doi:10.19562/j.chinasae.qcgc.2025.05.001

面向狭窄环境的安全泊车路径规划算法研究*

管家意[1],李 斌[1],周 傲[1],赵治国[1],林 巧[2],陈 广[1]
(1. 同济大学,上海 201804; 2. 易控智驾科技有限公司,北京 100083)

[摘要] 针对自动泊车系统中路径规划的安全性、实时性和可行性问题,本文提出一种基于混合动作空间约束强化学习的泊车路径规划算法。具体地,该算法利用混合动作空间强化学习框架将离散动作和连续参数相结合实现参数化轨迹规划,提高了规划路径的可执行性;在此基础上设计一种混合动作空间的约束强化学习算法实现安全策略优化,确保了泊车路径的安全性。此外,在模型训练过程中引入课程学习机制逐步引导策略探索,增强了模型训练稳定性和收敛速度。最后,在垂直车位和平行车位进行广泛的对比和消融实验,实验结果表明所提出的泊车路径规划算法在成功率、安全性和实时性等指标上均表现出色,且综合性能明显优于现有基线算法。

关键词:自动泊车;混合动作空间强化学习;路径规划;安全约束

Study on Safe Parking Path Planning Algorithm for Narrow Environment

Guan Jiayi[1], Li Bin[1], Zhou Ao[1], Zhao Zhiguo[1], Lin Qiao[2] & Chen Guang[1]
1. *Tongji University*, *Shanghai* 201804; 2. *EACON Technology Co.*, *Ltd.*, *Beijing* 100083

[**Abstract**] For safe and feasible path-planning in real time of autonomous parking system, a parking path planning algorithm based on constrained reinforcement learning with a hybrid action space is proposed in this paper. Specifically, the proposed algorithm employs a hybrid action space reinforcement learning framework that integrates discrete actions with continuous parameters to achieve parameterized trajectory planning, thereby enhancing the executability of planned paths. On this basis, a constrained reinforcement learning algorithm within the hybrid action space is designed to optimize safe policy execution, ensuring the safety of parking paths. Moreover, a curriculum learning mechanism is introduced during model training to guide exploration progressively, improving training stability and convergence speed. Finally, extensive comparative and ablation experiments are conducted on both perpendicular and parallel parking scenarios. The experimental results show that the proposed parking path planning algorithm outperforms existing state-of-the-art methods in terms of success rate, safety, and real-time performance, exhibiting superior overall effectiveness.

Keywords: autonomous parking; hybrid-action reinforcement learning; motion-planning; constraint optimization

前言

随着汽车保有量的逐年攀升,城市场景的车辆密度显著增加,狭窄拥挤的泊车环境加剧了泊车难度,这给新手驾驶员带来了泊车困扰[1-2]。自动泊车系统通过感知周围的环境并规划路径完成自主泊车任务,有望解决上述难题[3]。安全高效的泊车路径规划作为自动泊车系统的关键技术,得到国内外学者的广泛关注[4-5]。

* 国家重点研发计划项目(2024YFE0211000)、国家自然科学基金面上项目(62372329)、上海市科技创新行动计划社会发展科技攻关项目(23DZ1203400)、同济大学-Qomolo商用车自动驾驶联合实验室和小米青年学者基金资助。
原稿收到日期为 2025 年 01 月 14 日,修改稿收到日期为 2025 年 03 月 04 日。
通信作者:陈广,教授,博士,E-mail:guangchen@tongji.edu.cn。

经过多年的探索研究,自动泊车任务的路径规划已经取得阶段性成果。但是在狭窄和不确定的泊车环境中安全可行的路径规划仍然存在诸多挑战[6]。在解决泊车路径规划问题时,现有的方法主要包括:几何曲线法[7-8]、启发式搜索[9-10]和强化学习[11-14]等方法。几何曲线法依赖于先验知识和规则的设计,利用弧线和直线等几何曲线构造泊车路径[15],其中Reeds-shepp最早被用于解决泊车路径规划[7],随后大量的工作在此基础上引入几何约束和曲线拟合等策略提高算法规避周围障碍物的能力及路径的平滑性[16-18]。虽然这种几何曲线法在简单和规范的泊车场景具有不错的效果,但是其依赖于先验知识且在复杂拥挤的场景难以实现无碰撞的路径规划。

为提高泊车路径规划算法应对拥挤和复杂泊车场景的路径规划的安全性和鲁棒性,已开发启发式搜索法,由此通过结合车辆运动学特性和搜索策略,探索安全可行的路径,其中早期RRT算法通过结合车辆运动学特性,通过采样和代价计算,生成安全可行的路径[19-20],随后Hybrid A*及各种改进的启发式搜索算法被用于解决泊车场景的安全路径搜索[10,21]。这类方法虽然克服了对先验知识的依赖,并提高了路径规划策略对环境的鲁棒性,但是如何选择合适的分辨率以平衡规划能力和实时性阻碍了该类方法在现实场景中的应用,且在不确定性和复杂的场景下时间成本难以控制[9-10]。

另一方面,近年来强化学习在交互式决策规划方面表现突出[22-24]。不少研究尝试将强化学习与车辆运动学模型结合实现满足车辆运动学特性的泊车路径规划[25-27]。大量研究直接利用基于高斯分布的强化学习策略[28-30],实现安全泊车路径规划。此外,为提高交互式策略学习的训练效率,一些研究通过引入MCTS对离散最优动作的选择提高模型的训练效率[31-32]。虽然这些基于强化学习算法的路径规划策略提高了应对复杂和不确定环境的鲁棒性和实时性,但由于奖励的非单调性导致所规划的路径换挡过于频繁,且单纯的连续或离散动作所规划的路径可执行性较差,难以应用于实车场景[33-35]。

为解决上述狭窄、不确定性泊车场景所存在的安全性不足、可执行性差、鲁棒性低和实时性不足的路径规划问题,本文提出了一种混合动作约束强化学习(hybrid-action constrained reinforcement learning,HCRL)的泊车路径规划算法,实现了狭窄和不确定场景的泊车路径规划。本文的主要贡献如下。

首次引入混合动作空间强化学习将离散动作和连续参数相结合实现满足车辆动力学特性的参数化轨迹规划,提高了泊车路径规划的可执行性。

设计一种新颖的混合动作空间约束强化学习算法实现安全策略优化,确保了泊车策略所规划路径的安全性。

引入课程学习机制引导策略采样,提高了模型训练过程的稳定性和收敛速度。并且广泛的实验表明所提出的算法具有更高的安全性、可执行性和实时性。

1 理论基础及问题定义

1.1 泊车安全路径规划问题定义

自动泊车的安全路径规划问题旨在为车辆在复杂环境中找到一条安全、平滑且可执行的路径,使其能够从起始位置(x^0, y^0, δ^0)停靠至目标泊车位(x^t, y^t, δ^t)。该问题的核心在于确保车辆在有限的空间中规划一条满足车辆运动学特性的无碰撞的行驶路径。其优化目标的数学表征为

$$p_{x,y} = f(\boldsymbol{p}_0, \boldsymbol{p}_t, \boldsymbol{o}_c),\ \text{s.t.}\ p_{x,y} \in p_c \cap p_s \tag{1}$$

式中:$p_{x,y}$为安全泊车路径点集合;$\boldsymbol{p}_0$为起始坐标及航向;$\boldsymbol{p}_t$为目标库位的中心坐标及航向;$\boldsymbol{o}_c$为库位周围的车辆和障碍物信息;p_c为满足车辆运动学和动力学特性的路径点集合;p_s为无碰撞的路径点集合。(x^0, y^0)和(x^t, y^t)分别为车辆起始位置和目标位置的横纵坐标。δ^0和δ^t分别表示起始位置和目标位置的航向。从式(1)可知泊车安全路径规划问题涉及非线性约束优化,且具有较高的计算复杂度和环境不确定性。此外,在实际应用中规划的路径还须综合考虑路径的平滑性和泊车效率。

