

基于深度强化学习的自动驾驶车辆 与行人交互建模

胡伟超^{1,2}, 杨镇铭³, 于鹏程², 陈艳艳¹, 马社强³

(1. 北京工业大学 城市建设学部, 北京 100124; 2. 公安部道路交通安全研究中心, 北京 100062; 3. 中国人民公安大学 交通管理学院, 北京 100091)

摘要: 为满足自动驾驶车辆安全、高效地与行人进行交互, 保护行人安全, 本文使用多智能体深度确定性策略梯度算法建立自动驾驶车辆和人工驾驶车辆混行下的人车交互模型并求解交互策略, 使自动驾驶车辆能够在不依赖通信的前提下避免事故发生。将本文算法与其他基线算法对比, 在训练效果、碰撞率和通行效率方面均有显著提高, 同时将本文模型在不同风险等级的场景中进行实验, 结果表明: 随着行人行为噪声强度的增加, 两种车辆的通行效率降低, 而自动驾驶车辆的碰撞率出现先增加后降低的趋势, 在高噪声强度下自动驾驶车辆的避碰能力比人工驾驶车辆强, 更好地保护了行人的安全。

关键词: 交通运输系统工程; 自动驾驶车辆; 人车交互; 深度强化学习; 多智能体系统

中图分类号: U491 **文献标志码:** A **文章编号:** 1671-5497(2025)10-3180-09

DOI: 10.13229/j.cnki.jdxbgxb.20240017

Modeling interaction policy of autonomous vehicle and pedestrian based on deep reinforcement learning

HU Wei-chao^{1,2}, YANG Zhen-ming³, YU Peng-cheng², CHEN Yan-yan¹, MA She-qiang³

(1. Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, Beijing 100124, China; 2. Research Institute for Road Safety of the Ministry of Public Security, Beijing 100062, China; 3. School of Traffic Management, People's Public Security University of China, Beijing 100091, China)

Abstract: To facilitate safe and efficient interactions between Autonomous Vehicles (AVs) and pedestrians, this study employs the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm to establish a pedestrian-vehicle interaction model in a mixed traffic context that includes both autonomous and human-driving vehicles. This model formulates interaction strategies enabling AVs to avert accidents without the necessity of direct inter-vehicle communication. In comparison with several benchmark algorithms, the proposed algorithm demonstrates substantial improvements in terms of training efficacy, collision frequency reduction, and traffic capacity. Additionally, the robustness of the proposed model is assessed across varied risk scenarios. Findings reveal that as the intensity of pedestrian behavioral randomness, or behavioral noise rises, the duration of interaction delays of both vehicle categories

收稿日期: 2024-01-03.

基金项目: 国家重点研发计划项目(2020YFB1600304).

作者简介: 胡伟超(1987-), 男, 副研究员, 博士. 研究方向: 智能交通, 交通安全. E-mail: dfengwin@163.com

increases. Remarkably, the collision rate of AVs initially increases before declining, indicating an adaptive learning phase. Under conditions of elevated noise, AVs exhibit a superior capability for collision avoidance compared to human-driving vehicles, highlighting their enhanced resilience in chaotic urban traffic conditions. These outcomes underscore the potential of MADDPG-based frameworks to significantly contribute to safer, more efficient AV integration in mixed traffic scenarios.

Key words: engineering of transportation system; autonomous vehicle; interaction of vehicle and pedestrian; deep reinforcement learning; multi-agent system

0 引言

根据最新数据统计^[1],2022年我国因道路交通事故死亡的行人数量达18 088人,占道路交通事故死亡总人数的29.81%,因此对行人的保护至关重要。在现实中,虽然交通法规规定了人行横道处车辆有避让行人的义务,但是现实中驾驶人仍存在不避让行人的情况^[2],导致行人过街时需要与车辆进行博弈交互,以确定哪一方优先通行^[3]。随着人工智能、传感器、信息通讯等技术的发展,自动驾驶技术有望成为减少人车事故、保护行人安全的新方法^[4]。自动驾驶车辆(Autonomous vehicle, AV)配备了行人避碰系统^[5],通过实时评估车辆-行人的交互过程,以保护行人的安全。鉴于AV能够更好地保护行人安全,因此对AV与行人之间的交互研究非常重要^[4]。

在以往的研究中,Camara等^[6]用博弈论方法对人车交互过程进行建模,其中AV使用期望最大化(Expectation-maximum, EM)算法进行制动决策,降低了人车交互碰撞风险,但其场景较为简单,涉及变量较少。Chae等^[7]使用深度强化学习(Deep reinforcement learning, DRL)方法把车辆视为智能体,处理AV的人车交互过程,能够处理较多的变量及其复杂关系,但Chae等设计的AV仅包含简单的刹车任务,忽略了其他行为方式且AV通行效率不高。在此基础上,Papini等^[8]引入了速度控制,但仅涉及单个AV与行人的交互,未考虑多车环境的影响,而研究表明^[9,10]现实中车辆处于跟驰状与自由行驶状态下与行人交互时会采取不同策略。

在目前的研究中还存在以下方面的不足:①在AV与行人交互过程中,缺少多车状态下的人车交互策略研究,AV的人车交互策略应考虑在多智能体(Multi-agent)系统下求解;②自动驾驶技术发展需要持续迭代更新,势必经历一个漫

长的自动驾驶车辆与人工驾驶车辆(Human driving vehicle, HDV)混合行驶的过渡阶段^[11-14],在以往的AV-行人交互研究中往往忽略了HDV的交互影响。

