信号交叉口网联电动汽车自适应学习生态驾驶策略

庄伟超，丁昊楠，董昊轩，殷国栋，王　茜，周朝宾，徐利伟

（东南大学 机械工程学院，南京 211189）

摘　要：提出了一种面向信号交叉口的自适应学习生态驾驶策略。首先，搭建了电动汽车纵向动力学模型，建立了信号灯交叉路口的虚拟交通仿真环境；其次，以车辆能耗最小化与通行效率最大化为目标，耦合设计强化学习奖励函数，基于深度确定性策略梯度算法（DDPG）对车辆加速度进行实时控制与训练；最后，通过蒙特卡洛试验法，验证本文提出的强化学习生态驾驶策略在不同初始交通场景下的有效性与鲁棒性。仿真结果表明，相较于常规“加速-匀速-制动（ACB）”策略，本文提出的强化学习生态驾驶策略在单路口和多路口场景下均可有效提升通行效率和能量效率。同时，智能网联汽车数字孪生试验平台的多次实车试验表明，本文的强化学习算法控制效果良好，可以有效减少车辆路口等待时长，降低能耗同时提高通行效率。

关键词：车辆工程；网联电动汽车；生态驾驶；深度强化学习；信号交叉口；数字孪生

中图分类号：U469.72　　**文献标志码**：A　　**文章编号**：1671-5497(2023)01-0082-12

DOI：10.13229/j.cnki.jdxbgxb20210598

Learning based eco-driving strategy of connected electric vehicle at signalized intersection

ZHUANG Wei-chao，Ding Hao-nan，DONG Hao-xuan，YIN Guo-dong，
WANG Xi，ZHOU Chao-bin，XU Li-wei

（*School of Mechanical Engineering*，*Southeast University*，*Nanjing* 211189，*China*）

Abstract：A deep reinforcement learning based eco-driving strategy for connected electric vehicle（EV）was proposed to improve its energy efficiency at signalized intersection. Firstly，the dynamics of the EV is modelled，and the simulation environment of signalized intersection crossing scenario is established. Secondly，the reward function including multiple objectives is designed considering energy consumption reduction and travel efficiency improvement. The Deep Determinate Policy Gradient（DDPG）is developed to control the vehicle acceleration in continuous action space. Finally，a Monte Carlo simulation is

收稿日期：2021-06-26.

基金项目：国家杰出青年科学基金项目（52025121）；国家自然科学基金项目（51805081，51975118）；江苏省重点研发计划项目（BE2019004）.

作者简介：庄伟超（1990-），男，副教授，博士. 研究方向：车辆动力学与控制，智能网联汽车. E-mail：wezhuang@seu.edu.cn

通信作者：殷国栋（1976-），男，教授，博士. 研究方向：车辆动力学与控制，新能源与智能网联汽车，车辆节能与安全控制. E-mail：ygd@seu.edu.cn

conducted to verify the effectiveness and robustness of proposed method in different driving conditions. The simulation results show that the proposed strategy can improve the vehicle energy efficiency while ensuring travel efficiency in both single and multiple intersection scenarios, compared to a conventional accelerate-constant-brake strategy. In addition, a field test is conducted based on a developed connected automated vehicle digital twin platform. The experiment results show that the proposed reinforcement learning based eco-driving strategy has the potential to improve the vehicle energy efficiency and travel efficiency, simultaneously.

Key words: vehicle engineering; connected electric vehicle; eco-driving; deep reinforcement learning; signalized intersection; digital twin

0　引　言

受限于电池能源的技术限制,电动汽车能源利用成为热点研究领域[1]。为了提高分布式驱动电动汽车的经济性和续航里程,除了从汽车转矩优化分配研究电动汽车驱动效率特性[2],在不同驾驶场景下汽车的能量利用率上,基于V2X通信技术的车联网系统可以实现提前获取道路交通信息,并依据节能驾驶策略,可大幅提高能源利用率[3]。在城市环境中,汽车受信号灯影响的走走停停现象是其能耗的重要原因之一[4]。为提高城市场景下汽车的能量利用率,相关学者[5]提出了经济性路口通行策略(Eco-approach and departure, EAD)。EAD通过获取信号灯相位配时(Signal phase and timing, SPaT)、周围车辆状态、道路限速等信息,合理规划车速,避免怠速停车,减少汽车能耗并提高交通效率[6]。作为智能网联汽车节能优化研究[7],最为典型的EAD应用是绿波车速引导策略(Green light optimal speed advisory, GLSOA)[8],其依据信号灯SPaT和道路限速,优化不停车的建议车速区间,引导驾驶员以高效且节能的车速通过路口。

从控制算法的角度出发,生态驾驶策略可以分为基于规则的方法和基于优化的方法[9]。基于规则的方法针对信号灯动态调节控制车速,算法计算简单直观,应用广泛,Hao等[10]提出一种基于规则算法的协同经济性驾驶控制系统,适用于固定配时和自适应配时信号灯路口,但该方法依赖经验的总结,因此在实际应用中需要大量的标定与调参。基于优化理论的控制策略一般可分为解析优化法、数值计算优化法。庞特里亚金法[11]、伪谱法[12]等是典型的解析优化法,可求解能效最优车速,同时具有较高的计算效率,但其解可能为局部最优。动态规划算法(Dynamic programming, DP)可以依据特定环境条件计算全局最优解,Han[13]利用DP算法研究了燃油汽车与电动汽车的不同节能特性,求解混合动力汽车的最优控制律,获得了出色的燃油经济性。但DP算法对于状态空间较大、且多维复杂场景时计算效率较低,难以保证控制的实时性,针对动态交通系统存在鲁棒性问题,因此该算法无法广泛应用于实际动态交通场景。

近年来,强化学习方法因其场景适应性强、控制实时性好,被逐渐应用在车辆自动驾驶决策领域[14],基于连续动作空间的强化学习路口通行决策算法被初步挖掘。Wu等[15]采用深度确定性策略梯度算法(Deep determinate policy gradient, DDPG),以插电混合动力大巴为研究对象,在全程载客运营工况下针对乘客数量变化与车辆运行状态的变化场景,优化了连续状态空间下的电池与燃油能量分配,所提出的能量管理策略性能接近DP。Li等[16]考虑了地形对智能体决策影响,所提出的算法与DP策略相比,燃油经济性差距缩小至近6.4%。Zhou等[17]结合DQN与DDPG算法,通过多智能体协同控制,解决换道与路口通行横纵耦合决策问题,在不增加通行时间的基础上能耗提升了46%。Zhu等[18]提出了一种基于长短期记忆近端策略优化强化学习算法,利用循环网络优化混合动力汽车的能量管理策略,对比人类驾驶员行为提升了17%的燃油消耗。上述研究多考虑简化的静态交通场景,当道路环境发生变化时算法的适应性较差,导致上述方法在实际驾驶场景中使用率较低。从实际角度出发,在连续状态下,动态交通场景的车路协同驾驶策略需权衡通行效率与能量消耗多目标耦合综合优化。