1.2 混合动作空间的强化学习

强化学习通常建模为马尔科夫决策过程[36](Markov decision process, MDP),用五元组$(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{P}, r, \gamma)$表示,其中$\boldsymbol{S}$是状态空间,$\boldsymbol{A}$是动作空间,$\boldsymbol{P}$: $\boldsymbol{S} \times \boldsymbol{A} \times \boldsymbol{S} \to [0, 1]$是环境的状态转移矩阵,$r$: $\boldsymbol{S} \times \boldsymbol{A} \to R$是奖励函数,$\gamma \in (0, 1]$是折扣系数。强化学习的目标是通过最大化累计奖励学习一个从状态$\boldsymbol{s}_t \in \boldsymbol{S}$到动作$\boldsymbol{a}_t \in \boldsymbol{A}$的最优概率分布$\boldsymbol{\pi}(\cdot|\boldsymbol{s}_t)$。其最大化累计奖励的数学表征为

$$\boldsymbol{\pi}^* = \text{argmax}_{\boldsymbol{\pi}} \mathbb{E}_{\tau \sim \boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right] \tag{2}$$

式中:$\boldsymbol{\tau} = \boldsymbol{s}_0, \boldsymbol{a}_0, \cdots \sim \boldsymbol{\pi}$表示动作状态对的轨迹;

$\boldsymbol{\pi}^* \in \boldsymbol{\pi}(\cdot|\boldsymbol{s}_t)$是最大化累计奖励的最优策略。

标准强化学习的动作空间都是同一性质的动作如全是离散或全是连续动作。然而,实际应用中,被控对象需要同时输出离散和连续动作[37]。为解决这类问题混合动作空间的强化学习算法被提出。混合动作空间的强化学习算法是在马尔科夫的框架下,将动作空间$\boldsymbol{A}$扩展为同时包含离散动作和连续参数的动作空间$\hat{\boldsymbol{A}}$。因此HRL动作空间数学表征为

$$\boldsymbol{a}_t = \left[a_d, a_p\right] \in \hat{\boldsymbol{A}} \tag{3}$$

式中:a_d是离散动作;a_p是连续参数。

2　安全泊车路径规划算法

为解决式(1)方程所示的泊车安全路径规划问题,本文提出一种基于混合动作空间约束强化学习的安全路径规划算法,其基于混合动作空间强化学习框架,构建符合车辆运动学特性的参数化轨迹规划模型,并设计带约束的策略优化算法实现安全泊车路径规划。

2.1　安全泊车路径规划

2.1.1　安全泊车路径规划模型

根据式(1)方程所示的泊车路径规划问题的数学表征可知,为实现安全路径规划不仅须满足初始位姿$\boldsymbol{p}_0$和目标位姿$\boldsymbol{p}_t$的需求,同时还须满足车辆运动学特性$\boldsymbol{p}_c$以及避开周围障碍物$\boldsymbol{o}_c$。因此本文中将观测状态定义为$\boldsymbol{s}_t = (\boldsymbol{p}_0, \boldsymbol{p}_t, \boldsymbol{o}_c)$,并将安全泊车轨迹表示成一组参数化动作$p_{x,y} = \{\boldsymbol{a}_0, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_t\}$。为寻找最优泊车路径,通过最大化泊车过程的累计奖励;此外为避免危险碰撞,通过将安全成本约束在安全阈值范围内。基于上述泊车路径规划的原始需求,本文将泊车的最优安全路径规划问题转化为混合动作空间约束强化学习的安全策略优化问题,泊车的安全路径规划问题建模为

$$\begin{cases} \boldsymbol{\pi}_\theta^* = \underset{\boldsymbol{\pi}_\theta}{\operatorname{argmax}}\, \mathbb{E}_{\tau\sim\boldsymbol{\pi}_\theta}\left[\sum_{t=0}^{\infty}\gamma^t r(\boldsymbol{s}_t, \boldsymbol{a}_t)\right] \\ \text{s.t. } \mathbb{E}_{\tau\sim\boldsymbol{\pi}_\theta}\left[\sum_{t=0}^{\infty}\gamma^t c(\boldsymbol{s}_t, \boldsymbol{a}_t)\right] \leqslant \bar{c} \end{cases} \tag{4}$$

式中:$\boldsymbol{\pi}_\theta^*(\cdot|\boldsymbol{s}_t)$是满足安全约束的最优路径规划策略;$c(\boldsymbol{s}_t, \boldsymbol{a}_t)$通常也简化为$c_t$是与奖励$r(\boldsymbol{s}_t, \boldsymbol{a}_t)$对应的在状态$\boldsymbol{s}_t$下执行动作$\boldsymbol{a}_t$产生的安全成本;$\bar{c}$是安全成本阈值;$\boldsymbol{a}_t$是所规划路径的参数化动作。混合动作$\boldsymbol{a}_t$如式(3)所示,包括离散动作$a_d$和连续参数$a_p$。其中混合动作的离散动作$a_d$和连续参数$a_p$如图1所示,离散动作$a_d$包括左转、直行和右转,连续参数$a_p$是车辆前进或后退的弧长,当$a_p \geqslant 0$时则表示车辆前进,反之则表示车辆后退。注意这里采用车辆的最小转弯半径作为转向的固定半径,因此给定起点和一组离散动作a_d与连续参数a_p之后车辆的运动路径是唯一确定的。

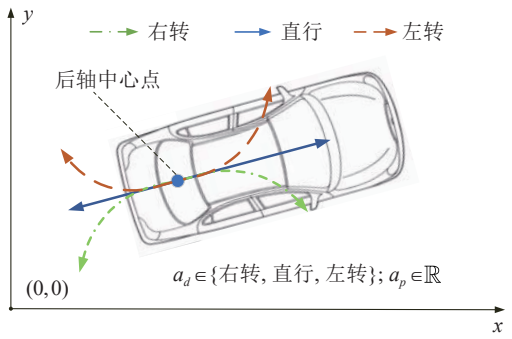


图1　泊车路径规划的参数化动作

2.1.2　基于混合动作的状态转移模型

低速工况下车辆的运动学方程能正确表征车辆位置的转换关系。基于上述的混合动作,构建以车辆后轴中心的位姿状态转移模型。按照离散动作分类,当车辆左转前进时即$a_d = 0$且$a_p \geqslant 0$时状态转移过程如图2所示。图中所示的变量(x_t, y_t)和(x_{t+1}, y_{t+1})分别表示车辆在t时刻和$t+1$时刻的坐标位置;δ_t和δ_{t+1}分别是车辆在t和$t+1$时刻的航向;β是车辆从t时刻和$t+1$时刻行驶圆弧所对应的弧度;R_0是车辆转弯半径;a_p和a_p'分别是车辆从t时刻到$t+1$时刻行驶的弧长和该弧线对应的弦长。根据图2所示的几何关系得到弦长a_p'与转弯半径R_0和转过弧度角β之间的关系数学表征:

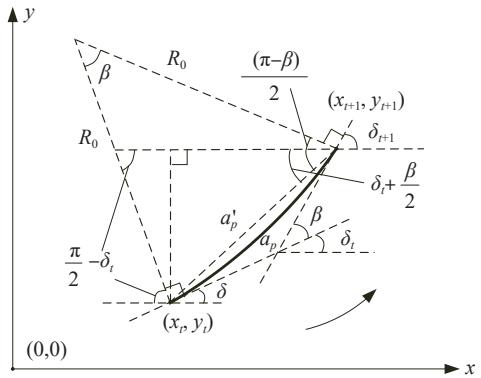


图2　车辆左转前进工况的状态转移示意图

$$a_p' = 2R_0 \sin\left(\frac{\beta}{2}\right) \tag{5}$$

基于上述几何关系和式(5)所示的数学关系，进一步得到车辆从t时刻到$t+1$时刻的状态转移方程：

$$\begin{cases} x_{t+1} = x_t + 2R_0 \sin\left(\dfrac{\beta}{2}\right)\cos\left(\delta_t + \dfrac{\beta}{2}\right) \\ y_{t+1} = y_t + 2R_0 \sin\left(\dfrac{\beta}{2}\right)\sin\left(\delta_t + \dfrac{\beta}{2}\right) \\ \delta_{t+1} = \delta_t + \beta \end{cases} \tag{6}$$

