

DOI:10.16544/j.cnki.cn43-1494/u.20230821002

文章编号:1674-599X(2026)02-0088-10

引用格式:罗归权,张兆磊,易可夫,等.基于深度强化学习的匝道合流区车辆自主换道方法[J].交通科学与工程,2026,42(2):88-97.

Citation: LUO Guiquan, ZHANG Zhaolei, YI Kefu, et al. Autonomous lane-changing in ramp merge area based on deep reinforcement learning [J]. Journal of Transport Science and Engineering, 2026, 42(2): 88-97.

基于深度强化学习的匝道合流区车辆自主换道方法

罗归权¹,张兆磊¹,易可夫²,唐峰¹

(1.长沙理工大学 交通学院,湖南 长沙 410114;2.长沙理工大学 汽车与机械工程学院,湖南 长沙 410114)

摘要:【目的】解决主线网联自动驾驶车辆(connected and autonomous vehicle, CAV)在匝道合流区的交通冲突问题。【方法】提出一种基于深度双重Q网络(double-deep Q network, DDQN)算法的CAV自主换道控制方法。首先,构建包含CAV、相邻车辆及匝道车辆的交通状态矩阵,并通过引入虚拟车辆解决状态空间矩阵维度缺失问题,实现状态矩阵维度的统一;然后,针对CAV换道行为对目标车道上游车流产生的负效应,在DDQN算法的奖励函数中引入上游车辆速度与紧急制动特征,实现对车辆自身效益与交通流整体状态的联合优化,提升换道过程的安全性;最后,针对单主线双车道的高速公路匝道合流场景,通过仿真试验验证所提方法的有效性。【结果】与经典的2013版换道模型(lane-changing model 2013, LC2013)相比,所提方法在保障CAV通行效率的前提下,将紧急制动次数减少了49.8%;与未考虑后车影响的基础DDQN算法相比,所提方法将紧急制动次数减少了15.8%。敏感性分析进一步证实,在不降低CAV行驶速度的前提下,增大安全权重可使CAV自身的紧急制动次数减少16.4%,并使跟随车辆的紧急制动次数减少46.6%。【结论】所提方法在保证交通效率的同时能够有效提升匝道合流区的交通安全水平。

关键词:智能交通;换道决策;深度强化学习;匝道合流区;自动驾驶车辆

中图分类号:U491.4

文献标志码:A

Autonomous lane-changing in ramp merge area based on deep reinforcement learning

LUO Guiquan¹, ZHANG Zhaolei¹, YI Kefu², TANG Feng¹

(1. School of Transportation, Changsha University of Science & Technology, Changsha 410114, China; 2. School of Vehicle and Mechanical Engineering, Changsha 410114, China)

Abstract: [Purposes] This paper aims to solve the traffic conflict of connected and autonomous vehicle (CAV) on the mainline in the ramp merge area. [Methods] An autonomous lane-changing method for CAV based on the double-deep Q network (DDQN) algorithm was proposed. First, a traffic state matrix incorporating CAV, adjacent vehicles, and on-ramp vehicles was established. Virtual vehicles were introduced to address missing dimensions in the state space, resulting in a unified matrix dimension. Second, to mitigate the negative impact of CAV lane changes on upstream traffic in the target lane, upstream vehicle speed and emergency braking characteristics were incorporated into the DDQN reward function. This enables joint optimization of individual vehicle performance and overall traffic conditions, improving lane-changing safety. Finally, the proposed method was validated

收稿日期:2023-08-21

基金项目:国家自然科学基金青年基金项目(52002036, 52302429);长沙理工大学研究生实践创新项目(CLSJXC22009);湖南省研究生科研创新项目(CX20220852);湖南省科学研究项目优秀青年项目(22B0325);教育部工程研究中心开放基金项目(kf220403)

通信作者:张兆磊(1995—),男,博士,主要从事交通运输规划与管理方面的研究工作。E-mail:913545170@qq.com

投稿网址:https://jtkxygc.csust.edu.cn/jtkxygc/home

by simulation experiments for a single-mainline two-lane freeway ramp merge scenario. [**Findings**] Compared with the lane-changing model 2013 (LC2013), the proposed method reduces the number of emergency braking events by 49.8% while ensuring CAV traffic efficiency. Compared with a baseline DDQN algorithm that does not consider the influence of following vehicles, the number of emergency braking is reduced by 15.8%. Sensitivity analysis further shows that, without reducing CAV driving speed, increasing the safety weight reduces emergency braking by 16.4% for CAVs and 46.6% for following vehicles. [**Conclusions**] The proposed method can effectively improve the traffic safety level of the ramp merge area while ensuring traffic efficiency.

Key words: intelligent transportation; lane-changing decision; deep reinforcement learning; ramp merge area; automated driving vehicle

中国高速公路匝道区域发生的交通事故约占总事故的30%^[1],其主要诱因是车辆合流过程中产生的交通冲突。近年来,随着智能交通系统的发展,智能辅助驾驶与自动驾驶技术已得到广泛研究与应用。这些技术为提升驾驶安全性与效率提供了新途径,也为解决匝道合流冲突开辟了全新视角。目前,匝道合流控制的研究主要集中在主线车辆与匝道车辆的协同协调^[2]以及主线车辆的实时可变限速策略^[3]等方面,然而这些方法多侧重于通过整体车流调控来缓解合流冲突。与此同时,随着深度学习与强化学习等数据驱动方法在交通领域的兴起,其相关应用已逐渐落地^[4-5],涵盖自动驾驶、交通状态预测及换道轨迹预测等多个方向。强化学习具备与环境持续交互的强大学习能力^[6],可有效应对复杂交通场景的分析需求,因此在交通领域的应用日益广泛。

目前,深度强化学习(deep reinforcement learning, DRL)在交通领域的研究与应用,主要集中于车辆驾驶决策、车辆驾驶操作两个层面。在驾驶决策层面,已有研究采用深度双重Q网络(double-deep Q network, DDQN)实现车辆自由换道决策^[7];另有研究利用多智体技术开展协同换道决策^[8],且该方法已被证实能够适配多种交通流状态。在驾驶操作层面,相关研究通过深度Q网络(deep Q network, DQN)对车辆进行横向换道控制^[9],该方法可高效训练出快速、安全的避障驾驶策略。本文重点关注车辆驾驶决策层面,具体研究网联自动驾驶汽车(connected and autonomous vehicles, CAV)在匝道合流场景下的换道决策问题。