综上所述,本文以智能网联电动汽车为研究

对象,针对城市工况单信号灯控路口经济性通行问题,设计车辆能耗与通行效率的多目标耦合奖励函数,基于强化学习算法设计路口通行速度优化控制策略;车辆通过获取当前时刻下环境道路与整车状态信息,并考虑动态交通约束下实时控制车辆安全高效地通过路口,实现能量与通行时间最优;然后,基于蒙特卡洛试验法,对随机路口距离、信号灯相位配时、单路口与多路口不同交通场景下的自适应生态驾驶进行仿真,以验证所提出的策略有效性与鲁棒性,仿真验证策略在多路口条件下在能耗与通行效率方面的优势。最后,搭建智能网联电动汽车强化学习数字孪生试验平台,对所提出的通行策略进行实车试验。

1 信号灯控路口经济通行问题

1.1 问题描述

在智能网联环境下,车载控制系统根据信号获取到的实施道路数据计算建议车速,并结合强化学习算法实时优化车速谱,从而实现满足通行效率与能量最优的车速轨迹,避免在信号灯路口的停车等待。

本文具体研究场景为,在一条道路长度为l的信号灯路口车辆以车速v_0进入路口,并在t_f时刻通过路口停车线,由本文的优化算法来实时计算出当前信号灯状态下车速轨迹,兼顾通行时间与能量消耗的多目标问题,信号灯控路口经济性驾驶问题如图1所示。

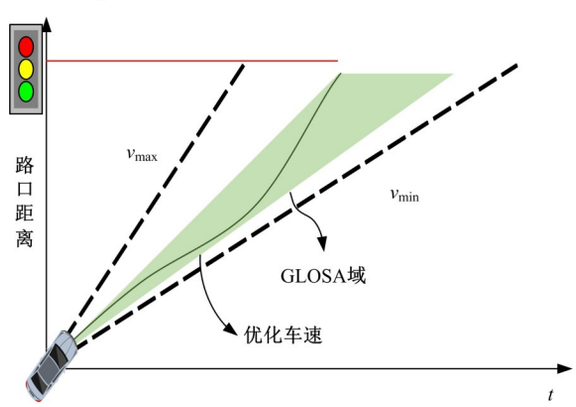


图1 信号灯交叉口经济性驾驶

Fig. 1 Economic driving at signalized intersections

1.2 最优控制问题构建

本文以分布式驱动电动汽车为研究对象,以加速度为控制量,由"时间-能耗"最优控制原理可以建立性能指标:

$$\begin{cases}\min J(x(t),u(t))=\int_0^{t_f}\left[\lambda+P_b(t)\right]\mathrm{d}t\\ \text{s.t. } l(0)=0,l(t_f)=S\\ v(0)=v_{ini}\\ u_{min}\leqslant u(t)\leqslant u_{max}\\ v_{min}\leqslant v(t)\leqslant v_{max}\end{cases}\tag{1}$$

式中:λ为通行时间的权重;$P_b(t)$为当前时刻下电池的功率;t_f为车辆当前路口所需要的时间;v_{ini}为初始车速;为保证车辆行驶过程中的舒适度,u_{min}和u_{max}分别为减速度与加速度的最小值与最大值;v_{min}和v_{max}分别为交通效率最小值与安全速度最大值,本文中v_{min}取20 km/h,v_{max}取60 km/h。

上述优化问题针对环境状态已知工况下可求最优解,然而在动态交通环境下需再次获取环境状态重新求解,增加不必要等待时间,因此控制实时性较差,难以适用于实际交通场景。本文采用的强化学习方法,其控制目标与控制问题相似,但目标函数表达方式不同,需要使用不同的术语表示相同的概念[19]。强化学习的目标是基于当前环境状态下找到最大化奖励值的策略。由马尔可夫决策过程(Markov decision process, MDP)可以将智能体决策过程以数学模型表示为状态S、动作A、状态转移概率P、回报值R以及折扣因子γ五个向量。根据转移概率P是否已知可进一步将智能体学习决策分为基于模型(Model based)和无模型(Model free),基于模型的强化学习则可以转化为最优控制问题,对于无模型而言,又可分策略迭代[20]、值迭代[21]与策略搜索[22]方法。本文基于无模型的强化学习方法考虑动态环境变量,即当车辆进入不同相位红绿灯路口时,无需重新计算最优值,而是根据训练好的策略计算适应不同环境的最佳通行策略。设定目标函数为:

$$\begin{cases}\max\left[\sum_{t=0}^{N}r_t(x_t,u_t)\right]\\ \text{s.t. } x_{t+1}=f_t(x_t,u_t,e_t)\\ u_t=\pi_t(\tau)\end{cases}\tag{2}$$

式中:N为智能体完成任务目标所需的步数;r_t为当前策略状态下获得的奖励值;f_t为车辆状态方程;$\tau_t=(u_1,u_2,\cdots,u_{t-1},x_0,x_1,\cdots,x_t)$为智能体所选的策略轨迹;$\pi_t=\tau_t$为控制策略。

1.3 车辆动力学模型

本文采用的无模型强化学习算法依据现实环境下的状态实时决策,其最大的优点在于无需建

立现实世界的模型，但考虑到训练的安全性以及整体效率，训练过程在虚拟环境下进行，因此需要建立车辆模型。

为了简化汽车在信号灯交叉路口处的行驶环境，对仿真进行了如下假设：①本文所研究的EV装备是4台独立轮毂电机，只考虑车辆纵向运动；②车辆在无坡度变化的平直路面上行驶，路面附着良好，轮胎无打滑。

依据车辆纵向动力学，可分析车辆在行驶过程中受到加速阻力、坡道阻力、滚动阻力以及空气阻力，建立如下表达式：

$$\delta m\frac{\mathrm{d}v}{\mathrm{d}t}=\frac{T_{\mathrm{m}}}{r_{\mathrm{d}}}-\frac{1}{2}\rho AC_{\mathrm{D}}v^{2}-mgf_{\mathrm{r}}\cos\varphi-mg\sin\varphi \tag{3}$$

式中：m为车辆的质量，kg；δ为汽车旋转换算系数；T_{m}为电机驱动转矩，N•m；r_{d}为车轮滚动半径，m；ρ为空气密度，kg/m^3；A为车辆迎风面积，m^2；C_{D}为空气阻力系数；g为重力加速度，m/s^2；f_{r}为滚动阻力系数；φ为地面坡度。

考虑本研究行驶工况为单车道行驶，4台轮毂电机负载相同，因此汽车每个轮毂电机的功率计算公式为：

$$P_{\mathrm{m}}=\frac{F_{\mathrm{m}}r_{\mathrm{d}}n}{9.55}\eta_{\mathrm{m}}^{-\mathrm{sign}(F_{\mathrm{m}})} \tag{4}$$