式中$\beta = a_p/R_0$是车辆左转的弧长对应转过的弧度。考虑到车辆左转有前进和后退两种工况，在2.1.1节已约定当车辆前进时弧长为正，后退时弧长为负，转过的弧度β和转过的弧长a_p同号，因此式(6)的状态转移关系仍适用于车辆左转后退工况的状态转移关系。

当车辆右转前进时a_d=2且$a_p \geqslant 0$时，车辆的状态转移过程如图3所示。参照左转工况的分析和推理，同理得到右转前进工况下车辆状态转移矩阵数学表征：

$$\begin{cases} x_{t+1} = x_t + 2R_0 \sin\left(\dfrac{\beta}{2}\right)\cos\left(\delta_t - \dfrac{\beta}{2}\right) \\ y_{t+1} = y_t + 2R_0 \sin\left(\dfrac{\beta}{2}\right)\sin\left(\delta_t - \dfrac{\beta}{2}\right) \\ \delta_{t+1} = \delta_t - \beta \end{cases} \tag{7}$$

同理因转过的弧度角β带矢量符号使式(7)同样适用于右转后退工况。此外，当车辆直行前进时即a_d=1且$a_p \geqslant 0$，其车辆的航向角不变即$\delta_{t+1} = \delta_t$，车辆坐标的变化通过车辆行驶的弧长向坐标轴上的投影来计算，因此车辆直行工况下其状态转移方程表征如下：

$$\begin{cases} x_{t+1} = x_t + a_p \cos(\delta_t) \\ y_{t+1} = y_t + a_p \sin(\delta_t) \\ \delta_{t+1} = \delta_t \end{cases} \tag{8}$$

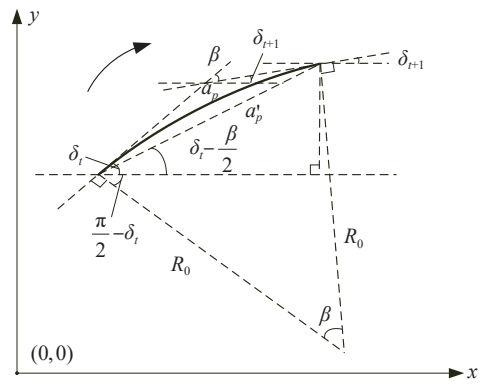


图3 车辆右转前进工况的状态转移示意图

同理直行路径长度a_p带矢量符号使式(8)同样适用于车辆直行后退工况。综上所述式(6)~式(8)为车辆后轴在执行参数化动作$\boldsymbol{a}_t$之后车辆后轴中心点的位姿状态转移方程。

2.2 混合动作空间的约束优化算法

根据式(4)所示的优化目标，泊车的安全路径规划问题转化成求解一个混合动作空间安全强化学习的策略优化问题。为解决该约束优化问题，本文提出一种混合动作空间的原策略优化算法。具体地，该算法包括两个步骤：约束评估和策略更新。其中，约束评估是评估当前策略是否满足安全约束，策略更新是根据约束评估的结果确定更新梯度的损失函数并更新策略。

2.2.1 约束评估

当前策略$\boldsymbol{\pi}_\theta$的平均安全成本与样本$\mathcal{D}$下所有轨迹的统计平均成本之间存在如下关系：$\mathbb{E}_{\tau\sim\boldsymbol{\pi}_\theta}\left[\sum_{t=0}^{\infty}\gamma^t c(\boldsymbol{s}_t,\boldsymbol{a}_t)\right] \approx \frac{1}{|\mathcal{D}|}\sum_{\tau\sim\mathcal{D}}\sum_{t=0}^{T}\gamma^t c_t$，因此在约束评估过程中直接采用当前策略$\boldsymbol{\pi}_\theta$采集的样本来估计成本，故式(4)中的约束条件改写成：

$$\mathcal{L}_c(\boldsymbol{\pi}_\theta) \approx \frac{1}{|\mathcal{D}|}\sum_{\tau\sim\mathcal{D}}\sum_{t=0}^{T}\gamma^t c_t \leqslant \bar{c} \tag{9}$$

式中：$\mathcal{D}$是$\boldsymbol{\pi}_{\theta_i}$采集的样本数据；$|\mathcal{D}|$表示样本集中包含的轨迹数量。

2.2.2 策略更新

在策略更新过程中，如果当前策略的成本满足成本约束$\mathcal{L}_c(\boldsymbol{\pi}_\theta) \leqslant \bar{c}$，则表明当前策略$\boldsymbol{\pi}_\theta$在安全解空间区域，此时只须通过最大化式(4)所示的累计奖励来获取最优回报的策略。受到近似梯度更新策略[38]的启发通过将式(4)所示的最大化累计奖励转化成为最大化当前策略优势函数，而混合动作空间包含离散动作和连续参数，因此最大化累计回报的损失函数定义为

$$\mathcal{L}(\theta) = \underset{\theta}{\operatorname{argmax}}\, E_{\boldsymbol{\pi}_{\theta_i}}\left[\boldsymbol{K}(\theta) A_r^{\boldsymbol{\pi}_{\theta_i}}\right] \tag{10}$$

式中$\boldsymbol{K}(\theta) = \boldsymbol{\pi}_\theta(\boldsymbol{a}|\boldsymbol{s})\left[\boldsymbol{\pi}_{\theta_i}(\boldsymbol{a}|\boldsymbol{s})\right]^{-1}$是当前策略$\boldsymbol{\pi}_\theta(\boldsymbol{a}|\boldsymbol{s})$与第$t$步策略$\boldsymbol{\pi}_{\theta_i}(\boldsymbol{a}|\boldsymbol{s})$的比值，奖励的优势函数包含离散动作和连续参数的优势函数即$A_r^{\boldsymbol{\pi}_\theta}(\boldsymbol{s},\boldsymbol{a}) = A_{r,d}^{\boldsymbol{\pi}_\theta}(\boldsymbol{s},\boldsymbol{a}) + A_{r,p}^{\boldsymbol{\pi}_\theta}(\boldsymbol{s},\boldsymbol{a})$，$A_{r,d}^{\boldsymbol{\pi}_\theta}(\boldsymbol{s},\boldsymbol{a})$和$A_{r,p}^{\boldsymbol{\pi}_\theta}(\boldsymbol{s},\boldsymbol{a})$分别表示离散动作和连续参数关于奖励的优势函数。

当式(9)所示的约束条件不满足成本约束时即$\mathcal{L}_c(\boldsymbol{\pi}_\theta) > \bar{c}$，则表明当前策略处于非安全的解空间区域，此时须通过最小化式(4)中的累计成本将策略重

新调整至安全解空间区域。同理利用优势函数最小化累计成本,因此最小化累计成本的损失函数定义为

$$\mathcal{L}(\theta) = \underset{\theta}{\operatorname{argmin}}\, E_{\boldsymbol{\pi}_{\theta_t}}\left[\boldsymbol{K}(\theta) A_c^{\boldsymbol{\pi}_{\theta_t}}\right] \tag{11}$$

式中 $A_c^{\boldsymbol{\pi}_{\theta_t}}(\boldsymbol{s},\boldsymbol{a}) = A_{c,d}^{\boldsymbol{\pi}_{\theta_t}}(\boldsymbol{s},\boldsymbol{a}) + A_{c,p}^{\boldsymbol{\pi}_{\theta_t}}(\boldsymbol{s},\boldsymbol{a})$, $A_{c,d}^{\boldsymbol{\pi}_{\theta_t}}(\boldsymbol{s},\boldsymbol{a})$ 和 $A_{c,p}^{\boldsymbol{\pi}_{\theta_t}}(\boldsymbol{s},\boldsymbol{a})$分别表示离散动作和连续参数关于成本的优势函数。为提高策略训练的稳定性,采用裁剪的方式约束策略更新的距离,因此式(10)和式(11)中策略系数$\boldsymbol{K}(\theta)$修正为$\hat{\boldsymbol{K}}(\theta)$:

$$\hat{\boldsymbol{K}}(\theta) = \min\left\{\boldsymbol{K}(\theta), clip\left(\boldsymbol{K}(\theta), 1-\epsilon, 1+\epsilon\right)\right\} \tag{12}$$

