径规划领域的研究主要应用于开阔水域及避碰,而对船舶自动靠泊的研究较少,且在船舶路径规划过程中大都集中于规划路径的长短,而对船舶自动靠泊所需的推力和时间研究很少,本研究的重点是使用双深度 Q 网络(double DQN, DDQN)算法在离散动作空间进行船舶自动靠泊路径规划,目的不是寻找船舶靠泊的最短路径,而是寻找一条考虑船舶动力特性并实现最小推力和最小时间的实用安全的靠泊路径,以此达到自动靠泊的安全性和经济性目的。不同于以往传统的奖励函数只考虑到目的地和碰撞惩罚的设计,本文充分考虑船舶靠泊过

程中距离、航向、碰撞、推力和时间等方面,改进了奖励函数,避免由于奖励稀疏而导致智能体在学习过程中出现停滞。并且设计了一个新的距离依赖奖励,即通过使用势场解决智能体在平衡距离和航向时产生的困惑,以此引导船舶到达泊位的同时,航向也满足要求,并完成减小推力和缩短时间的预期目标。在学习过程中考虑船舶运动学方程,实现了船舶的自动驾驶。

1 船舶数学模型与仿真系统构建

1.1 船舶模型

本文的研究对象是一艘名为 CSAD(cybership arctic drillship)的全驱动船模,其原型为挪威科技大学以 1:90 的比例缩放的北极钻井船。该船模一共包含 6 个推进器,推进器具体分布位置如图 1 所示。

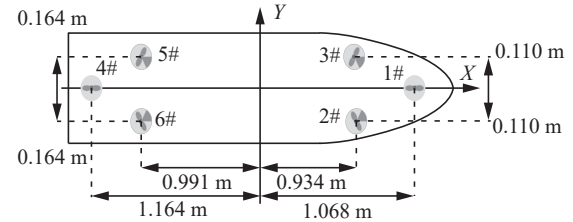


图 1 CSAD 船模的推进器分布

Fig. 1 Propeller distribution of CSAD ship model

由于船舶的近岸靠泊是在水平面上的低速运动,故仅考虑船体在横荡、纵荡和艏摇 3 个自由度方向上的运动。参考 Fossen 提出的三自由度模型^[21],船舶靠泊的数学模型为

$$\dot{\eta} = R(\psi)v$$

$$M_{RB}\dot{v} + M_A\dot{v}_r + C_{RB}(v)v + C_A(v_r)v_r + D(v_r)v_r = \tau_{thr} \quad (1)$$

式中: v 为船舶速度; v_r 为船舶相对于水流 v_c 的速度,即 $v_r = v - v_c$; M_A 和 M_{RB} 分别为刚体质量矩阵和附加质量矩阵; D 表示阻尼矩阵; C_{RB} 和 C_A 表示科里奥利力矩阵和向心力矩阵; τ_{thr} 为推进器施加的力。

假设水流速度缓慢,即 $\dot{v}_c \approx 0$, 则有 $\dot{v}_r \approx \dot{v}$, 然后可以把两个质量矩阵结合成 $M = M_{RB} + M_A$, 由于船舶低速航行,故忽略科里奥利力的影响。此外,引入一个已知的偏差 $b \in R^3$, 最后将式(1)简化表示船舶速度变化:

$$\dot{v}_r = M^{-1}(-D(v_r)v_r + \tau_{thr} + R(\psi)^T b)$$

式中 ψ 为航向。

1.2 船舶路径规划

路径规划是指为了完成特定的任务,为系统构建预先定义的路径。通常基于一些特定属性设计路径,这些属性取决于船舶和航行环境的系统

约束。例如, 需要考虑船舶的机动性和速度, 同时需要避免碰撞, 使船舶远离危险。

1.2.1 船舶推力方程

使船舶收敛并遵循一条或多条期望的参数化路径的实践被称为机动问题, 分为几何和动态两个子任务, 几何任务是使船舶几何上收敛于用连续标量变量参数化的路径的行为, 动态任务是通过时间、速度或加速度的分配来实现船舶沿着几何任务的路径期望的动态行为^[22]。

几何任务表示为

$$\lim_{t \rightarrow \infty} |p^n(t) - p_d^n(s(t))| = 0$$

$$p_d(s) = \begin{bmatrix} p_{d,x}(s) \\ p_{d,y}(s) \end{bmatrix} = \begin{bmatrix} a_1 s + b_1 \\ a_2 s + b_2 \end{bmatrix} \in \mathbb{R}^2$$