在算法方面,强化学习领域已有不少针对单智能体的成功算法,如深度Q网络^[15](Deep Q-Network, DQN)和深度确定性策略梯度^[16](Deep deterministic policy gradient, DDPG)等。但这些算法难以处理MA问题:不仅需要考个体行为的优化,还要考虑智能体之间的交互及其对整体性能的影响。在此背景下,OpenAI提出了多智能体深度确定性策略梯度(Multi-agent deep deterministic policy gradient, MADDPG)算法^[17]用以解决多智能体问题,利用中心化的critic学习器实现多智能体在无通信的条件下进行决策。为解决前文所述的不足,本文对AV与HDV混行的场景下的车辆-行人交互进行建模,使用MADDPG算法求解AV、HDV和行人三类智能体的交互策略,以指导AV能够避免碰撞的同时保持较高的通行效率。

1 算法框架

1.1 深度确定性策略梯度算法

深度强化学习结合了强化学习和深度神经网络,使智能体能够处理高维复杂的状态空间。其基本组成部分包括以下内容:“智能体”是与环境交互并根据策略采取行动的学习者;“策略”是将智能体状态映射到动作的函数;“环境”是智能体的外部系统,智能体与环境交互并从中获取状态信息和奖励;“奖励”是智能体在采取行动后获得的奖赏值,需要根据实际问题设置,比如车辆智能体发生碰撞或停滞不前会扣除奖励;“状态”是智能体所处的外部环境信息;“动作”是智能体在感知到外部环境后所要采取的行为;“动作价值函数”也称Q值函数,是智能体在某一策略下对未

来奖励的预期,用于评估智能体不同动作下可能的效果,并指导智能体学习;“轨迹”是描述智能体与环境互动的历史记录,储存在经验池中。

深度确定性策略梯度^[18]是结合了深度 Q 网络和 actor-critic 方法的深度强化学习算法,其中策略函数由 actor 神经网络生成,用来根据状态信息生成动作,动作价值函数由 critic 神经网络生成,用来评估 actor 的动作价值。在人车交互问题中,一个训练回合内智能体往往要与环境交互上百次,每次交互的动作空间又是连续的,无法直接计算未来所有动作的 Q 值,因此在 DDPG 中,actor 网络输出动作后,用 critic 网络估计 Q 值,该估计值用于更新 actor 网络的参数,随着训练的进行,critic 的估计值将逐渐逼近真实 Q 值。经验池会储存每个训练回合的历史轨迹,用于更新 critic 网络的参数,DDPG 工作流程如图 1 所示。在本文探讨的 AV、HDV 与行人交互场景涉及多个能够自主决策的行为主体,需充分考虑 MA 系统的特点。

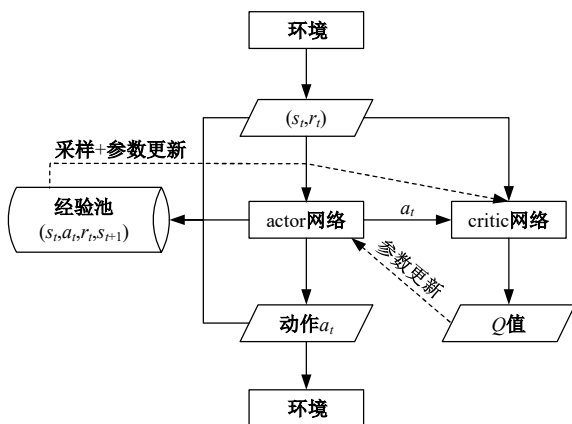


图 1 DDPG 工作流程图

Fig. 1 DDPG working flow chart

1.2 多智能体系统

Kiran 等^[19]总结了相关深度强化学习方法,以往的研究在处理多智能体(Multi-agent, MA)系统问题时,大多直接使用 DDPG 算法建模,如 Vasquez 等^[20]基于 DDPG 算法设计了多目标主动刹车系统。这类算法在解决 MA 系统问题时会出现训练环境不稳定的情况,训练效率低下,并且随着智能体数量的增多,方差急剧增大。强化学习的马尔科夫决策过程要求智能体所处的环境具有稳定性和可观测性,即能够准确计算每次交互后环境的状态变化,这样才能保证 Q 值预测的准确性。但在 MA 环境中,对其中的某一智能体来说,

其他智能体的行为都是环境的一部分,当智能体不能通信时,该智能体无法计算其他智能体对环境的影响,即在状态 s_t 下,智能体在执行动作 a_t 后, s_{t+1} 不能准确得出,因此不满足传统强化学习的收敛条件,导致方差大、训练效果不佳。以往在用 DDPG 算法解决 MA 问题时有两种思路:一是直接对每个智能体分别建立各自的动作、状态空间以及 actor-critic 网络,这样会导致上文提到的高方差问题;二是将所有智能体的动作(状态)建立在同一个动作(状态)空间中,相当于一个中央大脑对所有智能体集中控制,这种方法虽然能在一定程度上降低方差,但是要求所有智能体完全共享参数,并对全局环境处于全知状态,无法应用于现实问题中。

MADDPG 算法是一种改进的 DDPG 算法,可用于求解 MA 系统难题。该算法通过改进 actor-critic 网络,使智能体不依靠信息共享,仅通过自身观测的信息进行合作或竞争。MADDPG 算法将单一智能体的 DDPG 算法应用到 MA 系统中,克服了智能体状态变化不稳定、值函数方差高等缺陷。

MADDPG 的特点是集中式学习,分布式执行。其中, critic 网络用全局信息估计每个智能体的 Q 值,而 actor 网络只根据智能体的局部信息输出动作,即训练时使用全局信息优化神经网络参数,模型收敛后,智能体的决策不再依赖全局信息。既保证了训练时 Q 值预测的准确性,又满足了智能体不再依靠信息共享的现实情况。因此,在 MADDPG 中,所有智能体共用同一 critic 网络,但 actor 网络是每个智能体独有的。