式中ε是裁剪系数。将式(12)所示的修剪之后的$\hat{\boldsymbol{K}}(\theta)$代入式(10)和式(11)所示的损失函数中即为本混合动作空间的原约束优化算法的策略网络损失函数。

2.3　泊车安全路径规划实例

为便于理解所提出的算法,本文提供一个详细的算法实例。考虑到本文的混合动作空间的策略网络,须同时输出离散动作及动作的连续参数(弧长)。策略网络采用两个输出头,其中一个为离散动作的概率,另一个为对应动作的连续参数,详细的算法实例的伪代码如表1所示。其中θ_0、ϕ_0和φ_0分别表示初始策略网络、奖励值和成本值网络的网络参数,θ_t、ϕ_{t+1}和φ_{t+1}分别表示更新到第t步的策略网络、奖励值和成本值网络的网络参数。此外,为详细介绍该算法在泊车路径规划中的应用过程,本节详细介绍约束评估、课程学习机制及环境的奖励和成本函数的设计。

2.3.1　约束评估

在成本约束评估时考虑到当$\gamma \in (0,1]$时样本数据的带折扣累计成本满足以下条件:

$$\frac{1}{|\mathcal{D}|}\sum_{\tau\sim\mathcal{D}}\sum_{t=0}^{T}\gamma^t c_t \leqslant \frac{1}{|\mathcal{D}|}\sum_{\tau\sim\mathcal{D}}\sum_{t=0}^{T} c_t \tag{13}$$

式中T表示轨迹的长度。基于式(13)和式(9)所示的数学关系,本实例在约束评估过程中直接采用如式(14)所示的非折扣累计成本判定策略是否满足约束:

$$\frac{1}{|\mathcal{D}|}\sum_{\tau\sim\mathcal{D}}\sum_{t=0}^{T} c_t \leqslant \bar{c} \tag{14}$$

2.3.2　奖励及成本函数

为引导车辆从任意符合交通规则的位姿泊入目标车位,本文为泊车场景设计了引导车辆泊入目标车位的奖励函数及评估安全的成本函数。考虑到车辆成功泊车车位须同时考虑车辆相对于泊车位的位置偏差以及航向偏差,因此在泊车场景中本文的奖励函数定义为

$$r = k_{\mathrm{s}}\left(1 - \frac{d_{\mathrm{c}}}{d_{\mathrm{m}}}\right) + k_{\mathrm{r}}\left(1 - \frac{|\delta_{\mathrm{c}}|}{\pi}\right) + k_{\mathrm{w}} r_{\mathrm{w}} \tag{15}$$

式中:k_{s}、k_{r}和k_{w}分别表示距离偏差、航向偏差和完成泊车等奖励的加权系数;d_{c}表示车辆当前位置相对于目标位置的距离;d_{m}表示该泊车场景初始位置相对于目标位置的最大距离;δ_{c}表示车辆当前的航向角;r_{w}表示达到目标车位的离散奖励。

为避免规划的泊车路径与停泊的车辆及周围的障碍物发生危险碰撞,针对安全约束设计成本函数,并基于2.2节中设计的约束强化学习算法优化安全策略。考虑安全的泊车路径不能与周围停泊的车辆或者障碍物发生碰撞。另外,车辆控制存在误差,因此所规划的路径不仅不能与周围停泊的车辆发生碰撞,还要与周围车辆及障碍物保持一定的安全距离。基于上述考虑,将泊车场景的成本函数定义为

表1　HCRL安全泊车路径规划算法

算法1:混合动作安全泊车路径规划算法

1:输入:初始化策略网络参数θ_0;初始化奖励值函数网络参数ϕ_0;初始化成本值函数网络参数φ_0

2:for 每个回合 do:

3:通过策略$\boldsymbol{\pi}_{\theta_t}$在环境中采集多组轨迹$\mathcal{D}_t = \{(\boldsymbol{s}_t, \boldsymbol{a}_t, r_t, c_t)_i\}$

4:计算后续轨迹奖励和成本$\hat{R}_t$、$\hat{C}_t$

5:基于当前奖励值网络$V_{\phi_t}^r$计算奖励的优势函数$A_{r,d}^{\boldsymbol{\pi}_{\theta_t}}$和$A_{r,p}^{\boldsymbol{\pi}_{\theta_t}}$

6:基于当前成本值网络$V_{\varphi_t}^c$计算成本的优势函数$A_{c,d}^{\boldsymbol{\pi}_{\theta_t}}$和$A_{c,p}^{\boldsymbol{\pi}_{\theta_t}}$

7:基于式(7)估算轨迹成本值$\mathcal{L}_{\mathrm{c}}(\boldsymbol{\pi}_{\theta_t})$

8:如果$\mathcal{L}_{\mathrm{c}}(\boldsymbol{\pi}_{\theta_t}) \leqslant \bar{c}$条件满足则通过奖励优势函数更新策略:

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmax}}\, E_{\boldsymbol{\pi}_{\theta_t}}\left[\hat{\boldsymbol{K}}(\theta)(A_{r,d}^{\boldsymbol{\pi}_{\theta_t}} + A_{r,p}^{\boldsymbol{\pi}_{\theta_t}})\right]$$

9:反之则通过成本优势函数更新策略:

$$\theta_{t+1} = \underset{\theta}{\operatorname{argmin}}\, E_{\boldsymbol{\pi}_{\theta_t}}\left[\hat{\boldsymbol{K}}(\theta)(A_{c,d}^{\boldsymbol{\pi}_{\theta_t}} + A_{c,p}^{\boldsymbol{\pi}_{\theta_t}})\right]$$

10:通过均方误差拟合奖励和成本值函数:

$$\phi_{t+1} = \operatorname{argmin}_{\phi}\frac{1}{|\mathcal{D}_t|T}\sum_{\tau\sim\mathcal{D}_t}\sum_{t=0}^{T}\left(V_{\phi}^r(\boldsymbol{s}_t) - \hat{R}_t\right)^2$$

$$\varphi_{t+1} = \underset{\varphi}{\operatorname{argmin}}\frac{1}{|\mathcal{D}_t|T}\sum_{\tau\sim\mathcal{D}_t}\sum_{t=0}^{T}(V_{\varphi}^c(\boldsymbol{s}_t) - \hat{C}_t)^2$$

11:终止 for

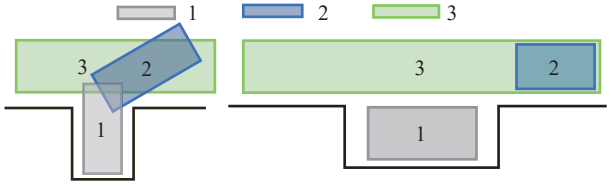
12:输出:策略网络$\boldsymbol{\pi}_{\theta}$

$$c = k_c c_c + k_d\left[I_{(\bar{d}_c^c - d_c^c)\geqslant 0}\cdot\frac{d_c^c}{\bar{d}_c^c}\right] \tag{16}$$

式中：k_c和k_d分别表示危险碰撞和危险距离的成本系数；c_c是产生碰撞的固定成本；d_c^c表示车辆相对于周围停泊的车辆或障碍物的距离；$\bar{d}_c^c$表示危险距离的阈值；$I_{(\bar{d}_c^c - d_c^c)\geqslant 0}$是符号函数，即当满足条件$(\bar{d}_c^c - d_c^c)\geqslant 0$时为1，否则为0。

2.3.3 课程学习引导

由于须避开周围停泊车辆及障碍物，导致可行解空间范围显著减小，同时奖励函数中的距离奖励与航向奖励之间存在一定冲突。这种特性增加了策略探索最优解的难度，从而对策略训练的稳定性和收敛速度产生不利影响。为解决该问题，本文在策略训练过程中引入课程学习机制，通过逐步增加泊车难度引导策略更新。如图4所示，为垂直和平行泊车的引导区域，其中区域对应的编号由1～3泊车难度逐步增加。在训练过程车辆从1～3区域生成随机位姿和航向，待模型训练完成之后，在测试过程中只从3区域产生随机位姿和航向的待泊车车辆。



