

doi:10.3969/j.issn.1673-9833.2024.01.009

自动驾驶路径优化的 RF-DDPG 车辆控制算法研究

焦龙飞, 谷志茹, 舒小华, 袁 鹏, 王建斌

(湖南工业大学 轨道交通学院, 湖南 株洲 412007)

摘 要: 针对自动驾驶车辆在行使中对目标路径跟踪精度不高、鲁棒性能较差等问题, 提出了一种深度确定性策略梯度 RF-DDPG (reward function-deep deterministic policy gradient) 路径跟踪算法。该算法是在深度强化学习 DDPG 的基础上, 设计 DDPG 算法的奖励函数, 以此优化 DDPG 的参数, 达到所需跟踪精度及稳定性。并且采用 aoplllo 自动驾驶仿真平台, 对原始的 DDPG 算法和改进的 RF-DDPG 路径跟踪控制算法进行了仿真实验。研究表明, 所提出的 RF-DDPG 算法在路径跟踪精度以及鲁棒性能等方面均优于 DDPG 算法。

关键词: 自动驾驶; 路径跟踪; 深度强化学习; 路径控制; DDPG 算法

中图分类号: TP273⁺.2

文献标志码: A

文章编号: 1673-9833(2024)01-0062-08

引文格式: 焦龙飞, 谷志茹, 舒小华, 等. 自动驾驶路径优化的 RF-DDPG 车辆控制算法研究 [J]. 湖南工业大学学报, 2024, 38(1): 62-69.

Research on RF-DDPG Vehicle Control Algorithm for Autonomous Driving Path Optimization

JIAO Longfei, GU Zhiru, SHU Xiaohua, YUAN Peng, WANG Jianbin

(College of Railway Transportation, Hunan University of Technology, Zhuzhou Hunnan 412007, China)

Abstract: In view of such flaws as low accuracy and poor robustness of target path tracking for autonomous vehicles in motion, a reward function-deep deterministic policy gradient (RF-DDPG) path tracking algorithm has thus been proposed. Based on the deep reinforcement learning DDPG, the algorithm designs the reward function of the DDPG algorithm for an optimization of the DDPG parameters so as to achieve the required tracking accuracy and stability. A simulation experiment has been conducted on the original DDPG algorithm and the improved RF-DDPG path tracking control algorithm based on the aoplllo autonomous driving simulation platform. The results show that the proposed RF-DDPG algorithm is characterized with an advantage over the DDPG algorithm in path tracking accuracy and robust performance.

Keywords: autonomous driving; path tracking; deep reinforcement learning; path control; deep deterministic policy gradient (DDPG) algorithm

收稿日期: 2023-06-13

基金项目: 国家自然科学基金区域联合基金资助重点项目 (U23A20385); 湖南省自然科学基金资助项目 (2022JJ50005); 湖南省研究生科研创新基金资助项目 (QL20230261); 湖南省教育厅科研基金资助项目 (23C0182)

作者简介: 焦龙飞, 男, 湖南工业大学硕士生, 主要研究方向为自动驾驶与强化学习, E-mail: 1012176179@qq.com

通信作者: 谷志茹, 女, 湖南工业大学副教授, 博士, 硕士生导师, 主要研究方向为智能网联系统, 电子系统设计, E-mail: guzhiru@hut.edu.cn

1 研究背景

目前,自动驾驶技术是人工智能领域的研究热点之一,而路径跟踪控制是自动驾驶系统中的核心问题。路径跟踪控制的目标是让车辆按照预定的路径行驶,并且尽可能地接近预定运行轨迹。路径跟踪控制方法主要分为基于模型和基于非模型两类。其中,基于模型的路径跟踪控制方法主要依赖于车辆的运动学和动力学模型,通过控制器输出控制信号来控制车辆运动,常见的基于模型的控制方法包括比例-积分-微分控制(PID控制)^[1-5]、模糊控制^[6-12]、模型预测控制(model predictive control, MPC)^[13-19]等。基于非模型的路径跟踪控制方法则不需要准确的车辆模型,而是通过感知和决策模块来实现控制,基于非模型的控制方法常见的是基于神经网络的控制方法^[20-23]等。

