

## 基于 TD3 的网联电动公交车生态驾驶优化方法

郝毅, 钱恒龙, 潘应久

(长安大学 汽车学院, 西安 710018)

**摘要:** 针对纯电动公交车在信号交叉口高能耗的问题, 提出了基于双延迟深度确定性策略梯度算法 (TD3) 的纯电动公交车信号交叉口生态驾驶优化方法。基于 SUMO 搭建仿真训练平台, 综合考虑能耗、通行效率、舒适性和安全性, 构建多目标优化的强化学习奖励函数; 基于 TD3 深度学习框架, 结合电动公交车在信号交叉口处的运行特征构建生态驾驶优化模型并进行参数训练; 以信号交叉口经典通行策略 GLOSA 为基准方法, 对提出的生态驾驶优化模型进行性能验证。结果表明, 相较于 GLOSA 策略, 提出的生态驾驶策略在信号交叉口 4 种典型场景下能耗分别降低了 9.82%、26.13%、19.00%、14.51%, 同时保证了车辆的安全、舒适和通行效率。

**关键词:** 生态驾驶策略; 网联电动公交车; 多目标优化; 深度强化学习; 信号交叉口

中图分类号: U491 文献标志码: A DOI: 10.3969/j.issn.2095-1469.2025.01.05

## A TD3-Based Eco-Driving Optimization Method for Connected Electric Buses

XI Yi, QIAN Henglong, PAN Yingjiu

(School of Automobile, Chang'an University, Xi'an 710018, China)

**Abstract:** To address the issue of high energy consumption in battery electric buses at signalized intersections, this paper proposes an eco-driving optimization model based on the Twin Delayed Deep Deterministic (TD3) policy gradient algorithm. First, a simulation training platform is developed using SUMO, which balances energy consumption, travel efficiency, comfort, and safety in a multi-objective optimized reinforcement learning reward function. Next, an eco-driving optimization model is created within the TD3 framework, tailored to the operational characteristics of electric buses at signalized intersections, and its parameters are trained. Finally, the performance of the proposed model is validated against the classic intersection passage strategy, Green Light Optimal Speed Advisory (GLOSA). The results show that the proposed eco-driving strategy reduces energy consumption by 9.82%, 26.13%, 19.00% and 14.51% in four typical intersection scenarios, while also maintaining vehicle safety, comfort, and travel efficiency.

**Keywords:** eco-driving strategy; connected electric bus; multi-objective optimization; deep reinforcement learning; signalized intersection

收稿日期: 2024-08-16 改稿日期: 2024-09-27 网络首发日期: 2024-10-21

基金项目: 国家自然科学基金项目 (52402417); 陕西省自然科学基金研究计划项目 (2023-JC-QN-0385)

参考文献引用格式:

郝毅, 钱恒龙, 潘应久. 基于 TD3 的网联电动公交车生态驾驶优化方法[J]. 汽车工程学报, 2025, 15(1): 38-48.

XI Yi, QIAN Henglong, PAN Yingjiu. A TD3-Based Eco-Driving Optimization Method for Connected Electric Buses [J]. Chinese Journal of Automotive Engineering, 2025, 15(1): 38-48. (in Chinese)



近年来，纯电动公交车因其低噪声、低污染等特性逐渐成为城市公交系统的首推车型。相比于传统燃油公交车，纯电动公交车存在续驶里程相对较短等问题，提高车辆能源利用率成为当下的研究热点。

在V2X (Vehicle to Everything) 环境中，车辆可获取周围车辆及道路交通环境信息，根据这些信息进行更优的决策使车辆以高效、节能的方式行驶<sup>[1]</sup>。相关学者利用V2X技术提出了大量生态驾驶优化方法，袁伟等<sup>[2]</sup>根据运动学关系将信号灯剩余时长划分为6种情况，综合考虑交叉口上下游路段设计节能驾驶策略。ZHANG Yali等<sup>[3]</sup>在V2X环境中，基于NSGA-II设计了加减速策略，建立信号交叉口不停车通行决策模型，与自然驾驶相比，该方法使能耗降低了21.11%，通行时间减少了16.1 s。刘显贵等<sup>[4]</sup>提出了一种基于前车速度预测的生态驾驶策略，通过预测前车的车速轨迹，使用动态规划算法，以油耗为目标函数，排放为约束构建节能策略。

随着人工智能的发展，结合强化学习和深度神经网络的深度强化学习 (Deep Reinforcement Learning, DRL) 等策略在交通系统优化方面取得了越来越多的研究成果。DRL将车辆视为智能体，智能体通过不断与环境进行交互，最终会选择最佳动作，以获得最大化的期望奖励。SHI等<sup>[5]</sup>将二氧化碳的排放量作为奖励值，利用Q-learning算法优化车辆的CO<sub>2</sub>排放。Q-learning算法是基于值的强化学习算法，不能在连续的状态空间下控制车辆，因此，容易出现局部最优。LIAN Renzong等<sup>[6]</sup>提出基于深度确定性策略梯度 (DDPG) 算法来优化混合动力汽车 (HEV) 的能耗，使车辆获得了更好的能源经济性。庄伟超等<sup>[7]</sup>使用DDPG算法，对电动汽车在信号交叉口的加速度进行训练，使车辆在单路口和多路口场景下均能提高能量效率和通行效率。BAI Zhengwei等<sup>[8]</sup>提出了一种混合强化学习框架，该框架将基于规则的策略和深度强化学习相结合，与最先进的基于模型的生态驾驶方法相比，可以减少12.70%的能耗，节省11.75%的旅行

时间。ZHOU Mofan等<sup>[9]</sup>将DQN算法与DDPG算法相结合，在不增加通行时间的前提下，将能耗降低了46%。

虽然上述研究为生态驾驶优化策略提供了重要的思路以及技术支持，但仍然存在不足。首先，在车辆行驶过程中无法忽视周围车辆的影响，当车流量较大时，车辆受到周围车辆影响，导致优化轨迹结果与安全性相冲突，使优化结果的安全性不足。如一些文献中<sup>[2-3, 5, 7]</sup>所提出的策略均未考虑路网中的其他车辆，忽视了车辆在行驶过程中的安全和跟车问题。其次，部分研究<sup>[7-9]</sup>集中于信号交叉口上游的节能驾驶策略，忽视了信号交叉口下游的速度策略。比如，车辆在信号交叉口因信号灯的限时而停车等待，当车辆重新启动后，虽然在交叉口上游存在加速行为，但最终车速较低，缺乏后续车速控制策略。此外，针对电动公交车而言，车速波动过大易造成乘客晕车，降低了乘客的乘坐舒适性，目前考虑电动公交车乘坐舒适性的生态驾驶策略相对较少。

