



基于深度强化学习的列车能量管理方法研究

胡恩泽, 陈 林, 许祖铭

(株洲中车时代电气股份有限公司, 湖南 株洲 412001)

摘 要: 柴油-电力混合动力列车被视为重载铁路低碳化进程中不可或缺的一种解决方案, 其能量管理策略需在高度动态的运行工况下, 始终满足柴油消耗最小化与电池健康状态高水平维持的要求。针对上述问题, 文章提出一种基于规则增强的深度强化学习 (rule-augmented deep reinforcement learning, RDRL) 方法。首先, 建立1.5 MW柴油-电力混合动力列车的高保真后向仿真模型, 其中发动机、发电机及牵引电机的效率图谱由台架试验获取, 电池采用内阻模型进行表征。其次, 将通过发动机最优制动比油耗曲线、牵引电池充放电特性以及离线动态规划 (dynamic programming, DP) 方法生成的动作轨迹作为3类先验知识, 编码为动作空间整形与奖励正则化项, 并嵌入深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 智能体中。然后, 采用优先经验回放与衰减的奥恩斯坦-乌伦贝格噪声对RDRL智能体进行训练, 提升学习效率与策略稳定性。最后, 通过仿真与半物理试验验证了所提方法的有效性。试验结果表明, 与基线 DDPG 算法和确定性规则型方法相比, 所提方法使列车油耗降低3.5%, 算法收敛速度提升38.7%, 以离线DP全局最优解为基准, 该方法的燃油经济最优性可达94.9%; 同时, 文章基于智轨电车平台进一步验证了该框架对氢燃料电池混合动力系统的良好迁移性, 为混合动力列车的动态能量管理策略提供了具备工程价值的解决方案。

关键词: 柴油-电力混合动力列车; 能量管理策略; 规则增强深度强化学习; 燃油经济性; 先验知识; 仿真

中图分类号: U267.1; TP183 **文献标志码:** A **文章编号:** 1000-128X(2025)06-0042-13

0 引言

交通运输行业是温室气体排放的主要来源之一^[1], 轨道交通作为传统的能耗大户, 其电气化改造已取得显著进展, 截至2025年, 我国电气化铁路里程占铁路总运营里程的比例已超过70%^[2-3]。然而, 内燃机车凭借其强耐久性、多场景适应性与高可靠性, 在复杂地形、极端气象条件下仍具有不可替代的作用^[4]。在可预见的未来, 仍将有相当规模的铁路机车采用柴油机动动力^[5], 针对内燃机车开展能量管理策略 (energy management strategy, EMS) 优化, 对提升铁路系统运营效率和实现环境可持续发展具有至关重要的意义^[6]。通过

结合柴油发动机与电力驱动系统的优势, 开发柴油-电力混合动力列车 (diesel-electric hybrid trains, DEHT) 为降低轨道交通对环境的影响提供了一种解决方案^[7], DEHT动力系统通过再生制动和发动机自动调节技术优化能源消耗^[8], 但是, 上述优势充分发挥的核心前提是设计并实施先进且高效的能量管理策略^[9-10]。

DEHT的能量管理策略主要分为3类: 基于规则的方法、基于优化的方法和基于学习的方法^[11]。基于规则的方法包括确定性与模糊逻辑系统, 此类方法具有简单、实时和易于实现等优势^[12-13], 但也存在依赖预设规则, 在变化的负载特性和行驶条件下表现欠佳, 达不到最优性能的问题^[14]。

基于优化的方法可分为离线优化和实时优化策略。

基金项目: 湖南省自然科学基金项目 (2025JJ50726)

通信作者: 胡恩泽, 男, 博士, 主要从事轨道交通能量调度与机器人集群调度方面的研究; E-mail: huez30@csrzc.com

引用格式: 胡恩泽, 陈林, 许祖铭. 基于深度强化学习的列车能量管理方法研究[J]. 机车电传动, 2025(6): 42-54. DOI:10.13890/j.issn.1000-128X.2025.06.005.

Citation: HU Enze, CHEN Lin, XU Zuming. Research on DRL-based energy management for trains[J]. Electric drive for locomotives, 2025(6): 42-54. DOI:10.13890/j.issn.1000-128X.2025.06.005.