(a) 垂直泊车场景 (b) 平行泊车场景

图4 训练过程中课程引导区域示意图

3 安全泊车路径规划实验验证

为验证泊车路径规划算法的有效性，利用Python和Gym开发了一个符合车辆运动学特性的仿真环境。并基于该仿真环境对所提出的路径规划算法进行了广泛的对比和消融实验。

3.1 泊车的仿真环境开发

考虑到车辆的运动学特性能准确表征车辆在泊车工况下路径规划过程中的状态转移特性。因此本文直接以参数化动作$\boldsymbol{a}_t = \left[a_d, a_p\right]$为车辆运动学模型的输入，车辆后轴中心位姿的状态转移关系如式(6)～式(8)所示。基于上述的状态转移方程，在Gym环境开发了一个泊车的仿真环境。该仿真环境参考实际的停车位尺寸和环境设计垂直和平行泊车位，并模拟8线雷达探测周围的环境。此外，仿真环境引入大量的随机值和噪声模拟泊车环境的不确定性。

图5和图6所示分别为垂直和平行泊车位仿真场景。图中所示的红色方框是已泊入车位的车辆，没有被红色方框覆盖的车位为可泊的车位。蓝色的射线是模拟的雷达，当射线接触到障碍物或者停车场边界墙则返回，蓝线的长度等价于车辆与障碍物的距离，绿色箭头表示车辆的航向，绿色方框表示待泊车的车辆。在仿真过程中车辆位置在车道和泊车位组成的连通区域随机生成，并且车辆的航向也引入随机值。此外，已停泊的车辆位姿也存在位置和航向的随机偏差。

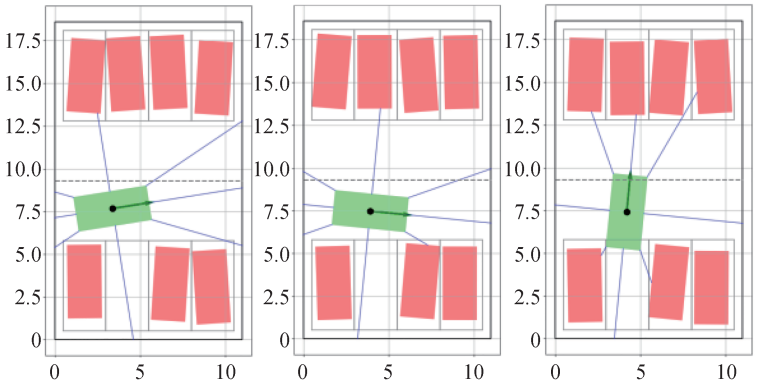


图5 垂直泊车位仿真环境可视化

3.2 实验的基线算法及评价指标

为客观评估HCRL路径规划算法的性能，本文在相同条件下与现有的泊车路径规划算法在多个客观评价指标下进行综合评估。

(1)基线算法：本文选用了几何曲线、Hybrid A*以及连续动作空间的强化学习算法PPO、SAC[39]。

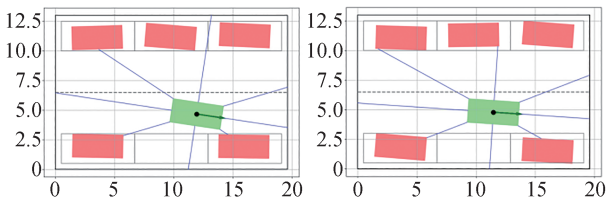


图6　平行泊车位仿真环境可视化

此外,本文中还将多个混合动作空间强化学习算法PADDPG、PDQN[40]、HPPO[41]等方法通过引入拉格朗日乘子改进为带约束优化的混合动作空间强化学习算法PADDPG-Lag、HPPO-Lag、PDQN-Lag-R。其中PADDPG-Lag和HPPO-Lag是在PADDPG和HPPO算法的基础上通过拉格朗日乘数法将泊车的带约束优化问题转化为无约束优化问题,PDQN-Lag-R算法是在PDQN的基础上直接通过拉格朗日乘子将奖励和成本组合成一个合成的奖励,并最大化合成之后的奖励来获取安全策略。

(2)评价指标:为客观评估泊车路径规划算法的安全性、实时性、易操控性,本文引入成功率S_r、成本C_s、耗时T_c、路径长度L_p、换挡次数N_g和曲率变化次数N_c。其中成功率S_r是重复测试多次,成功次数占总测试次数的百分比;成本C_s是泊车路径所产生的碰撞和危险距离成本;耗时T_c是完成一次从随机起点到目标库位的路径规划所消耗的时间;路径长度L_p是泊车路径的长度;换挡次数N_g是泊车路径所需的换挡次数;路径曲率变化次数N_c是所规划的路径矢量曲率变化次数。此外,为综合评估算法的性能,将评价指标加权得到一个综合评价指标S_o,该综合指标表征为

$$S_o = \sum_i^n k_i s_i \tag{17}$$

式中:s_i表示第i个评价指标的归一化评分;k_i为第i个评价指标的加权系数。利用线性归一化将所有指标的实际数值归一化到[0, 1]之间。指标归一化的数学表征为

$$s_i = \begin{cases} 1 - \dfrac{m_{i,\max} - m_i}{m_{i,\max} - m_{i,\min}}, & m_i \uparrow \\ \dfrac{m_{i,\max} - m_i}{m_{i,\max} - m_{i,\min}}, & m_i \downarrow \end{cases} \tag{18}$$

式中:$m_{i,\max}$和$m_{i,\min}$表示所有算法在第i评价指标下实际值的最大值和最小值;m_i是第i评价指标需要归一化的原始数值;$m_i \uparrow$表示该评价指标越大越好,$m_i \downarrow$表示该评价指标越小越好。

3.3　泊车的仿真实验

为验证所提出的安全泊车路径规划算法的效果,在垂直和平行车位进行了广泛的对比实验。此外,为保证实验评估的公平性,HCRL算法和基线算法的最小转弯半径均设置为5 m,这也符合车辆的运动学特性。

3.3.1　对比实验

常用的泊车路径规划算法与HCRL算法在垂直车位和平行车位泊车场景的实验结果如表2和表3所示。表中所示的结果为所有算法在3个不同随机种子下测试1 000次泊车规划所记录的均值和标准差。从表2中的结果可以看出,HCRL算法在保证路径规划成功率S_r的基础上,成本约束C_s和耗时T_c均明显低于其他算法,表明HCRL算法相比于其他算法具有更高的安全性和实时性。此外,路径长度L_p、换挡次数N_g和路径曲率变化次数N_c虽然略高于Reeds-shepp和Hybrid A*算法,但是非常接近最优值。并且综合评分S_o显著高于其他基线算法。表明HCRL算法在垂直泊车位场景下提供了具有竞争力的性能,并且具有更高的安全性和实时性。

表2　现有基线算法与HCRL算法在垂直车泊车位实验结果

方法	Reeds-Shepp	Hybrid A*	SAC	PPO	HCRL(Ours)
S_r/%↑	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
C_s↓	52.98±26.78	8.75±3.32	26.00±12.72	27.94±13.55	8.67±3.89
T_c/s↓	0.01±0.00	2.74±3.86	0.50±0.03	0.56±0.03	0.07±0.01
L_p/m↓	8.91±0.36	9.68±1.16	14.73±2.07	15.98±2.21	9.78±0.86
N_g↓	1.00±0.00	1.33±0.58	5.84±2.30	6.16±2.77	1.66±0.71
N_c↓	2.87±0.51	3.68±1.03	14.18±3.04	13.37±3.19	3.82±1.50
S_o↑	83.33	79.23	44.47	40.58	94.05

从表3所示的结果可以看出,HCRL算法在平行泊车位场景中,在安全成本C_s和耗时T_c等关键指标明显优于其他基线算法,并且HCRL的综合性能显著高于其他基线算法。表明HCRL算法在平行泊车