已有研究中,针对自动驾驶车辆在行使中对目标路径跟踪精度不高、鲁棒性能较差等问题,研究者们从基于模型的算法中寻找解决方法。如文献[3]提出了一种基于车辆运动学和动力学的模糊自适应PID控制方法,该方法根据预览理论规划下一个行驶路径。其首先利用车辆质心与所需路径预览点之间的位置关系计算横向偏差和航向偏差,然后利用模糊自适应PID控制器对误差的调整作用调整前轮角度。该方法虽然简单易行,但是在高要求控制场合的适应性和控制精度有限。文献[6]提出了一种基于动态双点预瞄策略的横纵向模糊控制方法,通过模糊控制来动态控制双点预瞄距离,进而控制车辆跟踪相应轨迹,但是模糊控制的效果受预瞄距离的改变影响较大。文献[11]提出了一种基于横向误差和航向误差的复合模糊控制方法,该方法通过指定相应的权重变量来调整两个模糊控制器的输出,采用积分补偿解决传统模糊控制稳态精度较低的问题。但是该方法在复杂道路中的路径跟踪精度较差。文献[14]提出了一种采用线性时变模型预测控制进行车辆轨迹跟踪的控制算法,相比于非线性控制,该方法具有全局最优解,且计算量更小。然而,该方法对车辆的建模要求较高,对非线性系统需要进行线性逼近,并且需要构造二次成本函数。同时,该方法对硬件存储空间和计算能力的要求较高,需要适当考虑计算资源的限制。文献[15]提出了建立模型预测控制和非线性动力学鲁棒控制(MPC-NDRC)框架,以提高轨迹跟踪性能。MPC-NDRC框架分为两个阶段:一是构建预测模型控制器,防止在线计算复杂模型带来的实时性能较差的问题;二是创建基于非线性动力学的鲁棒控制器,确保轨迹

跟踪性能和控制器模型精度。虽然该控制算法能稳定地跟踪路径,但是系统过于复杂,稳定性能较差。文献[20]采用了神经网络和模糊控制相结合的方法,通过控制方向盘转角来控制车辆的行驶方向,该方法的控制效果较为稳定,但是存在转向控制不够及时和跟踪误差较大等问题。

基于模型的控制算法在路径跟踪中需要依赖车辆模型,而车辆建模是一个复杂的过程,不仅需要考虑到机械结构、动力学特性、控制策略等多方面因素的影响,还需要考虑各种不确定因素的影响,因此建模的难度较大。基于非模型的控制算法,如神经网络控制,需要大量的车辆数据和环境数据进行神经网络训练,但是当前的技术条件很难保证环境数据的采集完整性。缺乏完整的环境数据会导致神经网络学习到不准确的信息,从而使得跟踪效果不佳。为了解决这个问题,科研工作者提出了有出色感知能力与决策能力的深度强化学习。深度强化学习方法是与环境不断地交互试错地学习,自主探索控制系统的最优行为。强化学习的算法包括SARSA^[24](state action reward state action)、Q-learning^[25]、DQN^[26](deep Q-learning)、DDPG^[27]等。SARSA是先建立一个Q表格,并通过与环境的交互来更新Q表格的状态,然后根据Q表格中的值采取动作,但是SARSA只能针对一些简单的游戏。Q-learning与SARSA类似,Q-learning的不同点是在更新Q表格时选择不同的策略,但本质上还是以表格的形式,Q-learning却是通过Q表格来选择最优。DQN是在Q-learning的基础上,通过引入神经网络代替Q表格,从而节省了软件空间,但是其不适用于连续空间。DDPG是一种利于深度函数逼近的策略,它可以在高维度和连续空间中应用,而前3种算法只适用于低维度、离散的行为空间中。但是在高维度、连续动作的自动驾驶中,DDPG的奖惩机制不能很好地设置。

综合上述分析可知,无论是基于模型还是非模型的路径跟踪控制算法,在路径跟踪过程中的表现都有所不足,主要表现在:

1) 基于模型的控制算法在路径跟踪过程中对模型的依赖程度比较高,但是车辆的建模比较困难,这会导致跟踪精度不佳。

2) 基于非模型的控制算法需要大量的车辆数据和环境数据以供神经网络学习,而环境数据采集的完整性难以满足,这导致跟踪效果比较差。

针对以上路径跟踪控制算法存在的弊端,课题组提出一种RF-DDPG路径跟踪控制算法,该算法既不依赖系统精准的数据模型,也不需要大量的环境数

据,仅通过车辆与期望路径的横向偏差、车辆的横向角速度等设计奖励函数和自适应权重系数,实现自动驾驶车辆的路径跟踪控制。

2 强化学习模型

2.1 马尔可夫决策过程

强化学习的本质是智能体与环境的交互过程,它可以被看作一个马尔可夫决策过程(Markov decision process, MDP)。MDP 是一个跟时间相关的序列决策过程,下一时刻的状态只取决于当前状态和动作。

MDP 定义了一个五元组 (S, A, R, P, γ) , 其中:

$S = \{s_1, s_2, s_3, \dots\}$, 代表车辆的状态;

$A = \{a_1, a_2, a_3, \dots\}$, 代表当前状态下智能体所输出的动作;

$R = \{r_1, r_2, r_3, \dots\}$, 代表当前状态下所输出动作的奖励, 具有滞后性;

$P = p[s_{t+1}, r_t | s_t, a_t]$, 代表当前状态下 s_t 输出动作 a_t 转移到下一个状态 s_{t+1} 并拿到奖励 r_t 的概率函数;