式中：F_{m}为电机作用于轮胎力，N；P_{m}为电机的功率，W；sign()为符号函数；η_{m}为电机效率，由图2电机特性效率MAP中查询可得。

当电机的功率已知时，忽略系统的电能损失，可计算出电池的功率：

$$P_{\mathrm{b}}=\sum_{i=1}^{4}P_{\mathrm{m}}^{i}\eta_{\mathrm{m}}^{-\mathrm{sign}(F_{\mathrm{m}})} \tag{5}$$

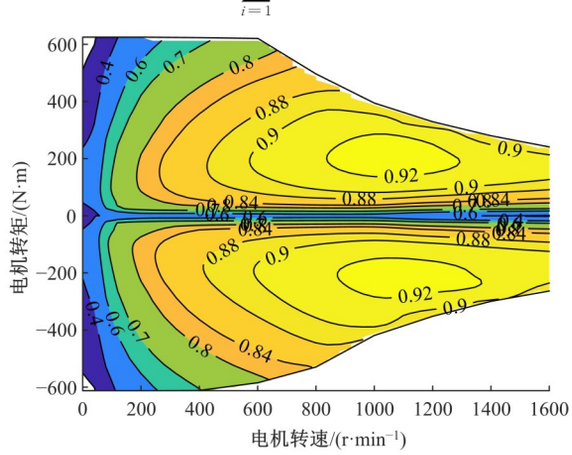


图2　电机效率和转矩-转速特性

Fig. 2　Characteristics of motor efficiency and torque-speed

2　基于强化学习的路口通行策略

基于第1节建立的车辆纵向动力学模型，本节建立了强化学习路口通行策略框架，其中包含信号灯状态表征函数与车辆-路状态信息，通行时间与能量效率多目标优化奖励函数，如图3所示。

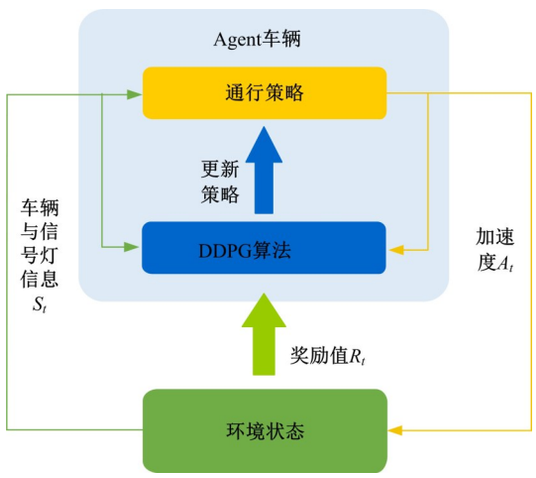


图3　基于强化学习的路口通行策略

Fig. 3　RL based intersection approaching strategy

2.1　环境信息与奖励函数

2.1.1　环境状态信息

为了准确描述路口环境信息，需要建立信号灯模型。本文将黄灯相位视为不可通行状态，信号灯相位设定采用了余弦函数来表述当前的状态，定义T_{state}来表述当前信号灯相位的通行状态，即当$T_{\mathrm{state}}\leqslant 0$时，表示当前的路口不可通行；当$T_{\mathrm{state}}>0$时，当前路口可通行。用$t_{\mathrm{r}}$，$t_{\mathrm{g}}$，$t_{\mathrm{y}}$分别表示红灯、绿灯和黄灯的相位时间，对于信号灯的相位周期描述可表示为：

$$T_{\mathrm{state}}=\begin{cases}\sin\left[\dfrac{\pi\left(t+t_{\mathrm{y}}\right)}{t_{\mathrm{r}}+t_{\mathrm{y}}}\right], \text{红灯}\\ \sin\left[\dfrac{\pi\left(t+t_{\mathrm{g}}-t_{\mathrm{r}}\right)}{t_{\mathrm{g}}}\right], \text{绿灯}\end{cases} \tag{6}$$

例如，当红色相位、绿色相位与黄色相位的时间分别为40、60、3 s时，则可用图4表示。

智能体的每一步决策都取决于环境对车辆的反馈作用，所以环境状态量的选择对智能体决策起到至关重要的作用。根据现实驾驶环境观测到的状态信息，本文选取9个状态信息，如表1所示，结合表中的状态信息，智能体可获得足够的环境信息。

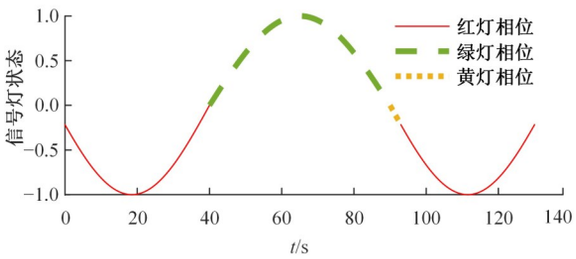


图4 信号灯相位状态模型

Fig. 4 Signal phase and state model

表1 环境状态输入

Table 1 Environment state input

输入量/单位	描述
$v/(\mathrm{m\cdot s^{-1}})$	车辆当前车速
$v_e/(\mathrm{m\cdot s^{-1}})$	与建议车速的差值
$\int\Delta v\mathrm{d}t$	累计误差值
$a/(\mathrm{m\cdot s^{-2}})$	当前时刻的加速度
l_t/m	当前时刻车辆到路口的距离
$\Delta E/\mathrm{kJ}$	每一步电池消耗的能量
$\int\Delta E\mathrm{d}t/\mathrm{kJ}$	累计能量消耗
$t_{\mathrm{rem}}/\mathrm{s}$	当前信号灯剩余时间
T_{state}	信号灯的状态

2.1.2 奖励函数

根据选取的状态信息与训练目标期望,奖励函数应当具有以下3个方面信息:

(1)根据任务要求,车辆应当以较快速度完成设定路程的行驶。当电动汽车静止不动时不需要消耗能量,该策略显然不可取。因此,为使智能体学习高效的通行策略,需对其设计参考车速。对于单信号灯控路口,电动汽车通过路侧设备获取到道路的实时信息,根据当前车辆距离路口的实时位置与信号灯的剩余时间,可以计算通过路口的最快建议车速。考虑动态交通环境下车辆每次进入路口时信号灯状态为随机值,定义若当前路口下为红灯时可以通行的最大建议车速为$v_{\mathrm{rec}}=\max\{l_t/t_{\mathrm{rem}},v_{\max}\}$;若当前路口下为绿灯时以当前车速加速至最大限速可通过,则$v_{\mathrm{rec}}=v_{\max}$;若当前最大车速不可通过时$v_{\mathrm{rec}}=0$,用公式表示为:

$$v_{\mathrm{rec}}=\begin{cases}\max\left\{\dfrac{l_t}{t_{\mathrm{rem}}},v_{\max}\right\},\text{红灯可行}\\ v_{\max},\text{绿灯可行}\\ 0,\text{其他}\end{cases}\tag{7}$$