换道决策方法主要分为基于规则的方法和基于数据驱动方法两大类。其中,在基于规则的方法中,研究人员利用Gipps安全换道模型^[10],将车辆换

道决策转化为最优车道序列的求解问题,进而实现了车辆换道决策优化。曹珊^[11]基于随机效用理论构建换道模型,该模型可对真实换道决策进行准确判断。为解决路径规划与路径跟踪决策算法中存在的决策与规划时滞问题,袁朝春等^[12]采用人工势场法,实现了车辆规划与决策的一体化。然而,基于规则的换道决策方法未充分考虑驾驶环境下人工驾驶车辆(human driven vehicle, HDV)驾驶行为的随机性,导致换道决策的时效性较差,无法实现实时最优决策。

基于数据驱动的决策方法中,XING等^[13]利用循环神经网络和长短期记忆网络,将车辆换道意图的预测精度提升至96.1%;另有研究通过贝叶斯网络对受控车辆周边车辆的轨迹进行精准预测,进而实现受控车辆的合理换道,但该模型缺乏对车辆换道控制相关内容的考虑^[14]。此外,基于数据驱动的决策方法的研究主要集中于利用换道成功数据对算法进行训练,未充分考虑换道失败或制动行为等场景,导致算法无法学习到不安全行为的特征,难以保证模型所做出的换道决策能够有效避免碰撞。

针对自动驾驶的现有研究,大多聚焦于控制车辆自身的运行状态,而较少关注其执行控制指令后对上游交通流产生的影响。针对这一研究缺口,本文提出一种适用于匝道合流场景且充分考虑上游车流影响的换道决策算法。具体工作如下:首先,针对车辆信息状态矩阵中信息维度缺失的问题,通过引入虚拟车辆的方式扩充状态空间维度,构建更完整的环境感知模型;其次,为抑制换道行为对目标车道上游车流造成的负效应,在DDQN算法的奖励函数中融合上游车辆速度与紧急制动特征,提升CAV换道安全性与通行效率;最后,通过仿真试验对所提算法进行验证。

1 深度强化学习算法

强化学习是一种数据驱动的学习方法,通常将模型建模为马尔可夫决策过程。强化学习智能体 (Agent) 在环境中执行一系列动作,以最大化长期累积回报。

将 DDQN 策略网络表示为 $Q(s_t, a_t; \theta)$, 其中 θ 为网络参数, a_t 为 t 时刻的动作, s_t 为 t 时刻的状态。Agent 在 t 时刻根据状态 s_t 采取动作 a_t 后, 依据状态转移概率分布 $p(s_{t+1}|s_t, a_t)$ 转移至下一状态 s_{t+1} , 并从环境中获得即时奖励 r_t 。强化学习算法通过优化网络, 使智能体获得最大的累积期望奖励。

DDQN 算法能够缓解 DQN 算法过度估计的缺陷^[15]。DDQN 引入两个结构相同的网络, 分别为动作选择网络 $Q_1(s_t, a_t; \theta_1)$ 与目标价值网络 $Q_2(s_t, a_t; \theta_2)$ 。Agent 在每个时间步 t 从环境中获取状态 s_t , 并依据式(1)选取相应动作 a'_{t+1} 。

$$a'_{t+1} = \arg \max Q(s_t, a_t; \theta_1) \quad (1)$$

计算奖励 r_{t+1} , 并利用式(2)得到目标函数值 y_t :

$$y_t = \begin{cases} r_t, & s_{t+1} \text{ 为终止状态} \\ r_t + \gamma Q(s_{t+1}, a'_{t+1}; \theta_2), & \text{其他} \end{cases} \quad (2)$$

式中: γ 为未来的 Q 在当前时刻的折扣因子。

采用经验回放机制, 通过式(3)计算损失函数 $L(\theta_1)$ 以更新网络参数 θ_1 。

$$L(\theta_1) = \frac{1}{m} \sum_{j=1}^m [y_j - Q_j(s, a; \theta_1)]^2 \quad (3)$$

式中: m 为最小批次的大小; y_j 为第 j 条数据的目标函数值; $Q_j(s, a; \theta_1)$ 为在状态 s 和动作 a 前提下第 j 条数据利用网络参数 θ_1 得到的网络。

算法输入的环境信息考虑三个模块, 包括 Agent 自身车辆信息 s_{cav} 、车辆传感器范围内邻近 Agent 的车辆信息 s_1 , 以及投影至主线道路上且位于 Agent 前方的匝道车辆信息 s_{on} 。因此, 整个状态空间可表示为 (s_{cav}, s_1, s_{on}) 。为完整描述车道信息, 输入信息按照当前车道、目标车道及匝道进行划分。为固定状态矩阵的维度并保障 Agent 安全, 当传感器范围内无车辆时, 在 Agent 传感器范围边界处设置与 Agent 速度相同的虚拟车辆。

当主线车道传感器范围内存在多辆车时, 仅选取与 Agent 相邻的 4 辆车的信息, 包括目标车道前车、目标车道后车、当前车道前车及当前车道后车。车辆传感器范围内邻近 Agent 的车辆信息 s_1 为:

$$s_1 = (s_{c,f}, s_{c,1}, s_{T,f}, s_{T,1}) \quad (4)$$

式中: $s_{c,f}$ 为当前车道后车状态信息; $s_{c,1}$ 为当前车道前车状态信息; $s_{T,f}$ 为目标车道后车状态信息; $s_{T,1}$ 为目标车道前车状态信息。

由于匝道车辆行驶速度较低, 仅考虑投影至 Agent 前方的匝道车辆信息, 将整个状态空间数组扁平化输入到 DDQN 网络中得到:

$$Q(s_t, a_t; \theta) = Q[(s_{cav}, s_1, s_{on}), a_t; \theta] \quad (5)$$

从经验回放池中抽取小批量数据进行训练, 单条数据可用包含 t 时刻状态 s_t 、动作 a_t 、奖励 r_t 及下一时刻状态 s_{t+1} 的组合 (s_t, a_t, r_t, s_{t+1}) 表示。利用式(3)计算该小批量数据的损失函数, 并基于损失函数更新网络参数 θ_1 。