γ 为折扣因子, 且 $\gamma \in [0, 1]$ 。

2.2 强化学习过程

强化学习过程示意图如图 1 所示。

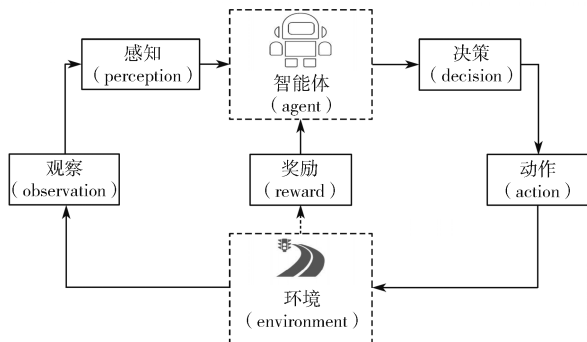


图 1 强化学习过程示意图

Fig. 1 Reinforcement learning process

在强化学习过程中,智能体在每个时间点根据当前的状态参数 S 给出动作 A , 然后进入下一个环境状态, 给出反馈回报 R 。然后在记忆池里就会记录一系列的数据 $(s_1, a_1, r_1, s_2, a_2, r_2, \dots, s_t, a_t, r_t)$, 并计算累计回报 G_t , 其计算式如下:

$$G_t = R_{t+1} + \gamma R_{t+2} + \dots = \sum_{k=0}^{+\infty} \gamma^k R_{t+k+1} \quad (1)$$

用 π 表示智能体的策略, 根据当前的状态 s 选择输出动作 a 的概率。

$$\pi(a, s) = p(A_t = a | S_t = s) \quad (2)$$

用价值函数 Q 表示当前状态下 s 采取动作 a 的价值:

$$Q(s, a) = E_{\pi}(G_t | A_t = a, S_t = s) \quad (3)$$

利用贝尔曼方程递归后的价值函数为

$$Q(s, a) = E_{\pi}(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) | A_t = a, S_t = s) \quad (4)$$

式 (3) (4) 中, $E(x)$ 为期望函数。

强化学习的核心任务, 就是不断地调整策略, 以使得奖励函数值最大化。在强化学习过程中, 智能体通过获取奖励函数值最大化来更新策略, 策略再给出下一步动作, 并且拿到奖励, 以此循环, 最终达到系统控制目标。

3 轨迹跟踪控制算法

3.1 深度确定性策略梯度算法

DDPG 是一种基于确定性策略梯度的无模型算法, 该算法基于 Actor-Critic 框架, 它可以被应用于连续行为空间中, 它由 Actor、Actor-target 和 Critic、Critic-target 网络构成。其中, Actor 网络的作用是根据环境反馈的状态 S 输出动作 A ; Critic 网络的作用是根据环境反馈的状态 S 和 Actor 对应的动作 A 输出 Q 值; Actor-target 网络和 Critic-target 网络的作用是提供网络的稳定性。网络首先固定自身的参数一段时间, 然后通过复制 Actor 网络和 Critic 网络的参数来更新自己的参数, DDPG 的算法原理框图如图 2 所示。

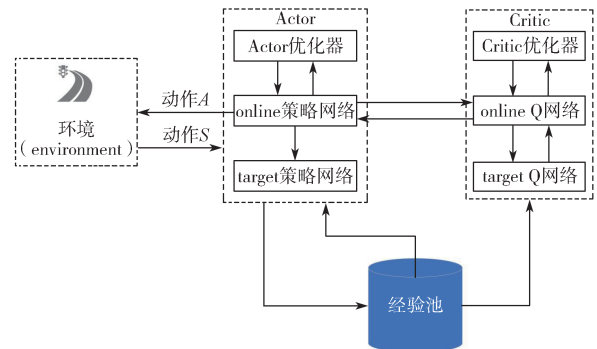


图 2 DDPG 算法原理框图

Fig. 2 DDPG algorithm block diagram

Actor 网络在状态观测量的基础上, 输出对应的决策行为, 并且将这些行为参数化为一个包含 n 维向量 θ 的策略 π 。

$$\pi(a | s, \theta) = p(A_t = a | S_t = s) \quad (5)$$

Actor 网络以策略梯度方法为基础进行更新, 通过策略梯度改进策略。

策略梯度 $\nabla_{\theta} J(\theta)$ 表达式为

$$\nabla_{\theta} J(\theta) = \left[\frac{\partial J(\theta)}{\partial \theta_1} \quad \frac{\partial J(\theta)}{\partial \theta_2} \quad \dots \quad \frac{\partial J(\theta)}{\partial \theta_n} \right]^T \quad (6)$$

随机性策略的策略梯度 $\nabla_{\theta} J(\theta)$ 表达式为

$$\nabla_{\theta} J(\pi_{\theta}) = E_{\pi_{\theta}} \left(\left[\nabla_{\theta} \lg \pi_{\theta}(s, a) \right] Q^{\pi}(s, a) \right). \quad (7)$$

式(6)(7)中: $J(\pi_{\theta})$ 为策略目标函数;

$\nabla_{\theta} \lg \pi_{\theta}(s, a)$ 为分数函数, 可表示为

$$\nabla_{\theta} \lg \pi_{\theta}(s, a) = (\nabla_{\theta} \pi_{\theta}(s, a)) / (\pi_{\theta}(s, a)). \quad (8)$$

