

DOI: 10.13718/j.cnki.xdzk.2022.03.002

考虑驾驶员特性的个性化跟驰控制策略研究

任玥¹, 邹博文², 尹旭³, 刘学高², 梁新成¹

1. 西南大学 工程技术学院, 重庆 400715; 2. 西南大学 人工智能学院, 重庆 400715;

3. 重庆长安汽车软件科技有限公司 智能控制室, 重庆 401120

摘要: 为提高车辆自主跟驰功能的个性化程度, 使之能适应不同驾驶员的驾驶风格, 提高乘员对自动驾驶功能的接受度, 提出了一种基于深度强化学习的个性化跟驰控制策略。首先基于模拟驾驶试验平台进行驾驶员在环试验, 获取真实驾驶员跟驰数据。根据车辆跟驰动力学, 建立了连续动作空间决策模型。构建了基于 Actor-Critic 的深度强化学习架构, 并综合考虑跟驰过程的安全性、舒适性和宜人性设计了奖励函数, 通过双延迟深度确定性策略梯度 (Twin Delayed Deep Deterministic Policy Gradient, TD3) 对决策模型进行训练。基于 CARLA 模拟器的仿真结果表明, 本研究提出的个性化跟驰控制策略在保证车辆自主跟驰过程稳定性和安全性的前提下, 其决策结果更接近驾驶员驾驶习性。

关键词: 自主跟驰; 个性化; 深度强化学习; 奖励函数

中图分类号: U461.8

文献标志码: A

文章编号: 1673-9868(2022)03-0012-08

开放科学(资源服务)标识码(OSID):



Study of Personalized Car-following Control Strategy by Considering the Driver Characteristics

REN Yue¹, ZOU Bowen², YIN Xu³, LIU Xuegao², LIANG Xincheng¹

1. School of Engineering and Technology, Southwest University, Chongqing 400715, China;

2. School of Artificial Intelligence, Southwest University, Chongqing 400715, China;

3. Department of Intelligent Control, Chongqing Changan Automobile Software Technology Co. Ltd., Chongqing 401120, China

Abstract: In order to improve the degree of personalization of the car-following control which makes it can adapt to the driving styles of different drivers, a personalized car-following control algorithm based on deep reinforcement learning is proposed in this paper. Firstly, the driving simulation platform was established to adopt the driver-in-loop experiment for obtaining the real driver following data. Then, based on the car-following dynamics, the continuous action space decision model was built. The deep reinforcement learning model was introduced with actor-critic architecture and the reward function was designed considering the safety, comfort, and driver behavior. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm was proposed to train the decision model. Finally, the results of simulation based on the CARLA

收稿日期: 2021-11-21

基金项目: 重庆市自然科学基金面上项目(cstc2020jcyj-msxmX0496); 中央高校基本业务费项目(SWU119021).

作者简介: 任玥, 博士, 讲师, 主要从事智能汽车决策与控制研究.

通信作者: 邹博文, 教授.

simulator demonstrated that the personalized car-following control strategy proposed in this paper is closer to the driving habits of human drivers on the premise of ensuring the stability and safety of the autonomous car-following process.

Key words: autonomous car-following; personalization; deep reinforcement learning; reward function

根据美国国家公路交通安全管理局(The National Highway Traffic Safe Administration, NHTSA)统计,超过90%的安全事故都是由于驾驶员失误造成的^[1].近年来,随着传感器的大范围普及和控制器算力的逐步提升,高级驾驶辅助系统(Advanced Driver Assistance System, ADAS)得到广泛应用.其中,自动紧急制动系统能够有效降低驾驶员负荷,减少因驾驶员疏忽导致的交通事故.以此功能为基础的车辆自主跟驰控制系统,也是车辆高度/完全自动驾驶系统的核心功能之一,对提高车辆安全性、舒适性、经济性有着重要的意义,近年来受到全球学者的关注和研究.

