第 10 卷 第 3 期	集	成	技	术	Vol. 10	No. 3
2021年5月	JOURNAL OF I	NTEGR	ATION	I TECHNOLOGY	May	2021

引文格式:

李卫,郑春花,许德州. 基于深度强化学习的燃料电池混合动力汽车能量管理策略研究 [J]. 集成技术, 2021, 10(3): 47-60.

Li W, Zheng CH, Xu DZ. Research on energy management strategy of fuel cell hybrid vehicles based on deep reinforcement learning [J]. Journal of Integration Technology, 2021, 10(3): 47-60.

基于深度强化学习的燃料电池混合动力汽车能量

管理策略研究

李 卫^{1,2} 郑春花^{1*} 许德州^{1,3}

¹(中国科学院深圳先进技术研究院 深圳 518055) ²(中国科学院大学 北京 100049) ³(中国矿业大学 徐州 221116)

摘 要 为提高燃料电池混合动力汽车的燃油经济性和燃料电池寿命,该文提出一种基于深度强化学 习(Deep Reinforcement Learning, DRL)的能量管理策略。该策略首先在 DRL 奖励信号中加入寿命因 子,通过降低燃料电池功率波动,起到延长燃料电池寿命的效果;其次,通过限制 DRL 的动作空间的 方法,使燃料电池系统工作在高效率区间,从而提高整车效率。在 UDDS、WLTC、Japan1015 三个标 准工况下进行了离线训练,并在 NEDC 工况下实时应用以验证所提出策略的工况适应性。仿真结果显 示,在离线训练中,所提出的策略可以快速收敛,表明其具有较好的稳定性。在燃油经济性方面,与 基于动态规划的策略相比,在 3 个训练工况下的差异仅为 5.58%、3.03% 和 4.65%,接近最优燃油经济 性;相比基于强化学习的策略,分别提升了 4.46%、7.26% 和 5.35%。与无寿命因子的 DRL 策略相比, 所提出的策略在 3 个训练工况下将燃料电池平均功率波动降低了 10.27%、47.95% 和 10.85%,这有利于 提升燃料电池寿命。在未知工况的实时应用中,所提出策略的燃油经济性比基于强化学习的策略提升 了 3.39%,这表明其工况适应性。

关键词 燃料电池混合动力汽车;能量管理策略;深度强化学习;寿命增强;动作空间限制;强化学习 中图分类号 U 469.72 文献标志码 A doi: 10.12146/j.issn.2095-3135.20210118001

收稿日期: 2021-01-18 修回日期: 2021-02-22

基金项目:深圳市海外高层次人才创新创业计划项目(KQJSCX20180330170047681);深圳无人驾驶感知决策与执行技术工程实验室计划项目 (Y7D004);深圳电动汽车动力平台与安全技术重点实验室计划项目

作者简介: 李卫,硕士研究生,研究方向为燃料电池汽车的智能能量管理策略;郑春花(通讯作者),副研究员,博士研究生导师,研究方向为新能源 汽车能量管理及燃料电池,E-mail: ch.zheng@siat.ac.cn;许德州,硕士研究生,研究方向为混合动力汽车能量管理策略。

Research on Energy Management Strategy of Fuel Cell Hybrid Vehicles Based on Deep Reinforcement Learning

LI Wei^{1,2} ZHENG Chunhua^{1*} XU Dezhou^{1,3}

¹(Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China) ²(University of Chinese Academy of Sciences, Beijing 100049, China) ³(China University of Mining and Technology, Xuzhou 221116, China) ^{*}Corresponding Author: ch.zheng@siat.ac.cn

Abstract In order to improve the fuel economy and fuel cell lifetime of fuel cell hybrid vehicles, this research proposes an energy management strategy based on deep reinforcement learning (DRL). The strategy first adds a lifetime factor to reward signal of DRL, the lifetime of fuel cell is extended by limiting the power fluctuation. Then, the fuel cell system works in a high efficiency range by limiting the action space of DRL, improving the efficiency of the entire vehicle. After offline training under UDDS, WLTC, and Japan1015, it is applied in real time under NEDC to verify the adaptability of the proposed strategy. The results show that the proposed strategy can converge quickly in offline training, which proves its stability. Compared with dynamic programming-based strategy, the fuel economy difference in training cycles is only 5.58%, 3.03% and 4.65%, which is close to the optimal, and the promotion is 4.46%, 7.26% and 5.35% compared with reinforcement learning-based strategy. Compared with the DRL-based strategy without a lifetime factor, the proposed strategy reduces the average power fluctuation by 10.27%, 47.95%, and 10.85% under training cycles, which is beneficial to improve the fuel cell lifetime. In the real-time application, the fuel economy of the proposed strategy is improved by 3.39% compared with the reinforcement learning-based strategy.