隐藏层激活函数采用修正线性单元 (ReLU)。神经网络结构参数见表 1。

表 1 神经网络结构参数

Table 1 Neural network structure parameters

层数	单元个数	激活函数
1	18	
2	64	ReLU
3	64	ReLU
4	2	Softmax

2 算法设计

本文采用开源交通仿真软件 SUMO, 通过 TraCI 接口进行二次开发, 构建深度强化学习环境 with 受控 CAV。鉴于深度强化学习基于马尔可夫决策过程, 本文将整个驾驶环境定义为状态空间与动作空间, 并设计合理的奖励函数, 整体框架如图 1 所示。

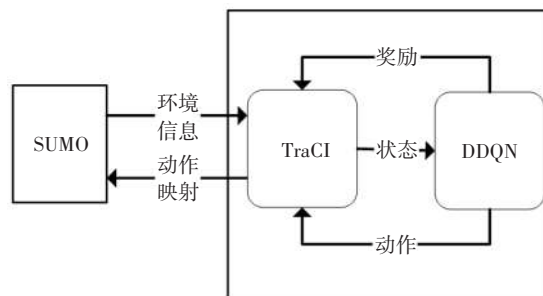


图 1 基于 DDQN 的仿真过程示意图

Fig. 1 Simulation process of DDQN

2.1 场景设置

匝道合流区仿真场景示意图如图 2 所示。本研

究场景中,主线为双车道,匝道为单车道,研究区域为汇入点上游500 m至下游200 m(-500~200 m)区间。选取该场景的原因如下:

1) 高速公路主线为多车道,匝道合流冲突主要

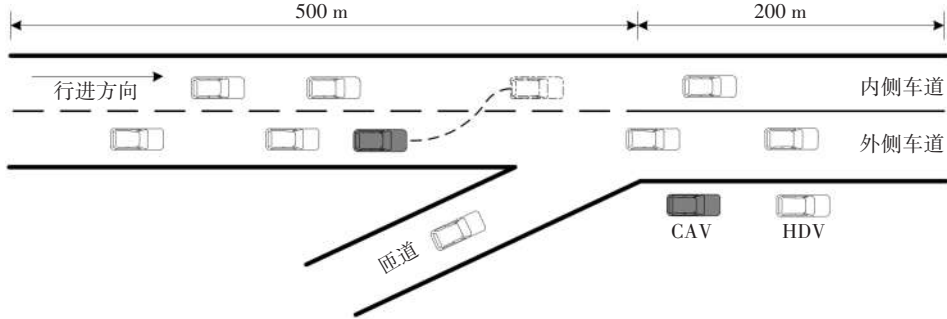


图2 匝道合流区仿真场景示意图

Fig. 2 Simulation scenario of ramp merge area

假设CAV行驶在主线外侧车道,可通过自身传感器及路侧单元获取其上下游300 m范围内所有车辆(包括HDV)的信息。为使算法能够解决主线车辆与匝道车辆的合流冲突,本文对单一匝道车辆设定特定驶入时刻,以此探讨匝道合流对主线车辆的影响,确保CAV能够学习到合理的换道决策。

2.2 状态空间

将状态空间划分为三部分,分别为CAV本车信息、匝道车辆信息以及传感器感知范围内邻近CAV的车辆信息。CAV本车信息包括速度、位置及所在车道,考虑到实际应用场景将其转换为相对于总量的相对量,CAV相对位置 $x_{c,r}$ 可定义为车辆位置与道路长度的比值:

$$x_{c,r} = \frac{x_{cav}}{L} \quad (6)$$

式中: L 为路段长度; x_{cav} 为CAV在当前道路的位置。

CAV相对车道 $l_{c,r}$ 可定义为CAV所处的车道与车道总数的比值:

$$l_{c,r} = \frac{l_{cav}}{n} \quad (7)$$

式中: n 为车道总数; l_{cav} 为CAV所处的车道。

CAV相对速度 $v_{c,r}$ 可定义为CAV行驶速度与路段限制速度的比值:

$$v_{c,r} = \frac{v_{cav}}{v_{max}} \quad (8)$$

式中: v_{max} 为路段限制速度; v_{cav} 为CAV行驶速度。

CAV环境信息 s_{cav} 可表示为:

$$s_{cav} = (x_{c,r}, l_{c,r}, v_{c,r}) \quad (9)$$

对于匝道车辆,因其整体车速较低,仅考虑主线道路上投影位置位于CAV前方的车辆。由于匝

道车辆与主线CAV在合流前分属于不同道路,匝道车辆的状态空间包含其与CAV的相对位置、相对车速,以及二者是否处于同一条道路、同一车道的相对信息。匝道车辆的相对位置可通过下式计算:

2) 匝道合流场景中,匝道车辆最终均通过加速车道汇入主线,因此将匝道设置为单车道。

道车辆与主线CAV在合流前分属于不同道路,匝道车辆的状态空间包含其与CAV的相对位置、相对车速,以及二者是否处于同一条道路、同一车道的相对信息。匝道车辆的相对位置可通过下式计算:

$$x_{r,r} = \frac{x_{r,on} - x_{cav}}{R} \quad (10)$$

式中: $x_{r,r}$ 为相对位置; $x_{r,on}$ 为匝道车辆投影到主线的位置; R 为传感范围。

匝道车辆车道相对信息 $l_{r,r}$ 为:

$$l_{r,r} = \begin{cases} 1, & l_{cav} = l_r \text{ 且 } w_{cav} = w_r \\ 0, & \text{其他} \end{cases} \quad (11)$$

式中: w_{cav} 为CAV道路序号; w_r 为匝道道路序号; l_r 为匝道车道序号。

匝道车辆相对速度 $v_{r,r}$ 可定义为匝道车辆本身速度和路段限制速度的比值:

$$v_{r,r} = \frac{v_r}{v_{max}} \quad (12)$$

式中: v_r 为匝道车辆速度。

匝道车辆的状态信息 s_r 可表示为:

$$s_r = (x_{r,r}, l_{r,r}, v_{r,r}) \quad (13)$$

传感器感知范围内的主线车辆信息包括:当前车道CAV后方车辆信息 $s_{c,f}$ 、当前车道CAV前方车辆信息 $s_{c,1}$ 、目标车道上投影位于CAV后方的车辆信息 $s_{T,f}$,以及目标车道上投影位于CAV前方的车辆信息 $s_{T,1}$ 。