式中 a_1 、 b_1 、 a_2 、 b_2 分别为多项式系数。

动态任务表示为

$$\lim_{t \rightarrow \infty} |\dot{s}(t) - v_s(s(t), t)| = 0$$

$$v_s(t, s) = \frac{u_d(t)}{|\bar{p}_d^s(s)|}$$

式中: $0 \leq s \leq 1$ 为路径变量; p^n 是船舶位置 p 的一阶到 n 阶导数; $u_d(t)$ 表示船舶在时刻 t 的期望速度; 期望航向 ψ_d 为生成路径相切的角度, 即 $\psi_d(s) = \angle p_d^s(s)$ 。机动问题的完整目标为生成一个控制力 τ , 使得 $\eta \rightarrow \eta_d$, 其中 $\eta_d(s) = [p_d(s) \psi_d(s)]^T$ 。定义以下参数:

$$z_1 = R(\psi)^T [\eta - \eta_d(s)]$$

$$z_2 = v - \alpha_1$$

$$w = \dot{s} - v_s(s, t)$$

式中 α_1 为定义的虚拟控制变量, 计算公式为

$$\alpha_1 = -K_1 z_1 + R(\psi)^T \eta_d^s v_s - k_1 z_1$$

其中: K_1 和 k_1 为调谐增益, z_1 和 z_2 分别为位置误差和速度误差。定义一个保持恒定切向量长度的更新律, 即单位切线梯度更新律 w 。

$$w = \mu \frac{\eta_d^s(s)^T}{|\eta_d^s(s)|} R(\psi) z_1, \mu \geq 0$$

则有 $\dot{s} = v_s(s, t) + w$

$$\sigma_1 = r \tilde{K}_1 S z_1 - \tilde{K}_1 v - r S R^T \eta_d^s(s) v_s(t, s) + R(\psi)^T \eta_d^s(s) v_s'(t, s)$$

式中: $\tilde{K}_1 = K_1 + k_1 I$, σ_1 为期望加速度, r 为速度, S 为斜对称矩阵。从而推导出船舶的推力方程:

$$\tau = -K_2 z_2 + D \alpha_1 - R(\psi)^T b + M(\sigma_1 + \alpha_1^s \dot{s})$$

其中 b 为测量偏差。

1.2.2 水流的影响

大多数情况下, 在靠泊过程中, 船舶会受到水流的影响, 水流会对船舶产生一定的作用力, 水流通常具有复杂的方向和速度^[23], 因此增加了靠泊过程的复杂性。但如果使用得当, 水流也可以

成为一种优势, 例如, 当水流方向和泊位平行时, 可以通过调整船舶方向, 使船的航向与泊位的夹角为锐角, 水流可以帮助船舶靠泊, 如图 2 所示。此外, 低速行驶时, 需要避免水流使船体接近障碍物。

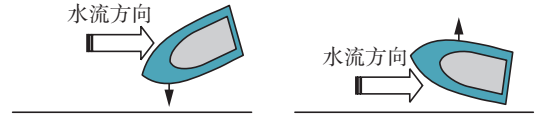


图 2 水流对船舶靠泊的影响

Fig. 2 Effect of water currents on ship berthing

2 DDQN 算法和奖励函数改进

本文研究的船舶自动靠泊是基于深度强化学习使船舶的控制系统学习最佳的靠泊机动。马尔可夫决策过程 (Markov decision process, MDP) 是解决强化学习相关问题的框架, 也是强化学习底层的数学模型^[24-25]。本文针对船舶运动特性设置状态空间、动作空间和奖励函数, 应用 DDQN 算法引导智能体获得最大累积奖励, 使船舶能够学习到完成自动靠泊的同时减小推力和缩短时间。

2.1 DDQN 算法

结合神经网络和 Q-learning 的 DDQN 算法是强化学习的一种算法, 可以解决 DQN 算法对不可见状态的估计误差过大而导致的高估问题, 这些高估是不可忽略的^[26-27]。因此本文采取 DDQN 算法, DDQN 算法的核心思想是构建两个网络, 分别用来估计动作和动作的价值, 因此用估计网络确定动作, 用目标网络确定该动作的价值, 通过防止对奖励的高估来获得更合适的策略, 与 DQN 相比可以解决更一般的问题^[28-30]。

在本文中, 船舶自动靠泊的 MDP 由 5 个要素 (S, A, P, r, γ) 定义。 S 表示状态空间; A 表示动作空间; P 表示智能体在选择动作后状态过渡到下一个状态的概率, 即状态转移函数; r 代表智能体所获得的奖励, 即智能体选择某一动作所获得的即时奖励, 可以是正的, 也可以是负的; γ 表示折扣因子。