对于车辆跟驰控制,其主要基于自车和前方障碍的运动信息进行碰撞风险评估.目前较为成熟的风险评估模型包括以 MAZDA 模型、Berkley 模型、NHTSA 模型为代表的的目标距离模型^[2-4]和以碰撞时间(Time to collision, TTC)为代表的的时间时距模型^[5-6].基于碰撞风险模型,Gerdes 等^[7]采用了一种基于多面滑模控制器对发动机输出扭矩和制动力矩进行控制,在跟车工况中,该方法能够精确地跟踪车辆期望速度并与前车保持合适的安全距离. Kim 提出了一种新型的时变参数自适应速度控制器,其控制发动机和制动力矩跟踪车辆期望速度,该方法有着较高的跟踪精度并对于外界扰动有良好的鲁棒性^[8].模型预测控制策略(MPC)由于能够系统地处理全局约束,在跟驰控制中有着较好的效果. Li 通过建立自适应巡航优化模型,平衡了在跟驰过程中车辆跟踪性能、燃油经济性和驾驶员期望响应的矛盾^[9].文献[10]采用高斯核函数描述了碰撞风险,并采用 MPC 控制器的优化目标,实现了车辆自适应巡航功能.

近年来随着人工智能和强化学习技术的快速发展,其越来越广泛地应用于决策和控制系统.罗颖等^[11]采用深度确定性策略梯度(DDPG)算法,结合屏障控制方法,实现了车辆低速跟驰控制.朱冰等^[12]考虑前车运动不确定性,采用基于 PPO 的深度强化学习方法,实现了车辆自主跟驰,并有效降低了在线计算量.虽然目前大多数跟驰控制策略已经更有效地避免碰撞,实现安全驾驶,但现有大多数控制策略是基于固定控制器参数的.而在实际驾驶中,由于不同驾驶员具有差异化的性格、驾驶技术、驾驶风格,而统一标定的控制策略无法满足不同驾驶员的驾驶习性,从而导致车辆自主跟驰功能的宜人性较差,乘员接受度低.管欣等^[13]引入驾驶人模型并提出了基于驾驶人最优预瞄加速度模型的自适应巡航系统,实现了不同风格的跟驰控制. Yi 等^[14]通过采集实际驾驶人数据,采用基于具有遗传因子的递归最小二乘算法实现了驾驶人特性参数,并应用于自适应巡航的起停控制系统.文献[15-16]采用强化学习方法进行跟驰建模,并在学习过程中考虑了驾驶员行为特性.

为使得车辆自主跟驰功能能够满足不同驾驶员需求,文献[17-18]在进行安全距离建模时考虑了驾驶员行为特性.文献[19]将驾驶员数据进行聚类分析,并基于驾驶风格辨识设计了不同的 ACC 控制器参数,提高了 ACC 系统人性化.文献[20-21]采用逆强化学习方法,直接从驾驶员数据中拟合出决策算法,实现了拟人化的跟驰决策.

针对不同驾驶员的驾驶习性,基于模拟驾驶试验采集驾驶员真实驾驶数据和深度强化学习的车辆个性化自主跟驰控制算法进行研究,以期通过设置奖励函数满足车辆跟驰过程的安全性、舒适性和宜人性,并采用改进 DDPG 算法实现车辆加速度自适应控制.

1 驾驶员数据采集

相较于实车试验,模拟驾驶试验由于其具有高安全、低成本、多工况、可重复性的特点,被广泛应用于驾驶员驾驶数据的采集,本研究通过驾驶员在环虚拟仿真试验获取驾驶员跟驰行为数据.驾驶行为采集系统如图1所示.

驾驶模拟试验在 Ubuntu 环境下进行.通过 CARLA 自动驾驶模拟器建立实时交通场景和环境车辆,采用罗技 G29 驾驶模拟器采集驾驶员方向盘转角、油门/制动踏板信号并传输到车辆动力学模型,输出主

车实际运动状态,再与仿真场景实时交互.针对跟驰工况,本研究设计了前车匀速行驶、减速行驶、随机变速行驶等工况,以模拟城市环境的车辆跟驰典型工况.在模拟驾驶试验中,采集主车与前车的相对距离、相对车速、主车车速以及主车加速度序列作为驾驶员跟驰数据集.