本文针对纯电动公交车在信号控制交叉口范围内的通行策略问题，利用从路侧设备获取信号相位与配时 (SPaT) 信息，考虑了前车对当前车的行驶策略影响，构建了综合考虑能耗、安全、舒适和通行效率的奖励函数；根据电动公交车进入交叉口范围时的信号灯状态划分为4种典型场景，基于TD3算法构建了考虑电动公交车“通过前+交叉口+通过后”整个过程的生态驾驶深度强化学习模型。

## 1 信号交叉口电动公交车节能驾驶问题

### 1.1 问题描述

在V2X条件下，电动公交车具有与交叉口信号灯及周边车辆通信的能力。在通信范围内，车辆获取交叉口SPaT和动态交通环境信息，利用生态驾驶算法优化车辆的行驶速度谱，从而实现车辆以一种低能耗、高效率的车速轨迹通过信号交叉口，最大可能地避免在信号交叉口停车等待或减少停车次数。

本文设定的场景如图1所示。假定信号交叉口通信范围为 $l$ ，纯电动公交车在 $t_0$ 时刻以车速 $v_i$ 进入

信号交叉口的通信范围, 此时车辆可获取交叉口 SPaT、道路限速、跟车距离、前车速度等信息, 根据这些动态交通信息, 兼顾车辆能耗、安全、通行效率及舒适性, 构建生态驾驶优化模型, 计算最优车速轨迹, 使车辆以最优驾驶模式在  $t_s$  时刻通过路口停车线, 最终在  $t_f$  时刻离开交叉口。

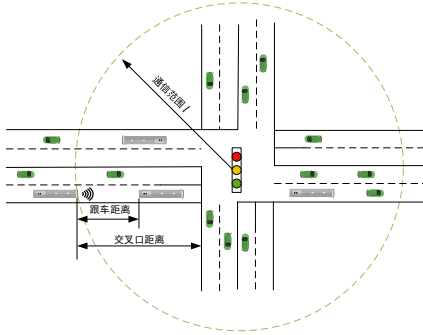


图1 网联公交车生态驾驶场景

## 1.2 最优控制模型构建

依据“时间-能耗”最优控制原理, 可以建立如下模型:

$$\begin{aligned} \min J(x(t), a(t)) &= T + W, \\ \text{s.t. } l(t_0) &= 0, l(t_f) = S, \\ v(t_0) &= v_{\text{ini}}, \\ v_{\min} &\leq v(t) \leq v_{\max}, \\ d_{\max} &\leq a(t) \leq a_{\max}. \end{aligned} \quad (1)$$

式中:  $W$  为车辆的能耗;  $T$  为通行时间;  $t_f$  为电动公交车通过交叉口所需时间;  $x(t)$  为  $t$  时刻车辆的位置;  $a(t)$  为  $t$  时刻车辆的加速度;  $v_{\text{ini}}$  为车辆到达交叉口信号通信范围边界处的初始速度; 考虑到通行效率及道路限速,  $v_{\min}$  和  $v_{\max}$  分别为交通效率最小值和速度最大值; 考虑到公交车乘客舒适性,  $d_{\max}$  和  $a_{\max}$  分别为最大减速度和最大加速度, 其中  $d_{\max} < 0$ 。

上述优化模型在环境状态已知条件下可求得最优解, 但是针对交通环境的动态变化需不断重复计算, 求解效率较低, 优化实时性差, 无法实现动态变化的交通环境下高效的自适应优化。针对这一问题, 本文采用深度强化学习算法进行优化求解。深度强化学习的控制目标与式 (1) 的控制模型相似, 其目标是在当前环境状态下找到最大化奖励值的策略, 但是目标函数的表达方式与式 (1) 的传统控

制模型不同, 需使用不同的术语进行表达<sup>[10]</sup>。深度强化学习通常将实际环境简化为马尔可夫决策过程 (Markov Decision Process, MDP), 其通过五元组  $(S, A, R, P, \gamma)$  描述智能体的优化决策过程。其中,  $S, A, R$  分别为智能体的状态空间、动作空间和奖励函数,  $P$  为状态转移概率,  $\gamma$  为折扣因子。从状态  $S_t$  开始, 智能体选择动作  $A_t$ , 智能体进入新状态  $S_{t+1}$ , 同时获得奖励值  $R_{t+1}$ 。智能体通过不断学习得到最佳策略使未来获得最大的奖励值。其中状态转移概率  $P$  的值决定了智能体的下一步动作, 根据  $P$  是否已知可划分为基于模型 (Model-Based) 和无模型 (Model-Free) 的强化学习决策方法<sup>[11]</sup>。基于模型的强化学习方法通过智能体和环境交互来对环境进行建模, 然后再利用这个模型做出动作规划或者策略选择; 而无模型强化学习通过智能体直接与环境交互学习价值函数或策略函数, 无需对环境建模, 可划分为策略迭代、值迭代与策略搜索 3 类<sup>[12]</sup>。考虑到交通环境信息的动态变化特性, 利用无模型强化学习方法进行生态驾驶策略优化, 即当公交车进入不同跟车情况和信号信息的场景时, 无需重新计算最优值, 只需根据训练好的策略自适应选取针对当前环境的最优行驶策略。则式 (1) 的最优控制模型可改写为:

$$\begin{aligned} \max \left[ \sum_{t=0}^N r_t(x_t, a_t) \right], \\ \text{s.t. } x_{t+1} &= f(x_t, a_t, e_t), \\ a_t &= \pi_t(\tau). \end{aligned} \quad (2)$$

式中:  $N$  为智能体 (电动公交车) 完成任务目标所需步数;  $r_t$  为当前策略状态下的奖励;  $\pi_t = (a_1, a_2, \dots, a_{t-1}, x_0, x_1, \dots, x_t)$  为智能体所选的每一步策略;  $\pi_t = \tau_t$  为控制策略。