Keywords fuel cell hybrid vehicle; energy management strategy; deep reinforcement learning; lifetime enhancement; action space limitation; reinforcement learning

Funding This work is supported by Shenzhen Science and Technology Innovation Commission (KQJSCX20180330170047681), Shenzhen Engineering Laboratory for Autonomous Driving Technology (Y7D004), and Shenzhen Key Laboratory of Electric Vehicle Powertrain Platform and Safety Technology

1 引 言

在环境污染和化石能源危机的大背景下, 世界各国相继针对燃油车颁布了严苛的排放法 规,与此同时也促进了新能源汽车产业的发展。 其中,燃料电池混合动力汽车(Fuel Cell Hybrid Vehicles, FCHVs)以其高效率、零排放的优点, 成为许多厂商和科研机构的重点研发对象^[1-2]。 由于燃料电池系统(Fuel Cell System, FCS)动态 响应较慢,一般使用电池或超级电容作为辅助动力,以适应剧烈变化的工况。为了合理分配不同动力源的输出功率,提升整车的燃油经济性,设计合理的能量管理策略尤为重要^[3]。

常见的能量管理策略主要分为基于规则和 基于优化两类^[4]。前者针对动力源工作特性和驾 驶工况设计控制规则,简单实用,但是不具备 最优性,控制效果依赖于专家经验^[5]。后者以动 态规划^[6]为代表,在已知工况下遍历所有可行 解,获得指定性能指标下的最优控制,由于要 求工况全局已知,难以实时应用,一般作为燃 油经济性的比较基准。其他基于优化的能量管 理策略,如基于庞特里亚金极小值原理^[7-8]、等 效燃油消耗最小^[9]等,工况适应性较差,限制了 算法的应用。近年来,随着人工智能技术的发 展,基于学习的策略成为新的研究热点。该方法 主要分为3类:基于神经网络、基于强化学习 (Reinforcement Learning, RL)和基于深度强化学 习(Deep Reinforcement Learning, DRL)。基于神 经网络的策略利用神经网络进行工况分类[10]、速 度预测^[11-12]、参数优化^[13]等。而与把神经网络作 为数据处理工具不同,强化学习利用试错机制, 直接从历史状态中学习得到优化控制策略^[14-18]。 Liu 等^[19]针对混合动力履带式车辆,设计了基于强 化学习的能量管理策略,结果显示,相比随机动 态规划策略,该策略具有较好的燃油经济性。Lin 等^[20]在基于强化学习的能量管理策略中,提出根 据余弦相似度来判断两个概率转移矩阵的差异, 当超过阈值时,更新当前矩阵,实时选择最适合 当前工况的策略,提高了整车经济性。基于强化 学习的策略结合了基于规则和基于优化的特点, 具有较好的工况适应性,但算法需要离散状态变 量和动作变量, 随着离散程度的增加, 算法复杂 度呈指数级别上升,影响其收敛性和稳定性^[21]。

为此,一些研究结合神经网络和强化学 习,将 DRL 引入到混合动力汽车的能量管理领 域,利用神经网络拟合值函数,避免了状态离 散化^[21-24]。Han 等^[25]在基于 DRL 的能量管理策 略中应用双层网络^[26]的方法,避免算法在离线 训练中陷入局部最优,在控制性能方面相比其他 基于学习的策略具有一定优势。Wu 等^[21]针对混 合动力公交,提出一种基于 DRL 的能量管理策 略,在固定路线和工况下进行的训练和应用结果 显示,所提出的策略比传统强化学习具有更好的 收敛性和燃油经济性。但对于非公交类车型,行 驶工况多变,需要在典型工况下训练,并在未知 工况下验证控制效果。上述基于 DRL 的能量策 略,控制对象均为油电式混合动力汽车,而对于 FCHV 来说,策略除了提高燃油经济性和维持电 池 SOC 以外,还应该考虑提升燃料电池寿命, 降低整车成本。

本研究针对 FCHV 提出一种基于 DRL 的能 量管理策略,通过在奖励信号中加入寿命因子, 限制 FCS 输出功率的波动,起到延长系统寿命 的作用;同时,通过限制动作区间的方法,使 FCS 工作在高效率区间,提高整体效率。在算法 角度,利用目标值网络等技巧加快算法收敛,提 高泛化性能。所提出的策略在 UDDS、WLTC 和 Japan1015 等标准工况下进行了离线训练,并在 NEDC 工况上进行了实时应用。仿真结果表明, 所提出的策略可以得到接近动态规划的最优燃油 经济性,大大减小了功率波动,具有较好的燃料 电池寿命增强效果。在验证工况下,所提出的策 略在燃油经济性、平均效率和平均功率波动方面 的表现都超过了基于强化学习的策略,表现出较 好的工况适应性。

2 控制模型的建立

本文使用的燃料电池混合动力汽车动力传动 系统结构如图1所示,主要由FCS、电池、DC/DC 转换器和电机等组成。整车和动力系统的主要参 数如表1所示。

2.1 燃料电池混合动力汽车的需求功率

当工况已知,可以根据车辆的动力学平衡方程计算出总需求功率 P_{req}:

 $P_{req} = (fMg \cos \alpha + 0.5 \rho_a A C_D v^2 + Mg \sin \alpha + \delta Ma)v$ (1) 其中, *f* 为滚动阻力系数; *M* 为整车质量; *g* 为重 力加速度; *α* 为坡度; ρ_a 为空气密度; *A* 为迎风面 积; *C*_D 为空气阻力系数; *v* 为车速; *δ* 为质量系 数; *a* 为车辆加速度。车辆的功率平衡方程为:





图 1 燃料电池混合动力汽车动力系统结构

Fig. 1 Powertrain configuration of fuel cell hybrid vehicle

表1 整车主要参数

Table 1 Fuel cell hydrid vehicle da	Table 1	cle data
-------------------------------------	---------	----------

类别	参数	数值
整车	整车质量(kg)	1 763
	滚动半径(m)	0.301
	滚动阻力系数	0.013
	空气阻力系数	0.24
	迎风面积(m²)	2.5
FCS	单元数量	400
	最大功率(kW)	89
	最大效率	0.59
电池	类型	Ni-MH
	最大输出功率(kW)	39
	总能量(kW·h)	7.8
电机	最大功率(kW)	107
	最大效率	0.9

 $P_{req} = (P_{fcs}\eta_{fcs}\eta_{conv} + P_{batt}\eta_{batt})\eta_{mot}$ (2) 其中, P_{fcs} 和 P_{batt} 分别为 FCS 和电池的输出功 率; η_{fcs} 、 η_{conv} 、 η_{batt} 和 η_{mot} 分别为 FCS、DC/DC 转换器、电池和电机的效率。根据需求功率,不 同的能量管理策略调整两个动力源的动力分配, 达到不同的控制效果。

2.2 燃料电池混合动力汽车的部件模型

2.2.1 燃料电池系统模型

作为 FCHV 的主要动力源, FCS 将氢气和氧 气反应的化学能转换为电能,其效率和燃料消耗 率均与功率相关,关系曲线如图 2 所示。由于能 量管理策略不涉及 FCS 的其他内部参数,因此不 对模型细节进行讨论。



Fig. 2 Efficiency and hydrogen consumption rate of the FCS

对于 FCHV 来说,燃料电池模块成本较高, 延长其使用寿命对提高整车的经济性具有重要 意义。已有的研究表明,加载工况和启停工况对 FCS 寿命影响较大^[27-28]。这两种工况下,输出功率 剧烈变化,电堆内部的温度、湿度等难以维持在 最佳区间,影响系统寿命。由于 FCS 寿命的准确 预测需要大量实测数据,限制了其广泛应用^[29-30], 因此本文利用平均功率波动定性分析策略的寿命 增强作用。

$$\frac{\Delta P_{\text{fcs}}}{\Delta P_{\text{fcs}}} = \frac{\sum_{t=1}^{N} \left| \left(P_{\text{fcs}}(t) - P_{\text{fcs}}(t-1) \right) \right|}{N}$$
(3)

其中, $\overline{\Delta P_{\text{fcs}}}$ 为平均功率波动,值越小表示系统输出功率越平滑,有利于寿命提升;N为当前工况长度。

2.2.2 电池模型

FCHV 中,电池作为辅助动力源,在加速阶 段快速提供动力,在减速中回收制动能量。本 文采用 Ni-MH 电池,并使用 PNGV (Partnership for a New Generation of Vehicles)^[31]对电池进行建 模,具体如图 3 所示。相比于简单的内阻模型, PNGV 模型考虑了电池反应中的极化现象,更好 地表征了电池的动态响应,提高了建模精度。

图 3 中, *V*_{oc} 为电池开路电压, *I*_{batt} 为电池电 流, *R*_{int} 为电池内阻, *I*_{pol} 为极化电流, *R*_{pol} 为极化 电阻, *C*_{pol} 为极化电容, *C*_b 是一个表征开路电压

2021年



图 3 电池的 PNGV 模型

Fig. 3 Equivalent circuit of the PNGV battery model

随着电流积分变化的电容。输出电压 U_{batt} 和功率 P_{batt} 可以表示为:

$$\begin{cases} U_{\text{batt}} = V_{\text{oc}} - I_{\text{batt}} \cdot R_{\text{int}} - I_{\text{pol}} \cdot R_{\text{pol}} - \frac{1}{C_{\text{b}}} \int I_{\text{batt}} \, dt \\ P_{\text{batt}} = U_{\text{batt}} \cdot I_{\text{batt}} \\ \frac{dI_{\text{pol}}}{dt} = \frac{\left(I_{\text{batt}} - I_{\text{pol}}\right)}{\tau} \end{cases}$$
(4)

其中, τ 为时间常数, $\tau = C_{pol} \cdot R_{pol}$ 。电池的 SOC 是表征电池状态的重要参数,其瞬时变化值 SOC 表示为电流和电池总容量 Q_{cap} 的比值:

$$\dot{SOC} = -I_{\text{batt}} / Q_{\text{cap}}$$
 (5)

频繁的过充或过放会降低电池的寿命,因此能量管理策略需要考虑将 SOC 维持在合理范围。如果忽略温度的影响,那么电池的内阻、极化电阻、开路电压与 SOC 的关系曲线如图 4 所示。图 4 显示,将 SOC 保持在[0.4, 0.75],有利于减小电池内阻,提高电池寿命。