取建议车速与当前规划的车速的差值v_e来描述智能体的效率,即被控车辆以尽可能大的速度到达设定的信号灯控路口,当误差越大,效率越低,因此相对惩罚也应当较高。

(2)应当保证车辆有相对较低的能耗。在最优控制问题中,目标函数包含在给定的时间点内达到目标的距离时使得整段路程的能量消耗最小,因此在每一个时间步长内,应考虑车辆所消耗的能量J_t。考虑电动汽车再生制动效能,回收的能量有部分会被损失掉,为了避免过多的制动导致能量的损失,当车辆采取制动时应当采取相对较小的惩罚。

(3)考虑智能体控制车辆的状态成本。为了防止车辆快速通过而闯红灯,需加上对车辆到达路口时的状态判定,若车辆闯红灯时,训练终止,并获得一次值为100的惩罚。另外,为了防止训练中智能体选择直接减速停车所获得的惩罚最小造成的局部收敛,选取电池的能耗与速度的误差两个信息,当二者的值在合理范围内,则给予正向奖励。考虑智能体输出的控制量因震荡引起的不舒适性,对过大的加速度控制量也要有相应惩罚。

综上所述,本研究的奖励函数设置为:

$$r_t=\left(\zeta J_t^2+\xi v_{\mathrm{e}}^2\right)+u_{t-1}^2+M_{\mathrm{v}}+M_{\mathrm{red}}\tag{8}$$

式中:ζ、ξ为权重系数;u_{t-1}为上一步长智能体输出的加速度;M_{v}、M_{red}为阶跃状态函数,分别定义为当车速误差在0.5 m/s内可获得奖励与车辆到达路口时闯红灯需得到的惩罚;用加号表示奖励;减号表示惩罚。

$$M_{\mathrm{v}}=\begin{cases}+1,e^2\leqslant 0.25\\ 0,e^2>0.25\end{cases}\tag{9}$$

$$M_{\mathrm{red}}=\begin{cases}-100,\text{闯红灯}\\ 0,\text{其他}\end{cases}\tag{10}$$

2.2 深度确定性策略梯度通行策略求解方法

基于DDPG交叉路口通行决策框架如图5所示,下面具体介绍求解方法。

对于随机策略$\pi_\theta=P[a|s,\theta]$,其含义为在状态为s时,动作符合参数为θ的概率分布,例如高斯策略;对于采用确定性策略$a=\mu_\theta(s)$,表示为在状态s下动作输出唯一,本文采用基于策略的深度确定性策略梯度算法,智能体在给定的驾驶环境状态s下,根据策略π采取加速度a所返回的价值记作$Q_\pi(s,a)$,通过迭代Q值来最大化奖励值。若用τ表示当前一组状态下的行为序列$s_0,u_0,\cdots,s_n,u_n$,$r(t)=\sum r(s_t,u_t)$表示在当前轨迹下的奖励,目标函数即为:

$$J(\theta)=E\left[\sum_{t=0}^{n}r(s_t,u_t)|\pi_0\right]=\sum_{\tau}P(\tau,\theta)r_t\tag{11}$$

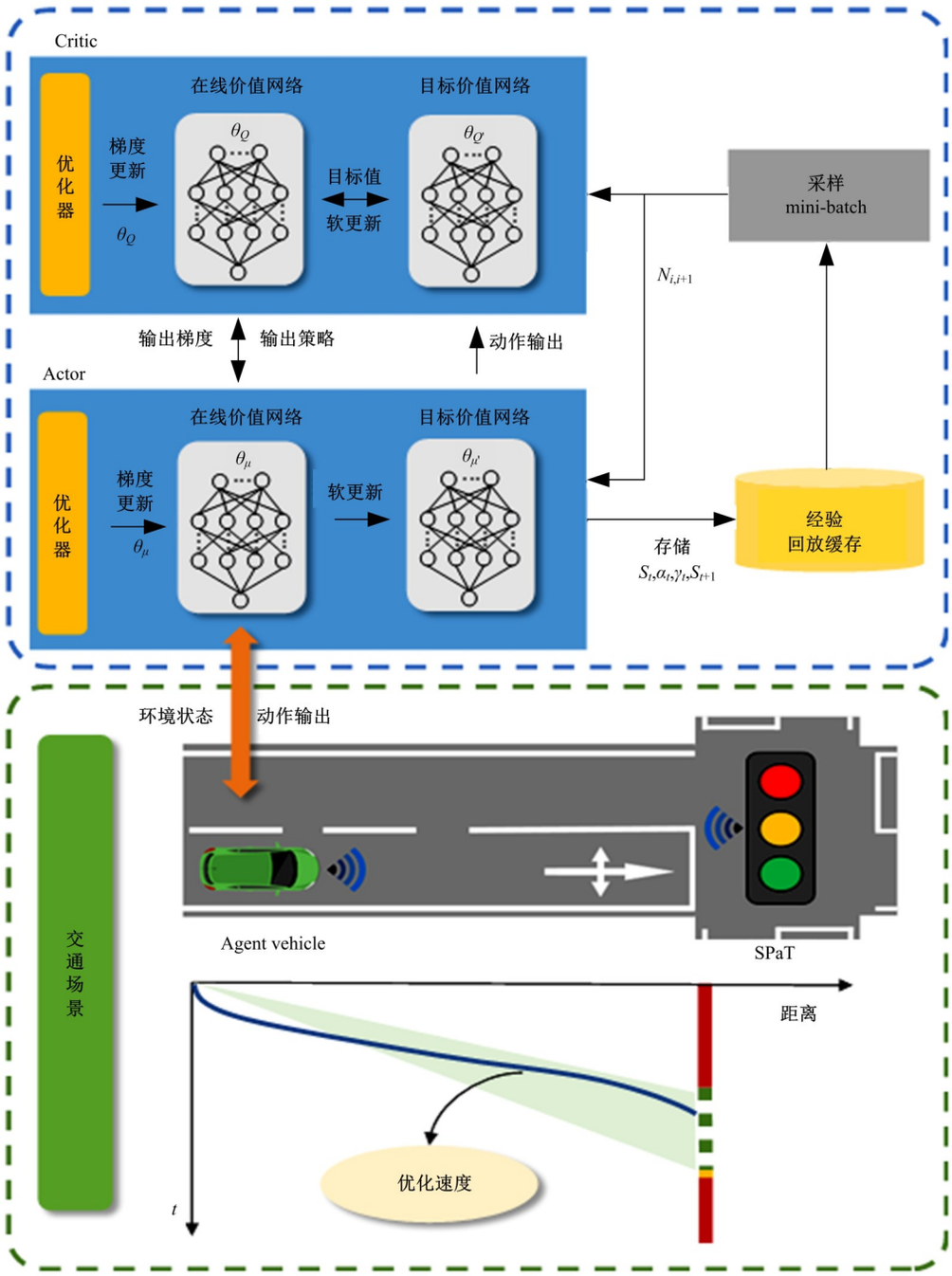


图 5　基于 DDPG 交叉路口通行决策框架

Fig. 5　DDPG-based intersection approaching strategy framework

式(11)表达含义为找到最优参数θ,使得$\max J(\theta)=\max[\sum P(\tau,\theta)r_t]$。采用策略梯度法,对参数$\theta$更新的公式为:

$$\theta'=\theta+\alpha\nabla_\theta J(\theta) \tag{12}$$

对目标函数式(12)求导可得:

$$\nabla_\theta J(\theta)=\nabla_\theta\sum_\tau P(\tau,\theta)r_t=$$

$$\nabla_\theta\sum_\tau P(\tau,\theta)\frac{\nabla_\theta P(\tau,\theta)r_t}{P(\tau,\theta)} \tag{13}$$