## 2 基于 TD3 算法的生态驾驶策略

针对上述问题, 本节建立了基于 TD3 的电动公交车信号交叉口生态驾驶策略框架, 如图 2 所示。当智能体 (电动公交车) 获取到周围的交通环境信息后会储存起来, 不断进行学习, 不断更新策略, 并根据策略输出动作 (加速度), 达到控制车辆行驶轨迹、实现生态驾驶的目的。

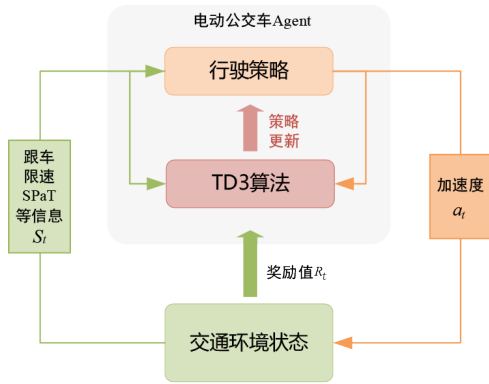


图2 基于TD3的生态驾驶策略框架

## 2.1 交通环境信息

电动公交车智能体基于当前环境观测所得到的信息做出正确的决策，因此，环境状态量的选择对于智能体的决策起着重要作用。根据电动公交车在实际行驶过程中可能会出现驾驶状态，提取11个状态信息来定义状态空间，分别为本车车速 $v_{ego}$ ；本车加速度 $a_{ego}$ ；前车车速 $v_{pre}$ ；前车加速度 $a_{pre}$ ；跟车距离 $D_{ego}$ ；车辆距离交叉口停止线的距离 $l$ ；信号灯当前状态 $T$ ，信号灯为绿灯时 $T=1$ ，黄灯或红灯时 $T=0$ ；当前信号灯剩余时间为 $t_{rem}$ ；下一次绿灯开启时间 $t_{g\_start}$ ；最小车速为 $V_{min}$ ；最大车速为 $V_{max}$ 。状态空间 $S_t$ 可表示为：

$$S_t = \left[ v_{ego}, a_{ego}, v_{pre}, a_{pre}, D_{ego}, l, T, t_{rem}, t_{g\_start}, V_{min}, V_{max} \right]。 \quad (3)$$

本研究以加速度作为智能体的输出。相关研究表明，乘客感到舒适的非紧急制动减速度应不大于 $1.47 \text{ m/s}^2$  [13]。YI等 [14] 研究认为，在非紧急制动情况下，以 $2.5 \text{ m/s}^2$ 减速度对于人类乘客来说是舒适的。在此基础上，综合考虑乘客舒适性，本文将车辆的加速度区间设置为 $[-2.0, 2.0] \text{ m/s}^2$ 。

## 2.2 奖励函数构建

本文旨在降低电动公交车智能体在通过交叉口时的能源消耗，同时要保证车辆通行效率、行驶安全及乘客的舒适性。因此，考虑以下5个方面构建奖励函数。

### 2.2.1 信号控制信息奖励函数

根据运动学基本公式 $v_t^2 - v_0^2 = 2ax$ ，在车辆刚进入信号范围时，记录初始车速 $v_{init}$ ，将最终车速

设为0，从而计算得出减速度 $a$ 。根据 $v_{init}$ 和 $a$ 实时计算车速 $v_t$ ，并将此车速与能不停车匀速通过信号交叉口的车速（ $v_{con} = l/t_{rem}$ ）进行比较，选取两者间较大值作为车速上限 $V_{max}$ ，车速下限 $V_{min}$ 则设计为在绿灯相位通过交叉口的最低限速。

当信号灯为红灯时，可推导其车速边界为：

$$\begin{cases} V_{max} = \max(v_t - at, \min(v_{max}, (l/t_{rem}))) \\ V_{min} = l/(t_{rem} + t_{g\_cycle}) \end{cases}。 \quad (4)$$

式中： $v_{max}$ 为道路限速； $t$ 为进入信号交叉口的时长； $l$ 为电动公交车距离信号交叉口的距离； $t_{rem}$ 为当前信号状态的剩余时间； $t_{g\_cycle}$ 为绿灯相位时长。

当速度在车速边界范围内时，公交车可以实现不停车的条件下通过交叉口。当公交车进入交叉口通信范围时，信号相位为绿灯，将出现两种状况：当绿灯时间过短时，车辆无法在当前绿灯相位通过；当绿灯时间充足时，车辆可以在当前相位通过停止线。针对两种不同的情景可推导出车速边界为：

$$\begin{cases} V_{max} = \begin{cases} \max(v_t - at, l/(t_{rem} + t_{r\_cycle})), & l > v_{max} \times t_{rem} \\ v_{max}, & l \leq v_{max} \times t_{rem} \end{cases} \\ V_{min} = \begin{cases} l/(t_{rem} + t_{cycle}), & l > v_{max} \times t_{rem} \\ l/t_{rem}, & l \leq v_{max} \times t_{rem} \end{cases} \end{cases} \quad (5)$$

式中： $t_{r\_cycle}$ 为红灯相位时长； $t_{cycle}$ 为信号周期时长。

当公交车行驶在交叉口下游时，本研究希望车辆能迅速恢复至一个兼顾能耗与交通效率的车速。袁伟等 [2] 经过大量数据分析后提出了纯电动公交车的生态驾驶车速 $v_{eco}$ ，其值确定为 $10 \text{ m/s}$  ( $36 \text{ km/h}$ )。根据此研究结果，当车辆位于交叉口下游时，车速边界可表示为：

$$\begin{cases} V_{max} = v_{max} \\ V_{min} = v_{eco} \end{cases}。 \quad (6)$$

利用式(4)、式(5)和式(6)可计算出速度边界。当车速不在推荐的速度范围内时，车辆可能会因为闯红灯而引发交通事故。为避免此类事故的发生，设置奖励函数。如果车速超出了 $[V_{min}, V_{max}]$ ，将给予负的奖励；当车速在边界速度范围内时，为提高通行效率，以速度为指标，给予正向奖励。由此可得信号控制信息奖励函数：

$$r_{\text{light}} = \begin{cases} -(\beta_1 + (v_{\text{ego}} - V_{\text{max}})^2), & v_{\text{ego}} > V_{\text{max}}, \\ v_{\text{ego}}, & V_{\text{min}} \leq v_{\text{ego}} \leq V_{\text{max}}, \\ -(\beta_2 + (v_{\text{ego}} - V_{\text{min}})^2), & v_{\text{ego}} < V_{\text{min}}. \end{cases} \quad (7)$$