2.2.3 电机模型

电机是 FCHV 的直接动力部件,可将 FCS 和电池的电能转换成机械能,驱动车辆前进,而 在减速时作为发电机回收制动能量,为电池充 电。电机的效率和最大充放电转矩如图 5 所示。

3 基于强化学习的能量管理策略

3.1 算法概述

强化学习主要包含智能体、环境、状态、动



Fig. 4 Influences of the battery SOC on different

battery parameters



图 5 电机效率图和最大转矩曲线

Fig. 5 Motor efficiency map and maximum torque curves 作和奖励函数等要素,其核心思想是智能体发出 动作,通过与环境交互不断地试错,而环境根据 反馈的奖励信号更新状态,改善策略直到收敛。

由于车辆工况的随机性导致需求功率随机变化, 故为了简化模型,一般将工况作为一阶马尔科夫 过程,计算出对应工况下的概率转移矩阵,而策 略在学习过程中根据该矩阵更新状态。基于强化 学习的能量管理策略框架如图6所示。







energy management strategy

基于强化学习的能量管理策略作为智能体, 接收当前状态 S, 和奖励 R, 产生动作 a, 作用于 环境,并根据概率转移矩阵更新状态和奖励函 数,不断循环直到策略收敛。常用的强化学习算 法是 Q 学习算法,算法目标是最大化智能体从环 境中获得的累计奖励的期望,利用贝尔曼方程可 以计算:

$$Q(S_{t}, a_{t}) = E\left[R_{t+1} + \gamma \max_{a_{t+1}} Q(S_{t+1}, a_{t+1}) | S_{t}, a_{t}\right] (6)$$

其中, $Q(S_{t},a_{t})$ 表示当前状态-动作对的值函数; E 为求期望; γ 为对未来值函数的折扣因子。Q 学习的更新规则如下:

 $Q(S_t, a_t) \leftarrow Q(S_t, a_t) + \eta \left[R_{t+1} + \gamma \max_{a_{1}, j} Q(S_{t+1}, a_{t+1}) - Q(S_t, a_t) \right] (7)$ 其中, η 为学习率。随着算法的不断迭代, 值函 数会逐渐收敛至最优值,最优控制策略 π^* 即是最 大化 Q 值函数的动作序列。

$$\pi^*(S) = \arg \max Q^*(S, a) \tag{8}$$

实际应用中会维护一个表格来保存每个状 态-动作对的 Q 值,这个表格称为 Q 表。但对于 高维状态和连续状态来说,状态离散化后 Q 表会 迅速增大, 计算负担呈指数级别增加, 算法难以 收敛。而利用神经网络等函数近似方法拟合 Q 值 函数,设计基于 DRL 的能量管理策略,避免对 状态的离散化,可以大大提高策略对高维状态的 处理能力。本文使用深度 Q 学习算法,利用参数 为 θ 的深度Q网络来拟合值函数,避免状态的离 散化。

$$Q(S_t, a_t; \theta) \approx Q(S_t, a_t) \tag{9}$$

为了提高算法性能,采用目标值网络的方 法,分别设计两个结构一样的网络:评价网络和 目标网络。其中,评价网络用于选择动作,更 新参数,每隔一段时间,复制参数到目标网络 $Q(S_t, a_t; \theta')$, 实现延时更新。这种方法降低了当 前 Q 值和目标 Q 值的相关性,提高了算法的稳 定性。算法目标是最小化损失函数 L(θ):

$$L(\theta) = E\left\{ \left[R + \gamma \max_{a_{t+1}} \mathcal{Q}\left(S_{t+1}, a_{t+1}; \theta^{-}\right) - \mathcal{Q}\left(S_{t}, a_{t}; \theta\right) \right]^{2} \right\} (10)$$

利用梯度下降算法,不断更新网络参数 θ , 直到学习收敛。为了平衡学习过程中"探索" 和"利用"的关系,采用 ε -greedy 算法时,策 略中有较小的概率 ε 随机选择动作,较大的概 率 $1-\varepsilon$ 选择最大化 *Q* 值的动作。学习初始阶 段 ε 较大,具有增强网络探索能力,而随着训练 的进行, ε逐步衰减, 以加快学习速度。本文使 用经验池回放^[32]的技巧,搜集多个状态-动作对 $e_t = (S_t, a_t, R_t, S_{t+1})$, 得到经验池 $D_t = \{e_1, e_2, \dots, e_t\}$, 算法学习时,直接从经验池随机取样。这种方 法可以平滑数据分布, 解决数据相关性过强的 问题,提高泛化能力。

3.2 基于深度强化学习的能量管理策略 3.2.1 关键概念

在基于 DRL 算法的能量管理策略中,系统 状态、动作空间和奖励信号等关键参数分别设置 如下。

系统状态如公式(11)所示,将车辆当前速 度、加速度、电池 SOC 作为系统状态。加速度 可由相邻时刻工况车速计算,根据电路平衡关系可以计算出 SOC 值。

$$S = \{v, a, SOC\}$$
(11)