2.2 DDQN 算法奖励函数的改进

奖励函数的改进会对强化学习智能体、动作、环境和状态产生影响, 具体表现为奖励函数直接影响智能体的行为, 会引导其朝着期望的方向学习; 奖励函数通过引导智能体做出能够获得更高奖励的动作, 影响动作选择策略; 在强化学习过程中, 通过智能体做出的动作与环境交互, 根据改进的奖励函数获得奖励值, 这可能会导致环境的动态变化; 改进的奖励函数可能会影响智

能体对状态空间的探索方式,能够更有效地利用重要的状态取得更大奖励。因此奖励函数能够直接影响船舶在靠泊过程中的表现水平以及能否学习到最优的靠泊策略。传统的奖励函数设计为

$$R_t = \begin{cases} C_1, & \text{到达泊位点} \\ C_2, & \text{发生碰撞} \\ 0, & \text{其他} \end{cases} \quad (2)$$

式中: C_1 通常为正数, C_2 通常为负数。在传统的奖励函数中,船舶只能在到达泊位或者碰撞时能够获得奖励外,在其他状态无法获取及时有效的反馈,即存在奖励稀疏问题。

本文在设计奖励函数时,考虑了船舶靠泊过程的距离、碰撞惩罚、航向、减小航行所需的推力和时间等。为使船舶到达码头,设计奖励 $r_t(\Delta_{\text{dist}})$, 定义船舶位置 p 和码头位置 p_{dock} 之间的距离为 d , 则奖励 $r_t(\Delta_{\text{dist}})$ 定义为 d 从当前路径点 p_0 到目标路径点 p_t 的减小量与调整常数 $C_{t,\text{dist}}$ 的乘积。

$$d = \|p_{\text{dock}} - p\|_2 = \sqrt{(p_{\text{dock},1} - p_1)^2 + (p_{\text{dock},2} - p_2)^2}$$

$$\Delta_{\text{dist}} = d_{p_0} - d_{p_t}$$

$$r_t(\Delta_{\text{dist}}) = C_{t,\text{dist}} \times \Delta_{\text{dist}}$$

根据船舶到达码头时与期望航向的偏移角度 ψ_{offset} 来设计奖励 $r(\psi_{\text{offset}})$ 。基于船舶靠泊最危险的角度 90° 开发奖励功能,对于 $\psi_{\text{offset}} \in [0^\circ 90^\circ]$, 根据从 $\psi_{\text{offset}} = 90^\circ$ (奖励最小为 0) 到 $\psi_{\text{offset}} = 0^\circ$ (奖励最大) 的线性增长定义奖励为

$$r(\psi_{\text{offset}}) = \frac{(90^\circ - \psi_{\text{offset}})}{C_{\psi_{\text{offset}}}}, \quad p \in A_{\text{dock}}$$

式中 $C_{\psi_{\text{offset}}}$ 是奖励函数的调节常数。

除了以正确的航向到达码头,本文还引入了减小推力的子目标,为了两者之间的权衡,本文提出由船舶进入目标停泊区域触发的终端奖励,将总推力的奖励表示为入口航向的奖励。从而得到到达对接状态的完整终端奖励:

$$r_{T,\text{dock}} = r_T(\tau, t)^{r_T(\psi_{\text{offset}})}, \quad p \in A_{\text{dock}}$$

式中: $r_T(\tau, t) = \frac{C_{T,\tau}}{1 + \int_0^{\tau(p_{\text{dock}})} \|\tau\|}$, $r_T(\psi_{\text{offset}})$ 为终端的 $r(\psi_{\text{offset}})$ 。

此外补充即时推力奖励,以提高智能体在整个过程中对推力减小的关注:

$$r_t(\tau, t) = \frac{C_{t,\tau}}{1 + \int_{t(p_0)}^{\tau(p_t)} \|\tau\|}$$

为了减小船舶沿着路径移动的时间。设计奖励最终时间奖励 $r_{T,t}(t)$ 和即时时间奖励 $r_{t,t}(t)$ 。

$$r_{T,t}(t) = \frac{C_{T,t}}{1 + t(p_{\text{dock}})}, \quad p \approx p_{\text{dock}}$$

$$r_{t,t}(t) = \frac{C_{t,t}}{1 + t(p_t) - t(p_0)}$$

式中: $t(p_0)$ 和 $t(p_t)$ 表示船舶在航路点 p_0 和 p_t 处的时间; $C_{T,t} > 0$ 和 $C_{t,t} > 0$ 为调整常数,这可以使智能体倾向于做出减小时间的行动。

到目前为止,奖励函数能引导船舶,以理想的角度到达码头,并在过程中减小推力和缩短时间。为了不让智能体产生困惑,不知道在减小距离和减小 ψ_{offset} 之间重视什么,本文提出一个新的距离依赖奖励,通过使用势场特征,在整个过程中返回所需的航向指示,依赖奖励表示为

$$r_t(\psi_{\text{offset}}, d) = C_{t,\text{dist}} \cdot e^{-d} \cdot r(\psi_{\text{offset}})$$