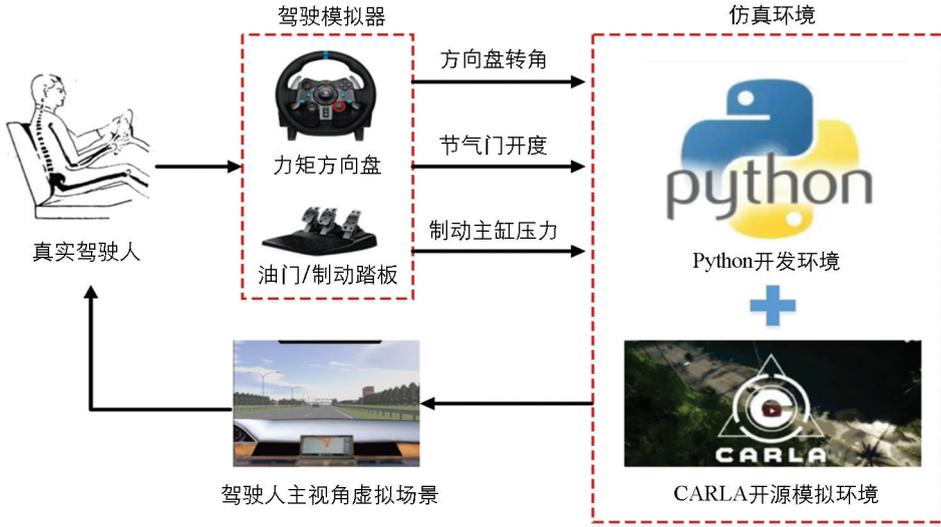


图 1 驾驶行为采集系统

2 个性化跟驰控制策略

2.1 车辆跟驰动力学建模

本研究中仅考虑车辆纵向运动,忽略车辆转向时侧向运动对纵向运动的耦合作用,建立车辆跟驰模型,如下式所示:

$$\begin{aligned} v_h(k+1) &= v_h(k) + a_h(k)\Delta T \\ v_r(k) &= v_l(k) - v_h(k) \\ d(k+1) &= d(k) + v_r(k)\Delta T \end{aligned} \quad (1)$$

式中, v_h 和 a_h 分别为主车车速和加速度, v_l 为前车车速, v_r 为相对车速, d 为相对距离. ΔT 为采样间隔. 其中, 主车加速度满足车辆纵向动力学方程:

$$a_h = \frac{F_t - F_b - F_f - F_w - F_i - F_\delta}{M} \quad (2)$$

其中, M 为车辆质量, F_t , F_b 分别为车辆的驱动/制动力. F_f , F_w , F_i , F_δ 分别为车辆的滚动阻力、空气阻力、坡度阻力和加速阻力^[22].

在车辆跟驰控制过程中,选择主车车速、相对速度和相对距离作为状态空间,主车加速度作为动作空间,可表示为:

$$\begin{aligned} S &= [v_h, v_r, d] \\ A &= [a_h] \end{aligned} \quad (3)$$

在车辆实际运动过程中,其纵向加速度受到附着力限制,因此,为避免在学习过程产生不合理决策,设置动作空间取值范围为 $-6 \text{ m/s}^2 \leq a_h \leq 6 \text{ m/s}^2$.

2.2 奖励函数

在车辆跟驰过程中,不仅需要保证车辆的安全性,还需尽可能提高乘员舒适性.同时,还应使得跟驰控制策略更符合实际驾驶员行为特性,因此,车辆自主跟驰是一个多目标控制问题.在 k 时刻,定义速度跟踪奖励函数 R_1 为:

$$R_1 = \omega_1 (v_h(k) - v_d)^2 \quad (4)$$

其中 v_d 为期望车速,由驾驶员启动自动跟驰功能时定义.

定义车辆跟踪奖励函数 R_2 为:

$$R_2 = \omega_2 e^{a \cdot d(k)^2} \quad (5)$$

其中 a 为形状因子,表示为:

$$a = \frac{1}{v_r(k) + \rho}$$

ρ 为距离安全系数,以避免 $v_r(k)$ 为 0 时 a 无法计算^[10].

定义舒适性奖励函数 R_3 为:

$$R_3 = \omega_3 (a_h(k) - a_h(k-1))^2 \quad (6)$$

舒适性奖励函数旨在避免车辆在短时间内产生较大的加速度变化对乘员造成较大的冲击度从而导致的舒适性恶化.

定义个性化奖励函数 R_4 为:

$$R_4 = \omega_4 (a_h(k) - \tilde{a}_h(k))^2 \quad (7)$$

其中 $\tilde{a}_h(k)$ 是在驾驶员跟驰数据序列中通过油门或制动踏板给车辆施加的实际加速度,个性化奖励旨在驱使自主跟驰策略动作选择与实际驾驶员行为更加接近^[23].