动作空间:输出动作设置为 FCS 输出功率。 为了提高学习效率,同时使 FCS 在高效率区间工 作,根据图 3 所示的效率数据,限制动作输出区 间在[2 kW,40 kW],并以 1 kW 作为间隔,离散 化表示为:

 $A = \{2, 3, 4, 5, \cdots, 36, 37, 38, 39, 40\}$ (12)

奖励信号:奖励函数用于评估向环境施加动 作后,环境给予智能体的反馈,协助调整网络 参数。一般将奖励设置为瞬时燃油消耗和电池 SOC 相关的函数。为了最大化奖励,算法会倾 向于减小燃油消耗,并尽量维持 SOC,防止电 池过充或过放,以提高电池寿命。但在 FCHV 中,FCS 成本较高,故能量管理策略应考虑提升 系统寿命。为此,在奖励信号中引入寿命因子, 奖励信号设置如下:

 $R = -\left[\alpha \frac{\mathrm{d}f_{\mathrm{tel}}}{\mathrm{d}t} + \beta (SOC_{\mathrm{ref}} - SOC)^2 + \lambda \left| P_{\mathrm{fcs}}(t) - P_{\mathrm{fcs}}(t+1) \right| \right] (13)$

其中, α 、 β 、 λ 分别是奖励信号中燃油经济性、 电池 SOC 维持、相邻时刻功率波动的权重因子; $\frac{df_{uel}}{dt}$ 为瞬时氢气消耗; SOC_{ref}为 SOC 初始值。

3.2.2 算法设计

本文基于 DRL 的能量管理策略算法如表 2 所示。

评价网络和目标网络具有相同的结构,均由 3 个隐含层组成,其中激活函数采用线性激活单 元,每一层由 200 个整流单元全连接组成。算法 主要参数设置如表 3 所示。

本文提出的基于 DRL 的能量管理策略框架 如图 7 所示。策略分为离线训练和在线应用两部 分。首先使用标准工况训练神经网络,将学习后 的参数拷贝到车载控制器中;在现实工况中,控 制器根据车速等状态直接输出动作,实现在线应

表 2 基于 DRL 的能量管理策略流程

Table 2 Steps of DRL-based energy management strategy

算法: 基于 DRL 的能量管理策略算法流程
初始化经验池 D ,容量为 N ,训练轮次 episodes 为 M ,单轮迭代
$\sqrt{3}$ Π_{i} 可始化评价网络和目标网络参数 θ 和 θ' ,两个网络结构和参数均
-致;
or episode = $1:M$ do
for $t = 1:T$ do
获得 t 时刻状态 S _t ;
以概率 ε 随机选择动作 a_r ;
否则选择动作 $a_i = \operatorname{argmax} Q(S_i, a_i; \theta);$
智能体对环境施加动作 a_i ,观察奖励 R_i ,状态更新至 S_{i+1} ;
將序列 (S_i, a_i, R_i, S_{i+1}) 存进经验池 D 中;
从 D 中选择最小样本集数量 n 的状态动作对;
for $i = 1:n$ do
if $t = T$, 当前循环结束, $y_i = R_i$
else $y_i = R_i + \gamma \max Q(S_{i+1}, a; \theta)$
利用梯度下降法更新网络,损失函数为 L(θ)
每隔 C 步,将评价网络的参数复制到目标网络: $\theta'=\theta$
end for
end for
nd for

表 3 DRL 算法参数设置

Table 3Parameters setting of DRL algorithm

参数	数值
训练轮次 episodes	1 200
记忆池容量 N	10 000
最小样本集样本数 n	64
学习率 η	0.001
折扣率 γ	0.9
权重因子 α	200
权重因子 β	1
权重因子 λ	0.000 035

用。新产生的驾驶数据会不断地进入训练工况, 参与网络训练,进一步提高工况适应性。

4 仿真结果

在 UDDS、WLTC 和 Japan1015 三个标准工 况下对基于 DRL 的策略进行离线训练,并实时 应用到 NEDC 工况,4 种工况曲线如图 8 所示。



图 7 基于 DRL 的能量管理策略框图





图 8 不同工况的车速曲线

Fig. 8 Velocity curves of different driving cycles

为了衡量策略的控制效果,分别实现了基于动 杰规划、基于 DRL(无寿命因子)、基于 RL 的策 略,并从收敛性、燃油经济性、寿命增强作用、 工况适应性等方面对策略性能进行评价。

4.1 收敛性

为了评估算法的学习效果,得到稳定的控制 效果, 需要对 DRL 的收敛性进行验证。通过对 每轮循环的总奖励进行累加,并根据工况长度计 算平均奖励,绘制3个离线训练工况下的平均奖 励曲线,如图9所示。可以看出,本文算法在不 同工况下训练到 400 轮左右时, 单轮平均奖励基 本稳定,表明算法具有较好的收敛性。