该奖励表明偏移角在远离码头时不重要,但随着距离的减小呈指数增长。智能体因此可以在靠泊过程的早期优先考虑距离,在接近泊位时优先考虑航向。这个优先级通过奖励优先级对 $\alpha_{\psi_{\text{offset}}}$ 和 $\alpha_{\Delta_{\text{dist}}}$ 进一步调整。

最后,由于本文的目标是开发足够安全的路径,当发生碰撞时,定义负奖励 $r_{T,\text{col}} = -C_{T,\text{col}}$, 其中 $C_{T,\text{col}}$ 是一个调整常数。所有的路径规划常数如表 1 所示。

表 1 用于路径规划的调整常数
Table 1 Adjustment constants used for path planning

描述	符号	值
距离变化常数	$C_{t,\text{dist}}$	10
靠泊角度偏移常数	$C_{\psi_{\text{offset}}}$	5
总推力常数	$C_{T,\tau}$	5000
推力常数	$C_{t,\tau}$	15
碰撞惩罚常数	$C_{T,\text{col}}$	1000
优先常数 $r_t(\Delta_{\text{dist}})$	$\alpha_{\Delta_{\text{dist}}}$	0.9
优先常数 $r_t(\psi_{\text{offset}}, d)$	$\alpha_{\psi_{\text{offset}}}$	0.2
终端时间奖励常数	$C_{T,t}$	100000
连续收获时间奖励常数	$C_{t,t}$	1

因此完整的奖励函数变为

$$R = \begin{cases} r_{T,\text{dock}} + r_{T,t}, & p_t \in A_{\text{dock}} \\ r_{T,\text{col}}, & \eta \notin S \\ (\alpha_{\Delta_{\text{dist}}} \cdot r_t(\Delta_{\text{dist}}) + \alpha_{\psi_{\text{offset}}} \cdot r_t(\psi_{\text{offset}}, d)) + \\ r_t(\tau, t) + r_{t,t}(t), & \text{其他} \end{cases} \quad (3)$$

3 基于改进 DDQN 的靠泊方案设计

本文将 DDQN 算法应用到船舶自动靠泊路径规划,并设计了不同水流速度下的对比实验和相同水流速度不同算法的对比实验。

实验的奖励函数采用本文提出的改进奖励函数,如式 (3) 所示。状态空间用 p_0 、 p_t 两个变量来表述。其中 p_0 是船舶出发的路径点, p_t 为船舶当

前路径点。船舶的动作空间包含 3 个动作,如图 3 所示,分别为向前、向左转弯和向右转弯。设置 ε 为 0.3,用以 DDQN 算法的平衡探索和利用,设置折扣因子 γ 为 0.9,学习率 α 为 0.3。通过此靠泊方案的设计可以完成船舶靠泊过程的路径规划。

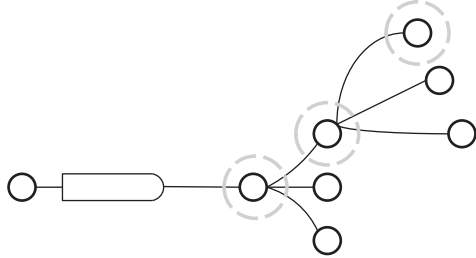


图 3 船舶的动作空间

Fig. 3 Action space of the ship

船舶靠泊路径规划的训练过程包括初始条件为 $p_{0,0} = [0 \ 0]^T$, $p_{t,0} = [0.5 \ 1]^T$ 以及位于 $p_{0,0}$ 的船舶,航向与结果段相切,速度为 0,即 $\eta_0 = [0 \ 0 \ \pi/4]^T$, $v_0 = [0 \ 0 \ 0]^T$ 。设置泊位点为 $[5,5]$ 。引入恒定的参考速度为 $u_{\text{const}} = 3 \text{ cm/s}$ 。训练的集数为 1000 次。在训练中,船舶成功靠泊为一次成功的路径规划,通过式 (2) 计算船舶靠泊所需的推力,并记录时间。图 4 为基于 DDQN 的船舶自动靠泊路径规划流程。