表3 现有基线算法与HCRL算法在平行车泊车位实验结果

方法	Reeds-Shepp	Hybrid A*	SAC	PPO	HCRL(Ours)
S_r/%↑	1.00±0.00	0.83±0.12	1.00±0.00	1.00±0.00	1.00±0.00
C_s↓	65.49±29.72	10.70±4.66	16.22±7.27	18.02±7.82	7.92±3.76
T_c/s↓	0.01±0.00	23.93±7.20	0.31±0.06	0.32±0.08	0.06±0.02
L_p/m↓	8.80±0.34	9.34±0.84	11.02±1.65	11.64±1.81	9.96±0.92
N_g↓	2.00±0.00	2.26±1.14	2.60±0.34	2.72±1.16	2.31±1.01
N_c↓	3.82±1.38	4.51±1.94	13.29±2.53	12.56±2.25	4.69±1.22
S_o↑	83.33	55.46	53.80	48.14	84.45

场景下具有更高的综合性能,并且在安全性和实时性表现优异。

常用的泊车路径规划算法与HCRL算法在垂直车位和平行车位泊车场景的实验结果的归一化得分如图7所示。从图中可以看出,HCRL算法在关键性指标成功率S_r、成本C_s、耗时T_c等方面相比于其他基线算法具有显著的优势,并且在垂直车位泊车场景,路径长度L_p、换挡次数N_g和路径曲率变化次数N_c等指标也非常接近最优值,虽然在平行车位泊车场景中路径长度L_p、换挡次数N_g和路径曲率变化次数N_c等指标略低,但是也提供了具有竞争力的性能。此外,从图示结果可以看出,HCRL算法各项指标相比于其他基线算法相对均衡,没有非常明显的弱项,这也表明HCRL算法相比于其他基线算法具有更加均衡的性能表现。

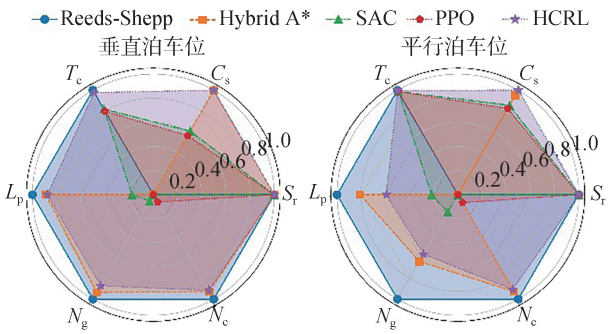


图7 垂直和平行位泊车场景评价指标归一化得分

混合空间的安全泊车路径规划算法与HCRL算法在垂直车位和平行车位的泊车路径规划的实验效果如表4和表5所示。其中,成本阈值统一设置为10。从表中结果可以看出,HCRL算法在成功率S_r和成本C_s等关键指标均优于其他带约束的混合动作空间的安全强化学习算法。此外,耗时T_c和路径长度L_p也相比于其他的算法具有明显的优势。表明HCRL算法相比于其他改进的混合动作空间安全强化学习的泊车路径规划算法具有更好的安全性、实时性和可执行性。并且在综合评分S_o上明显优于其它混合动作空间的安全强化学习算法。因此表明HCRL算法在垂直和平行泊车场景下相比于其他混合动作空间的安全强化学习算法具有更佳的综合性能。

表4 基于现有算法改进的安全规划算法与HCRL算法在垂直车泊车位实验结果

方法	PADDPG-Lag	HPPO-Lag	PDQN-Lag-R	HCRL(Ours)
S_r/%↑	0.98±0.01	0.99±0.01	1.00±0.00	1.00±0.00
C_s↓	11.33±5.10	10.04±4.86	12.58±5.77	8.67±3.89
T_c/s↓	0.10±0.04	0.09±0.05	0.11±0.05	0.07±0.01
L_p/m↓	10.30±1.05	12.67±1.49	11.08±0.88	9.78±0.86
N_g↓	1.56±0.53	1.74±0.89	1.67±0.70	1.66±0.71
N_c↓	3.61±1.36	4.14±1.86	3.95±1.64	3.82±1.50
S_o↑	75.61	77.54	89.05	94.05

表5 基于现有算法改进的安全规划算法与HCRL算法在平行车泊车位实验结果

方法	PADDPG-Lag	HPPO-Lag	PDQN-Lag-R	HCRL(Ours)
S_r/%↑	0.99±0.01	1.00±0.00	0.97±0.01	1.00±0.00
C_s↓	12.09±4.65	10.31±4.30	13.17±5.83	7.92±3.76
T_c/s↓	0.07±0.02	0.06±0.01	0.07±0.03	0.06±0.02
L_p/m↓	10.62±1.69	10.03±1.37	11.87±1.82	9.96±1.14
N_g↓	2.67±1.68	2.43±1.23	2.91±1.75	2.31±1.01
N_c↓	6.29±4.26	5.03±0.96	6.67±6.95	4.69±1.22
S_o↑	67.23	79.97	57.15	84.45

3.3.2 实验结果可视化

为进一步展示HCRL与基线算法的路径规划效果,本文中可视化了基线算法和HCRL算法在随机的泊车场景下所规划的泊车路径。垂直车位和平行车位泊车场景下不同算法规划的泊车路径可视化效果如图8和图9所示。从图8可以看出,Reeds-Shepp所规划的路径与周围停泊的车辆存在明显的重叠,因此所规划的路径产生较大的安全成本,表明

所规划的路径与停泊的车辆存在碰撞。Hybrid A*算法所规划的路径避开了周围停泊的车辆,但是所消耗的时间明显高于其他算法。基于SAC和PPO算法在连续动作空间所规划的路径虽然经过多次航向调整,但是仍然与周围停泊的车辆发生碰撞。HCRL算法所规划的路径通过换挡调节方向不仅保证所规划的路径完美避开周围停泊的车辆,而且用相对较短的时间完成整条路径的规划。此外,从图9所示的平行泊车场景的路径规划结果可以看出,HCRL算法仍然以较低的时间消耗,实现了安全路径的规划。上述结果表明,HCRL泊车算法能实时高效地为不同泊车场景规划出安全的泊车路径。