在确定性策略中, $a = \mu_{\theta}(s)$, 确定性策略的梯度是随机性策略梯度方差趋近于0的特殊形式, 其梯度表达式为

$$\nabla_{\theta} J(\mu_{\theta}) = E_{\mu_{\theta}} (\nabla_{\theta} \mu_{\theta}(s) Q^{\mu}(s, a) | \mu_{\theta}(s)). \quad (9)$$

在智能车辆轨迹跟踪控制的学习过程中, Actor 神经网络的输入是观测到的环境状态变量, 如位置、角度、速度等, 其输出是根据策略做出的决策, 如方向盘转角和油门刹车等。与此同时, Critic 基于行为价值函数的方法, 其输入变量是状态和行为, 输出变量是回报值。在学习过程中, Critic 用估计的价值函数作为更新 Actor 函数的基准, 同时评价 Actor 的策略。Actor-Critic 方法的优点在于 Critic 通过价值函数提供了更准确的评估, 从而改进了 Actor 策略, 使其更加优化。此外, Actor-Critic 方法不仅可以使 Critic 更新 Actor 策略, 还可以使用 Actor 更新 Critic 的价值函数, 可以更好地评估行为价值。

在实践中, 使用如下贝尔曼方程更新 Critic 的价值函数:

$$Q'(s, a) = Q(s, a) + \alpha [R(s, a) + \gamma \max_a Q(s, a) - Q(s, a)]. \quad (10)$$

式中: α 为学习率;

Q' 为新的价值函数。

Actor 网络更新参数 θ 采用链式求导得出。

$$\nabla_{\theta} J(\theta) = \frac{1}{n} \sum_{i=1}^n [\nabla Q(s_i, a_i) \nabla \pi_{\theta}(a_i | s_i)]. \quad (11)$$

Critic 网络更新参数 w 取期望值和实际值的均方误差, 即

$$J(w) = \frac{1}{n} \sum_{i=1}^n (Q(s'_i, a'_i, w') - Q(s_i, a_i, w))^2. \quad (12)$$

式中, $Q(s', a', w')$ 为 Critic-target 网络计算得到的目标值。

3.2 奖励函数设计

奖励函数的好坏是影响模型结果的关键因素。对于单一任务的智能体都有着明确的奖励目标, 故应做到让奖励值最大化。但是在处理复杂的自动驾驶任务中, 很难单一明确奖励目标, 故本论文拟通过组合的方式来设计奖励函数。

1) 路径跟踪能力。本研究中, 设计用车辆的质心位置 y_i 与期望轨迹 y_j 的横向距离来描述车辆的跟踪精度。

$$R_1 = \Delta_y = |y_i - y_j|. \quad (13)$$

跟踪精度误差与允许误差之比为 Δ_1 , 且

$$\Delta_1 = \Delta_y / 0.3. \quad (14)$$

2) 速度。 $R_2 = V_x \cos(\theta)$, 其中 $V_x \cos(\theta)$ 为车辆沿期望路径方向的速度, 在有限的时间和安全的条件下, 希望快速完成驾驶任务。

3) 车辆稳定性。车辆稳定性主要是由车辆的横摆角速度和质心侧偏角来体现。横摆角速度常采取实际横摆角速度 ω_p 与期望横摆角速度 ω_t 的差值进行描述。

$$R_3 = \Delta_{\omega} = |\omega_p - \omega_t|. \quad (15)$$

式中:

$$\omega_t = \min(|\omega_{des}|, \omega_d) \text{sgn}(\delta). \quad (16)$$

其中: ω_d 为横摆角速度上限;

ω_{des} 为稳态转向下的横摆角速度, 且 $\omega_{des} = G_{\omega zss} \times \delta$, 其中 $G_{\omega zss}$ 为横摆角速度的稳态增益;

δ 为转向盘的角度。

横摆角速度误差与期望角速度之比为 Δ_2 , 且

$$\Delta_2 = \Delta_{\omega} / \omega_{t0}. \quad (17)$$

同样, 质心侧偏角采取实际质心侧偏角 β_p 与期望质心侧偏角 β_t 的差进行描述。

$$R_4 = \Delta_{\beta} = |\beta_p - \beta_t|, \quad (18)$$

$$\beta_t = \min(|\beta_{des}|, \beta_d) \text{sgn}(\delta). \quad (19)$$

式中: β_d 为质心侧偏角上限;

β_{des} 为稳态转向下的质心侧偏角, 且 $\beta_{des} = G_{\beta zss} \times \delta$, 其中 $G_{\beta zss}$ 为质心侧偏角的稳态增益。

质心侧偏角误差与期望质心侧偏角之比为 Δ_3 , 且

$$\Delta_3 = \Delta_{\beta} / \beta_t. \quad (20)$$

4) 转向平稳性。转向的平稳性代表方向盘的震荡程度, 这里引入变异系数 R_5 进行描述, 且

$$R_5 = C_v = \sigma / \bar{\theta}. \quad (21)$$

式中: σ 为方向盘转角的标准差;