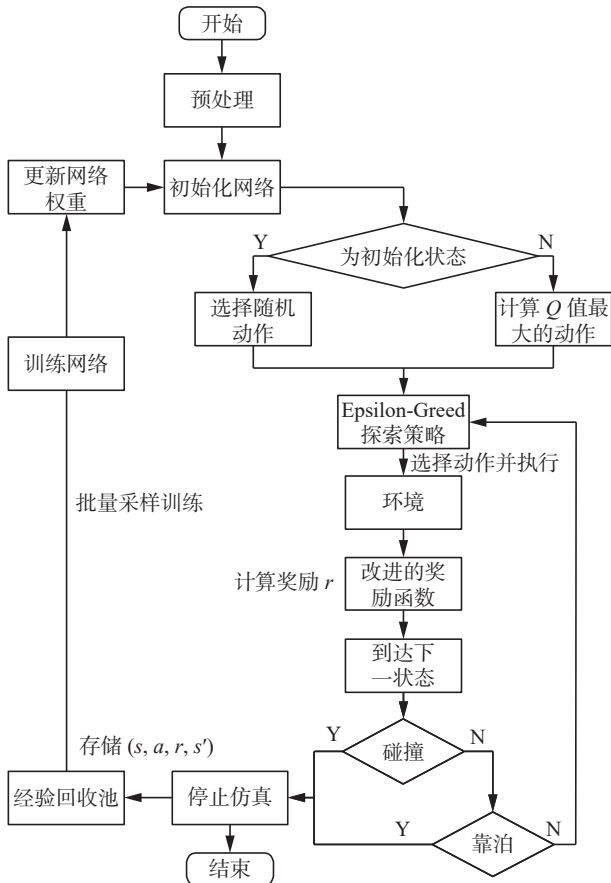


图 4 基于 DDQN 的船舶路径规划流程

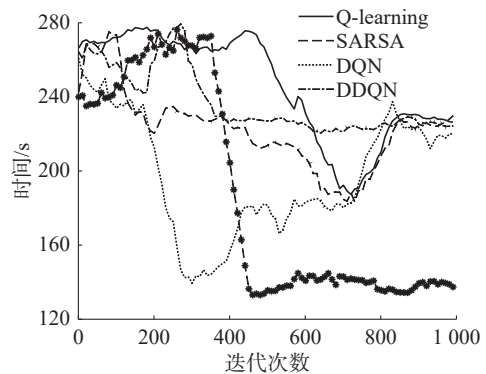
Fig. 4 Flowchart of ship path planning based on DDQN

4 仿真验证和对比分析

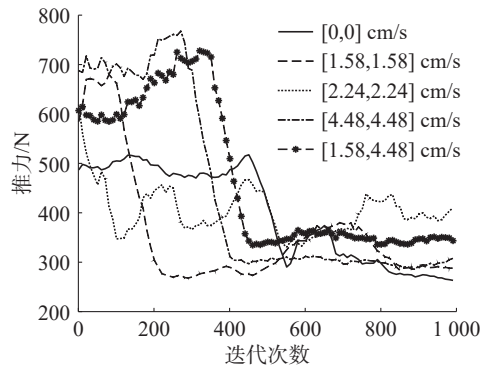
本文首先进行在不同水流速度下采用 DDQN 算法进行船舶自动靠泊的实验。然后在相同水流速度下,分别采用 DDQN 算法、DQN 算法、Q-learning 算法和 SARSA (state action reward state action) 算法,对船舶成功靠泊所需的平均推力和平均时间进行对比。

4.1 不同水流速度下的结果

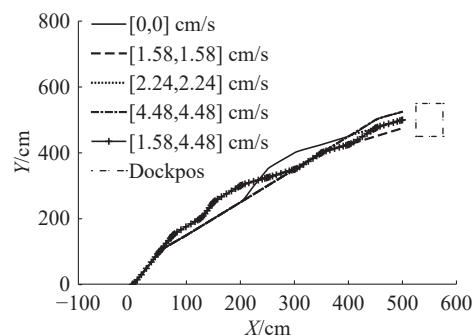
在不同水流速度环境下,研究船舶采用 DDQN 算法船舶的靠泊过程。本文分别取水流速度为 $[0,0]$ 、 $[1.58,1.58]$ 、 $[2.24,2.24]$ 、 $[4.48,4.48]$ 、 $[1.58,4.48] \text{ cm/s}$ 展开实验。对得到的数据进行平滑处理后,画出推力和时间随着训练迭代次数的变化以及靠泊路线图,如图 5 所示。



(a) 时间对比



(b) 推力对比



(c) 船舶路径规划线路

图 5 不同水流速度下的实验对比

Fig. 5 Experimental comparison chart of different water flow velocities

从图 5 可以看出,在不同的水流环境下,采用 DDQN 算法,均可以完成船舶安全靠泊的任务,靠泊航向满足要求,并且可以实现减小推力和缩短时间的预期目标。记录 15 次独立重复实验中船舶成功靠泊所需的平均时间和平均推力,如表 2 所示。