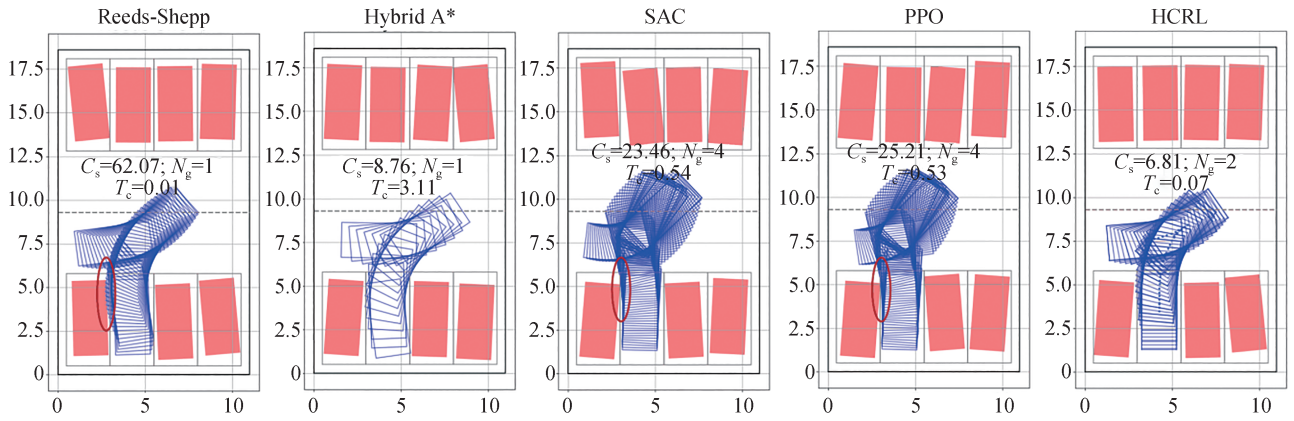


图8　常用泊车路径规划算法与HCRL在垂直泊车位可视化效果

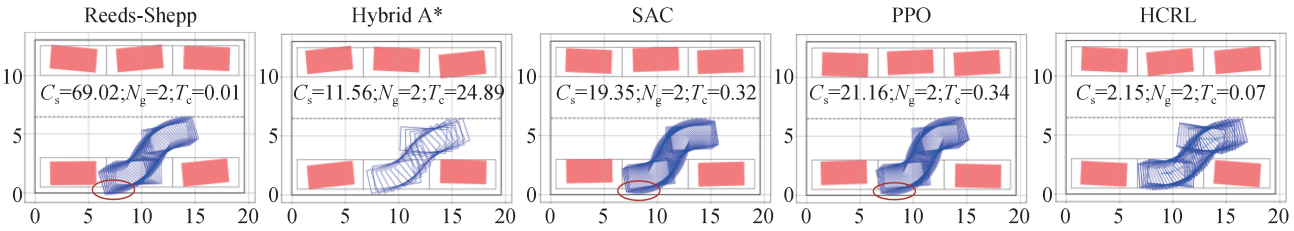


图9　常用泊车路径规划算法与HCRL算法在平行泊车位可视化效果

为验证HCRL算法在狭窄场景中的路径规划效果,进一步测试了存在障碍物和停车场靠墙车位的路径规划效果。狭窄泊车位场景的泊车路径如图10所示。从图中可以看出,HCRL算法在狭窄的泊车场景中也能高效地探索到安全的可执行泊车路径。表明HCRL算法在狭窄车位,仍然能完成安全泊车路径规划。

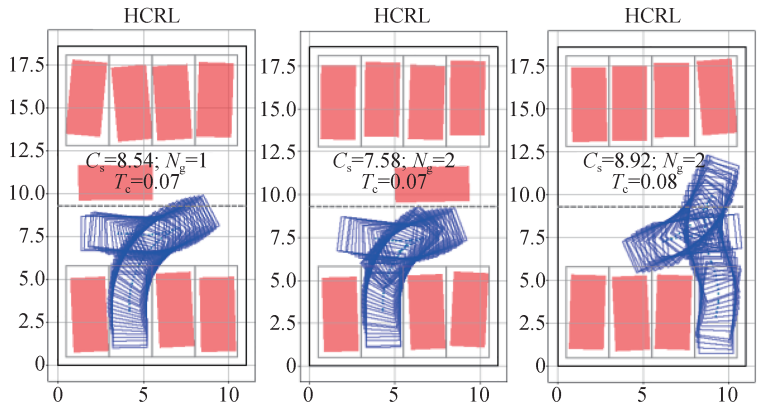


图10　HCRL算法在狭窄的垂直泊车位路径可视化效果

3.3.3 消融实验

为进一步验证HCRL算法中各模块的对算法的贡献,本文对成本约束和课程学习进行了消融实验。实验结果如图11和图12所示。此外,为方便标注图例,采用HCRL_curr表示不用采用课程引导的HCRL算法,HCRL_cost表示不进行约束优化的HCRL算法。图中的训练曲线是多个随机种子在多次训练过程中记录的均值和方差。从图11可以看出,不考虑成本约束的HCRL算法和HCRL算法的成功率最终均能达到100%,但是HCRL算法将成本约束在成本阈值以下,而不考虑成本的HCRL算法收敛之后成本远高于成本阈值,表明HCRL算法通过成本约束,能保证路径规划策略在垂直泊车场景满足安全约束。此外,图中还展示了不采用课程学习的HCRL算法在不同随机种子下规划的成功率和安全成本方差较大,表明没有课程引导的HCRL算法容易陷入局部最优,甚至在部分随机种子下难以规划出可行路径。

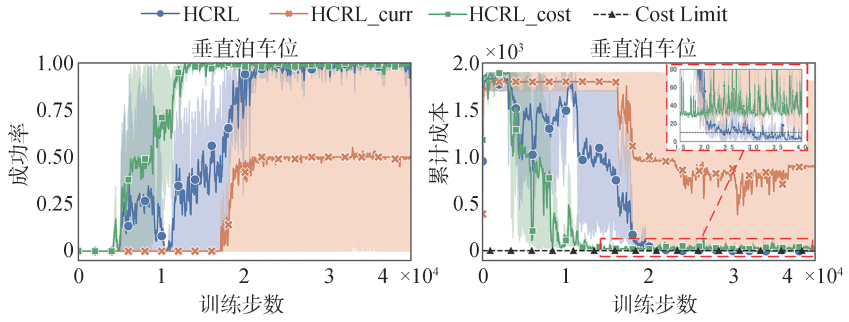


图11 HCRL在垂直泊车场景消融实验的测试曲线

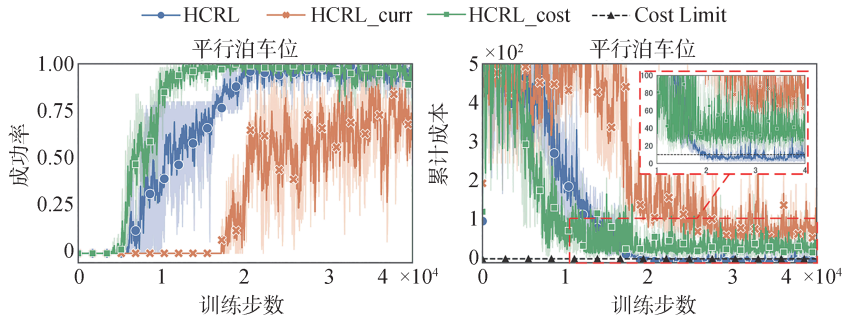


图12 HCRL在平行泊车场景消融实验的测试曲线

如图12所示的HCRL算法在平行泊车场景的消融实验的测试曲线同样可以看出,虽然引入成本约束的HCRL算法减缓了泊车策略训练过程的收敛速率,但是通过成本约束保证了规划路径的安全性。此外,不利用课程学习引导的HCRL算法收敛速度明显减慢,甚至在本文设置的训练周期内还未收敛,这也表明没有课程引导的HCRL算法探索到有效解的概率明显降低。

综合实验结果表明,安全约束策略和课程学习在HCRL算法中均发挥着关键作用。其中课程学习有效提高了训练效率和算法收敛的稳定性,而安全约束策略显著提升了算法所规划路径的安全性。两者的结合使得HCRL算法在泊车任务中能够以更高的成功率和更低的安全成本完成规划目标,验证了所设计算法的合理性和各策略模块的重要性。

4 结论

本文中提出一种HCRL的自动泊车路径规划算法,针对泊车场景路径规划的安全性、可执行性和实时性等问题进行了深入研究。基于混合动作空间强化学习框架将离散动作和连续参数相结合,实现了满足车辆运动学特性的参数化轨迹规划,提高了所规划路径的可执行性。并设计了一种混合动作空间的约束强化学习算法,实现混合动作空间的安全策略优化,确保所规划路径的安全性。此外,引入课程学习机制逐步引导策略探索和学习,提升了算法的训练效率和收敛速度。仿真实验结果表明,HCRL

算法在成功率、安全成本、实时性等关键指标均表现出优异性能,并且综合性能显著优于现有的基线算法。此外消融实验验证了安全约束策略和课程学习对路径规划的安全性和算法的稳定性具有显著的提升作用。综合实验结果表明HCRL算法能够在狭窄的泊车任务中高效地规划出安全、平滑、可行的泊车路径,具有较强的鲁棒性和适应性,为端到端泊车系统的设计和应用提供了理论支持和技术保障。

参考文献

[1] ZHAO C, LIAO F, LI X, et al. Macroscopic modeling and dynamic control of on-street cruising-for-parking of autonomous vehicles in a multi-region urban road network [J]. Transportation Research Part C: Emerging Technologies, 2021,128: 103176.

[2] BOCK F, DI MARTINO S, ORIGLIA A. Smart parking: using a crowd of taxis to sense on-street parking space availability [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(2): 496-508.