每辆车的信息包含其与CAV的相对位置、相对速度及相对车道信息。主线车辆的相对位置 x_r 可表示为:

$$x_r = \frac{x - x_{cav}}{R} \quad (14)$$

式中: x 为主线车辆在当前路段的位置。

主线车辆相对车道信息 l_r 可表示为:

$$l_r = l - l_{cav} \quad (15)$$

式中: l 为车辆所在的车道。

主线车辆相对速度信息 v_r 可表示为:

$$v_r = \frac{v - v_{cav}}{v_{max}} \quad (16)$$

式中: v 为车辆当前时刻车速。

当探测器在当前车道或目标车道没有探测到前后车辆时,为保证车辆安全可设置虚拟车辆。虚拟车辆与CAV的距离为探测器的范围(300 m)且速度与CAV保持一致,从而确保环境状态矩阵的大小。传感范围内主线车辆的环境信息 s_r 可表示为:

$$s_r = (s_{c,1}, s_{c,f}, s_{t,1}, s_{t,f}) \quad (17)$$

2.3 动作函数

CAV的换道行为决策 $a_t = \{0, 1\}$ 中,1表示车辆向内侧变道,0表示车辆向外侧变道。本研究为使CAV学习到更优策略,引入随机概率机制决定是执行神经网络输出的动作,还是采取随机动作。

2.4 奖励函数

奖励函数综合考虑了车辆的行驶效率与交通安全。由于CAV在执行换道时仅考虑自身的行驶效率与安全,未顾及换道行为对上游交通流的影响,因此换道后会对上游交通流产生负面影响。为解决该问题,本研究在奖励函数中融入了针对目标车道上游交通影响的奖惩因子。奖励函数一包含CAV速度奖励、换道惩罚与紧急制动惩罚;奖励函数二则在此基础上,增加了跟随车辆速度奖励与跟随车辆紧急制动惩罚,各函数具体说明见表2。

奖励函数一($R_{1,t}$):

$$R_{1,t} = \omega_1 R_{c,v} - \omega_2 \alpha_1 P_{lc} - \omega_3 \alpha_2 P_{emmm} \quad (18)$$

奖励函数二($R_{2,t}$):

$$R_{2,t} = \omega'_1 R_{c,v} - \omega'_2 \alpha_1 P_{lc} - \omega'_3 \alpha_2 P_{emmm} - \omega'_4 \alpha_3 P_{f,emmm} + \omega'_5 R_{f,v} \quad (19)$$

式中: $\omega_1, \omega_2, \omega_3, \omega'_1, \omega'_2, \omega'_3, \omega'_4$ 和 ω'_5 为算法可调权重,其中 $\omega_1 = \omega'_1 + \omega'_5, \omega_3 = \omega'_3 + \omega'_4, \omega_2 = \omega'_2; \alpha_1, \alpha_2, \alpha_3$ 为参数。

设置算法可调权重的目的是充分权衡驾驶安全性与交通效率。当 ω_1 被调至较高水平时,CAV速度奖励占主导,车辆会采取更激进的策略以获得更高车速;当 ω_2 被调至较高水平时,交通安全占主导,CAV会避免频繁换道,更倾向于保持当前车道,采取更为保守的驾驶策略; ω_3 被调至较高水平可有效

表2 奖励函数说明

Table 2 Description of reward function

分类	描述	符号
CAV速度奖励 $R_{c,v}$	CAV行驶速度 v_{cav} 与路段限制车速之比	$R_{c,v} = \frac{v_{cav}}{v_{max}}$
跟随车辆速度奖励 $R_{f,v}$	跟随在CAV后方车辆的速度 v_f 与路段限制速度之比	$R_{f,v} = \frac{v_f}{v_{max}}$
换道惩罚 P_{lc}	为保障交通安全,避免不安全换道策略与驾驶不适感,需防止车辆频繁换道	$P_{lc} = 1$
CAV紧急制动惩罚 P_{emmm}	为避免CAV与匝道车辆发生碰撞,同时引导车辆采取安全驾驶策略,对紧急制动行为进行惩罚,惩罚值取紧急制动加速度 $a_{cav,emmm}$ 绝对值的负值	$P_{emmm} = - a_{cav,emmm} $
跟随车辆紧急制动惩罚 $P_{f,emmm}$	为降低换道行为对CAV上游车流产生的负面影响,当CAV上游车辆发生紧急制动时,对其实施惩罚,惩罚值取上游车辆紧急制动加速度 $a_{f,emmm}$ 绝对值的负值	$P_{f,emmm} = - a_{f,emmm} $

减少CAV与匝道车辆在合流区的冲突,使CAV表现出更强的安全意识。

当奖励函数二中加入反映换道对目标车道交通影响的奖惩因子后,CAV在执行换道时会考虑其对目标车道上游车流的负面影响。当 ω'_4 和 ω'_5 提升时,CAV会更多考虑整体交通流的情况,采取更为安全的驾驶策略。当CAV执行换道行为时, $\alpha_1=1$;当CAV执行紧急制动时, $\alpha_2=1$;当上游车辆执行紧急制动时, $\alpha_3=1$ 。

为了CAV能够学习到更多经验,包括成功和失败(紧急制动)的经验,本文采用 ϵ -贪心策略来确定选择动作的概率。该算法的探索概率随着循环次数的增加而逐渐减小,因此训练初期能够学习到更多不同的经验,从而获得更好的训练效果。

3 案例分析

3.1 仿真场景参数设置

本文利用交通仿真软件SUMO搭建交通驾驶仿真模型,并分别完成交通场景、车辆控制算法与仿真参数的设置。在交通场景中,匝道合流区主线上游长度为500 m,下游长度为200 m,匝道长度为300 m。车辆速度服从高斯分布,其中匝道车辆平均速度为70 km/h,主线外侧车道车辆平均速度为80 km/h,内侧车道车辆平均速度为100 km/h。在车辆控制算法方面,所有车辆纵向控制均采用仿真软

件自带的IDM模型。CAV的期望车头时距为1.0 s, HDV的期望车头时距为1.5 s。考虑到驾驶员驾驶行为具有随机性,为每个HDV施加随机噪声;而CAV可精准控制行驶速度与方向,其随机扰动设为0.0。横向控制模型方面,所有HDV均采用仿真软件自带的2013版换道模型(lane-changing model 2013,LC2013),具体控制参数见表3。