2.2.2 驾驶安全奖励函数

由 Krauss 跟驰模型<sup>[15]</sup>可知,跟驰车辆保持安全行驶应满足:

$$L(V_f) + V_f t < L(V_p) + D. \quad (8)$$

式中:  $V_f$ 和 $V_p$ 分别为当前车辆和前车的速度;  $t$ 为驾驶员反应时间;  $D$ 为两车的车头间距;  $L(V_f)$ 和 $L(V_p)$ 为当前车辆和前车的制动距离。

根据式(8)可计算得出安全的跟车距离。为保证驾驶安全,当前公交车与前车的距离需大于最小安全距离。因此,当车辆与前车的距离小于安全距离时,本文设定为车辆将获得一次值为 $\beta_3$ 的惩罚,记为 $r_{\text{safe}} = -\beta_3$ 。

2.2.3 能耗奖励函数

能耗奖励函数以车辆从初始位置到达目标位置整段路程的能量消耗最小为目标,因此,应考虑每一个单位步长内车辆所消耗的电能 $W_t$ ,本文以SUMO中的电动汽车能耗模型为基础构建能耗奖励函数:

$$r_{\text{energy}} = -W_t. \quad (9)$$

2.2.4 舒适性奖励函数

相比高振幅和较短持续时间的车速变动,低振幅和较长持续时间的车速变动更舒适,则针对车辆的绝对加速度,奖励函数可表示为:

$$r_{\text{comfort}} = -a_{\text{ego}}^2. \quad (10)$$

2.2.5 惩罚函数

为避免车辆的不当行为,需设置惩罚函数以避免违规行为的发生。综合考虑碰撞前车、超速和闯红灯,构建的惩罚函数表示为:

$$r_{\text{penalty}} = p_{\text{light}} + p_{\text{safe}}. \quad (11)$$

其中:

$$p_{\text{light}} = -\beta_4, \text{ run the red light}. \quad (12)$$

$$p_{\text{safe}} = -\beta_5, \quad v_{\text{ego}} > v_{\text{max}} \text{ or } D_{\text{ego}} \leq 0. \quad (13)$$

综合以上构建的各部分奖励函数,可得综合奖励函数:

$$r = \alpha_1 r_{\text{light}} + \alpha_2 r_{\text{safe}} + \alpha_3 r_{\text{energy}} + \alpha_4 r_{\text{comfort}} + \alpha_5 r_{\text{penalty}}. \quad (14)$$

式中:  $\alpha_i (i = 1, \dots, 5)$ 为权重系数。

2.3 基于TD3的生态驾驶优化模型

TD3是 Actor-Critic (AC) 框架下的一种强化学习算法,通过设置双Q值网络,解决过高估计Q值的问题,通过延迟更新策略,提高策略在训练过程中的稳定性。基于TD3算法的生态驾驶模型框架如图3所示。

TD3算法中共有6个网络,分别为 Actor网络 $\mu(S|\theta^\mu)$ 、两个Critic网络 $Q_1(S, A|\theta^{Q_1})$ 和 $Q_2(S, A|\theta^{Q_2})$ 、Target-Actor网络 $\mu'(S|\theta^{\mu'})$ 、两个 Target-Critic网络 $Q'_1(S, A|\theta^{Q'_1})$ 和 $Q'_2(S, A|\theta^{Q'_2})$ ,  $\theta^\mu$ 、 $\theta^{Q_1}$ 、 $\theta^{Q_2}$ 、 $\theta^{\mu'}$ 、 $\theta^{Q'_1}$ 、 $\theta^{Q'_2}$ 分别为6个网络的网络参数。

在训练过程中,定期从经验缓存池随机采样若

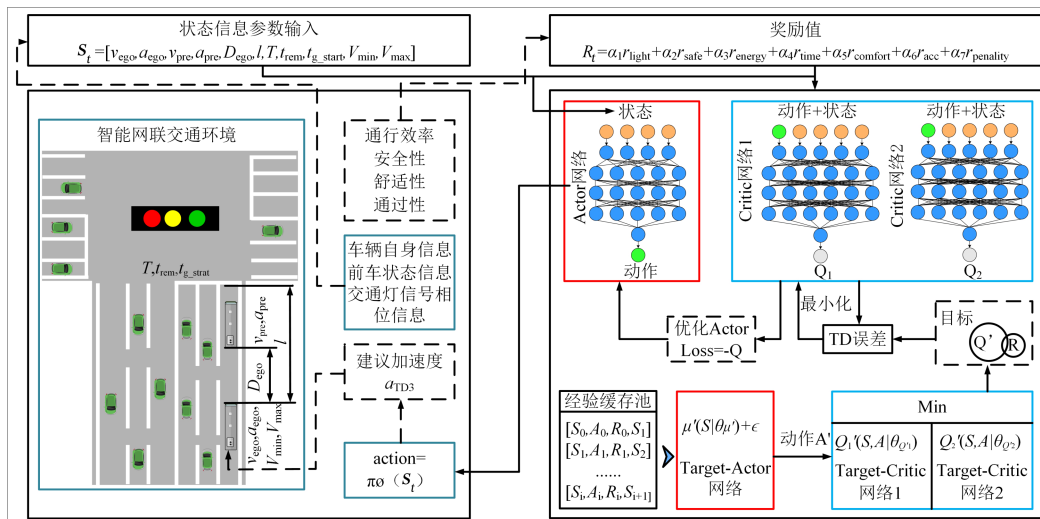


图3 基于TD3的生态驾驶模型框架

干组元组数据  $(S_t, A_t, R_t, S_{t+1})$ ，其中  $S_t$  为观测状态， $A_t$  为动作， $R_t$  为奖励值， $S_{t+1}$  为下一步动作。将采样的数据元组的动作价值与时序-差分目标相对应：