式(1)为智能体 i 的策略梯度,通过梯度上升(或下降)来更新 actor 网络:

$$\nabla_{\theta_i} J(\mu_i) =$$

$\mathbb{E}_{s, a \sim D} [\nabla_{\theta_i} \mu_i(a_i | o_i) Q_i^{\mu}(x, a_1, \dots, a_n) |_{a_i = \mu_i(o_i)}]$ (1)
式中: θ_i 为 actor 网络参数; μ_i 为第 i 个智能体的策略; D 为回放缓冲区(replay buffer); Q 为动作价值函数, a_i 为智能体 i 的动作; o_i 为智能体 i 的局部观测; $\mathbb{E}_s$ 为状态 s 的子集; $\mu_i(a_i | o_i)$ 为智能体 i 的策略函数; x 是所有环境信息,可能包括状态之外的有效信息。

式(2)为目标值与实际值的误差方程,通过最小化误差来更新 critic 网络:

$$L(\theta_i) = \mathbb{E}_{x, a, r, x'} [(Q_i^{\mu}(x, a_1, \dots, a_n) - y)^2]$$

$$y = r_i + \gamma Q'_i(x', a'_1, \dots, a'_n) |_{a'_i = \mu'_i(o_i)} \quad (2)$$

式中: θ_i 为 critic 网络参数; μ'_i 为目标网络智能体 i 的动作函数; r_i 为智能体 i 在状态 s_i 下的奖励; γ 为折扣因子。

2 实验设计

2.1 智能体状态空间及动作空间设计

2.1.1 HDV 状态空间及动作空间

状态空间指智能体所有可能的状态集合, actor 网络利用状态空间传入的参数进行决策。文献[21]回顾了以往驾驶行为建模的研究, 本文参考传统建模方法所用到的变量设计了智能体的状态空间, 包括前车速度、主车速度和主车加速度等变量。此外, 为了使车辆在安全和效率之间权衡, 状态空间还加入了与风险、效率和场景有关的变量。考虑到多车环境下, 车辆处于跟驰状态和自由行驶状态下的让行策略不同^[9,10], 本文分别为跟驰和自由行驶两种不同状态下的车辆设计不同的状态空间。式(3)为 HDV 的状态空间:

$$s_t^{\text{mv}} = \begin{cases} [TTC_t, |v_t^{\text{lv}}|, LV_{\text{type}}, |v_t^{\text{hdv}}|, |a_t^{\text{hdv}}|, R_{\text{time}}, \Delta x_t^{\text{rel}}, L_{\text{num}}, F_t, HF], & \text{当跟车时} \\ [TTC_t, |v_t^{\text{pe}}|, |v_t^{\text{hdv}}|, |a_t^{\text{hdv}}|, R_{\text{time}}, \Delta x_t^{\text{rel}}, L_{\text{num}}, PDTC_t, b_{\text{str}}, b_{s-\text{str}}], & \text{当自由行驶时} \end{cases} \quad (3)$$

式中: s_t^{mv} 为 t 时刻 HDV 的状态; TTC_t 为 t 时刻的碰撞时间; $|v_t^{\text{lv}}|$ 为 t 时刻前车的速度; LV_{type} 为前车类型, 包括 HDV 和 AV 两种; $|v_t^{\text{hdv}}|$ 为 t 时刻 HDV 智能体的速度; $|a_t^{\text{hdv}}|$ 为 t 时刻 HDV 智能体的加速度; R_{time} 为驾驶人反应时间, 用于模拟人类驾驶行为的滞后性, 本文设为 1.7 s ^[22]; Δx_t^{rel} 为 t 时刻智能体与前车或行人的相对距离; L_{mun} 为智能体所在车道; F_t 为智能体跟车时长; HF 为人为因子设为 1, 在所用仿真环境中用于得到相对固化的策略, 使其不易受多智能体策略的影响; $|v_t^{\text{pe}}|$ 为 t 时刻行人速度; $PDTC_t$ 为行人过街剩余距离; b_{str} 为街道宽度, $b_{s-\text{str}}$ 为行人行走方向。HDV 的动作空间参考文献[23]设计, 由典型的加速度值构成的集合 $a_t^{\text{hdv}} \in \{-9.8, 5.8, -3.8, 3, 1, 0\} \text{ m/s}^2$ 。

2.1.2 AV 状态空间及动作空间

AV 状态空间与 HDV 状态空间类似, 如式(4)所示:

$$s_t^{\text{cav}} =$$

$$\begin{cases} [TTC_t, |v_t^{\text{lv}}|, LV_{\text{type}}, |v_t^{\text{cav}}|, |a_t^{\text{cav}}|, \Delta x_t^{\text{rel}}, L_{\text{num}}, F_t], & \text{当跟车时} \\ [TTC_t, |v_t^{\text{pe}}|, |v_t^{\text{cav}}|, |a_t^{\text{cav}}|, \Delta x_t^{\text{rel}}, L_{\text{num}}, PDTC_t, b_{\text{str}}, b_{s-\text{str}}], & \text{当自由行驶时} \end{cases} \quad (4)$$

式中: s_t^{cav} 为 t 时刻 AV 智能体的状态; $|v_t^{\text{cav}}|$ 为 t 时刻智能体的速度; $|a_t^{\text{cav}}|$ 为 t 时刻智能体的加速度。动作空间 U_{vehicle} 为智能体加速度的集合: $a_t^{\text{cav}} \in \{-9.8, 5.8, -3.8, 3, 1, 0\} \text{ m/s}^2$ 。