表3 车辆控制参数设置

Table 3 Vehicle control parameter settings

车辆类型	控制方法	纵向控制	横向控制	车头时距/s	速度噪声
CAV	DDQN	IDM	DDQN	1.0	0.0
CAV	基准控制	IDM	LC2013	1.0	0.0
HDV	不控制	IDM	LC2013	1.5	0.5

本文算法采用基于经验回放的离线训练方式,在网络训练前,先通过初始化后的网络进行决策并采集数据,将经验池(大小为2 000)填充完毕(约需80次循环);正式训练时,将决策结果存入经验池,每次从中抽取128组数据用于更新网络参数。 ϵ -贪心策略可在探索与利用之间实现平衡,其探索概率随着循环次数的增加逐渐减小。为优化网络参数,采用随机梯度下降优化算法,初始学习率设为0.01,目标网络参数每100步更新一次,具体参数见表4。

表4 DRL超参数

Table 4 DRL hyperparameters

超参数	值
经验池	2 000
最小批次	128
学习率	0.01
网络更新率	100
探索率	0.01
ϵ 的衰减系数	10

3.2 结果分析

在训练过程中,前2 000步采用初始化网络对车辆进行决策,且不更新网络参数,该阶段的主要目的是让经验回放池充分采集学习样本。2 000步(约80次循环)后,网络开始更新,奖励函数逐渐趋于稳定。相较于随机选择策略,本文方法可获得更高的奖励值,且紧急制动次数显著减少。网络训练过程如图3所示。

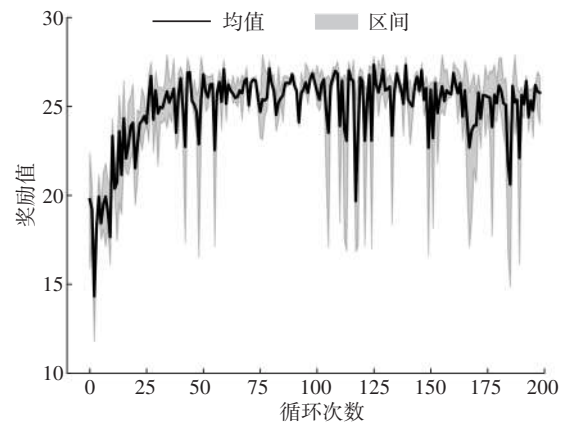
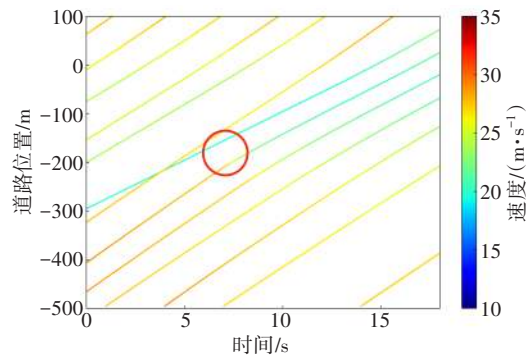


图3 网络训练过程

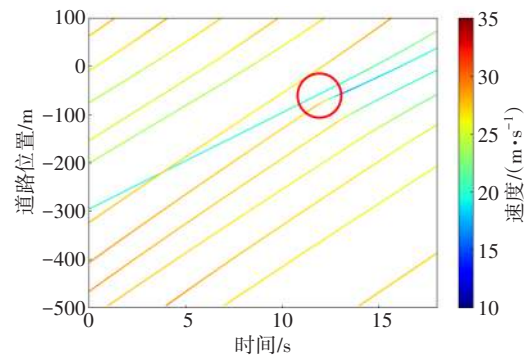
Fig. 3 Training process

为验证本文算法的稳定性与有效性,通过设置车辆驶入概率,测试了不同交通流量下换道决策算法的性能,仿真结果如图4所示。

由图4可知,考虑后车影响的换道算法能够在距离目标车道后方车辆更远的位置完成换道操作,并在匝道上游区域提前完成换道,而非在匝道合流点执行换道。



(a) 考虑后车



(b) 未考虑后车

图4 内侧车道车辆与控制车辆时空轨迹图

Fig. 4 Space-time trajectories of inside lane vehicle and control vehicle

考虑后车影响的换道算法通过充分考量换道行为对后方车流的影响,选择更为合理的时机与位

置完成换道,在避免合流冲突的同时,降低了对目标车道上游交通流的干扰。

奖励函数对比如图5所示。由图5可知,当目标车道流量为300~1 200 辆/h(本文每小时车辆数指每小时标准小客车当量)时,相较于未考虑后车影响的算法,考虑后车影响的算法可获得更高的奖励,且离群点更少;而当目标车道流量在1 500 辆/h以上时,未考虑后车影响的换道算法会取得更高的

奖励。基础模型的奖励函数分布更为集中,但由于CAV无法获取匝道车辆信息,在发生紧急制动时会得到更低的奖励值。考虑后车影响的换道算法在决策时会兼顾对后车的影响,因此决策更为谨慎,奖励值分布也更集中。低流量条件下,CAV会执行更多换道操作以获取更高奖励;当流量超过1 500 辆/h时,目标车道可利用的换道间隙减少,车辆决策趋于保守,对应的奖励值随之降低。