结合式(4)–(7),在 k 时刻个性化跟驰控制策略奖励函数表示为:

$$R_k = \begin{cases} -100 & d \leq d_{\text{safe}} \text{ 或 } v_h \leq 0 \text{ 或 } v_h \geq v_{\text{lim}} \\ 10 - \sum \omega_i R_i & \text{其他} \end{cases} \quad (8)$$

其中 d_{safe} 为最小安全距离, v_{lim} 为道路最大限速, ω_i 为各项奖励的归一化权重系数. 在训练过程中,当车辆发生不合理运动时(与前车碰撞、车速为负或超过限速),给予智能体一个较大的惩罚. 可以看出,奖励函数中权重系数 ω_i 的取值直接决定了智能体的跟驰效果,在实际训练过程中,先通过调试 $\omega_1 \sim \omega_3$ 使车辆拥有较合理的客观跟驰效果,再调节 ω_4 使得智能体决策结果能更与实际驾驶员特性更为接近.

2.3 基于 TD3 的跟驰控制策略

强化学习基于马尔科夫过程,通过状态、动作、奖励和状态转移函数描述智能体与环境的动态过程. 作为一种试错学习方法,强化学习控制智能体选择不同动作,通过与环境不断地进行交互尝试,并通过环境给出的奖励来判断动作的优劣,最终学习到一个最优策略,使累计回报的期望最大化.

由于车辆跟驰行为是一个连续过程,对于传统的 DQN(deep q-learning)、SARSA 等算法需要将连续状态空间离散化,易造成维数灾难. 深度确定性策略方法(deep deterministic policy gradient, DDPG)采用基于 Actor-Critic 框架,结合了 DQN 基于价值和策略梯度(policy gradient)基于策略的优势,通过对价值网络的时序差分更新与动作策略的梯度下降实现对连续动作空间问题的强化学习^[24].

DDPG 中的 actor 网络为策略网络,表示为 $\mu(s; \theta)$,其代表个性化跟驰控制策略. Critic 网络为价值网络,表示为 $q(s, a; \omega)$,价值网络是动作价值函数 $Q_\pi(s, a)$ 的近似,价值网络用于评价策略的好坏. 动作价值函数可表示为:

$$Q_\pi(s, a) = E(R_{k+1} + \gamma R_{k+2} + \gamma^2 R_{k+3} + \dots | S_k = s, A_k = a) \quad (9)$$

其中 γ 为奖励衰减系数.

对于策略网络,将其输出的动作和当前状态作为价值网络的输入,此时价值网络可以表示为 $q[s, \mu(a|s); \omega]$,其值越高,说明策略越好,故定义策略网络目标函数为:

$$J(\theta) = E_s[q(s, \mu(s; \theta); \omega)] \quad (10)$$

策略网络学习即为求最大化问题 $\max_{\theta} J(\theta)$,可以通过梯度上升进行求解. 从经验回放数组中任意抽取一个状态 s_j ,并定义:

$$g_j = \nabla_{\theta} q[s, \mu(s_j; \theta); \omega] \quad (11)$$

式(11)可以看做是 $\nabla_{\theta} J(\theta)$ 的无偏估计. 计算梯度,并基于学习率 β 即可实现策略网络参数 θ 的更新:

$$\theta_{\text{new}} = \theta_{\text{old}} + \beta \cdot \nabla_{\theta} \mu(s_j; \theta) \cdot \nabla_a q(s, a; \omega) \quad (12)$$

而对于价值网络,同样基于从经验回放数组中抽取的 (s_j, a_j, r_j, s_{j+1}) ,让价值网络通过历史数据和当前策略网络进行预测:

$$\hat{q}_j = q(s_j, a_j; \omega) \quad (13)$$

$$\hat{q}_{j+1} = q(s_{j+1}, \mu(s_{j+1}; \theta); \omega) \quad (14)$$

则 TD 目标可表示为:

$$\hat{y}_j = R_j + \gamma \cdot \hat{q}_{j+1} \quad (15)$$

定义价值网络损失函数为:

$$L(\omega) = \frac{1}{2} [q(s_j, a_j; \omega) - \hat{y}_j]^2 \quad (16)$$

可通过梯度下降更新价值网络参数,如下式所示:

$$\nabla_{\omega} L(\omega) = (\hat{q}_j - \hat{y}_j) \cdot \nabla_{\omega} q(s_j, a_j; \omega) \quad (17)$$

$$\omega_{\text{new}} = \omega_{\text{old}} - \alpha \cdot \nabla_{\omega} L(\omega) \quad (18)$$