动态规划(dynamic programming, DP)等离线优化方法在理论上可获得最优解, 但计算时间太长且需预知完整行驶工况, 难以应用于实时场景^[15]。等效燃油消耗最小化策略和庞特里亚金最小值原理是两种最常见的实时优化方法^[16-19]。等效燃油消耗最小化策略通过调整等效系数来最小化总体成本函数, 从而优化能量消耗^[20]; 庞特里亚金最小值原理用于求解非线性两点边值问题, 为即时功率管理提供了稳健的框架^[21]。这些实时策略的有效性在很大程度上取决于对未来行驶工况及关键参数的预测精度^[22], 也面临着计算负荷高、对存储与处理能力要求高以及实时数据处理复杂等挑战^[23]。

基于学习的方法, 特别是采用深度强化学习(deep reinforcement learning, DRL)的方法, 代表了DEHT能量管理领域的一种变革性技术。DRL算法能够从复杂环境中学习最优策略, 无需显式模型或完整的系统知识, 可动态适应运行工况的变化, 并通过试错持续优化其决策能力^[24-26]。这种自适应性使DRL特别适用于管理混合动力系统中复杂的能量动态, 而传统方法在此类场景中往往面临挑战^[27-29]。近年来, DRL在混合动力系统能量管理策略领域取得了显著进展。文献[30]提出了一种基于改进的深度确定性策略梯度算法(deep deterministic policy gradient, DDPG)的能量管理策略, 面向电池和超级电容混合动力电动汽车, 通过引入局部敏感哈希与优先经验回放机制, 增强了探索能力与学习效率; 然而, 当应用于更大规模、更复杂的系统时, 仍需进一步优化内存占用与训练时间。文献[31]针对燃料电池混合动力电动汽车, 提出了一种基于规则和DDPG方法的退化自适应能量管理策略, 该策略能够适应燃料电池堆的不同健康状态水平, 但需要更好地处理多能源集成并适应复杂运行工况。文献[32]针对网联混合动力电动公交车, 提出了一种结合离线预训练与在线微调, 显著提升了燃油经济性; 但其方法在实时驾驶条件下的适应性与鲁棒性缺乏有效验证。文献[33]提出了一种基于安全深度强化学习的约束最优控制方案, 用于混合动力电动汽车能量管理。该方法通过将基于神经网络的多约束优化策略与基于规则的约束系统相结合, 在优化燃油经济性的同时避免非理性控制信号。然而, 在处理更高维度的状态与动作空间时, 该方法的优化学习效率与更新效率较低。

综上所述, 上述研究在各自领域取得了重要进展, 但将其应用于柴油-电力混合动力列车时, 仍需要针对系统的复杂性、运行环境及能量需求进行优化, 包括管理更大规模系统、集成多元能源、在列车特定条件下可靠运行等。为解决这些问题, 本文提出一种规则增强的深度强化学习(rule-augmented deep reinforcement learning, RDRL)框架, 将预定义规则融入学习过

程提供初始引导, 并通过融入专家知识与运行启发式规则, 增强DDPG系统的鲁棒性和可靠性^[34]。本文的主要研究工作将集中于以下3个方面:

①框架设计。本文提出一种基于规则增强的深度强化学习框架, 将预定义的规则逻辑集成到DDPG架构中, 增强系统主动管理能量的能力, 从而在燃油效率和运行可靠性方面实现显著提升, 与传统DRL及基于规则的方法相比, 其燃油消耗量可减少3.5%。

②鲁棒性验证。通过结合真实世界数据与极端场景下的仿真数据, 验证RDRL在DP先验引导下对环境变化的快速响应能力, 展现出较强的鲁棒性与适应性, 且在多项指标中优于现有系统。

③迁移性测试。所提框架具备良好的通用性, 除柴油-电力混合动力列车外, 在氢燃料电池自主式轨道快运(autonomous rail rapid transit, ART)系统等应用中, 具备潜在的价值。