由于上述目标函数与策略函数均在连续可微的状态下,由策略梯度定理[23]可得随机策略梯度的计算公式为:

$$\nabla_\theta J(\theta)=E_{S\sim\rho_\pi,a\sim\pi_\theta}\left[\nabla_\theta\log\pi_\theta(s,a)Q^\pi(s,a)\right] \tag{14}$$

当策略$a=\mu_\theta(s)$确定时,梯度计算公式即可改为:

$$\nabla_\theta J(\theta)=E_{S\sim\rho^\mu}\left[\nabla_\theta\mu_\theta\nabla_\theta Q^\mu(s,a)|_{a=\mu_\theta(s)}\right] \tag{15}$$

由于确定性策略是由当前环境下产生的动作

策略是固定唯一的,无法进行探索环境,所以需采用异策略(Off-policy)进行学习。通过深度Q网络强化学习方法(Deep Q-Network,DQN)的经验回放(Memory replay)和独立目标网络(Independent target network),采用"动作-评估(Actor-critic, AC)"框架,动作网络用来调整θ的值,而评估网络则用来逼近值函数$Q_\omega(s,a)\approx Q_\pi(s,a)$,其中$\omega$为待估计的参数,AC算法的更新公式为:

$$\begin{cases}\delta_t=r_t+\gamma Q^{\omega'}\left[s_{t+1},\mu_{\theta'}(s_{t+1})-Q^{\omega}(s_t,a_t)\right]\\ \omega_{t+1}=\omega_t+\alpha_\omega\delta_t\nabla_\omega Q^{\omega}(s_t,a_t)\\ \theta_{t+1}=\theta_t+\alpha_\theta\nabla_\theta\mu_\theta\nabla_a Q^{\omega}(s_t,a_t)\\ \omega'=\tau\omega(1-\tau)\omega'\\ \theta'=\tau\theta+(1-\tau)\theta\end{cases}\tag{16}$$

综上所述,图5为基于DDPG算法的灯控路口经济性驾驶策略框架,其中环境状态包含了车辆自身状态信息与路侧设施提供的路口距离和SPaT信息,上层智能体控制器根据传感器获取到的信息输入AC网络,通过在虚拟仿真环境训练计算决策的梯度用来更新网络参数,并根据训练得到的网络在策略部署时的环境中计算出最佳通行策略。

3 仿真结果与分析

本节通过采集真实道路交通场景数据,采用MATLAB/Simulink建立仿真模型,验证本文提出方法在巡航车速计算速度和节能效率的优势。仿真使用电脑配置Intel Core i7-10700K@3.8 GHz的CPU和32 GB的RAM。

3.1 仿真场景与参数设置

本文采集了实际路况下的信号灯相位数据以及交通路口长度信息,选取南京市雨花台区宁丹大道一段长约500 m的道路,路面平直,具备仿真测试道路的理想条件。该路段交叉口的信号灯相位信息为红灯时长55 s,绿灯时长56 s,黄灯时长3 s。为了模拟车辆路口时的随机性,随机初始化速度的变化范围为35~45 km/h。

取车辆参数:质量m=1800 kg,重力加速度g=9.8 m/s^2,空气阻力系数C_D=0.3,迎风面积A=2.5 m^2,旋转质量换算系数δ=1.1,滚动阻力系数f=0.012。考虑乘坐的舒适度,车辆的加减速度区间取[-3,3] m/s^2,取随机初始车速为35~45 km/h。取奖励函数中的参数ζ、ξ分别为2×10^{-3}、5×10^{-4},其余的强化学习参数的设置如表2所示。

表2 DDPG算法仿真参数设置

Table 2 Parameter setting of DDPG algorithm

参数	数值	参数	数值
目标平滑更新因子	10^{-3}	单次训练最大步数	10^3
折扣系数γ	0.99	总仿真时间/s	10^2
mini-batch	32	停止训练奖励值	10^3
经验回放缓存	10^{-6}	随机噪声方差	0.6
采样时间t/s	0.1	随机噪声衰退率	10^{-5}

3.2 仿真结果

3.2.1 训练结果

为了验证算法的稳定性,取随机种子对本文提出的策略多次训练,4次训练过程累计奖励变化如图6所示。

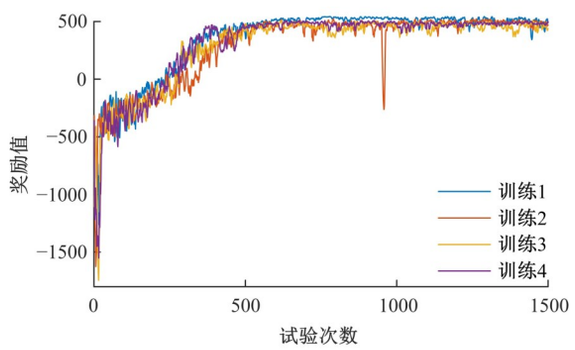


图6 DDPG训练过程累计奖励变化

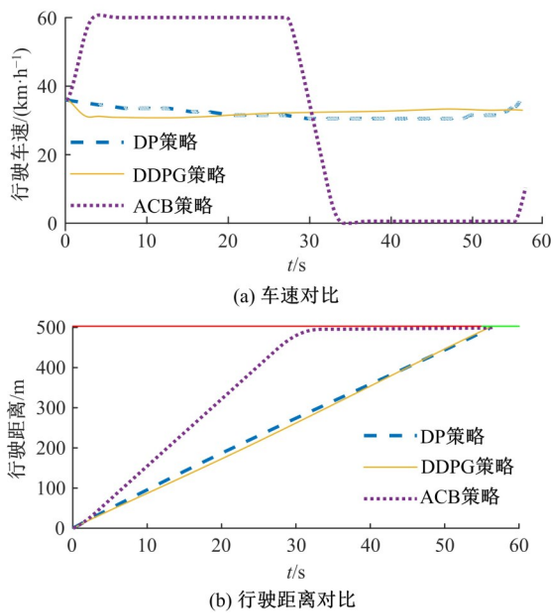
Fig. 6 Accumulated reward changes during DDPG training