对于式(15)和式(16), TD 目标是通过价值网络的估计进行计算的,而损失函数再次使用同样的价值网络对 TD 目标进行拟合,存在自举(Bootstrapping)的问题,容易导致对动作价值函数的高估或低估.为提升 DDPG 性能,使用 TD3 算法进行改进.此处采用两个价值网络和一个策略网络:

$$q(s, a; \omega_1), q(s, a; \omega_2), \mu(s; \theta) \quad (19)$$

再建立与 3 个网络对应的、并与各网络结构完全相同的目标网络:

$$q(s, a; \omega_1^-), q(s, a; \omega_2^-), \mu(s; \theta^-) \quad (20)$$

采用目标策略网络计算动作,并添加噪声:

$$\hat{a}_{j+1}^- = \mu(s_{j+1}; \theta^-) + \xi \quad (21)$$

其中 ξ 表示策略网络噪声,并服从截断正态分布,其可以提高算法的鲁棒性,同时又可保证噪声不会过大.

将式(15)的 TD 目标重新定义为:

$$\hat{y}_j = \min\{\hat{y}_{j,1}, \hat{y}_{j,2}\} \quad (22)$$

$$\hat{y}_{j,1} = R_j + \gamma \cdot q(s_{j+1}, \hat{a}_{j+1}^-; \omega_1^-)$$

$$\hat{y}_{j,2} = R_j + \gamma \cdot q(s_{j+1}, \hat{a}_{j+1}^-; \omega_2^-)$$

其中 $\hat{y}_{j,1}, \hat{y}_{j,2}$ 即为用两个不同目标价值神经网络估计的 TD 目标.在训练过程中,目标神经网络采用软更新的方法.

各神经网络间迭代关系如图 2 所示.

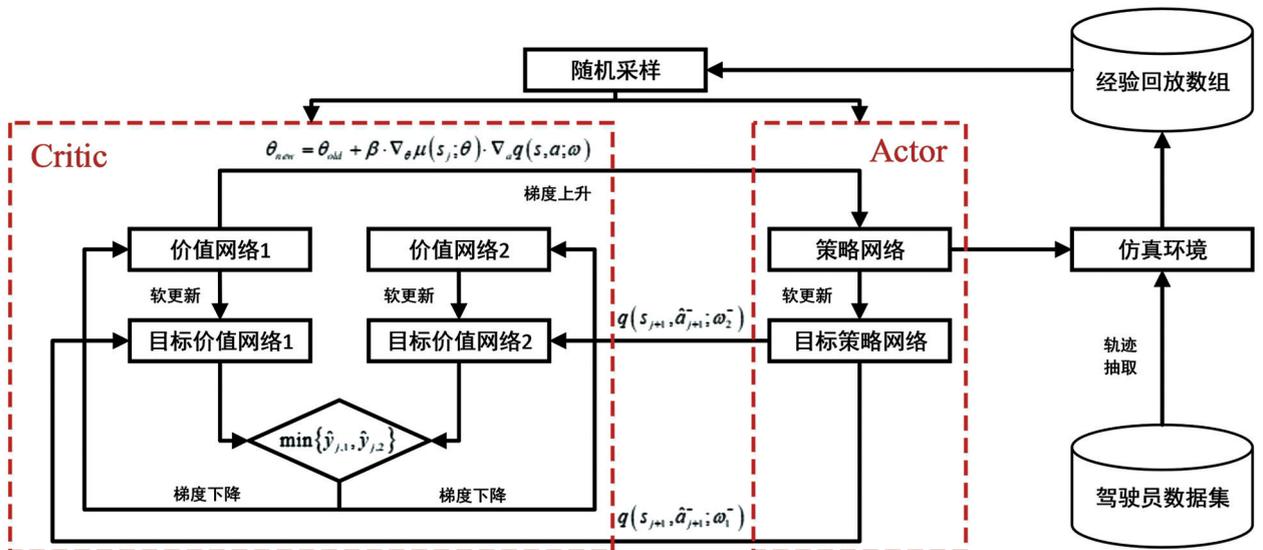


图 2 基于 TD3 深度强化学习的网络结构

为确保价值网络对策略网络打分的可靠程度,减少策略网络的更新程度,在训练过程中,每一轮更新

一次价值网络参数,间隔两轮更新一次策略网络参数.

基于 TD3 的个性化跟驰控制策略训练伪代码如表 1 所示.