2.1.3 行人智能体策略

考虑到现实中驾驶人存在不避让行人的情况^[2], 行人需要寻找合适的车辆间隙完成过街^[24], 根据行人过街时的 TTC 来定义车辆间隙是否满足行人的过街要求, 进而定义行人步行策略^[25], 如式(5)所示:

$$\text{Agent}_t^{\text{pe}} = \begin{cases} \text{通行}, & \text{当 } TTC > 3 \text{ s} \\ \text{等待}, & \text{当 } TTC \leq 3 \text{ s} \end{cases} \quad (5)$$

行人的初速度从 $v_{\text{pe}} \in \{1.15, 1.39, 1.48, 1.54, 1.56\} \text{ m/s}$ ^[26] (城市道路中行人的典型步行速度)中随机抽样, 若不满足过街条件 ($TTC \leq 3 \text{ s}$), 行人至少等待车辆通过人行横道 4 m 后再考虑是否通行^[23]。

2.2 奖励函数

总体的奖励函数 R' 由行人奖励 r_t^{pe} 和车辆奖励 r_t^{ve} 组成, 如式(6)所示:

$$R' = r_t^{\text{pe}} + r_t^{\text{ve}} \quad (6)$$

本文从安全和效率两方面来考虑智能体的策略, 要求在较低噪声强度下车辆智能体能够在避免碰撞的前提下尽快完成交互过程。行人奖励函数 r_t^{pe} 由式(7)定义:

$$r_t^{\text{pe}} = -a - \begin{cases} \delta, & \text{碰撞发生时} \\ 0, & \text{无碰撞时} \end{cases} \quad (7)$$

式中: a 为常数惩罚项; δ 为碰撞惩罚项。为确保智能体的通行效率, 此处通过设置常数惩罚项 $a = 0.01$ 使智能体每经过一个训练步长都会得到一定的惩罚, 同时考虑到现实中车辆与行人碰撞的后果远大于车辆因避让行人造成的延误, 此处设置碰撞惩罚项 $\delta = 10$ 。

车辆奖励函数 r_t^{ve} 由式(8)定义:

$$r_t^{\text{ve}} = -b - \begin{cases} \partial, & \text{碰撞发生时} \\ 0, & \text{无碰撞时} \end{cases} - \begin{cases} \phi, & \text{超速时} \\ 0, & \text{未超速时} \end{cases} \quad (8)$$

式中: b 为常数惩罚项, $b = 0.01$; ∂ 为碰撞惩罚

项, $\partial = 10$; ϕ 为超速惩罚项, 因实验并未对车辆的动力传动系统建模, 为保证速度输出不会超过车辆动力传动的上限并使车辆遵守交通规则, 此处设置超速惩罚项 $\phi = 0.05$ (必要时车辆能够超过设置的限速)。

2.3 仿真设置与策略训练

本文在 Highway-env 仿真环境中创建场景并对算法进行训练。实验中车长为 5 m, 车宽为 2 m, 车辆初始速度从 15~50 km/h 中随机抽样, 道路宽度为 10 m, 选择距离交互位置 100 m 路段进行结果输出。

为充分保障行人安全, 仿真中在判断人车相撞时增加了 0.5 m 的安全余量, 即车辆与行人距离不足 0.5 m 时, 认为碰撞发生:

$$x^{\text{hdv}} - x^{\text{pe}} < 0.5 \text{ 或 } x^{\text{av}} - x^{\text{pe}} < 0.5 \quad (9)$$

式中: x^{av} 、 x^{hdv} 和 x^{pe} 分别为仿真中 AV、HDV 和行人的位置, 满足不等式(9)时则判定碰撞发生。

车辆的目标位置为超过人行横道后的 10 m 处, 行人目标位置为道路对面。其他训练超参数的设置如下: 折扣因子为 0.8, 小批量为 128, 更新

因子为 0.9, 最大步长为 2 000, 学习率为 0.000 1, 学习率衰减值为 0.001, 隐藏层神经元 64 个, 优化器为 Adam。训练回合开始时, AV 和 HDV 在起点的 3 条车道内随机生成, 行人在距离路缘一定范围内生成。当所有智能体达到目标位置或仿真步长达到 2 000 时回合判定结束。

3 结果分析

3.1 算法比较

本文将 MADDPG 与 DDPG、DQN 作对比, 选择奖励(得分)、回合平均速度、碰撞率作为评价指标, 比较模型的学习效率与效果。奖励是强化学习的重要评估指标, 能够反映智能体的行为效果。为使结果更直观, 本文把奖励值百分化至 0~100 的得分数值, 结果如图 2 所示。本文方法在 3 000 次训练后得到了近似收敛的策略, 而 DDPG 和 DQN 由于训练的不稳定性导致其在 2 000 次训练回合后仍未收敛, 且训练得分方差非常大。MADDPG 集中式的 critic 网络使智能体在相同的训练次数下获得的得分明显高于 DDPG 和 DQN。