从图6中可以看出,经过3次训练的奖励值均收敛至同一水准,这说明了策略已达到奖励值的最大化,策略算法稳定。其中,智能体在前50次训练中表现较差,由于处于策略探索阶段,智能体通过策略梯度尝试不同的驾驶策略,因此获得奖励较低。在50次至500次训练中,智能体根据获得的奖励信号不断调整策略,逐渐向最小梯度方向不断收敛策略,并最终在500次训练后逐渐趋于稳定。为了避免训练陷入局部最优,在趋于稳定的同时,智能体可能会多次采用策略探索,根据小范围初始速度状态加速度的输出有所不同,奖励值在稳定域附近振荡直至达到训练次数的最大值时训练终止。

3.2.2 策略性能对比

为了验证所提出的DDPG通行策略能耗提升的性能,选取了DP策略与ACB策略进行对比。DP策略选取车辆的初始车速为36 km/h,末端车速约束同为36 km/h,匀速策略的速度值来自DDPG策略仿真后的全程速度的均值,"加速-匀

速-制动（Accelerate-constant-brake，ACB）[9]”策略采用PI控制器，从初始速度加速至最大道路限速值，经匀速行驶至路口，当到达路口为红灯时则需减速停车等待绿灯通行。4种不同策略下的速度对比与行驶距离仿真结果对比图如图7所示。



(a) 车速对比

(b) 行驶距离对比

图7　不同策略通行仿真结果

Fig. 7　Simulation results of different strategies

由图7(a)中可以看出，本文所提出的策略在初始立即减速至33 km/h，并在该速度指引下几乎保持匀速通过路口，这表明智能体学会了以较短时间内快速跟踪平稳通过路口的推荐车速；而DP策略则选择分段减速策略，并在48 s后加速至指定末端车速约束附近。由图7(b)可以看出，DP与DDPG策略均能在绿灯相位开始处通过路口，而ACB策略则需要有15 s左右的停车等待时间。从整体上看，两种策略反映在速度谱上的差距较小。

由表3的对比数据可以看出，在相同的道路条件下，DP策略所消耗的总能量最低，其次是DRL策略，表现最差的为ACB策略。结合图7中的仿真结果，可以看出本文策略接近DP策略，以ACB策略为基准，本文策略在总的能耗表现上提升37.3%，相较于DP算法降低了2.7%。

从电动车的节能机理方面分析，当车速越高时，受空气阻力与加速阻力的能耗越高，因此为了节省能耗，应当以较小的车速通过路口。而对于DP算法事先得知红绿灯相位信息，离线规划全局最优车速。此外由上文可知，奖励函数包含了与建议车速差值信息和自车的能耗信息，在这两个信息引导下，DDPG策略的节能效果接近全局最优策略。

表3 不同策略能耗对比

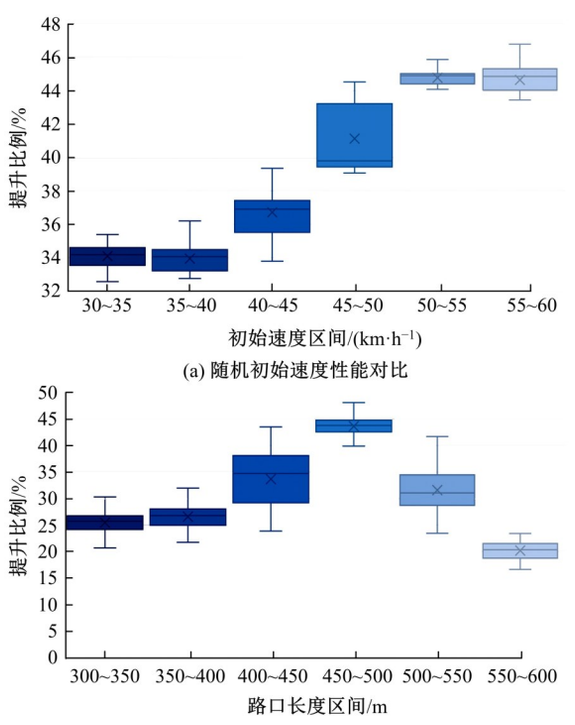
Table 3　Comparison of energy consumption of different strategies

策略	ACB	DP	DDPG
通过时间/s	56.4	56.0	56.0
初始速度/(km·h^{-1})	36.0	36.0	36.0
末端速度/(km·h^{-1})	10.4	31.4	36.4
电池能耗/kJ	291.4	213.1	230.3
动能变化/kJ	−82.5	−21.5	1.8
总能量/kJ	373.9	234.6	228.5
提升比例	DRL vs. ACB：37.3%		
	DRL vs. DP：−2.7%		

3.2.3　单交叉口随机仿真实验

为了验证本文提出的策略在多种随机工况下的性能，采用蒙特卡洛试验法选取初始车速为30～60 km/h、随机路口长度为300～600 m的两种不同场景进行仿真验证，各随机场景进行100次实验，同一变量下总计仿真600次，结果与ACB策略进行对比，根据结果统计了提升比例的分布情况，如图8所示。

由图8(a)中可以看出，在不同的初始速度



(a) 随机初始速度性能对比

(b) 随机路口长度性能对比

图8　DDPG策略与ACB策略随机验证性能对比

Fig. 8　Comparison of stochastic performance validation between DDPG and ACB strategy

下,相对于ACB策略,本文策略能量效率平均提升幅度在39.2%左右,且随着初始速度的增加,提升的比例增高。在接近道路最大限速时,在本文设定的优化场景下提升的性能有限,其性能表现应与全局优化结果近似,与3.2.2节部分结论一致。结合4.2.1的仿真结果对比分析,本文策略对车速优化较为平稳。

由图8(b)可以看到,对于不同的道路长度下所提出的策略对比ACB策略有较大幅度的提升,在450~500 m路口长度之间达到最高水平,该区间内也是训练时道路的初始值,平均能量效率提升30.2%。在400~450 m区间与500~550 m内性能提升比例波动幅度较大,这是因为该处的值远离了训练时的初始路口长度设定值,性能受到小幅度影响,但是从整体的节能效果来看,对比ACB策略仍有较大优势。相较于DP算法,当车辆进入不同的道路环境时需要重新计算获取最佳通行速度谱,DDPG算法在单路口下的动态交通场景下有较强的适应性。

由以上分析可以得出,本文所提出的算法不仅在对不同初始车速下的随机场景下有着较好的能量效率提升比例,对于不同路口长度下同样可以获得较大的提升。

3.3 多交叉口仿真实验

为了验证本文提出的策略在多路口下的可行性,本文选取5个含有随机信号相位配时的灯控路口来模拟真实的道路,路口长度分别选取为360、480、545、475、640 m,总长度为2500 m,用来代表较短、短、一般、较长和长5种路口环境。当车辆通过路口时,系统自动将获取下一路口的信息。策略能耗对比见表4,仿真的加速度、速度与行驶距离如图9所示。