表 1 基于 TD3 的个性化跟驰控制策略伪代码

算法: 基于 TD3 的个性化跟驰控制策略

初始化策略网络参数 θ , 价值网络参数 ω_1, ω_2 及经验回放数组 D

初始化目标策略网络参数和目标价值网络参数 $\theta^- = \theta, \omega_1^- = \omega_1, \omega_2^- = \omega_2$

开始循环

从驾驶员跟驰数据集中抽取一段跟驰数据, 采集状态 s , 根据策略网络选择动作 $a = \mu(s) + \epsilon$ 并执行

采集奖励 R , 下一步状态 s' , 终止条件判定逻辑值 d

将 (s, a, r, s', d) 存入经验回放数组 D

if $d = 1$

重置环境状态, 重新抽取跟驰轨迹.

从经验回放数组随机采样

按式(21)计算 TD 目标

按式(17)更新价值网络参数

每隔 t 轮

按式(11)更新策略网络参数

更新目标网络参数

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

$$\omega_i^- = \tau\omega_i + (1-\tau)\omega_i^-$$

结束循环

3 仿真与分析

为验证本研究提出的跟驰控制策略的有效性, 并体现个性化程度, 本节选取两种控制算法进行仿真对比, 个性化跟驰控制策略训练方法如表 1 所示. 未考虑驾驶员习性的跟驰控制策略采用同样的网络架构, 也未考虑式(7)所示的奖励函数, 训练过程中也未采用实际驾驶员数据. 选择一段跟驰片段进行对比, 仿真采样间隔为 0.1 s. 前车以 18 km/h 的初速度前进, 随后进行先加速再减速的随机变速运动, 前车车速变化如图 3 所示.

主车初始速度与前车同为 18 km/h, 两车初始相对距离为 20 m, 跟驰期望车速定义为 60 km/h. 两种跟驰控制策略的对比以及驾驶员实际驾驶数据如图 4—图 6 所示.