$\bar{\theta}$ 为方向盘转角的平均值。

3.3 自适应权重设计

路径跟踪精度和车辆的稳定性能对自动驾驶路径跟踪控制的影响比较大。当两者不能同时满足时, 要确定先处理差距大的那一个指标。本研究设计了自适应权重系数, 当跟踪精度误差百分比大于稳定性误差百分比时, 跟踪精度的奖励函数权重就会加大, 相反亦然。

跟踪精度权重系数为

$$C_1 = 0.5 + e^{A_1} / (e^{A_1} + e^{A_2 + A_3}); \quad (22)$$

稳定性权重系数为

$$C_2 = 0.5 + e^{A_2 + A_3} / (e^{A_1} + e^{A_2 + A_3}). \quad (23)$$

跟踪精度权重系数和稳定性权重系数满足如下表达式:

$$C_1 + C_2 = 2. \quad (24)$$

自动驾驶车辆在训练过程中,会出现两种情况:正常行驶和超出车道。正常行驶的奖励函数已经设计完成,超出车道的情况在这里统一设置为 0。则奖励函数表达式如下:

$$R = \begin{cases} R_2 - C_1 R_1 - C_2 (R_3 + R_4) - R_5, & \text{正常行驶;} \\ 0, & \text{超出车道。} \end{cases} \quad (25)$$

4 仿真测试及验证

为了评估本研究中所提出自动驾驶车辆控制方法的优劣,接下来将在 apollo 仿真平台上搭建模型,并且对智能车辆在轨迹跟踪过程中采用基于 RF-DDPG 算法和基于原始的 DDPG 算法进行仿真与分析。本文所提出的 RF-DDPG 算法基于 Actor-Critic 网络结构,其中 Actor 网络以策略梯度方法更新,根据策略的梯度将策略向更好的方向进行优化。Actor 网络的输入为观测量(位置、角度和速度等),输出为控制信号(方向盘转角和油门刹车)。Critic 网络则基于行为价值函数,输入变量为状态和行为,输出变量为回报值,用于评估策略的优劣。与传统的 DDPG 算法奖励函数不同,本文的 RF-DDPG 算法采用了一种新的奖励函数,这一改进使得算法更加具有鲁棒性和泛化能力。原始 DDPG 算法和 RF-DDPG 算法的评价方式如图 3 所示。

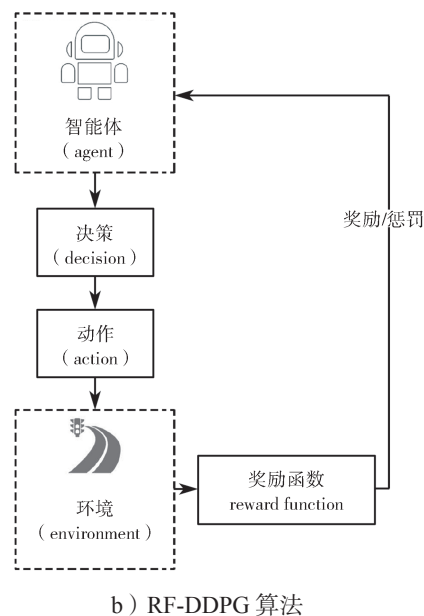
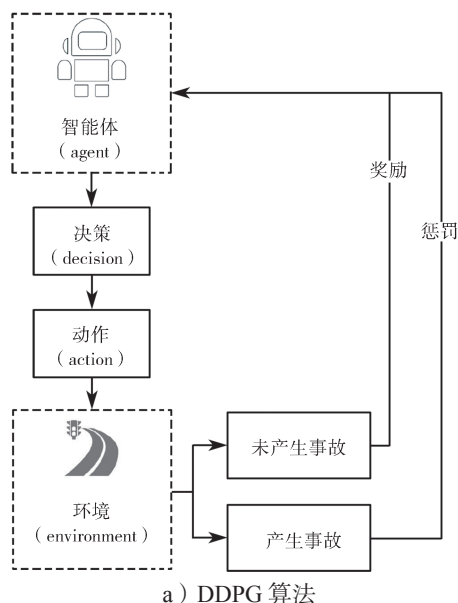


图 3 两种算法的评价方式

Fig. 3 Evaluation method of the two algorithms

图 3a 是原始评价算法,可以看出,原始算法对智能车的评价方式只区分了未产生事故和产生事故,训练出来的效果难以达到智能车辆对路径跟踪精度的要求。图 3b 是改进后的评价,利用组合方式设计奖励函数,使得评价更加合理,训练后的控制效果也更加精确。