表4 策略能耗对比

Table 4 Comparison of energy consumption

策略	通行时间/s	能耗/kJ	提升比例/%
ACB	290.7	1751	—
DRL	289.9	1201	31.4

由图9可以看出,在上述设定的道路条件下所提出的方法同样可以根据当前环境状态信息进行实时控制,且在当通过一个路口后,车辆可快速调节自身车速以适应新的交通状态。对比ACB策略,所提出的方法在所有的信号灯路口处没有停止等待。

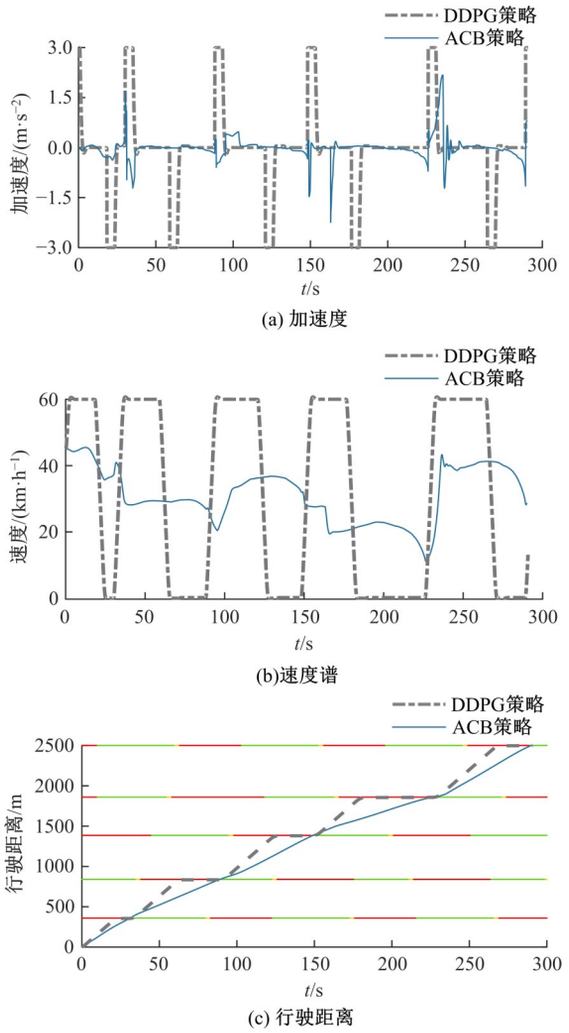


图9 多路口工况下验证

Fig. 9 Validation of multi-intersections

4 数字孪生试验平台与实车试验

为了验证上述策略在实际中的应用情况,搭建了智能网联汽车强化学习数字孪生试验平台,架构如图10所示。试验平台主要分为两层:上层包含PC上位机,接收来自下层的车辆实时数据和数字孪生系统中的信号灯状态,通过MATLAB/Simulink实时计算车辆的加速度,输出车速控制指令信号,并将车辆状态位置等信息发送至数字孪生系统中;下层包括下位机采集实时信号,通过ROS操作系统与上位机实时交互,并接收来自上位机发送的车速控制指令,实时控制车辆速度,从而实现路口通行的试验验证。

本节选取一段长为400 m的试验道路,道路如图11所示。在数字孪生系统中,信号灯的相位配时与仿真部分的时间设置一致,即红灯55 s,绿灯56 s,黄灯3 s,为清晰显示强化学习行驶模式

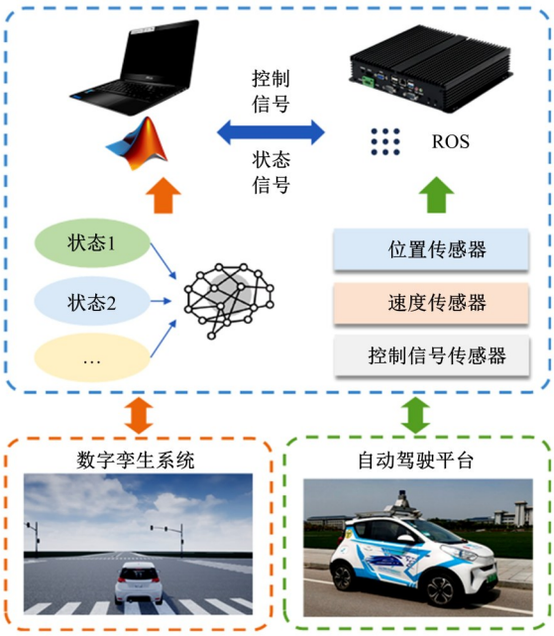


图 10　智能网联汽车数字孪生试验平台架构

Fig. 10　Architecture of digital twin platform for CAV

下虚拟与现实距离控制误差和车速变化情况，根据试验相关结果分别绘制了距离误差与车速误差的变化曲线，如图 12 所示。

图 11　试验选取道路

Fig. 11　Experimental road selection

由图 12(a)中可以看到，试验车辆在初始路段紧跟随强化学习算法计算的控制车速，并在 18 s 左右后达到通行建议的车速，而此时的实际车速将在接下来 20 s 内小幅度增加。随后在 35 s 左右进行小幅度减速，并跟随建议车速。

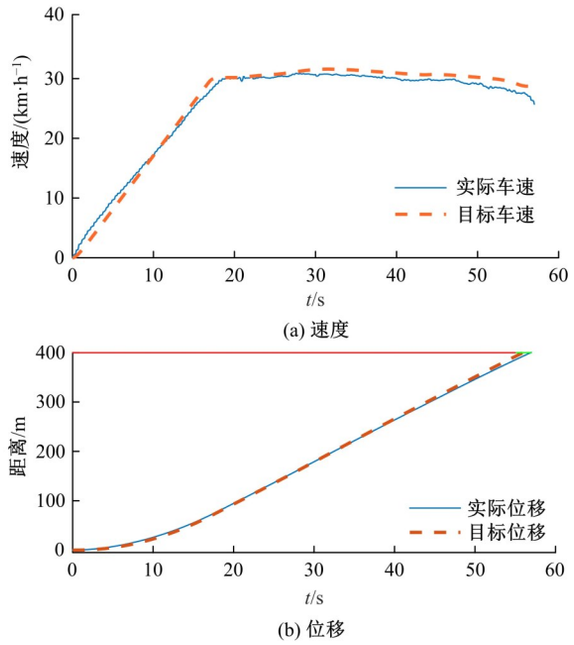


图 12　数字孪生实车试验结果

Fig. 12　Experimental results of digital-twin system

从图 12(b)中可以看到，试验在 57 s 时，数字孪生系统中的虚拟车辆到达交叉路口，仿真停止，此时实际车辆已在信号灯相位转为绿灯后通过路口。考虑到试验平台的通信延迟等影响，在通过路口前，实际记录的车速与强化学习实时规划的车速最大误差绝对值的平均值为 0.581 km/h，且距离误差绝对值的平均值为 2.702 m，在试验误差允许范围内，因此试验有效。所以，从整体的控制效果来看，试验完成了控制目标，且在绿灯相位间通过了路口，避免了停车等待的时间，从而提升了交通的效率。