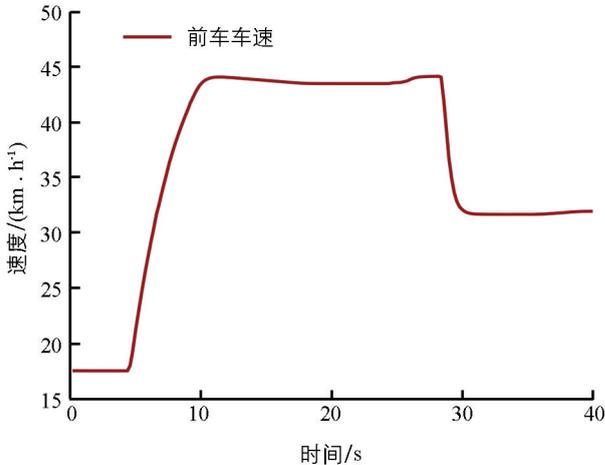


图 3 前车车速变化

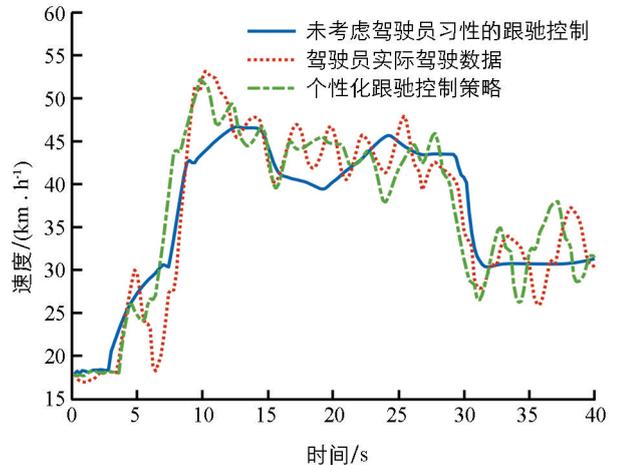


图 4 主车跟驰速度对比结果

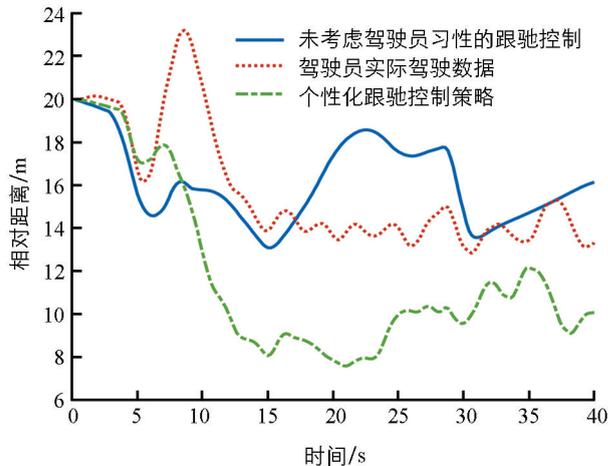


图 5 跟驰相对距离对比结果

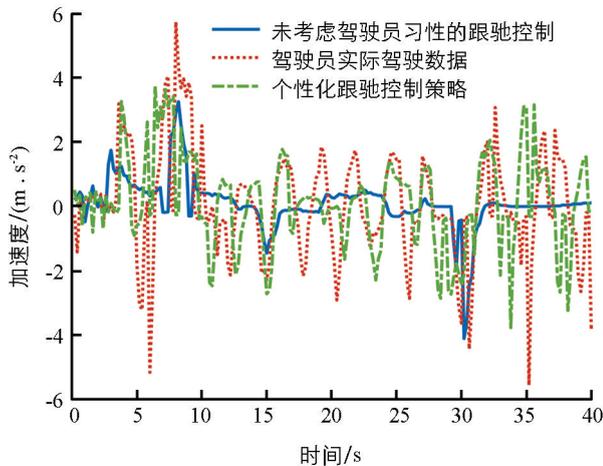


图 6 主车加速度对比结果

图 4—图 6 分别为跟驰过程中主车车速、相对距离以及加速度变化。可以看出, 无论是否考虑驾驶员习性, 基于深度强化学习的跟驰控制算法均能稳定、安全地实现自主跟驰功能。未考虑驾驶员习性的跟驰控制策略仅以客观跟随性、安全性和舒适性作为奖励函数, 因此在整个跟驰过程中, 无论是主车车速和加速度变化均较平稳。车辆加速度在大部分时间内维持在 $-1 \sim 1 \text{ m/s}^2$ 之内, 均方根值 0.86 m/s^2 , 最小跟驰距离为 13.7 m 。从实际驾驶数据可以看出, 该驾驶员驾驶风格偏“激进”类型, 在实际跟驰过程中, 频繁加减速, 导致车速波动较大, 加速度变化也更频繁, 车辆最大加速度绝对值达到 5.6 m/s^2 , 均方根值为 1.73 m/s^2 , 而整个跟驰过程的车辆相对距离也更近。当考虑了该驾驶员的驾驶习性后, 本研究提出的个性化跟驰控制策略相较未考虑驾驶员习性的跟驰控制策略更加“激进”, 车速波动和加速度波动更剧烈, 和实际驾驶数据趋势更加接近, 加速度均方根值为 1.57 m/s^2 。另外, 从相对距离可以看出, 个性化跟驰控制策略的跟驰距离更近, 最小跟驰距离达到了 7.8 m 。因此, 本研究提出的自主跟驰控制策略通过驾驶员实际驾驶数据, 实现了一定程度的个性化驾驶, 相较基于客观目标奖励函数的跟驰控制策略, 更能迎合驾驶员的驾驶习性和偏好。但从另一方面可以看出, 由于本研究提出的个性化跟驰控制策略并未对驾驶员的驾驶技能专业程度进行评价, 仅考虑了驾驶员的驾驶风格, 因此实际决策结果与参考驾驶员驾驶的熟练度相关。当参考驾驶员的驾驶能力较差时, 基于学习的决策难以使得车辆跟驰过程的平顺性、舒适性达到最优。

4 结论

本研究针对车辆自主跟驰过程中的宜人性问题, 首先基于 CARLA 模拟器搭建了模拟驾驶试验平台, 通过设计城市跟驰仿真场景, 获取了驾驶员实际驾驶数据。然后建立了车辆跟驰动力学模型, 并综合考虑车辆跟驰过程的安全性、舒适性、跟随性和驾驶员习性, 设计了相应的奖励函数。最后结合驾驶员试驾数据, 采用了改进 DDPG 算法对自主跟驰控制策略进行了训练。仿真结果表明, 本研究提出的个性化跟驰控制策略能够实现稳定、安全地车辆跟驰控制。同时, 通过在奖励函数中添加驾驶员偏好特性, 能够有效地提高自主跟驰控制的个性化程度, 使得决策结果更倾向于驾驶员的驾驶行为特性。

参考文献:

- [1] YI K, WOO M, KIM S H, et al. An Experimental Investigation of a CW/CA System for Automobile using Hardware in the Loop Simulation [C]. San Diego: Proceedings of the American Control Conference, 1999.
- [2] YIZHEN Z, ANTONSSON E K, GROTE K. A New Threat Assessment Measure for Collision Avoidance Systems [C]. Toronto: Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference, 2006.
- [3] AOUBE G S, LUDERS B D, LEE K, et al. Threat Assessment Design for Driver Assistance System at Intersections [C]. Madeira Island: Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, 2010.