算法改进前后自动驾驶车辆的航向角偏差曲线、横摆角速度偏差曲线、质心侧偏角差曲线,分别如图 4~6 所示。

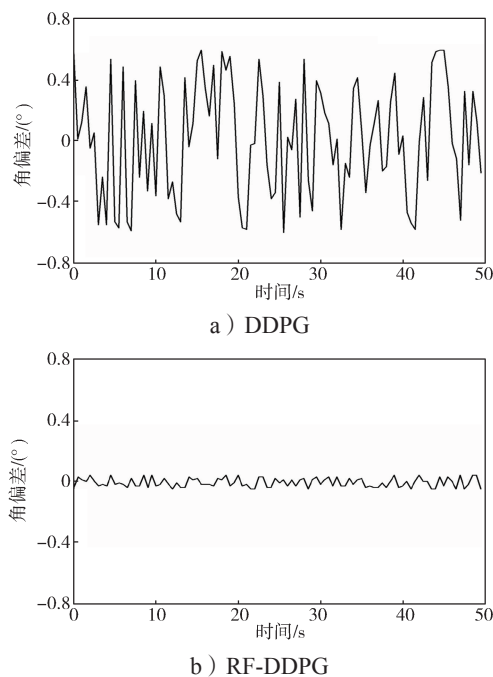


图 4 算法改进前后自动驾驶车辆的航向角偏差曲线
Fig. 4 Heading angle deviation curves of autonomous vehicles before and after an algorithm improvement

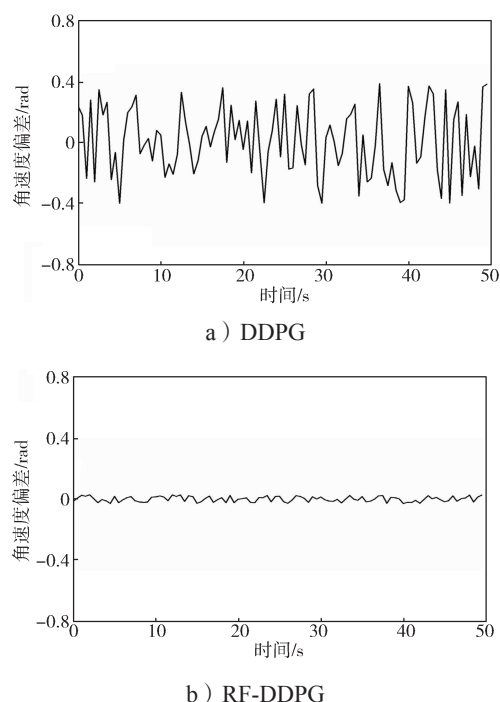


图5 算法改进前后自动驾驶车辆的横摆角速度偏差曲线
Fig. 5 Yaw rate deviation curves of the autonomous vehicle before and after an algorithm improvement

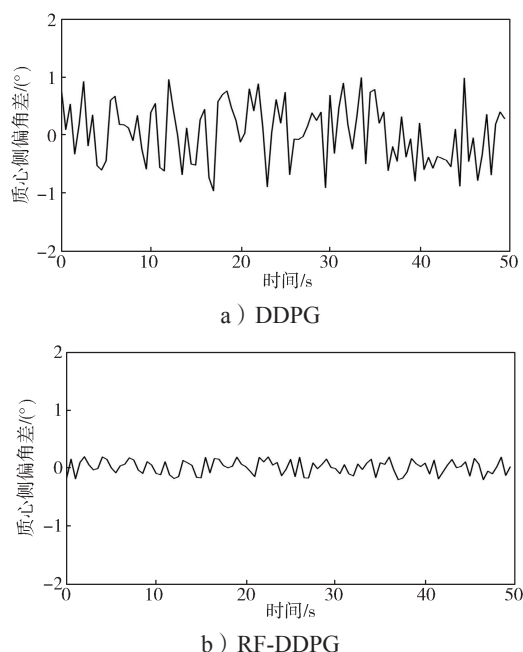


图6 算法改进前后自动驾驶车辆的质心侧偏角差曲线
Fig. 6 Centroid sideslip angle difference of autonomous vehicles before and after an algorithm improvement

从图4~6可以看出,采用RF-DDPG控制算法的自动驾驶车辆在实验过程中的稳定性能明显比采用DDPG算法车辆的稳定性能高,控制过程更加合理。这不仅验证了本研究对于算法策略输出的改进效果,也说明了本研究的改进方式在仿真环境中拥有良好的泛化性。

图7是采用RF-DDPG算法与DDPG算法的自动驾驶车辆的横向误差对比图。

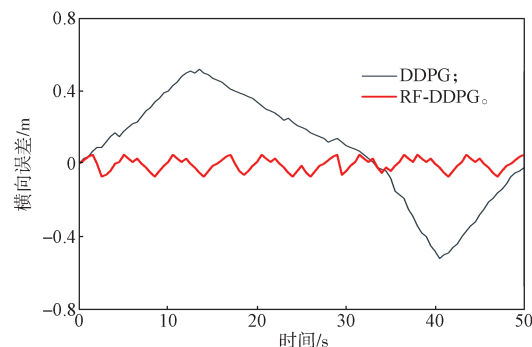


图7 算法改进前后自动驾驶车辆的横向误差对比图
Fig. 7 Lateral error comparison chart of autonomous vehicles before and after an algorithm improvement

由图7所示采用RF-DDPG算法与DDPG算法的自动驾驶车辆的横向误差对比图,同样可以直观地看出,RF-DDPG控制算法在跟踪精度性能方面表现为比DDPG算法的跟踪精度更高,控制过程更加合理。