表 2 平均推力和平均时间
Table 2 Average thrust and average time

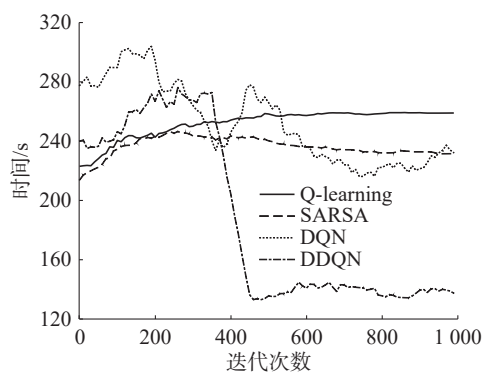
水流速度/(cm/s)	推力/N	时间/s
[0,0]	334.687	248.215
[1.58,1.58]	335.266	246.244
[2.24,2.24]	332.572	250.739
[4.48,4.48]	333.400	245.319
[1.58,4.48]	364.319	252.485

从表 2 可以看出,由于水流速度增幅不高,当水流速度增大时,船舶成功靠泊时所需的推力变化并不大,当水流速度为 [0,0]、[1.58,1.58]、[2.24,2.24]、[4.48,4.48] cm/s 时,推力大约为 334 N,时间相差也不大,为 248 s 左右。当水流速度为 [1.58,4.48] cm/s 时,由于水流在 Y 轴方向的速度大于 X 轴方向的速度,船舶靠泊成功所需的平均推力变大到 364.319 N,平均时间也需要 252.485 s。基于这种情况研究当水流速度为 [1.58,4.48] cm/s 时,采用不同算法船舶成功靠泊所需的平均时间和平均推力。

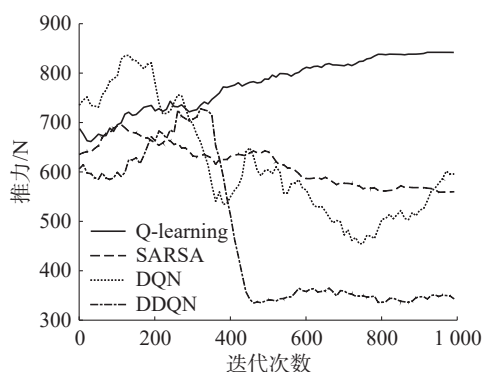
4.2 采用不同算法的对比结果

取水流速度为 [1.58,4.48] cm/s 时,对比采用 Q-learning、SARSA、DQN 和 DDQN 等算法船舶靠泊的结果,如图 6 所示。从图 6 看出,采用不同的强化学习算法,均可以在船舶靠泊的同时,靠泊航向满足要求,并完成减小推力和时间的预期目标。

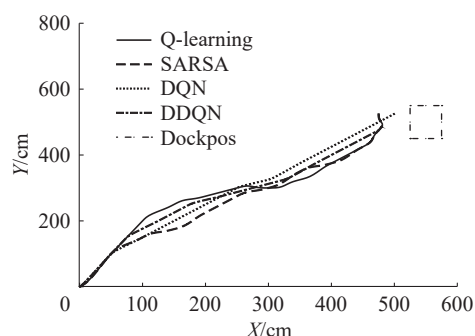
当水流速度为 [1.58,4.48] cm/s 时,将采用 DDQN 算法与 DQN 算法、Q-learning 算法以及 SARSA 算法时进行 15 次独立重复实验的船舶成功靠泊的平均推力和平均时间进行对比,如表 3 所示。从表 3 可以看出,本文采用的 DDQN 算法,所需的推力最小,具体为比 DQN 算法小 80.202 N,比 Q-learning 算法小 241.941 N,比 SARSA 算法小 234.614 N。所需的时间与采用 SARSA 算法和 Q-learning 算法时的时间大体相同,且比 DQN 算法短 6.655 s。因此本文采用的 DDQN 算法在船舶自动靠泊路径规划方面优于 DQN 算法、Q-learning 算法和 SARSA 算法。



(a) 时间对比



(b) 推力对比



(c) 船舶路径规划线路

图 6 采用不同算法结果对比

Fig. 6 Comparative chart of results using different algorithms

表 3 采用不同算法的平均时间和平均推力

Table 3 Average time and average thrust using different algorithms

算法	DDQN	DQN	Q-learning	SARSA
推力/N	364.319	444.521	606.260	598.933
时间/s	252.485	259.140	251.098	252.543

5 结束语

本文针对传统船舶自动靠泊路径规划算法无法回收历史数据用于在线训练和学习,导致算法的准确率较低、适用性差和规划的路径不光滑等问题,提出了一种基于 DDQN 算法的船舶自动靠泊的路径规划方法,DDQN 算法是在 DQN 算法