- [4] MASUMI N, RAKSINCHAROENSAK P, NAGAI M. Study on Forward Collision Warning System Adapted to Driver Characteristics and Road Environment [C]. Seoul: Proceedings of the 2008 International Conference on Control, Automation and Systems, 2008.
- [5] ZHANG R, LI K, HE Z, et al. Advanced Emergency Braking Control Based on a Nonlinear Model Predictive Algorithm for Intelligent Vehicles [J]. Applied Sciences, 2017, 7(5): 504.
- [6] 李霖, 朱西产, 董小飞, 等. 自主紧急制动系统避撞策略的研究 [J]. 汽车工程, 2015, 37(2): 168-174.
- [7] GERDES J C, HEDRICK J. Vehicle Speed and Spacing Control Via Coordinated Throttle and Brake Actuation [J]. Control Engineering Practice, 1997, 5(11): 1607-1614.
- [8] KIM H, KIM D, SHU I, et al. Time-varying Parameter Adaptive Vehicle Speed Control [J]. IEEE Transactions on Vehicular Technology, 2016, 65(2): 581-588.
- [9] LI S, LI K Q, RAJAMANI R, et al. Model Predictive Multi-objective Vehicular Adaptive Cruise Control [J]. IEEE Transactions on Control Systems Technology, 2011, 19(3): 556-566.
- [10] REN Y, ZHENG L, YANG W, LI Y N. Potential Field Based Hierarchical Adaptive Cruise Control for Semi-autonomous Electric Vehicle [J]. The Proc. IMechE, Part D: Journal of Automobile Engineering, 2019, 233(10): 2479-2491.
- [11] 罗颖, 秦文虎, 翟金凤. 基于改进 DDPG 算法的车辆低速跟驰行为决策研究 [J]. 测控技术, 2019, 38(9): 19-23.
- [12] 朱冰, 蒋渊德, 赵健, 等. 基于深度强化学习的车辆跟驰控制 [J]. 中国公路学报, 2019, 32(6): 53-60.
- [13] 管欣, 王景武, 高振海. 基于最优预瞄加速度决策的汽车自适应巡航控制系统 [J]. 吉林大学学报(工学版), 2004, 34(2): 190-193.
- [14] YI K, MOON I K. A Driver-adaptive Stop-and-go Cruise Control Strategy [C]. Taipei: IEEE International Conference on Networking, Sensing and Control, 2004.
- [15] ZHU M, WANG X, WANG Y. Human-like Autonomous Car-following Model with Deep Reinforcement Learning [J]. Transportation Research Part C: Emerging Technologies, 2018, 97: 348-368.
- [16] CHEN X, ZHAI Y, LU C, et al. A Learning Model for Personalized Adaptive Cruise Control [C]. Redondo Beach: 2017 IEEE Intelligent Vehicles Symposium(IV), 2017.
- [17] MARTINEZ J J, CANUDAS C. A Safe Longitudinal Control for Adaptive Cruise Control and Stop-and-go Scenarios [J]. IEEE Transactions on Control Systems Technology, 2007, 15(2): 246-258.
- [18] 边明远. 考虑驾驶员个体特性的汽车安全行驶间距模型 [J]. 中国机械工程, 2010, 21(12): 1502-1506.
- [19] 陈涛, 郭丛帅, 李旭川, 等. 适应驾驶人的个性化自适应巡航控制策略 [J]. 重庆理工大学学报(自然科学), 2021, 35(7): 1-9.
- [20] JIANG Y, DENG W, WANG J. Studies on Drivers' Driving Styles Based on Inverse Reinforcement Learning [N]. SAE Paper, 2018-01-06(12).
- [21] 唐明弘. 基于深度强化学习的拟人化自适应巡航控制算法设计 [D]. 长春: 吉林大学, 2020.
- [22] 余志生. 汽车理论: 第6版 [M]. 北京: 机械工业出版社, 2018.
- [23] 杨威. 高速跟车工况下智能汽车个性化驾驶决策方法研究 [D]. 重庆: 重庆大学, 2020.
- [24] 裴晓飞, 莫烁杰, 陈祯福, 等. 基于 TD3 算法的复杂交通环境自动驾驶汽车换道研究 [J]. 中国公路学报, 2021, 34(11): 246-254.

责任编辑 欧宾