图 9 DRL 算法在训练工况下的平均奖励

Fig. 9 Average reward of DRL algorithm in training cycles 4.2 燃油经济性

离线训练中,将提出的策略与基于动态规划 和基于 RL 的策略进行对比, FCS 输出功率和电 池输出功率如图 10~12 所示, SOC 曲线如图 13 所示。在相同工况下,不同策略对两个动力源的 分配不同,进而产生了不同的控制效果。如基于 RL 的策略倾向于更大幅度地使用两个动力源, SOC 曲线波动最大; 基于动态规划的策略则更 加平滑地分配动力, 使 SOC 保持在一个较高的 水平: 而基于 DRL 的策略控制特性介于前两者 之间,限制 FCS 的功率在高效率区间,其平均 效率较高,有利于提升燃油经济性和燃料电池 寿命。由于在同一工况下不同策略的最终 SOC 存在差异, 需要对 SOC 进行修正, 将最终电力 消耗等效为氢耗。表 4 为 3 种策略的等效燃油 经济性结果。在 UDDS、WLTC 和 Japan1015 工况下,基于 DRL 的策略与基于动态规划策 略在燃油经济性上的差异分别为 5.58%、3.03% 和 4.65%,达到了接近最优的结果;在基于 RL 的策略基础上,分别提升了 4.46%、7.26% 和 5.35%。以上结果表明,基于 DRL 的策略在提 高燃油经济性和维持电池 SOC 方面具有较好的 效果。

4.3 寿命增强作用

将基于 DRL 和基于 DRL (无寿命因子)的策 略对应的平均功率波动大小进行比较,结果如 表 5 所示。在 3 个训练工况下,与无寿命因子的 策略相比,本文提出的策略分别将功率波动降低 了 10.27%、47.95% 和 10.85%, 其他性能如平均 效率、燃油经济性基本相当甚至略有提升。这表 明在典型工况下,考虑寿命增强的策略将平均功 率波动均降低了10%以上,有利于提升燃料电池 的寿命。

4.4 工况适应性分析

基于 DRL 的策略离线训练好后,在线应用 于未知的 NEDC 工况,验证其工况适应性。将仿 真结果与基于动态规划和基于 RL 的策略进行比 较。其中,基于动态规划将 NEDC 作为已知工 况计算最优燃油经济性;基于 RL 的策略在 3 个 训练工况下离线训练,在验证工况下根据 Q 表 实时控制。验证工况下, FCS 和电池输出功率如 图 14 所示, SOC 曲线如图 15 所示, 等效燃油经 济性如表 6 所示。在验证工况下, 与基于 RL 的 策略相比,本文提出的策略将燃油经济性提高了 3.39%,与最优燃油经济性差异在10%左右,具 有明显的效率优势和减小功率波动的作用,表明 该策略对工况的适应性。



图 10 UDDS 工况下的功率输出





图 11 WLTC 工况下的功率输出







Fig. 12 Power output under Japan1015

5 讨论与分析

略,控制对象主要是油电式混合动力汽车,以减 少氢气消耗和维持电池 SOC 为控制目标^[21,33-35]。 对于 FCHV 来说,燃料电池成本较高,故提高

56

目前基于 DRL 的混合动力汽车能量管理策



图 13 训练工况下不同策略的 SOC 曲线

Fig. 13 SOC curves of different strategies under training cycles

表 4 训练工况下不同策略的燃油经济性比较

T. I.I. 4	T	· · · · · · · · · · · · · · · · · · ·	
I anie 4	HILL PCONOMV	of different strategie	s under training eveles
	r uci cconomy	or uniterent strategic	s under training cycles
	•	8	

工况	策略	最终 SOC	平均效率	等效氢耗(kg/100 km)	燃油经济性差异(%)
UDDS	基于动态规划	0.677 9	0.492 9	0.607 3	
	基于 RL	0.671 6	0.464 0	0.671 1	10.51
	基于 DRL	0.677 5	0.580 3	0.641 2	5.58
WLTC	基于动态规划	0.584 8	0.516 6	0.806 4	
	基于 RL	0.496 0	0.574 7	0.895 8	11.09
	基于 DRL	0.535 1	0.576 5	0.830 8	3.03
Japan1015	基于动态规划	0.687 3	0.497 9	0.522 6	
	基于 RL	0.681 3	0.476 5	0.577 8	10.56
	基于 DRL	0.689 9	0.577 5	0.546 9	4.65

表 5 训练工况下的平均功率波动

Table 5 Average power fluctuation under training cycles

工况	策略	最终 SOC	平均效率	等效氢耗(kg/100 km)	平均功率波动(kW)	相对降低(%)
UDDS	DRL(无寿命因子)	0.679 1	0.577 6	0.625 6	1.205 3	
	DRL	0.677 5	0.580 3	0.622 1	1.081 5	10.27
WLTC	DRL(无寿命因子)	0.514 1	0.576 5	0.860 2	0.999 1	
	DRL	0.535 1	0.576 5	0.830 8	0.520 0	47.95
Japan1015	DRL(无寿命因子)	0.688 6	0.577 2	0.550 1	0.472 1	
	DRL	0.689 9	0.577 5	0.546 9	0.420 9	10.85