表1为采用DDPG和RF-DDPG控制算法的不同跟踪控制值结果对比,由表1中的数据可以得出,RF-DDPG算法的跟踪控制值均优于DDPG算法的对应值。

表1 算法改进前后的跟踪控制值对比

Table 1 Comparison of tracking control values before and after an algorithm improvement

参 数	不同算法的跟踪控制值	
	DDPG	RF-DDPG
横向误差绝对值最大值/m	0.52	0.07
横向误差绝对值平均值/m	0.22	0.03
角偏差绝对值最大值/(°)	0.60	0.05
角偏差绝对值平均值/(°)	0.29	0.03
角速度偏差绝对值最大值/rad	0.40	0.03
角速度偏差绝对值平均值/rad	0.22	0.01

5 结语

本文以智能车辆为研究对象,采用基于强化学习的方法研究了车辆在跟踪轨迹时的最优控制问题,提出了一种深度确定性策略梯度RF-DDPG路径跟踪算法,旨在优化车辆的跟踪精度和运行稳定性。该算法在深度强化学习DDPG的基础上,设计了智能车辆在轨迹跟踪时的奖励函数和自适应权重系数,从而优化了RF-DDPG的参数。控制器以车辆当前的位置、速度、跟踪路径信息和航向角为输入,输出转向盘转角和油门刹车。并在仿真平台上测试了采用本文提出的算法和基于原始DDPG算法的智能车辆轨

迹跟踪效果。仿真结果表明,相比于基于原始 DDPG 的强化学习方法,本文提出的基于 RF-DDPG 的强化学习方法在跟踪精度和控制效果方面有了显著提高,并保证了车辆行驶过程的安全性和稳定性。

为进一步探究智能车辆轨迹跟踪问题,课题组将继续进行轨迹规划研究,以期将跟踪控制策略应用于所规划的轨迹中,并对轨迹跟踪策略进行仿真验证。在此基础上,进一步完善 RF-DDPG 算法,提高其控制精度和鲁棒性。本研究对于智能车辆的自动驾驶和智能交通系统的发展具有重要意义,有望为实现车辆安全行驶和交通流畅提供有效的技术支持。

参考文献:

- [1] MASOUMI A, HOURALI M, MOHTARAMI A. Designing an Intelligent Fuzzy PID Controller for Controlling and Directing Military Unmanned Vehicles[C]//2017 7th International Conference on Computer and Knowledge Engineering (ICCKE). Mashhad: IEEE, 2017: 161-166.
- [2] ZHENG S, HUANG Z Y, ZHANG E H. Finite-Time Disturbance Observer Based Parameter Self-Tuning PID Path-Following Controller Design[C]//2021 IEEE International Conference on Unmanned Systems (ICUS). Beijing: IEEE, 2021: 735-740.
- [3] ZHANG J Q, ZHANG T, LI G, et al. Path Following Control Based on Fuzzy Adaptive PID for Unmanned Vehicle[C]//2020 4th CAA International Conference on Vehicular Control and Intelligence (CVCI). Hangzhou: IEEE, 2021: 502-507.
- [4] HUANG Y F, TIAN Z S, JIANG Q, et al. Path Tracking Based on Improved Pure Pursuit Model and PID[C]//2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT). Weihai: IEEE, 2021: 359-364.
- [5] ELTAYEB A, RAHMAT M F, MOHAMMED ELTOUM M A, et al. Trajectory Tracking for the Quadcopter UAV Utilizing Fuzzy PID Control Approach[C]//2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE). Khartoum: IEEE, 2021: 1-6.
- [6] 刁勤晴, 张雅妮, 朱凌云. 双预瞄点智能车大曲率路径的横纵向模糊控制[J]. 中国机械工程, 2019, 30(12): 1445-1452.
DIAO Qinqing, ZHANG Yani, ZHU Lingyun. A Lateral and Longitudinal Fuzzy Control of Intelligent Vehicles with Double Preview Points for Large Curvature Roads[J]. China Mechanical Engineering, 2019, 30(12): 1445-1452.
- [7] WANG Y Y, JIANG B, WU Z G, et al. Adaptive Sliding Mode Fault-Tolerant Fuzzy Tracking Control with Application to Unmanned Marine Vehicles[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(11): 6691-6700.
- [8] MA H X, CHEN M, FENG G, et al. Disturbance-Observer-Based Adaptive Fuzzy Tracking Control for Unmanned Autonomous Helicopter with Flight Boundary Constraints[J]. IEEE Transactions on Fuzzy Systems, 2023, 31(1): 184-198.
- [9] DONG Z P, BAO T, ZHENG M, et al. Heading Control of Unmanned Marine Vehicles Based on an Improved Robust Adaptive Fuzzy Neural Network Control Algorithm[J]. IEEE Access, 2019, 7: 9704-9713.
- [10] WANG X Y, FU M Y, YANG Y, et al. Lateral Control of Autonomous Vehicles Based on Fuzzy Logic[C]//2013 25th Chinese Control and Decision Conference (CCDC). Guiyang: IEEE, 2013: 237-242.
- [11] LI Y B, YU J L, GUO X N, et al. Path Tracking Method of Unmanned Agricultural Vehicle Based on Compound Fuzzy Control[C]//2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). Chongqing: IEEE, 2020: 1301-1305.
- [12] GONG Y, LIU Y, TANG Z M. Path Tracking of Unmanned Vehicle Based on Parameters Self-Tuning Fuzzy Control[C]//2013 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems. Nanjing: IEEE, 2013: 52-57.
- [13] JI J, KHAJEPOUR A, MELEK W W, et al. Path Planning and Tracking for Vehicle Collision Avoidance Based on Model Predictive Control with Multiconstraints[J]. IEEE Transactions on Vehicular Technology, 2017, 66(2): 952-964.
- [14] GUTJAHR B, GRÖLL L, WERLING M. Lateral Vehicle Trajectory Optimization Using Constrained Linear Time-Varying MPC[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(6): 1586-1595.
- [15] SONG J R, TAO G, ZANG Z, et al. Nonlinear Dynamics Based Trajectory Tracking Robust Control of Unmanned Ground Vehicle[C]//2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI). Nanjing: IEEE, 2022: 1-6.
- [16] XIN P, WANG Z W, SUN H T, et al. Model Predictive Control of Unmanned Mine Vehicle Trajectory Tracking[C]//2021 40th Chinese Control Conference (CCC). Shanghai: IEEE, 2021: 4757-4762.
- [17] ZHANG C Y, CHU D F, LIU S D, et al. Trajectory Planning and Tracking for Autonomous Vehicle Based on State Lattice and Model Predictive Control[J]. IEEE