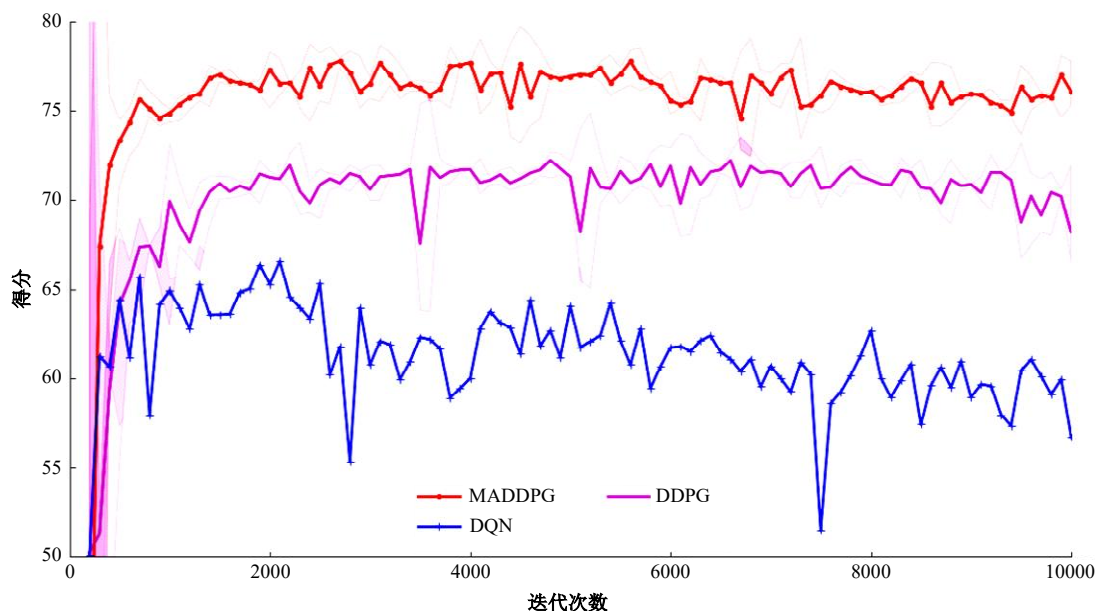


图 2 训练奖励对比

Fig. 2 Training reward comparison

图 3 为训练后车辆的速度变化及滚动误差对比, 反映了车辆的通行效率和车内乘客的舒适性。本文方法在人车交互任务中保持了较高的平均速度, 车辆在避让行人时速度不会损失太多。相比于 DDPG 和 DQN 的表现, MADDPG 的速度滚动误差更小, 也就是减速幅度更小, 不会因为剧烈的

加减速给车内乘客带来不良体验。因此, MADDPG 训练的 AV 能够在保护行人的同时快速完成交互任务且速度变化比 DDPG 和 DQN 更平缓。

为模拟现实中行人行为的随机性、不确定性, 此处行人的状态空间中增加噪声处理。此时行

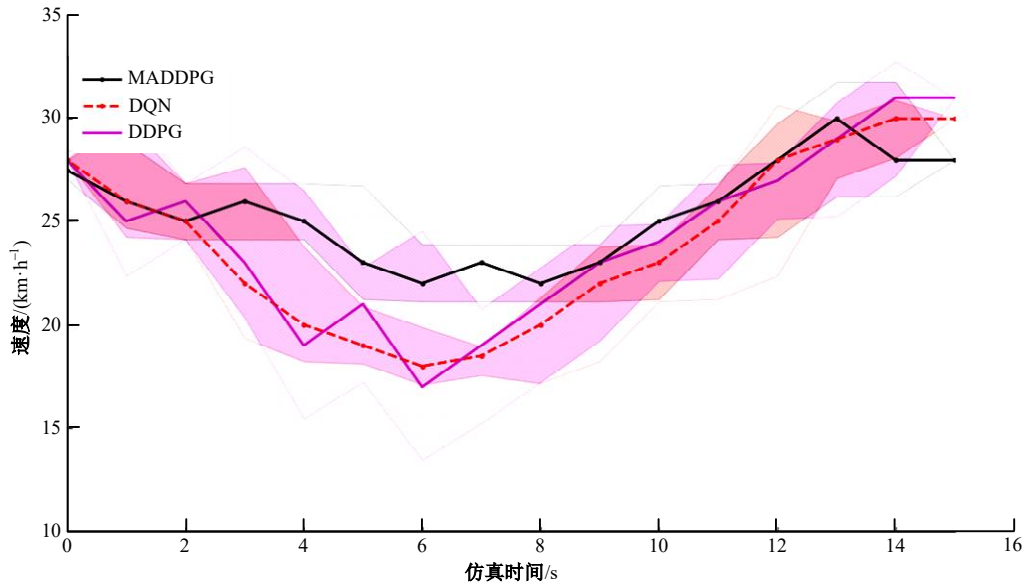


图 3 速度对比

Fig. 3 Velocity comparison

人对环境状态的观测值是真实值与随机动态噪声变量的乘积,如式(10)所示:

$$z_t = (1 + n_t) \cdot s_t \quad (10)$$

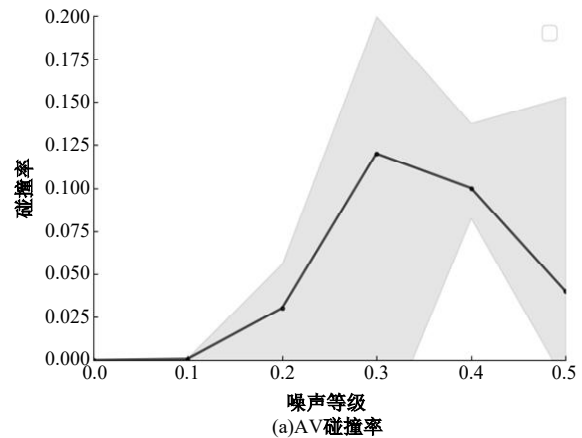
式中: z_t 为智能体的观测状态值; s_t 为智能体的实际状态值; n_t 为遵循高斯分布 $n_t \sim N(0, \alpha^2)$ 的噪声变量, α 为噪声强度。当 α 较高时,行人因获取带偏差的信息,因此不会按照预定的策略过街,从而使行人的行为带有一定的随机性。