其寿命对降低整车成本具有重要意义^[28]。本文针 对 FCHV 设计了一种基于 DRL 的燃料电池寿命 增强型能量管理策略。仿真结果显示,该策略在 收敛性、燃油经济性、寿命增强作用和工况适应 性上都具有较好的性能。得益于目标值网络等技

3 期

巧,相比已有研究^[22],所提出的策略奖励信号 收敛更加稳定。燃油经济性方面,所提出的策略 达到了基于动态规划的 95% 左右,相比之下, Wu 等^[21]离线训练时经济性仅达到了基于动态规 划的 89%, Han 等^[25]的研究结果则为 93%。分析



图 14 验证工况下的功率输出

Fig. 14 Power output under validation cycle

表 6 验证工况下的燃油经济性比较

Table 6	Fuel econom	y of different	strategies	under	validation	cycle
						•

策略	最终 SOC	平均效率	平均功率波动(kW)	等效氢耗(kg/100 km)	燃油经济性差异(%)
基于动态规划	0.639 0	0.532 8	0.89	0.637 3	
基于 RL	0.635 1	0.526 9	2.01	0.731 5	14.78
基于 DRL	0.638 2	0.571 6	0.53	0.706 7	10.89





validation cycle

原因: 主要是本研究设计了合理的奖励信号权重 因子,并使用了经验回放等技巧。另外,本策略 使用动作区间约束来提高整车效率。Lian 等^[35]在 油电式混合动力汽车上也采用类似的方法,针对 发动机设置了最优工作曲线,最终平均燃油经济 性达到基于动态规划的 94% 左右。本文的创新 点主要体现在: 通过在 DRL 的奖励信号中加入 燃料电池寿命因子,有效降低了燃料电池功率波 动,提升了其使用寿命;限制 FCS 动作输出在 高效率区间,提升了整车效率;策略离线训练后 在线应用到未知工况的结果表明,所提出的策略 具有较好的适应性。未来的研究内容将主要集中 在两方面:(1)设计连续控制的 DRL 能量管理策 略,避免控制动作的离散化误差;(2)使用硬件 在环或实车试验对策略进行验证,以更加贴近真 实环境。

6 结 论

本文针对 FCHV,设计了基于 DRL 的寿命 增强型能量管理策略,在奖励信号中加入寿命因 子,降低 FCS 功率波动,并且限制燃料电池输 出功率在高效率区间,提高整车效率。仿真结果 显示,所提出的策略在不同工况均能收敛,表明 其具有较好的稳定性。在 3 个训练工况下,其 燃油经济性接近动态规划的最优燃油经济性, 比基于 RL 的策略分别提升了 4.46%、7.26% 和 5.35%,并将功率波动降低了 10% 以上。实时应 用中,所提出的策略在燃油经济性、寿命增强效 果、平均效率方面均优于基于 RL 的策略,表明 本文策略具有较好的工况适应性。

参考文献

- Bernay C, Marchand M, Cassir M. Prospects of different fuel cell technologies for vehicle applications [J]. Journal of Power Sources, 2002, 108(1-2): 139-152.
- [2] Chalk SG, Patil PG, Venkateswaran S. The new generation of vehicles: market opportunities for fuel cells [J]. Journal of Power Sources, 1996, 61(1-2): 7-13.
- [3] Ahmadi S, Bathaee SMT, Hosseinpour AH. Improving fuel economy and performance of a fuel-cell hybrid electric vehicle (fuel-cell, battery, and ultra-capacitor) using optimized energy management strategy [J]. Energy Conversion and Management, 2018, 160: 74-84.
- [4] Han L, Jiao XH, Zhang Z. Recurrent neural network-based adaptive energy management control strategy of plug-in hybrid electric vehicles considering battery aging [J]. Energies, 2020, 13(1): 202.
- [5] 汪亚飞,田国富,刘忠旭,等.电动汽车复合储能系统能量管理策略及其快速控制原型验证 [J].集成技术,2020,9(1):66-79.
 Wang YF, Tian GF, Liu ZX, et al. A hybrid energy storage system and its energy management strategy verification for electric vehicle applications [J]. Journal of Integration Technology, 2020, 9(1): 66-79.
- [6] Lei ZZ, Cheng D, Liu YG, et al. A dynamic control strategy for hybrid electric vehicles based on parameter optimization for multiple driving cycles and driving pattern recognition [J]. Energies, 2017, 10(1): 54.
- [7] Hou C, Ouyang MG, Xu LF, et al. Approximate Pontryagin's minimum principle applied to the energy management of plug-in hybrid electric

vehicles [J]. Applied Energy, 2014, 115: 174-189.