$$A' = \mu'(S_{t+1}|\theta^\mu) + \epsilon, \epsilon \sim \text{clip}(N(0, \sigma), -c, c)。(15)$$

$$y = R + \gamma \min_{m=1, 2} Q'_m(S_{t+1}, A'|\theta^{Q'_m})。(16)$$

$$q = Q_1(S_t, A_{\text{new}}|\theta^{Q_1})。(17)$$

式中： $\epsilon$  为 Critic 网络更新噪声，Target 网络与对应的 Actor 和 Critic 网络结构相同，随着 Actor 和 Critic 网络参数的更新而软更新：

$$\begin{cases} \theta^{Q'_m} = \tau\theta^{Q'_m} + (1 - \tau)\theta^{Q'_m}, (m = 1, 2), \\ \theta^\mu = \tau\theta^\mu + (1 - \tau)\theta^\mu。 \end{cases} (18)$$

Critic 网络通过最小 TD 误差实现更新：

$$\min Q_m((S_t, A_t|\theta^{Q'_m}) - y)^2, (m = 1, 2)。(19)$$

利用 Actor 网络计算  $S_t$  下的新动作，然后利用 Critic 网络计算新动作的评估值  $q_{\text{new}}$ ，如式 (20) 所示，通过梯度上升算法，最大化  $q_{\text{new}}$  更新 Actor 网络。

$$q_{\text{new}} = Q_1(S_t, \mu(S_t|\theta^\mu)|\theta^{Q_1})。(20)$$

### 3 仿真结果与分析

#### 3.1 仿真场景与参数

选择一辆某品牌 10 m 纯电动公交车作为研究对象，使用 SUMO 搭建仿真环境，以 Python 作为算法编程语言，通过 SUMO 的 Traci 接口来实现对车辆的控制。所涉及车辆选型等参数见表 1，图 4 为仿真车辆的电机 MAP 图。

为模拟信号交叉口的随机性，控制车辆以 30~40 km/h 的随机速度在随机时刻进入信号交叉口的通信范围内，信号交叉口的车流量为 1 000 车/h，算法参数及模型结构分别见表 2 和表 3。

#### 3.2 训练结果与分析

为验证算法的稳定性，本文对所提出的策略进行多次训练，训练过程的奖励轨迹如图 5 所示。由于 TD3 算法添加了噪声，在探索过程中会产生大量的边界动作，在本研究中表现为车辆停止不动直到此回合结束。因此，智能体在前 20~30 轮的训练过程中，所获得的奖励值一直都特别低。随着智能

表 1 主要仿真参数

参数	值
最大电池容量 $E_{\text{max}}/\text{Wh}$	48 300
车辆质量 $m/\text{kg}$	12 400
电机内阻 $R_m/\Omega$	0.12
迎风面积 $A/\text{m}^2$	7.6
空气阻力系数 $C_D$	0.67
滚动阻力系数 $f$	0.012
转矩常数 $K_T$	10.06
传动比 $G$	6.33
传动系统效率 $\eta_T$	0.9
能量回收效率 $\eta_r$	0.8
最高车速 $v_{\text{max}}/(\text{km/h})$	40
最低车速 $v_{\text{min}}/(\text{km/h})$	13
最大加速度 $a_{\text{max}}/(\text{m/s}^2)$	2.0
最大减速度 $d_{\text{max}}/(\text{m/s}^2)$	2.0
交叉口上游通信范围 $l_1/\text{m}$	300
交叉口下游路段长度 $l_2/\text{m}$	100
绿灯相位时长 $t_{\text{g\_cycle}}/\text{s}$	40
红灯相位时长 $t_{\text{r\_cycle}}/\text{s}$	60
仿真步长 $D/\text{s}$	1

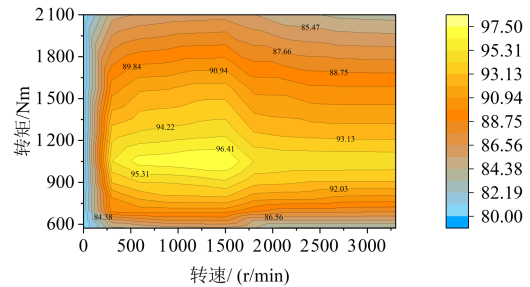


图 4 电机 MAP 图

体的进一步探索寻优，逐渐学习到怎样获得较大的奖励值，但由于其训练轮次较少，奖励值较之前有所升高但奖励函数曲线波动幅度仍较大。随着训练回合数的增加，智能体逐渐学习到了最优行驶策略，累计奖励值开始逐渐稳定，但由于添加了探索噪声，累计奖励值会在小范围内波动。总体而言，随着学习轮次的增加，累计的奖励值呈现出上升趋势，并逐渐收敛。训练结果表明，设计的奖励函数能使智能体（电动公交车）学习到最优的生态驾驶策略。

表 2 算法参数

参数	值
学习率	0.000 3
折现系数	0.99
重放缓冲区容量	100 000
样本批次大小	256
勘探噪声	0.1
Target 网络更新速率	0.005
Critic 网络更新噪声	0.2
策略更新延迟频率	2
$\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$	1, 1, 0.5, 3, 1
$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$	10, 10, 10, 50, 50

表 3 模型结构

	Actor		Critic1 和 Critic2	
	神经元数量	激活函数	神经元数量	激活函数
输入层	11	ReLU	12	ReLU
隐藏层	48	ReLU	128	ReLU
隐藏层	48	ReLU	128	ReLU
隐藏层	48	ReLU	64	ReLU
隐藏层	48	ReLU	64	ReLU
输出层	1	Tanh	1	Linear