加载收敛后的模型在有噪声的场景下指导车辆进行多次人车交互任务,统计 MADDPG、DDPG 和 DQN 算法下的车辆碰撞率。在噪声强度 $\alpha_{pe}=0.1$,MADDPG 的碰撞率(0.07%)明显低于 DDPG(2.6%)和 DQN(13.0%),本文方法在行人行为不确定时仍保障了行人的过街安全,也反映了本文方法具有一定的泛化能力。

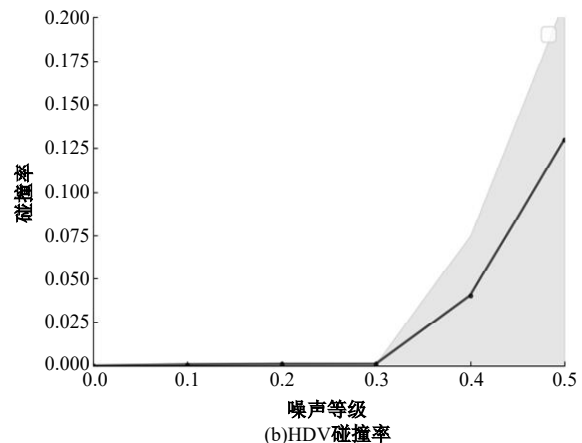
3.2 人车交互安全与效率评估

此部分评估智能体对安全风险和效率的平衡能力,风险指标为碰撞率,效率的指标为智能体的延误时间。此处在不同 α_{pe} 下用收敛后的模型进行实验并统计碰撞率,如图 4 所示。在较低的噪声强度下,经过训练的车辆智能体能够避免碰撞。随着噪声强度的增加,行人将不再完全按照预设定的策略行走,有概率在 $TTC < 3$ s 时过街,当距离过近以致车辆来不及完全制动时将发生碰撞,因此 AV/HDV 的碰撞率和延误时间有增大的趋势。比起 HDV,AV 更易受到多智能体策略的影响,因此其碰撞率较高,但当噪声强度继续增加

时,AV 的碰撞率出现了下降,该现象源于行人过街策略趋向于“不计后果”,行人智能体期望自动驾驶车辆能采取措施避免碰撞。结合后文的 AV 交互延误时间可知,在较高的噪声强度下 AV 为



(a) AV 碰撞率



(b) HDV 碰撞率

图 4 AV/HDV 碰撞率

Fig. 4 AV/HDV collision rate

获取较高的奖励而选择过于保守的策略导致延误时间大大增加。这种现象与文献[27]中提到的“人车演化博弈”现象一致,与较低的噪声强度相比,AV 更容易适应这种近乎确定性的策略。

此处计算了各智能体在不同风险强度下的延误时间,并把延误时间定义为智能体完成人车交互任务所需时间与智能体在不与其他智能体交互的情况下单独按原来的路线通过场景所需的时间之差,如式(11)所示:

$$D = T_{\text{agent}} - T_{\text{alone}} \quad (11)$$

式中: D 为人车交互延误; T_{agent} 为智能体在交互环境下到达目标位置所需时间; T_{alone} 为智能体单独完成任务所需时间。

如图 5 所示,当风险等级从 0.0 增加到 0.5 时,AV 与 HDV 的延误时间分别增加了 21% 和 13%,行人的延误时间分别降低了 19% 和 11%。

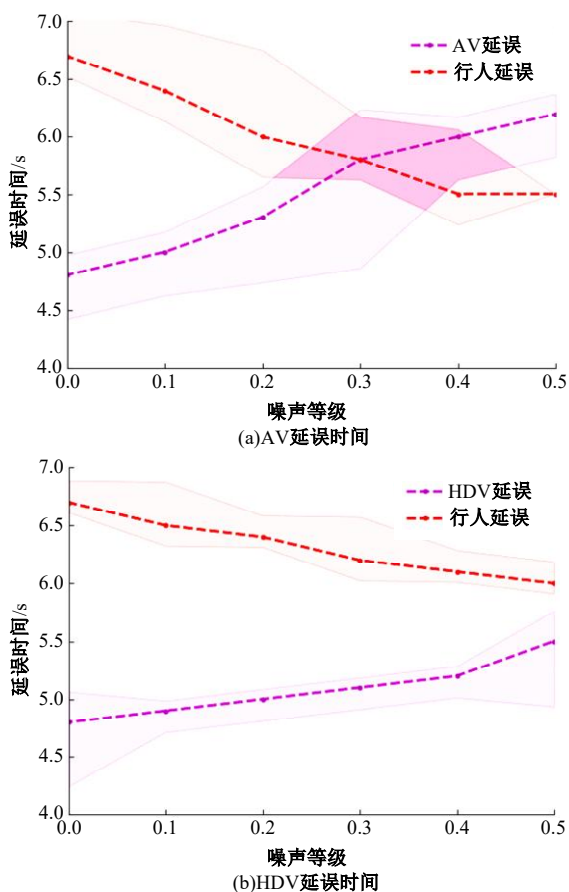


图 5 AV/HDV 延误时间

Fig. 5 AV/HDV delay

随着风险的不增加,AV 和 HDV 在人车交互中的策略越来越保守,因此两类车辆的延误时间均有增加趋势,而行人的延误时间反而减少。从图 5 中还可观察到 HDV 延误时间比 AV 延误时间

短,结合上文中对碰撞率的结果可得,HDV 因反应时间的限制不能快速反应行人行为的随机性变化,导致其策略变化不明显,虽然延误时间较低,但碰撞率较高,因此 HDV 在处理高风险下安全与是效率的取舍问题上不如 AV 表现好。