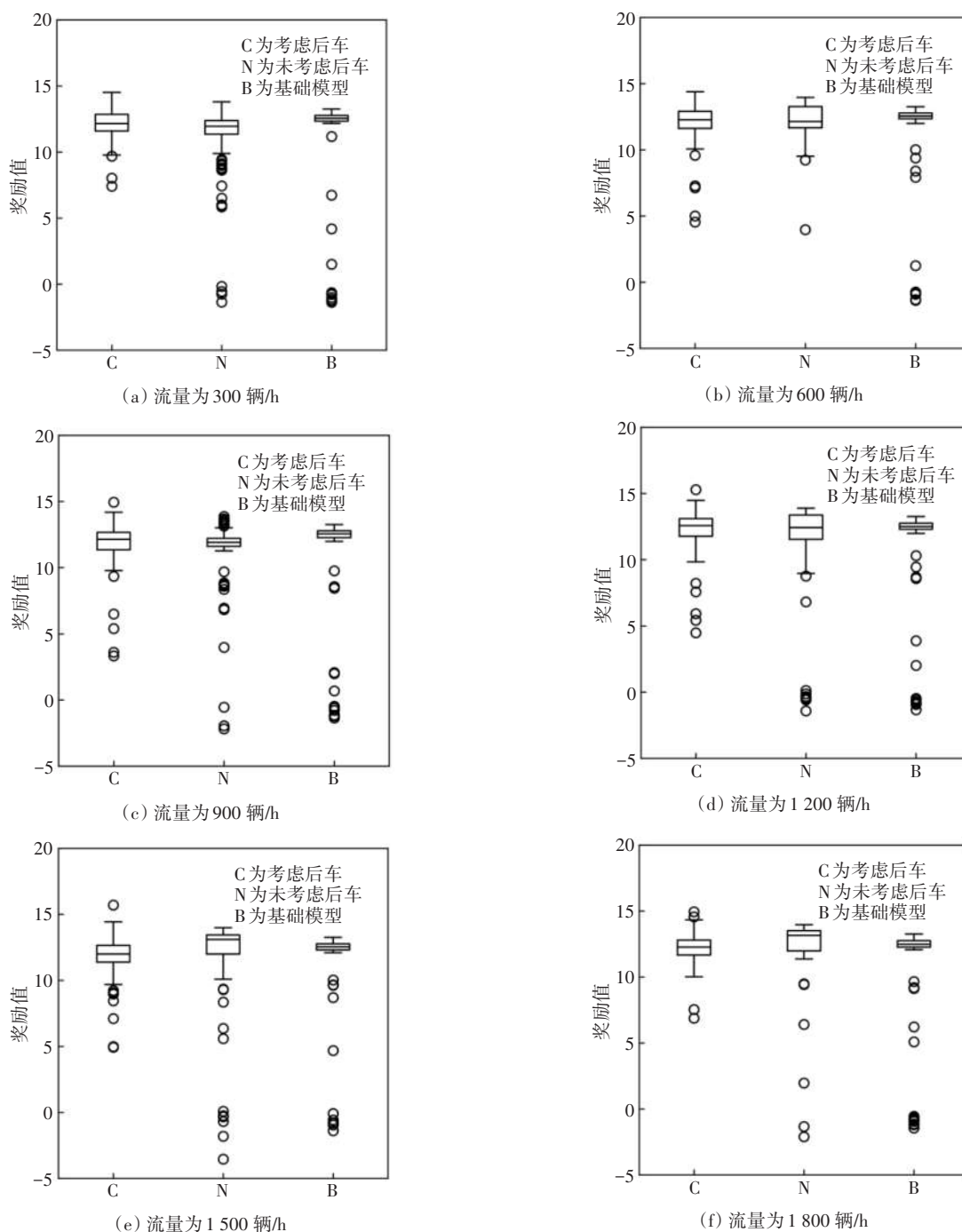


图5 奖励函数对比

Fig. 5 Comparison of reward functions

CAV 紧急制动对比如图 6 所示。不同流量下 CAV 速度对比见表 5。由图 6 和表 5 可知,与 LC2013 模型相比,本文所提算法可使 CAV 紧急制动次数减少 49.8%,且 CAV 速度差异小于 1.0%,这说明该算法在保障行车安全的同时不会显著降低 CAV 行驶速度。相较于未考虑后车影响情况,考虑后车影响时,CAV 紧急制动次数减少 15.8%,速度差异同样小于 1.0%。这是因为在内侧车道交通量较小时,车辆可利用的换道间隙较大,使得车辆规避合流冲突的机会减少。本文所提考虑后车影响的 DDQN 换道算法可做出更优的换道决策,有效减少匝道合流区域的交通冲突。

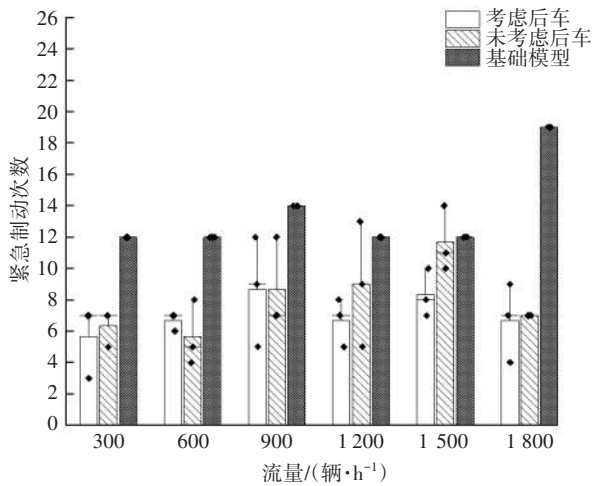


图 6 CAV 紧急制动对比

Fig. 6 Comparison of CAV emergency braking

表 5 不同交通量下 CAV 速度对比

Table 5 Comparison of CAV speed under different traffic flows

流量/(辆·h ⁻¹)	速度/(m·s ⁻¹)		
	LC2013	未考虑后车	考虑后车
300	22.23	22.45	22.47
600	22.26	22.46	22.45
900	22.18	22.34	22.41
1 200	22.21	22.25	22.35
1 500	22.21	22.24	22.33
1 800	22.01	22.32	22.31
均值	22.18	22.34	22.39

3.3 敏感性分析

本文探究奖励函数中安全权重与效率权重对车辆决策的影响,主要以车速为核心评价指标。安全指标选取紧急制动次数及换道次数。

3.3.1 权重分析

本文所述权重,特指奖励函数中效率项与安全

项的权重。效率指标采用车辆行驶速度与路段限速的比值表征;安全指标则兼顾 CAV 自身的换道行为、紧急制动行为,以及后方跟随车辆的紧急制动状态。

试验中权重设置如下:以安全为首要目标时,紧急制动权重设为 0.50、速度奖励权重设为 0.25、换道惩罚权重设为 0.25;采用均衡权重配置时,紧急制动权重设为 0.33、速度奖励权重设为 0.33、换道惩罚权重设为 0.34;以效率为首要目标时,紧急制动权重设为 0.25、速度奖励权重设为 0.50、换道惩罚权重设为 0.25。试验结果如图 7~10 所示。