[3] XIE M, ZHANG X, WU Z, et al. A shared parking optimization framework based on dynamic resource allocation and path planning[J]. Physica A: Statistical Mechanics and Its Applications, 2023, 616:128649.

[4] KHALID M, WANG K, ASLAM N, et al. From smart parking towards autonomous valet parking: a survey, challenges and future works[J]. Journal of Network and Computer Applications, 2021, 175: 102935.

[5] 胡杰, 朱令磊, 陈瑞楠, 等. 狭小车位平行泊车路径规划方法研究[J]. 汽车工程, 2022, 44(7): 1040-1048.
HU J, ZHU L, CHEN R, et al. Research on path planning method for parallel parking in narrow spaces [J]. Automotive Engineering, 2022, 44(7): 1040-1048.

[6] QIN Z, CHEN X, HU M, et al. A novel path planning methodology for automated valet parking based on directional graph search and geometry curve [J]. Robotics and Autonomous Systems, 2020, 132:103606.

[7] REEDS J, LAWRENCE S. Optimal paths for a car that goes both forwards and backwards [J]. Pacific Journal of Mathematics, 1990,145(2): 367-393.

[8] 胡杰, 张敏超, 徐文才, 等. 自动驾驶车辆的平行泊车轨迹规划[J]. 汽车工程, 2022, 44 (3): 330-339.
HU J, ZHANG M, XU W, et al. Trajectory planning for parallel parking of autonomous vehicles [J]. Automotive Engineering, 2022, 44(3): 330-339.

[9] CHEN G, HOU J, DONG J, et al. Multi-objective scheduling strategy with genetic algorithm and time-enhanced A* planning for autonomous parking robotics in high-density unmanned parking lots [J]. IEEE/ASME Transactions on Mechatronics, 2020, 26(3): 1547-1557.

[10] SHENG W, LI B, ZHONG X. Autonomous parking trajectory planning with tiny passages: a combination of multistage hybrid A-star algorithm and numerical optimal control [J]. IEEE Access, 2021,9:102801-102810.

[11] DU Z, MIAO Q, ZONG C. Trajectory planning for automated parking systems using deep reinforcement learning [J]. International Journal of Automotive Technology, 2020, 21(4): 881-887.

[12] ZHANG P, XIONG L, YU Z, et al. Reinforcement learning-based end-to-end parking for automatic parking system[J]. Sensors. 2019, 19(18): 3996.

[13] SONG S, CHEN H, SUN H, et al. Data efficient reinforcement learning for integrated lateral planning and control in automated parking system[J]. Sensors, 2020, 20(24):7297.

[14] CHAN K H, MUSTAPHA A, JUBAIR M A. Comparative analysis of loss functions in TD3 for autonomous parking [J]. Journal of Soft Computing and Data Mining, 2024,5(1):1-4.

[15] 高强, 陆洲, 段晨东, 等. 汽车垂直泊车路径规划与路径跟踪研究[J]. 汽车工程, 2021, 43 (7): 987-994,1012.
GAO Q, LU Z, DUAN C, et al. Research on path planning and path tracking for vertical parking of vehicles [J]. Automotive Engineering, 2021, 43(7): 987-994,1012.

[16] LI B, WANG K, SHAO Z. Time-optimal maneuver planning in automatic parallel parking using a simultaneous dynamic optimization approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(11): 3263-3274.

[17] 胡文, 谭运生, 康龙云, 等. 基于驾驶员经验的自动泊车规划算法研究 [J]. 汽车工程, 2019, 41 (12): 1394-1400,1415.
HU W, TAN Y S, KANG L Y, et al. Study on automatic parking planning algorithm based on driver's experience [J]. Automotive Engineering, 2019, 41(12): 1394-1400,1415.

[18] 张家旭,王晨,赵健. 面向狭小平行泊车位的路径规划与跟踪控制[J]. 吉林大学学报(工学版), 2021, 51(5): 1879-1886.
ZHANG J, WANG C, ZHAO J. Path planning and tracking control for narrow parallel parking spaces [J]. Journal of Jilin University (Engineering and Technology Edition), 2021, 51(5): 1879-1886.

[19] DOLGOV D, THRUN S, MONTEMERLO M, et al. Path planning for autonomous vehicles in unknown semi-structured environm-ents [J]. The International Journal of Robotics Research, 2010, 29(5):485-501.

[20] JHANG J H, LIAN F L. An autonomous parking system of optimally integrating bidirectional rapidly-exploring random trees and parking-oriented model predictive control [J]. IEEE Access, 2020, 8: 163502-163523.

[21] KARAMAN S, FRAZZOLI E. Sampling-based algorithms for optimal motion planning[J]. The International Journal of Robotics Research, 2011,30(7): 846-894.

[22] GUAN J, SHEN L, ZHOU A, et al. POCE: primal policy optimization with conservative estimation for multi-constraint offline reinforcement learning [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 26243-26253.

[23] LIU H X, FENG S. Curse of rarity for autonomous vehicles[J]. Nature Communications, 2024,15(1):4808.

[24] TAKEHARA R, GONSALVES T. Autonomous car parking system using deep reinforcement learning [C]. 2021 2nd International Conference on Innovative and Creative Information Technology (ICITech). IEEE, 2021: 85-89.

[25] SHI J, LI K, PIAO C, et al. Model-based predictive control and reinforcement learning for planning vehicle-parking trajectories for vertical parking spaces[J]. Sensors, 2023, 23 (16): 7124.

[26] WU Y, WANG L, LU X, et al. Reinforcement learning-based autonomous parking with expert demonstrations [C]. 2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI). IEEE, 2023: 1-6.

[27] DEN HENGST F, FRANÇOIS-LAVET V, HOOGENDOORN M, et al. Planning for potential: efficient safe reinforcement learning[J]. Machine Learning, 2022, 111(6):2255-2274.

[28] TIONG T, SAAD I, TEO K T, et al. Autonomous valet parking with asynchronous advantage actor-critic proximal policy optimization [C]. 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2022: 334-340.

[29] JUNZUO L, QIANG L. An automatic parking model based on deep reinforcement learning [J]. In Journal of Physics: Conference Series, 2021,1883(1).

[30] TANG X, YANG Y, LIU T, et al. Path planning and tracking control for parking via soft actor-critic under non-ideal scenarios [J]. IEEE/CAA Journal of Automatica Sinica, 2023, 11 (1): 181-195.

[31] WÖHLKE J, SCHMITT F, VAN HOOF H. Hierarchies of planning and reinforcement learning for robot navigation [C]. 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 10682-10688.

[32] YUAN Z, WANG Z, LI X, et al. Hierarchical trajectory planning for narrow-space automated parking with deep reinforcement learning: a federated learning scheme [J]. Sensors, 2023, 23 (8):4087.

[33] CAI L, GUAN H, ZHOU Z Y, et al. Parking planning under limited parking corridor space [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(2):1962-1981.

[34] LIU W, LI Z, LI L, et al. Parking like a human: a direct trajectory planning solution [J]. IEEE Transactions on Intelligent Transpor-tation Systems, 2017,18(12):3388-3397.

[35] SOUSA B, RIBEIRO T, COELHO J, et al. Parallel, angular and perpendicular parking for self-driving cars using deep reinforcement learning [C]. 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC). IEEE, 2022: 40-46.

[36] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor [C]. International Conference on Machine Learning, PMLR. 2018: 1861-1870.

[37] HAUSKNECHT M, STONE P. Deep reinforcement learning in parameterized action space [C]. International Conference on Learning Representations, 2015.

[38] WANG Y, HE H, TAN X. Truly proximal policy optimization [C]. Uncertainty in Artificial Intelligence, 2020:113-122.

[39] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft actor-critic algorithms and applications [C]. Proceedings of the 35th International Conference on Machine Learning, 2018.

[40] XIONG J, WANG Q, YANG Z, et al. Parametrized deep Q-networks learning: reinforcement learning with discrete-continuous hybrid action space [J]. arXiv preprint arXiv:1810. 06394. 2018.

[41] FAN Z, SU R, ZHANG W, et al. Hybrid actor-critic reinforcement learning in parameterized action space [C]. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, 2019: 2279-2285.