3.3 AV 行为分析

此处选取了两个 AV 智能体,生成于不同的车道上,进行人车交互实验,统计其速度变化,如图 6、图 7 所示。AV1 和 AV2 经历了先减速后加速再减速的过程,当任务开始时,AV1 和 AV2 在距离人行横道 100 m 处识别到了行人并开始减速,AV 会在以往的训练中得到经验,即经过一段时间的减速后,即使不再继续减速也不会与行人发生碰撞,为缩短任务时间,AV 在接下来的路程中适当加速,最后在人行横道前进行第二次减速。



图 6 智能体交互示意图

Fig. 6 Agent interaction diagram

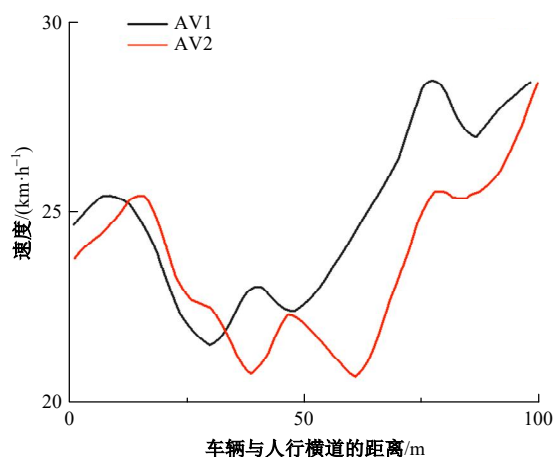


图 7 AV-行人交互过程中 AV1 和 AV2 的速度变化

Fig. 7 Velocity change of AV1 and AV2

4 本文模型的优点和局限性

综上所述,本文模型优点如下:

(1)在模型训练方面,MADDPG 集中训练、分散执行的框架使其在处理类似人车交互的 MA 系统问题上表现出更优越的性能,策略收敛速度明显优于 DDPG 和 DQN 算法。3 个算法都是利用历史经验 (s_t, a_t, r_t, s_{t+1}) 优化神经网络,而 DDPG

和DQN训练的智能体无法计算其他智能体对环境的影响,相同(s_t, a_t)下得出不同的 r_t 和 s_{t+1} ,且差别较大,增加了收敛难度。虽然当前研究在有限智能体规模下取得了成功,但是仍难以应对大规模的智能体系统,尤其是在高度复杂的实际交通环境中。

(2)在策略对比方面,MADDPG在人车交互任务中的碰撞事故率明显低于DDPG和DQN算法,能够更好地保护行人过街安全。同时,MADDPG在人车交互过程中的速度变化波动比其他算法稳定,能够带给乘客更好的体验感。3种算法都是只利用智能体的局部信息进行决策,但MADDPG的critic网络给智能体提供类似“大局观”的引导(critic评价网络由全局信息的历史经验得到),使其能够获得更好的人车交互策略。

(3)在车辆安全-效率分析方面,随着场景风险增加,AV、HDV训练的碰撞率有上升趋势,延误时间也不断增加,相反行人的延误时间有减少趋势。

(4)在行人行为随机性的影响方面,在较低的噪声水平下,AV更易受到多智能体策略的影响,其碰撞率较高,而在高噪声水平下,行人“不计后果”的过街行为对AV来说更容易适应,其策略随着噪声水平增高逐渐保守,降低了碰撞率但交互延误时间增加较大。

(5)在AV行为分析方面,AV在与行人交互过程中大致经历了减速-加速-减速的过程。任务开始时,AV识别行人后开始减速,充分减速后,AV判断接下来的路程中不会发生碰撞,因此开始加速,最后在人行横道前10~15 m,处第二次减速。AV通过对自身速度的精确控制使其在避免碰撞的同时缩短任务时间。

本文通过对AV/HDV-行人交互策略建模,为AV的行人避让系统提供了良好思路,在保护行人安全的同时提高了车辆通行的效率。但还存在以下局限性:

(1)MADDPG算法虽然在数量有限的智能体场景中解决了MA系统智能体状态不稳定的问题,但是随着智能体规模进一步扩大,如包含数十个具有自主决策能力的行人、不同流行的机动车和非机动车等,critic网络的输入维度急剧增长,可能会出现模型难以收敛的问题。

(2)由于实验场景较为简单,未考虑到其他

流向的车辆和非机动车的影响,因此还未能完全还原现实中的复杂情况。

5 结束语

基于MADDPG算法训练智能体的人车交互策略,改进了AV的行人避让决策,使车辆和行人能够既安全又快速地通过人行横道区域,同时避免AV对信息共享的依赖。本文模型还存在一定的局限性,在未来的工作中,将考虑使用更先进的训练方法,考虑如何在复杂场景下降低输入维度,以处理维度快速上升的问题;同时本文场景较为简单,忽略了视线遮挡、行人群体性行为等现实中存在的现象,未来也将考虑更真实的场景仿真。

参考文献:

- [1] 公安部交通管理局. 中华人民共和国道路交通事故统计年报(2022年度)[R/OL]. [2024-05-22]. <https://data.stats.gov.cn/index.htm>
- [2] 刘荣, 王凤兰, 吕良东. 基于改进复制动态演化博弈模型的行人与机动车冲突[J]. 科学技术与工程, 2020, 20(30): 12486-12491.
Liu Rong, Wang Feng-lan, Lyu Liang-dong. Game model of pedestrian-vehicle conflict based on improved replication dynamic evolution[J]. Science Technology and Engineering, 2020, 20(30): 12486-12491.
- [3] Gupta S, Vasardani M, Winter S. Negotiation between vehicles and pedestrians for the right of way at intersections[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(3): 888-899.
- [4] Kalatian A, Farooq B. Deepwait: pedestrian wait time estimation in mixed traffic conditions using deep survival analysis[C]//IEEE Intelligent Transportation Systems Conference(ITSC), Auckland, New Zealand, 2019: 2034-2039.
- [5] Schratte M, Hartmann M, Watzenig D. Pedestrian collision avoidance system for autonomous vehicles[J]. SAE International Journal of Connected and Automated Vehicles, 2019, 2(12): 279-293.
- [6] Camara F, Romano R, Markkula G, et al. Empirical game theory of pedestrian interaction for autonomous vehicles[C]//Proceedings of Measuring Behavior, Manchester, UK, 2018: 238-244.
- [7] Chae H, Kang C M, Kim B D, et al. Autonomous braking system via deep reinforcement learning[C]//IEEE 20th International Conference on Intelligent

- Transportation Systems(ITSC), Shanghai, China, 2017: 1-6.
- [8] Papini G P R, Plebe A, Da Lio M, et al. A reinforcement learning approach for enacting cautious behaviours in autonomous driving system: safe speed choice in the interaction with distracted pedestrians[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(7): 8805-8822.
- [9] Schroeder B J, Roupail N M. Event-based modeling of driver yielding behavior at unsignalized crosswalks[J]. *Journal of Transportation Engineering*, 2011, 137(7): 455-465.
- [10] Zhao J, Malenje J O, Wu J, et al. Modeling the interaction between vehicle yielding and pedestrian crossing behavior at unsignalized midblock crosswalks[J]. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2020, 73: 222-235.
- [11] 张健, 李青扬, 李丹, 等. 基于深度强化学习的自动驾驶车辆专用道汇入引导[J]. *吉林大学学报: 工学版*, 2023, 53(9): 2508-2518.
- Zhang Jian, Li Qing-yang, Li Dan, et al. Merging guidance of exclusive lanes for connected and autonomous vehicles based on deep reinforcement learning[J]. *Journal of Jilin University(Engineering and Technology Edition)*, 2023, 53(9): 2508-2518.
- [12] 秦严严, 王昊, 王炜. 智能网联环境下的混合交通流LWR模型[J]. *中国公路学报*, 2018, 31(11): 147-156.
- Qin Yan-yan, Wang Hao, Wang Wei. LWR model for mixed traffic flow in connected and autonomous vehicular environments[J]. *China Journal of Highway and Transport*, 2018, 31(11): 147-156.
- [13] Becker F, Axhausen K W. Literature review on surveys investigating the acceptance of automated vehicles[J]. *Transportation*, 2017, 44(6): 1293-1306.
- [14] Elliott D, Keen W, Miao L. Recent advances in connected and automated vehicles[J]. *Journal of Traffic and Transportation Engineering(English Edition)*, 2019, 6(2): 109-131.
- [15] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning[J/OL]. [2023-03-10]. <https://arxiv.org/abs/1312.5602>.
- [16] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning[J/OL]. [2023-12-10]. <https://arxiv.org/pdf/1509.02971>.
- [17] Lowe R, Wu Y, Tamar A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments[J]. *Advances in Neural Information Processing Systems*, 2017, 30:1-16.
- [18] Franois-Lavet V, Henderson P, Islam R, et al. An introduction to deep reinforcement learning[J]. *Foundations and Trends® in Machine Learning*, 2018, 11(3-4):219-354.
- [19] Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: a survey[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(6): 4909-4926.
- [20] Vasquez R, Farooq B. Multi-objective autonomous braking system using naturalistic dataset[C]//*IEEE Intelligent Transportation Systems Conference(ITSC)*, Auckland, New Zealand, 2019: 4348-4353.
- [21] 王殿海, 金盛. 车辆跟驰行为建模的回顾与展望[J]. *中国公路学报*, 2012, 25(1): 115-127.
- Wang Dian-hai, Jin Sheng. Review and outlook of modeling of car following behavior[J]. *China Journal of Highway and Transport*, 2012, 25(1): 115-127.
- [22] Willis A, Gjersoe N, Havard C, et al. Human movement behaviour in urban spaces: implications for the design and modelling of effective pedestrian environments[J]. *Environment and Planning B: Planning and Design*, 2004, 31(6): 805-828.
- [23] Trumpp R, Bayerlein H, Gesbert D. Modeling interactions of autonomous vehicles and pedestrians with deep multi-agent reinforcement learning for collision avoidance[C]//*IEEE Intelligent Vehicles Symposium (IV)*, Beijing, China, 2022: 331-336.
- [24] 王辉, 秦华, 冉令华, 等. 无交通信号路口行人过街的人车交互过程研究[J]. *科学技术与工程*, 2023, 23(28):12275-12281.
- Wang Hui, Qin Hua, Ran Ling-hua, et al. Human vehicle interaction process of pedestrian crossing at no traffic signal intersection[J]. *Science Technology and Engineering*, 2023, 23(28): 12275-12281.
- [25] Schmidt S, Faerber B. Pedestrians at the Kerb-recognising the action intentions of humans[J]. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2009, 12(4): 300-310.
- [26] Dean B A K. Grammatical design and crowd behaviour: a study of factors that influence human movement in urban spaces[C]//*Proceedings of the 10th International Conference on Computer Aided Architectural Design Research in Asia*, New Delhi, India, 2005:648-650.
- [27] Millard-Ball A. Pedestrians, autonomous vehicles, and cities[J]. *Journal of Planning Education and Research*, 2018, 38(1): 6-12.