由图 7~10 可知,在目标车道交通量恒定条件下,随着奖励函数中安全权重占比提升,不同车流情况下 CAV 与跟随车辆的紧急制动次数均呈下降趋势。高安全权重情况下,CAV 车速未出现明显降低,但其紧急制动次数降幅达 16.4%;跟随车辆紧急制动次数减少 46.6%,车速同步提升 4.0%。

当增大效率(速度)权重时,CAV 与跟随车辆的车速均无显著变化。究其原因:效率权重占比过高时,车辆会频繁触发换道行为以追求更高行驶车速,但匝道合流场景交通环境复杂,此类换道并不能使 CAV 车速得到显著提升,反而会产生大量不安全的换道操作,对整体交通流造成明显负面影响。

当增大奖励函数中的安全权重时,CAV 自身紧急制动频次有所降低,同时跟随车辆的紧急制动行为也得到显著抑制。这是由于安全权重增大后,CAV 主动减少了为提升自身通行效率而发起的不合理换道,有效降低了对周边交通流的干扰。该结果表明:匝道合流区域的主线车辆控制应优先保障行车安全,以此实现车辆在合流区段的平稳、安全汇入。

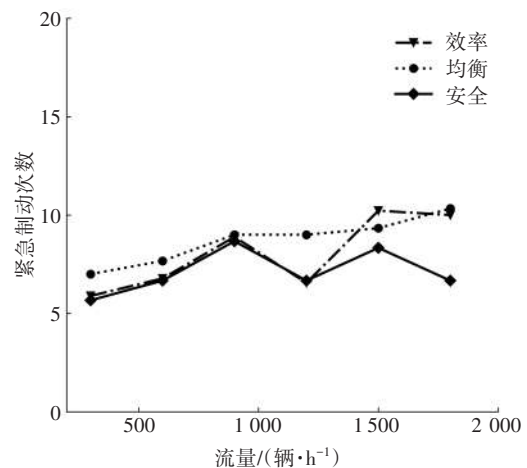


图 7 CAV 紧急制动

Fig. 7 CAV emergency braking

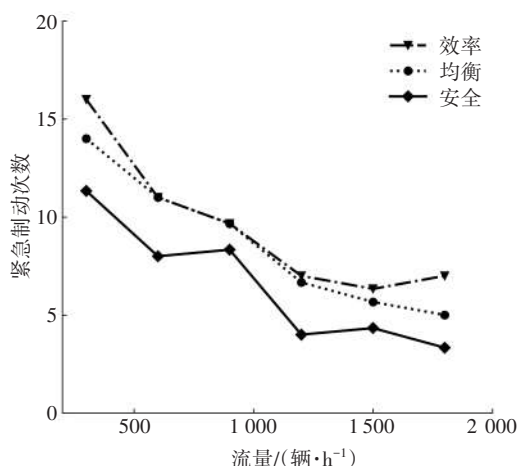


图8 跟随车辆紧急制动

Fig. 8 Following vehicle emergency braking

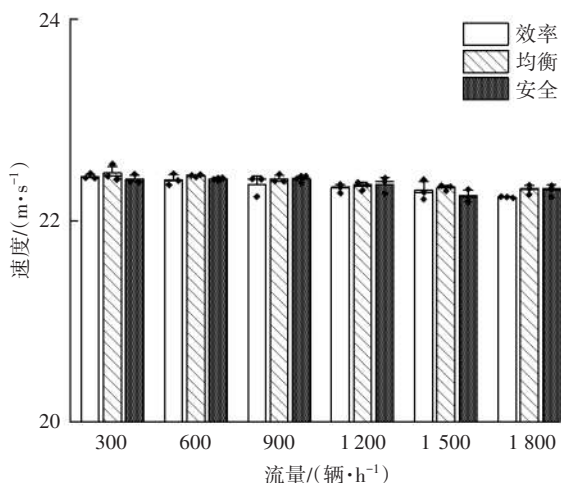


图9 CAV速度

Fig. 9 CAV velocity

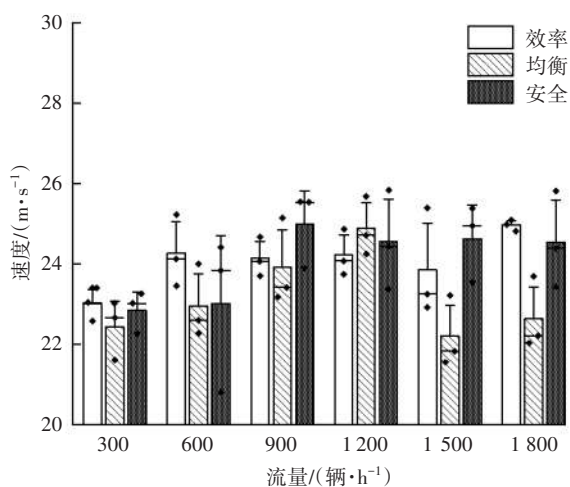


图10 跟随车辆速度

Fig. 10 Following vehicle velocity

3.3.2 目标车道流量分析

目标车道交通量对CAV换道行为具有显著影响。本文将目标车道定义为主线内侧车道,通过调

控该车道交通量可改变换道间隙;试验设置流量为300、600、900、1 200、1 500、1 800辆/h。

由图7~8可知:高流量工况下,相较于其余权重策略,高安全权重的换道策略可显著降低CAV紧急制动次数,每100次仿真中制动次数均低于8次。随着目标车道流量增大,不同权重配置下跟随车辆的紧急制动次数均呈下降趋势,且高安全权重策略始终维持在较低制动水平。

当目标车道流量低于900辆/h时,CAV紧急制动次数无明显差异;流量升至1 200辆/h时,每100次仿真中跟随车辆紧急制动次数可降至5次以下。在300~1 800辆/h全流量区间,高安全权重奖励函数均表现出更优的控制效果。究其机理:目标车道交通量增大会压缩可用换道间隙,使CAV主动减少不必要换道,进而降低跟随车辆的紧急制动频次。在高安全权重策略下,跟随车辆紧急制动次数由11.3次降至3.3次。这是由于高车流条件下换道间隙受限,决策模型会主动规避无效换道,在保障通行效率的同时,有效提升了匝道合流区段的行车安全性。