综上所述，本文所提出的路口通行策略可以充分利用环境信息，合理控制车辆的加速度，使车辆速度保持平稳通过路口，且无需停车等待，有效降低制动次数，从而提高了电机工作的效率，在实现高效路口通行的同时降低了电动汽车的能量消耗。此外，本节建立的数字孪生平台为安全的强化学习提供了保障性试验，同时也为今后进一步的安全研究打下基础。

5　结束语

针对智能网联电动汽车在信号灯控路口处的通行问题，利用网联信息环境下交通状态信息获取的便利性，结合车辆自身的信息，基于深度强化学习算法构建了车辆通行策略，基于深度确定性策略梯度算法训练智能体掌握高效节能的通行策

略。仿真结果表明,所提出的通行策略在单交叉路口中保证快速通行路口的前提下,能保持较好的经济性,相较于ACB策略,电池的能量效率提升了37.3%以上,且节能效果接近DP。在保持了良好的控制效果下,本文所提出的算法在不同初始场景与环境下同样适用,且针对动态信号灯多路口场景下均能保证策略的经济性与通行高效性。此外,在本文的试验中搭建了智能网联汽车数字孪生试验平台,对所提出的强化学习算法实时规划的车速进行了跟踪测试,试验结果表明跟踪效果良好,可以达到通行时间最优的控制效果,提升了能耗效率。

参考文献:

[1] 《中国公路学报》编辑部. 中国汽车工程学术研究综述·2017[J]. 中国公路学报, 2017, 30(6): 1-197.
Editorial Department of China Journal of Highway. Review on China's automotive engineering research progress: 2017[J]. China Journal of Highway and Transport, 2017, 30(6): 1-197.

[2] 徐兴,陈特,陈龙,等. 分布式驱动电动汽车转矩节能优化分配[J]. 中国公路学报, 2018, 31(5): 183-190.
Xu Xing, Chen Te, Chen Long, et al. Optimal distribution of torque energy saving for distributed drive electric vehicles[J]. China Journal of Highway and Transport, 2018, 31(5): 183-190.

[3] Sivak Michael, Schoettle Brandon. Eco-driving: strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy[J]. Transport Policy, 2012, 22: 96-99.

[4] 付锐,张雅丽,袁伟. 生态驾驶研究现状及展望[J]. 中国公路学报, 2019, 32(3): 1-12.
Fu Rui, Zhang Ya-li, Yuan Wei. Research status and prospects of eco-driving[J]. China Journal of Highway and Transport, 2019, 32(3): 1-12.

[5] Ye Fei, Hao Peng, Qi Xue-wei, et al. Prediction-based eco-approach and departure at signalized intersections with speed forecasting on preceding vehicles [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(4): 1378-1389.

[6] 金辉,张俊. 智能车经济性起步车速规划研究[J]. 汽车工程, 2020, 42(2): 270-277.
Jin Hui, Zhang Jun. Research on economic speed planning of intelligent vehicle for starting stage[J]. Automotive Engineering, 2020, 42(2): 270-277.

[7] 洪金龙,高炳钊,董世营,等. 智能网联汽车节能优化关键问题与研究进展[J]. 中国公路学报, 2021,34(11):306-334.
Hong Jin-long, Gao Bing-zhao, Dong Shi-ying, et al. Key problems and research progress of energy saving optimization for intelligent connected vehicles[J]. China Journal of Highway and Transport, 2021, 34(11):306-334.

[8] Stahlmann Rainer, Möller Malte, Brauer Alexej, et al. exploring GLOSA systems in the field: technical evaluation and results[J]. Computer Communications, 2018, 120: 112-124.

[9] 陈浩, 庄伟超, 殷国栋, 等. 网联电动汽车信号灯控路口经济性驾驶策略[J]. 东南大学学报:自然科学版, 2021, 51(1): 178-186.
Chen Hao, Zhuang Wei-chao, Yin Guo-dong, et al. Eco-driving control strategy of connected electric vehicle at signalized intersection[J]. Journal of Southeast University (Natural Science Edition), 2021, 51(1): 178-186.

[10] Hao Peng, Wu Guo-yuan, Boriboonsomsin Kanok, et al. Eco-approach and departure (EAD) application for actuated signals in real-world traffic[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(1): 30-40.

[11] 解少博,辛宗科,李会灵,等. 插电式混合动力公交车电池配置和能量管理策略协同优化的研究[J]. 汽车工程, 2018, 40(6): 625-631,645.
Xie Shao-bo, Xin Zong-ke, Li Hui-ling, et al. A study on coordinated optimization on battery capacity and energy management strategy for a plug-in hybrid electric bus[J]. Automotive Engineering, 2018, 40(6): 625-631,645.

[12] Li Chun-ming, Zhang Tao, Sun Xiao-xia, et al. Connected ecological cruise control strategy considering multi-intersection traffic flow[J]. IEEE Access, 2020, 8: 219378-219390.

[13] Han Ji-hun, Vahidi Ardalan, Sciarretta Antonio. Fundamentals of energy efficient driving for combustion engine and electric vehicles: an optimal control perspective[J]. Automatica, 2019, 103: 558-572.

[14] 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.
Zhu Bing, Jiang Yuan-de, Zhao Jian, et al. A car-following control algorithm based on deep reinforcement learning[J]. China Journal of Highway and Transport,2019,32(6):53-60.

[15] Wu Yuan-kai, Tan Hua-chun, Peng Jian-kun, et al. Deep reinforcement learning of energy management

with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus [J]. Applied Energy, 2019, 247: 454-466.

[16] Li Yue-cheng, He Hong-wen, Khajepour Amir , et al. Energy management for a power-split hybrid electric bus via deep reinforcement learning with terrain information[J]. Applied Energy, 2019, 255: 113762.

[17] Zhou Mo-fan, Yu Yang, Qu Xiao-bo. 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, 2019, 21(1): 433-443.

[18] Zhu Zhao-xuan, Gupta Shobhit, Gupta Abhishek, et al. A deep reinforcement learning framework for eco-driving in connected and automated hybrid electric vehicles[J/OL]. [2021-01-13]. https://arxiv. org/pdf/2101. 05372. pdf.

[19] Bertsekas Dimitri. Reinforcement Learning and Optimal Control[M]. Belmont: Athena Scientific, 2019.

[20] Xu Xin, Zuo Lei, Li Xin, et al. A reinforcement learning approach to autonomous decision making of intelligent vehicles on highways[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2018, 50(10): 3884-3897.

[21] Pi Jian-zong. A reinforcement learning framework for autonomous eco-driving[D]. The Ohio State University, 2020.

[22] Schulman John, Wolski Filip, Dhariwal Prafulla, et al. Proximal policy optimization algorithms[J/OL]. [2017-04-23]. https://arxiv. org/pdf/1707. 06347. pdf.

[23] Silver David, Lever Guy, Heess Nicolas, et al. Deterministic policy gradient algorithms[C]// International Conference on Machine Learning, Detroit, USA, 2014: 387-395.