- [8] Zou Y, Liu T, Sun FC, et al. Comparative study of dynamic programming and Pontryagin's minimum principle on energy management for a parallel hybrid electric vehicle [J]. Energies, 2013, 6(4): 2305-2318.
- [9] Han JH, Park YJ, Kum D. Optimal adaptation of equivalent factor of equivalent consumption minimization strategy for fuel cell hybrid electric vehicles under active state inequality constraints [J]. Journal of Power Sources, 2014, 267: 491-502.
- [10] Zhou DM, Al-Durra A, Gao F, et al. Online energy management strategy of fuel cell hybrid electric vehicles based on data fusion approach [J]. Journal of Power Sources, 2017, 366: 278-291.
- [11] Xiang CL, Ding F, Wang WD, et al. Energy management of a dual-mode power-split hybrid Electric vehicle based on velocity prediction and nonlinear model predictive control [J]. Applied Energy, 2017, 189: 640-653.
- [12] Liu JC, Chen YZ, Zhan JY, et al. An on-line energy management strategy based on trip condition prediction for commuter plug-in hybrid electric vehicles [J]. IEEE Transactions on Vehicular Technology, 2018, 67(5): 3767-3781.
- [13] Xie SB, Hu XS, Qi SW, et al. An artificial neural network-enhanced energy management strategy for plug-in hybrid electric vehicles [J]. Energy, 2018, 163: 837-848.
- [14] Zou Y, Liu T, Liu DX, et al. Reinforcement learning-based real-time energy management for a hybrid tracked vehicle [J]. Applied Energy, 2016, 171: 372-382.
- [15] Chen Z, Li L, Hu XS, et al. Temporal-difference learning-based stochastic energy management for plug-in hybrid electric buses [J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(6): 2378-2388.
- [16] Liu T, Wang B, Yang CL. Online Markov Chainbased energy management for a hybrid tracked vehicle with speedy Q-learning [J]. Energy, 2018, 160: 544-555.
- [17] 郑春花,李卫.强化学习在混合动力汽车能量管

理方面的应用 [J]. 哈尔滨理工大学学报, 2020, 25(4): 1-11.

Zheng CH, Li W. Research on the application of reinforcement learning in hybrid electric vehicle energy management [J]. Journal of Harbin University of Science and Technology, 2020, 25(4): 1-11.

- [18] Xiong R, Cao JY, Yu QQ. Reinforcement learningbased real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle [J]. Applied Energy, 2018, 211: 538-548.
- [19] Liu T, Zou Y, Liu DX, et al. Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle [J]. IEEE Transactions on Industrial Electronics, 2015, 62(12): 7837-7846.
- [20] Lin XY, Zhou BH, Xia YT. Online recursive power management strategy based on the reinforcement learning algorithm with cosine similarity and a forgetting factor [J]. IEEE Transactions on Industrial Electronics, 2021, 68(6): 5013-5023.
- [21] Wu JD, He HW, Peng JK, et al. Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus [J]. Applied Energy, 2018, 222: 799-811.
- [22] Hu Y, Li WM, Xu K, et al. Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning [J]. Applied Sciences, 2018, 8(2): 187.
- [23] Kong HF, Yan JP, Wang H, et al. Energy management strategy for electric vehicles based on deep Q-learning using Bayesian optimization [J]. Neural Computing and Applications, 2020, 32(18): 14431-14445.
- [24] Li YC, He HW, Peng JK, et al. Deep reinforcement learning-based energy management for a series hybrid electric vehicle enabled by history cumulative trip information [J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 7416-7430.
- [25] Han XF, He HW, Wu JD, et al. Energy management based on reinforcement learning with double deep Q-learning for a hybrid electric tracked vehicle [J].

Applied Energy, 2019, 254: 113708.

- [26] Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double Q-learning [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2016.
- [27] Pei PC, Chen HC. Main factors affecting the lifetime of proton exchange membrane fuel cells in vehicle applications: a review [J]. Applied Energy, 2014, 125: 60-75.
- [28] Zheng CH, Xu GQ, Park YI, et al. Prolonging fuel cell stack lifetime based on Pontryagin's minimum principle in fuel cell hybrid vehicles and its economic influence evaluation [J]. Journal of Power Sources, 2014, 248: 533-544.
- [29] Chen HC, Pei PC, Song MC. Lifetime prediction and the economic lifetime of proton exchange membrane fuel cells [J]. Applied Energy, 2015, 142: 154-163.
- [30] Pei PC, Chang QF, Tang T. A quick evaluating method for automotive fuel cell lifetime [J]. International Journal of Hydrogen Energy, 2008, 33(14): 3829-3836.
- [31] Daowd M, Omar N, van Mierlo J, et al. Extended PNGV battery model for electric and hybrid vehicles [J]. International Review of Electrical Engineering, 2011, 6(3): 1264-1278.
- [32] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning [Z/OL]. arXiv Preprint arXiv:1312.5602, 2013.
- [33] Zhao P, Wang YZ, Chang N, et al. A deep reinforcement learning framework for optimizing fuel economy of hybrid electric vehicles [C] // 2018
 23rd Asia and South Pacific Design Automation Conference, 2018: 196-202.
- [34] Wu YK, Tan HC, Peng JK, 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.
- [35] Lian RZ, Peng JK, Wu YK, et al. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle [J]. Energy, 2020, 197: 117297.