的基础上构建两个网络,分别用来估计动作和动作的价值,从而解决 DQN 算法的高估问题,同时,有针对性地设计状态空间、动作空间,并改进奖励函数以解决奖励稀疏问题,随后将算法置于不同的水流速度环境中进行训练。实验结果表明,本文采用的 DDQN 算法均可以在靠泊成功的同时缩短时间和减小推力。随后将 DQN、Q-learning、SARSA 等 3 种算法与 DDQN 算法在相同水流速度下进行对比实验,结果表明 DDQN 算法成功靠泊所需的推力和时间最小。

参考文献:

- [1] 刘志林,苑守正,郑林焄,等.船舶自动靠泊技术的发展现状和趋势[J]. *中国造船*, 2021, 62(4): 293–304.
LIU Zhilin, YUAN Shouzheng, ZHENG Linhe, et al. Development status and trend of ship automatic berthing technology[J]. *Shipbuilding of China*, 2021, 62(4): 293–304.
- [2] 李国帅,张显库,张安超.智能船舶靠泊技术研究热点与趋势[J]. *中国舰船研究*, 2024, 19(1): 3–14.
LI Guoshuai, ZHANG Xianku, ZHANG Anchao. Research hotspots and tendency of intelligent ship berthing technology[J]. *Chinese journal of ship research*, 2024, 19(1): 3–14.
- [3] 包政凯,朱齐丹,刘永超.满秩分解最小二乘法船舶航向模型辨识[J]. *智能系统学报*, 2022, 17(1): 137–143.
BAO Zhengkai, ZHU Qidan, LIU Yongchao. Ship heading model identification based on full rank decomposition least square method[J]. *CAAI transactions on intelligent systems*, 2022, 17(1): 137–143.
- [4] WANG Shaobo, ZHANG Yingjun, LI Lianbo. A collision avoidance decision-making system for autonomous ship based on modified velocity obstacle method[J]. *Ocean engineering*, 2020, 215: 107910.
- [5] LIU Xinzhaoh, REN Junsheng, WANG Hui. Optimal energy trajectory planning and control for automatic ship berthing[C]//2020 39th Chinese Control Conference. Shenyang: IEEE, 2020: 1420–1425.
- [6] GAO Xuanyu, DONG Yitao, HAN Yi. An optimized path planning method for container ships in Bohai Bay based on improved deep Q-learning[J]. *IEEE access*, 2023, 11: 91275–91292.
- [7] GUO Siyu, ZHANG Xiuguo, ZHENG Yisong, et al. An autonomous path planning model for unmanned ships based on deep reinforcement learning[J]. *Sensors*, 2020, 20(2): 426.
- [8] HE Zhibo, LIU Chenguang, CHU Xiumin, et al. Dynamic anti-collision A-star algorithm for multi-ship encounter situations[J]. *Applied ocean research*, 2022, 118: 102995.
- [9] LIU Chenguang, MAO Qingzhou, CHU Xiumin, et al. An improved A-star algorithm considering water current, traffic separation and berthing for vessel path planning[J]. *Applied sciences*, 2019, 9(6): 1057.
- [10] CHEN Chen, CHEN Xianqiao, MA Feng, et al. A knowledge-free path planning approach for smart ships based on reinforcement learning[J]. *Ocean engineering*, 2019, 189: 106299.
- [11] 周治国,余思雨,于家宝,等.面向无人艇的 T-DQN 智能避障算法研究[J]. *自动化学报*, 2023, 49(8): 1645–1655.
ZHOU Zhiguo, YU Siyu, YU Jiabao, et al. Research on T-DQN intelligent obstacle avoidance algorithm of unmanned surface vehicle[J]. *Acta automatica sinica*, 2023, 49(8): 1645–1655.
- [12] SILVA JUNIOR A G D, SANTOS D H D, NEGREIROS A P F, et al. High-level path planning for an autonomous sailboat robot using Q-learning[J]. *Sensors*, 2020, 20(6): 1550.
- [13] 邢博闻,张昭夷,王世明,等.基于深度强化学习的多无人艇协同目标搜索算法[J]. *兵器装备工程学报*, 2023, 44(11): 118–125.
XING Bowen, ZHANG Zhaoyi, WANG Shiming, et al. Multi-USV cooperative target search algorithm based on deep reinforcement learning[J]. *Journal of ordnance equipment engineering*, 2023, 44(11): 118–125.
- [14] 曹景祥,刘其成.基于深度强化学习的路径规划算法研究[J]. *计算机应用与软件*, 2022, 39(11): 231–237.
CAO Jingxiang, LIU Qicheng. Research on path planning algorithm based on deep reinforcement learning[J]. *Computer applications and software*, 2022, 39(11): 231–237.
- [15] 闫皎洁,张镔石,胡希平.基于强化学习的路径规划技术综述[J]. *计算机工程*, 2021, 47(10): 16–25.
YAN Jiaojie, ZHANG Qieshi, HU Xiping. Review of path planning techniques based on reinforcement learning[J]. *Computer engineering*, 2021, 47(10): 16–25.
- [16] 邓修朋,崔建明,李敏,等.深度强化学习在机器人路径规划中的应用[J]. *电子测量技术*, 2023, 46(6): 1–8.
DENG Xiupeng, CUI Jianming, LI Min, et al. Application of deep reinforcement learning in robot path planning[J]. *Electronic measurement technology*, 2023, 46(6): 1–8.
- [17] LI Lingyu, WU Defeng, HUANG Youqiang, et al. A path planning strategy unified with a COLREGS collision avoidance function based on deep reinforcement learning and artificial potential field[J]. *Applied ocean research*, 2021, 113: 102759.