- Intelligent Transportation Systems Magazine, 2019, 11(2): 29–40.
- [18] LEI Z L, TANG X M. Trajectory Tracking for Unmanned Vehicle Based on Model Predictive Control Using Linear Matrix Inequalities[C]//2020 Chinese Automation Congress (CAC). Shanghai: IEEE, 2021: 698–703.
- [19] WANG M, WU Y F, HUANG F Y. Trajectory Tracking of Unmanned Vehicle Based on Improved Model Predictive Control[C]//2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES). Nanning: IEEE, 2021: 57–60.
- [20] 张炳力, 李子龙, 沈 干, 等. 基于模糊神经网络的智能汽车轨迹跟踪研究 [J]. 汽车工程, 2019, 41(8): 953–959.
ZHANG Bingli, LI Zilong, SHEN Gan, et al. A Research on Path Tracking of Intelligent Vehicle Based on Fuzzy Neural Network[J]. Automotive Engineering, 2019, 41(8): 953–959.
- [21] OKAMOTO K, TSIOTRAS P. A Comparative Study of Data-Driven Human Driver Lateral Control Models[C]//2018 Annual American Control Conference (ACC). Milwaukee: IEEE, 2018: 3988–3993.
- [22] VALLON C, ERCAN Z Y, CARVALHO A, et al. A Machine Learning Approach for Personalized Autonomous Lane Change Initiation and Control[C]//2017 IEEE Intelligent Vehicles Symposium (IV). Los Angeles: IEEE, 2017: 1590–1595.
- [23] CHEN Y F, LI X F. Iterative Learning Model Predictive Control for Lateral Control of Autonomous Vehicles[C]//2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS). Chengdu IEEE, 2022: 911–916.
- [24] 吴少波, 杨薛钰. 基于 Sarsa 算法的交通信号灯控制方法 [J]. 信息与电脑 (理论版), 2021, 33(6): 49–51.
WU Shaobo, YANG Xueyu. Traffic Signal Light Control Method Based on Sarsa Algorithm[J]. China Computer & Communication, 2021, 33(6): 49–51.
- [25] 钱 信, 吕成伊, 宋世杰. 基于优化 Q-learning 算法的机器人路径规划 [J]. 南昌大学学报 (工科版), 2022, 44(4): 396–401.
QIAN Xin, LÜ Chengyi, SONG Shijie. Robot Path Planning Based on Optimizing Q-Learning Algorithm[J]. Journal of Nanchang University (Engineering & Technology), 2022, 44(4): 396–401.
- [26] 李奇儒, 耿 霞. 基于改进 DQN 算法的路径规划方法研究 [J/OL]. 计算机工程: 1–11. [2023–06–22]. DOI:10.19678/j.issn.1000-3428.0066348.
LI Qirui, GENG Xia. Research on Path Planning Method Based on Improved DQN Algorithm[J/OL]. Computer Engineering: 1–11. [2023–06–22]. DOI:10.19678/j.issn.1000-3428.0066348.
- [27] 杨青青, 陈 剑, 彭 艺. 基于 DDPG 的无人机轨迹规划及功率控制算法 [J]. 北京邮电大学学报, 2023, 46(3): 43–48.
YANG Qingqing, CHEN Jian, PENG Yi. Unmanned Aerial Vehicle Trajectory Planning and Power Control Algorithm Based on Deep Deterministic Policy Gradient[J]. Journal of Beijing University of Posts and Telecommunications, 2023, 46(3): 43–48.

(责任编辑: 廖友媛)