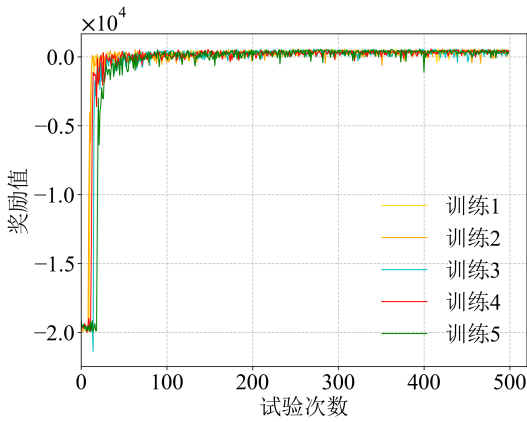


图 5 训练过程中的累计奖励值变化

3.3 不同场景下生态驾驶策略优化结果

根据公交车到达交叉口通信范围时信号灯的灯时信息，将交叉口划分为4种场景：

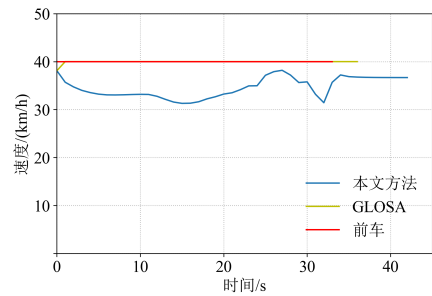
- 1) 绿灯时间充足，车辆可在此次绿灯相位通过；
- 2) 绿灯剩余时间较短，车辆无法在本次绿灯相位通过；

- 3) 红灯剩余时间较长，需要车辆减速停车；
- 4) 红灯剩余时间过短，车辆无需减速，可在绿灯相位通过。

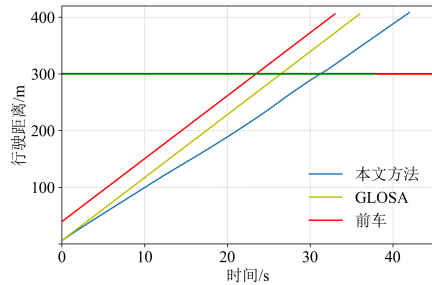
以典型的交叉口引导系统 GLOSA 作为基准算法，对比验证在 4 种不同的交通场景下车辆的节能效果、通行效率和舒适性。本文引入指标 RC 表征公交车乘客的舒适性：

$$RC = \frac{\sum_{i=1}^{t_f} (a_{ego})^2}{t_f} \quad (21)$$

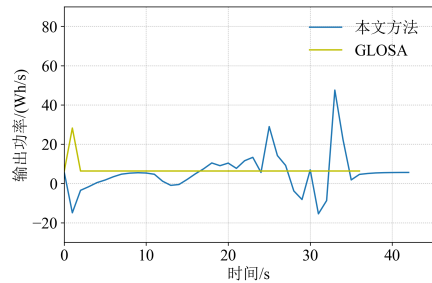
在交通场景 1 中，两种策略下的车辆行驶轨迹及输出功率变化如图 6 所示，指标计算结果见表 4。



(a) 速度轨迹对比



(b) 行驶轨迹对比



(c) 电池输出功率对比

图 6 场景 1 生态驾驶策略对比

该场景下，当前交通信号为绿灯，但剩余时间有 38 s，车辆以 38.10 km/h 的速度进入信号范围，此时绿灯剩余时间充足。GLOSA 系统选择加速至

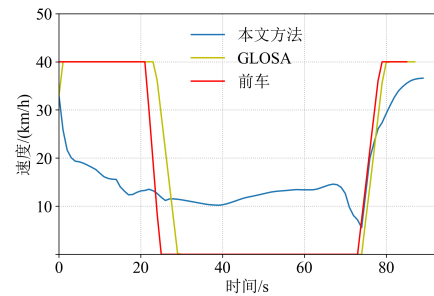
限速后匀速行驶，本文策略则选择先减速后加速，由图 5 可知，两种策略均可使车辆在绿灯相位通过停止线且与前车保持适当的距离。由表 4 可知，由于减速原因，本文策略控制的车辆通行时间与舒适性降低，但其在节能方面取得了较好的效果，能耗降低 9.82%。

表 4 场景 1 生态驾驶算法结果对比

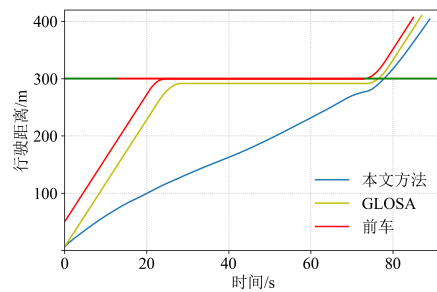
评价指标	GLOSA	本文方法
通行时间/s	37	43 (+16.22%)
平均速度/(m/s)	10.81	9.30 (-13.97%)
能耗/Wh	255.75	230.64 (-9.82%)
舒适性/(m <sup>2</sup> /s <sup>4</sup> )	0.007	0.006 (-28.57%)

在交通场景 2 中，两种策略下的车辆行驶轨迹及输出功率变化如图 7 所示，指标计算结果见表 5。该场景下，车辆以 32.96 km/h 的速度进入信号范围内，此时的信号灯相位为绿灯，剩余时间只有 13 s，无法在本次绿灯相位通过。当绿灯剩余时间小于车辆到达停车线的时间时，GLOSA 会控制车辆加速，由图 6 可知，车辆已经加速至最高车速，但依旧无法在绿灯窗口通过，当车辆即将抵达停车线时，车辆会停车等待较长时间。利用本文提出的生态驾驶策略，公交车进入信号范围内后，会减速至较低速度，而后以较低速度近似匀速行驶，通过停止线后加速至 36.61 km/h，然后选择匀速行驶。在整个过程中避免了停车等待，能耗降低 26.13%，舒适性提高 60%，通行时间仅增加 2 s。

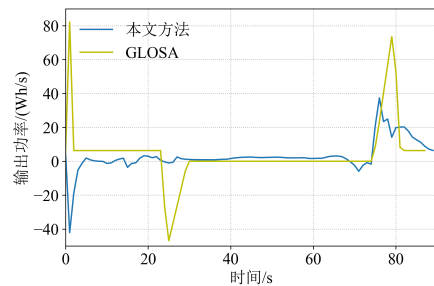
在交通场景 3 中，两种策略下的车辆行驶轨迹及电池输出功率变化如图 8 所示，指标计算结果见表 6。该场景下，车辆以 31.83 km/h 的速度进入通信范围，此时红灯剩余时间为 51 s，剩余时间较长，需要减速行驶。由图 7 中的行驶轨迹可知，GLOSA 策略和本文策略均能保证车辆在绿灯相位通过。GLOSA 策略以较小的减速度控制车辆减速，但受前方车辆排队的影响最终仍需停车等待。利用本文提出的策略，车辆进行持续的减速行驶，最终不停车通过交叉口。在通过交叉口后，两种策略控制下的车辆进行加速，GLOSA 策略控制的车辆最