- [18] ZHANG Xinyu, WANG Chengbo, LIU Yuanchang, et al. Decision-making for the autonomous navigation of maritime autonomous surface ships based on scene division and deep reinforcement learning[J]. *Sensors*, 2019, 19(18): 4055.
- [19] ZHOU Xinyuan, WU Peng, ZHANG Haifeng, et al. Learn to navigate: cooperative path planning for unmanned surface vehicles using deep reinforcement learning[J]. *IEEE access*, 2019, 7: 165262–165278.
- [20] OZTURK U, AKDAG M, AYABAKAN T. A review of path planning algorithms in maritime autonomous surface ships: navigation safety perspective[J]. *Ocean engineering*, 2022, 251: 111010.
- [21] MCCUE L. Handbook of marine craft hydrodynamics and motion control[bookshelf[J]. *IEEE control systems magazine*, 2016, 36(1): 78–79.
- [22] SKJETNE R. The maneuvering problem[EB/OL]. (2005–03–15)[2024–01–03]. https://www.researchgate.net/publication/236651127_The_Maneuvering_Problem.
- [23] SONG Zhaofeng, ZHANG Jinfen, WU Da, et al. A novel path planning algorithm for ships in dynamic current environments[J]. *Ocean engineering*, 2023, 288: 116091.
- [24] GAO Ning, QIN Zhijin, JING Xiaojun, et al. Anti-intelligent UAV jamming strategy via deep Q-networks[J]. *IEEE transactions on communications*, 2020, 68(1): 569–581.
- [25] DEGRIS T, PILARSKI P M, SUTTON R S. Model-Free reinforcement learning with continuous action in practice[C]// 2012 American Control Conference. Montreal: IEEE, 2012: 2177–2182.
- [26] HIGO Y, SAKANO M, NOBE H, et al. Development of trajectory-tracking maneuvering system for automatic berthing/unberthing based on double deep Q-network and experimental validation with an actual large ferry[J]. *Ocean engineering*, 2023, 287: 115750.
- [27] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning[C]//Proceedings of the AAAI conference on artificial intelligence. Arizona: AAAI Press, 2016: 2094–2100.
- [28] ZHANG Fei, GU Chaochen, YANG Feng. An improved algorithm of robot path planning in complex environment based on double DQN[M]//Lecture Notes in Electrical Engineering. Singapore: Springer Singapore, 2021: 303–313.
- [29] YANG Xiaofei, SHI Yilun, LIU Wei, et al. Global path planning algorithm based on double DQN for multi-tasks amphibious unmanned surface vehicle[J]. *Ocean engineering*, 2022, 266: 112809.
- [30] ZHANG Jiaqi, JIAO Xiaohong, YANG Chao. A double-deep Q-network-based energy management strategy for hybrid electric vehicles under variable driving cycles[J]. *Energy technology*, 2021, 9(2): 2000770.

作者简介:



李康斌, 硕士研究生, 主要研究方向为船舶自动靠泊和路径规划。E-mail: 422698152@qq.com。



朱齐丹, 教授, 博士生导师, 黑龙江省自动化学会常务理事, 主要研究方向为智能机器人技术及应用、智能控制系统设计、图像处理与模式识别。主持国家自然科学基金项目、工信部高技术船舶专项项目、科技部国际合作项目等课题 30 项。研究成果获国家科技进步二等奖 1 项、国防科技进步一等奖 3 项、军队科技进步一等奖 1 项、黑龙江省科技进步二等奖 3 项, 发明专利授权 20 项、软件著作权授权 5 项, 出版专著 4 部, 发表学术论文 200 篇。E-mail: zhuqidan@hrbeu.edu.cn。



牟进友, 博士研究生, 主要研究方向为船舶自主靠泊和智能船舶。E-mail: mujinyou96@163.com。