4 结论

1) 为解决高速公路主线合流区CAV与匝道车辆的合流冲突问题,本文提出一种基于DDQN的匝道合流区车辆自主换道方法。该方法将上游车流交通状态以相对量形式融入奖励函数,克服了传统深度强化学习换道决策仅关注CAV自身、未考虑换道行为对上游交通流产生负效应的缺陷,在保障交通运行效率的前提下可有效提升匝道合流区的交通安全水平。

2) 将所提算法与LC2013及未考虑上游交通状态的换道算法进行对比分析,结果表明:本文所提基于DDQN的换道算法,与LC2013相比,计算得到的紧急制动次数减少49.8%,CAV速度差异小于1.0%;与未考虑后方车辆影响的换道算法相比,计算得到的紧急制动次数减少15.8%;相较于侧重效率的权重设置,在安全权重较高的情况下,紧急制动次数减少16.4%,速度降幅小于1.0%,CAV后方跟随车辆紧急制动次数减少46.6%,跟随车辆运行速度提升4.0%。

在后续研究工作中,研究团队可借助存储系统与路侧单元,将周边车辆的历史行驶信息融入车辆

决策过程;通过融合长期历史信息与短期实时信息,使智能车辆制定更为合理的行驶决策,进而及时规避潜在的交通拥堵与交通事故;进一步考虑不同匝道交通量对车辆换道决策的影响,分析仿真场景与真实场景之间的误差,并解决真实训练与校准数据不足的问题。未来,本研究将采用更贴近实际的车辆数据对算法进行训练与优化,以构建更真实、合理的换道决策模型。

参考文献(References):

- [1] 管晓伟. 基于交通冲突技术的平面交叉口安全评价研究[D]. 北京: 北京交通大学, 2007.
GUAN Xiaowei. Intersection safety evaluation based on traffic conflict technique[D]. Beijing: Beijing Jiaotong University, 2007.
- [2] HU X W, SUN J. Trajectory optimization of connected and autonomous vehicles at a multilane freeway merging area [J]. *Transportation Research (Part C: Emerging Technologies)*, 2019, 101: 111-125. DOI: 10.1016/j.trc.2019.02.016.
- [3] VRBANIĆ F, IVANJKO E, KUŠIĆ K, et al. Variable speed limit and ramp metering for mixed traffic flows: a review and open questions[J]. *Applied Sciences*, 2021, 11(6): 2574. DOI: 10.3390/app11062574.
- [4] 邓兰, 吴义虎. 基于异步优势强化学习的交通信号控制策略[J]. *交通科学与工程*, 2023, 39(3): 110-117. DOI: 10.16544/j.cnki.cn43-1494/u.20220314001.
DENG Lan, WU Yihu. Traffic signal control strategy based on asynchronous advantage actor-critic algorithm [J]. *Journal of Transport Science and Engineering*, 2023, 39(3): 110-117. DOI: 10.16544/j.cnki.cn43-1494/u.20220314001.
- [5] 朱海峰. 基于Q-learning的单点信号配时方案选择算法[J]. *公路与汽运*, 2022, 38(1): 44-47. DOI: 10.3969/j.issn.1671-2668.2022.01.012.
ZHU Haifeng. Single point signal timing scheme selection algorithm based on Q-learning [J]. *Highways & Automotive Applications*, 2022, 38(1): 44-47. DOI: 10.3969/j.issn.1671-2668.2022.01.012.
- [6] VINCENT F L, PETER H, RIASHAT I, et al. An introduction to deep reinforcement learning [J]. *Foundations and Trends in Machine Learning*, 2018, 11(3-4): 219-354. DOI: 10.1561/22000000071.
- [7] DONG J Q, CHEN S K, LI Y J, et al. Space-weighted information fusion using deep reinforcement learning: the context of tactical control of lane-changing autonomous vehicles and connectivity range assessment [J]. *Transportation Research (Part C: Emerging Technologies)*, 2021, 128: 103192. DOI: 10.1016/j.trc.2021.103192.
- [8] CHEN D, HAJIDAVALLOO M R, LI Z J, et al. Deep multi-agent reinforcement learning for highway on-ramp merging in mixed traffic [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(11): 11623-11638. DOI: 10.1109/TITS.2023.3285442.
- [9] QI X W, LUO Y D, WU G Y, et al. Deep reinforcement learning enabled self-learning control for energy efficient driving [J]. *Transportation Research (Part C: Emerging Technologies)*, 2019, 99: 67-81. DOI: 10.1016/j.trc.2018.12.018.
- [10] 杨达, 吕蒙, 戴力源, 等. 车联网环境下自动驾驶车辆车道选择决策模型[J]. *中国公路学报*, 2022, 35(4): 243-255. DOI: 10.3969/j.issn.1001-7372.2022.04.020.
YANG Da, LYU Meng, DAI Liyuan, et al. Decision-making model for lane selection of automated vehicles in connected vehicle environment [J]. *China Journal of Highway and Transport*, 2022, 35(4): 243-255. DOI: 10.3969/j.issn.1001-7372.2022.04.020.
- [11] 曹珊. 城市道路车辆换道模型及换道影响研究[D]. 武汉: 华中科技大学, 2009.
CAO Shan. Study on the model and impact of lane-changing vehicles for city road [D]. Wuhan: Huazhong University of Science and Technology, 2009.
- [12] 袁朝春, 翁烁丰, 何友国, 等. 基于改进人工势场法的路径规划决策一体化算法研究[J]. *农业机械学报*, 2019, 50(9): 394-403.
YUAN Chaochun, WENG Shuofeng, HE Youguo, et al. Integration algorithm of path planning and decision-making based on improved artificial potential field [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(9): 394-403.
- [13] XING Y, LV C, WANG H J, et al. An ensemble deep learning approach for driver lane change intention inference [J]. *Transportation Research (Part C: Emerging Technologies)*, 2020, 115: 102615. DOI: 10.1016/j.trc.2020.102615.
- [14] WANG J H, ZHANG Z, LU G Q. A Bayesian inference based adaptive lane change prediction model [J]. *Transportation Research (Part C: Emerging Technologies)*, 2021, 132: 103363. DOI: 10.1016/j.trc.2021.103363.
- [15] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016, 30(1): 1-13. DOI: 10.1609/aaai.v30i1.10295.

(责任编辑:熊怡;校对:石月珍)