(a) 速度轨迹对比



(b) 行驶轨迹对比



(c) 电池输出功率对比

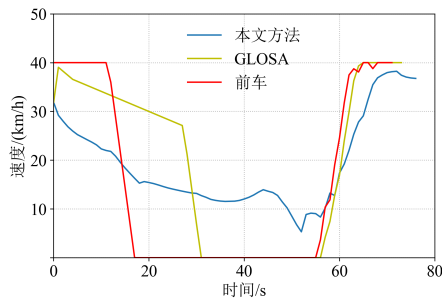
图 7 场景 2 生态驾驶策略对比

表 5 场景 2 生态驾驶算法结果对比

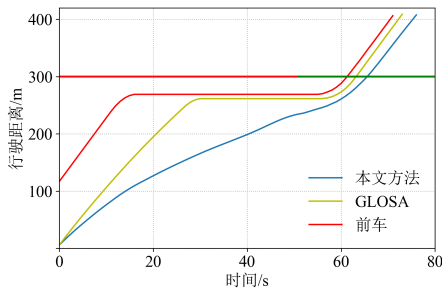
评价指标	GLOSA	本文方法
通行时间/s	88	90 (+2.27%)
平均速度/(m/s)	4.55	4.44 (-2.41%)
能耗/Wh	370.12	273.39 (-26.13%)
舒适性/(m <sup>2</sup> /s <sup>4</sup> )	0.50	0.20 (-60.00%)

终加速至最高限制车速，而本文策略控制车辆加速至 36.76 km/h 左右后匀速行驶。由表 6 可知，在整个过程中，本文方法相较于 GLOSA 其能耗降低 19%，舒适性提高 71.43%，通行时间增加 2 s，平均速度降低 0.14 m/s。

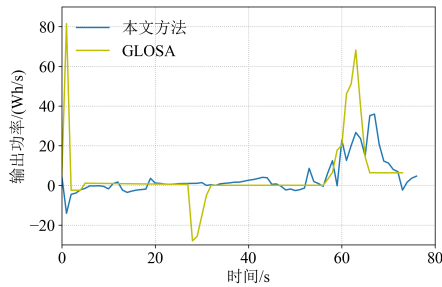
在交通场景 4 中，两种策略下的车辆行驶轨迹



(a) 速度轨迹对比



(b) 行驶轨迹对比



(c) 电池输出功率对比

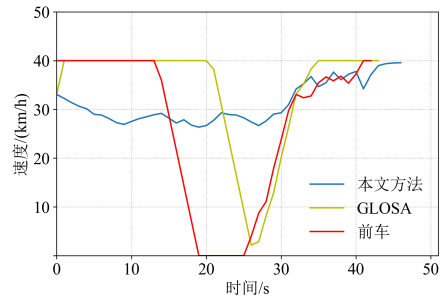
图 8 场景 3 生态驾驶策略对比

表 6 场景 3 生态驾驶算法结果对比

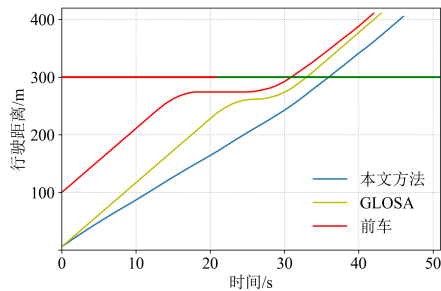
评价指标	GLOSA	本文方法
通行时间/s	75	77 (+2.70%)
平均速度/(m/s)	5.33	5.19 (-2.63%)
能耗/Wh	345.29	279.69 (-19.00%)
舒适性/(m <sup>2</sup> /s <sup>4</sup> )	0.49	0.14 (-71.43%)

及电池输出功率变化如图 9 所示, 指标计算结果见表 7。该场景下, 车辆以 33.06 km/h 的速度进入通信范围, 此时信号状态为红灯, 剩余时间为 21 s。红灯剩余时间较短, 即使全速行驶也可不停车通过交叉口。但由图 8 可知, 受红灯影响, 部分排队车辆无法在信号灯转绿后立刻加速离开, 因此, 造成

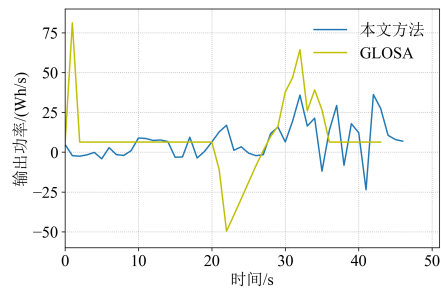
了 GLOSA 策略控制的车辆在接近停止线时, 仍需减速以避免撞到前车。本文方法在公交车进入信号范围内后减速至较低速度, 在车辆通过停止线后加速至 39.56 km/h 左右, 然后匀速行驶。整个过程中, 本文策略的能耗降低 14.51%, 舒适性提高 86.66%, 通行时间增加 4 s, 平均速度降低 0.79 m/s。



(a) 速度轨迹对比



(b) 行驶轨迹对比



(c) 电池输出功率对比

图 9 场景 4 生态驾驶策略对比

表 7 场景 4 生态驾驶算法结果对比

评价指标	GLOSA	本文方法
通行时间/s	43	47 (+9.30%)
平均速度/(m/s)	9.30	8.51 (-8.49%)
能耗/Wh	367.60	314.25 (-14.51%)
舒适性/(m <sup>2</sup> /s <sup>4</sup> )	0.90	0.12 (-86.66%)

通过上述分析可知,本文所提出的生态驾驶策略在不同场景下均可保证车辆与前车保持安全距离,并且于绿灯相位通过信号交叉口。在通过交叉口后,能加速至合理的车速并匀速行驶,符合实际情况。在车辆通过整个信号交叉口的过程中,本文策略在保证通行效率的前提下显著提高了能源利用效率,同时也提高了乘客的舒适性,实现了动态场景下电动公交车在信号交叉口通信范围内的生态驾驶。

#### 4 结论

1) 针对电动公交车如何最优通过信号交叉口的的问题,本文综合考虑安全、能耗、效率和舒适性多个目标,构建综合奖励函数,基于TD3深度强化

学习算法框架构建电动公交车生态驾驶优化模型。

2) 构建的模型可在保证安全、效率及舒适性的情况下,有效降低纯电动公交车通行交叉口区域的能量消耗。其中,相较于GLOSA策略,在第1类场景下能耗降低9.82%,第2类场景下能耗降低26.13%,第3类场景下能耗降低19.00%,第4类场景下能耗降低14.51%。

3) 构建的生态驾驶优化策略与传统只考虑控制车辆行驶、忽视复杂的交通环境的方法不同,综合考虑了交通环境信息的动态变化,并将车辆通过交叉口后的加速阶段纳入优化范围。针对不同的动态交通环境,实现不同的行驶策略,在保证安全和高效行驶的前提下有效降低了能量消耗,同时有效减少了急加速和急减速工况,提升了乘客的乘坐舒适性。

#### 参考文献 (References)

- [1] 李兵兵,庄伟超,董昊轩,等.车-路-云融合的智能混合动力汽车节能驾驶技术研究进展与挑战[J].汽车工程学报,2023,13(4):456-469.  
LI Bingbing, ZHUANG Weichao, DONG Haoxuan, et al. Eco-Driving Strategy for Intelligent Hybrid Electric Vehicles with Vehicle-Road-Cloud Fusion: Research Progress and Challenges [J]. Chinese Journal of Automotive Engineering, 2023, 13 (4) : 456-469. (in Chinese)
- [2] 袁伟,张雅丽,王虹霞,等.纯电动公交车交叉口节能驾驶策略[J].中国公路学报,2021,34(7):54-66.  
YUAN Wei, ZHANG Yali, WANG Hongxia, et al. Energy-Saving Driving Technique for Pure Electric Buses in Intersection [J]. China Journal of Highway and Transport, 2021, 34(7) : 54-66. (in Chinese)
- [3] ZHANG Yali, FU Rui, GUO Yingshi, et al. Eco-Driving Strategy for Connected Electric Buses at the Signalized Intersection with a Station [J]. Transportation Research, Part D: Transport and Environment, 2024, 128: 104076.1-104076.21.
- [4] 刘显贵,洪经纬,王晖年,等.基于前车速度预测的信号交叉口汽车生态驾驶控制策略[J].安全与环境学报,2021,21(6):2743-2750.  
LIU Xiangui, HONG Jingwei, WANG Huinian, et al. Vehicle Eco-Driving Control Strategy Based on Speed Prediction of the Front Vehicle at the Signalized Intersection [J]. Journal of Safety and Environment, 2021, 21(6) : 2743-2750. (in Chinese)
- [5] SHI J, QIAO F, LI Q, et al. Application and Evaluation of the Reinforcement Learning Approach to Eco-Driving at Intersections under Infrastructure-to-Vehicle Communications [J]. Transportation Research Record, 2018, 2672 (25) : 89-98.
- [6] LIAN Renzong, PENG Jiankun, WU Yuankai, et al. Rule-Interposing Deep Reinforcement Learning Based Energy Management Strategy for Power-Split Hybrid Electric Vehicle [J]. Energy, 2020, 197: 117297.1-117297.9.
- [7] 庄伟超,丁昊楠,董昊轩,等.信号交叉口网联电动汽车自适应学习生态驾驶策略[J].吉林大学学报(工学版),2023,53(1):82-93.  
ZHUANG Weichao, DING Haonan, DONG Haoxuan, et al. Learning Based Eco-Driving Strategy of Connected Electric Vehicle at Signalized Intersection [J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(1) : 82-93. (in Chinese)
- [8] BAI Zhengwei, HAO Peng, SHANGGUAN Wei, et al. Hybrid Reinforcement Learning-Based Eco-Driving Strategy for Connected and Automated Vehicles at Signalized Intersections [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23 (9) : 15850-15863.
- [9] ZHOU Mofan, YU Yang, QU Xiaobo. Development of an Efficient Driving Strategy for Connected and Automated Vehicles at Signalized Intersections: A Reinforcement

- Learning Approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(1):433-443.
- [10] HEUILLET A, COUTHOUIS F, DÍAZ-RODRÍGUEZ N. Explainability in Deep Reinforcement Learning [J]. Knowledge-Based Systems, 2021, 214: 106685.1-106685.14.
- [11] ZHOU Quan, LI Ji, SHUAI Bin, et al. Multi-Step Reinforcement Learning for Model-Free Predictive Energy Management of an Electrified Off-Highway Vehicle [J]. Applied Energy, 2019, 255: 113755.1-113755.12.
- [12] LI Guofa, YANG Yifan, LI Shen, et al. Decision Making of Autonomous Vehicles in Lane Change Scenarios: Deep Reinforcement Learning Approaches with Risk Awareness [J]. Transportation Research Part C: Emerging Technologies, 2022, 134: 103452.1-103452.18.
- [13] HOBEROCK L L. A Survey of Longitudinal Acceleration Comfort Studies in Ground Transportation Vehicles [R]. Washington DC: Council for Advanced Transportation Studies, 1976.
- [14] YI K S, CHUNG J T. Nonlinear Brake Control for Vehicle CW/CA Systems [J]. IEEE/ASME Transactions on Mechatronics, 2001, 6(1):17-25.
- [15] KRAUSS S, WAGNER P, GAWRON C. Metastable States in a Microscopic Model of Traffic Flow [J]. Physical Review E, 1997, 55(5):5597-5602.

### 作者简介



郝毅 (2000-), 男, 陕西西安人, 硕士研究生, 主要研究方向为汽车生态驾驶控制技术。

E-mail: 2022122054@chd.edu.cn

### 通信作者



潘应久 (1990-), 男, 山东章丘人, 博士, 讲师, 主要研究方向为智能汽车节能驾驶与安全。

E-mail: panyingjiu@chd.edu.cn