1 柴油-电力混合动力列车建模

1.1 系统架构构建

柴油-电力混合动力列车的系统架构如图1所示。柴油发动机驱动发电机, 将化学能转化为电能, 经整流器将交流电(alternating current, AC)转换为直流电(direct current, DC)。直流母线作为核心功率分配枢纽, 一方面向牵引逆变器供电, 通过逆变器将DC逆变为AC驱动牵引电机; 另一方面通过DC-DC变换器将能量送至电池系统, 实现高效存储。此外, DC-DC变换器支持能量双向流动, 可在发电机输出功率较低或处于节能模式时, 由电池为直流母线提供电能。列车详细参数如表1所示。

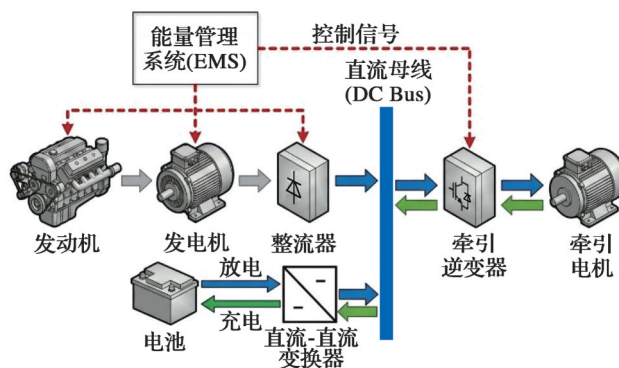


图1 DEHT体系结构

Fig. 1 Architecture of the DEHT

1.2 功率需求模型

本节将在文献[35]的研究基础上, 构建1.5 MW DEHT的高保真仿真模型, 用于EMS的训练与性能评估。模型基于纵向力平衡方程, 依据给定的运行速度曲线逆向推算出机车的实时功率需求, 如式(1)~式(5)

表1 车辆参数

Table 1 Key technical parameters of the train

类型	参数项目	参数值
列车	整备质量/t	155
	滚动阻力系数	0.001 6
	空气阻力系数	0.93
	正向迎风面积/m ²	12.238
发动机	最大功率/kW	2 500
	最大转矩/(N·m)	19 894
	最高转速/(r·min ⁻¹)	1 200
牵引电机	最大功率/kW	520
	最大转矩/(N·m)	1 655
	最高转速/(r·min ⁻¹)	5 000
发电机	最大功率/kW	300
	最大转矩/(N·m)	637
	最高转速/(r·min ⁻¹)	1 200
电池	容量/(kW·h)	1 250
	电压/V	1 500

所示。列车总阻力 F 由滚动阻力 F_f 、空气阻力 F_w 、坡道阻力 F_i 以及加速过程中的惯性力 F_a 4部分组成,具体计算公式如下:

$$F = F_f + F_w + F_i + F_a \quad (1)$$

$$F_f = mgf \quad (2)$$

$$F_w = \frac{1}{2} \rho \cdot A_f \cdot C_D \cdot v^2 \quad (3)$$

$$F_i = mg \cdot i \quad (4)$$

$$F_a = ma \quad (5)$$

式中: mg 为作用于DEHT的重力; f 为滚动阻力系数; ρ 为空气质量密度; A_f 为正向迎风面积; C_D 为空气阻力系数; v 为不考虑风速影响的车辆纵向速度; i 为线路梯度; m 为车辆质量; a 为加速度。

在传动系统建模中,假设多级齿轮箱的总传动比为 r_{total} (同时包含柴油机与电机的传动比),则该值可分解为若干级传动比的乘积,总传动比的定义为

$$r_{\text{total}} = r_{\text{gear1}} \times r_{\text{gear2}} \times \cdots \times r_{\text{gearn}} \quad (6)$$

式中: r_{gearn} 为第 n 级齿轮的传动比。

最终的牵引输出由柴油机与电机输出叠加得到,公式为

$$T_{\text{wheel}} = (T_{\text{engine}} \eta_{\text{engine}} + T_{\text{motor}} \eta_{\text{motor}}) \cdot r_{\text{motor}} \quad (7)$$

式中: T_{wheel} 为轮周处的转矩; T_{engine} 为发动机产生的转矩; T_{motor} 为电机产生的转矩; η_{engine} 、 η_{motor} 分别为两种动力源对应传动系统的效率; r_{motor} 为电机侧总传动比。

速度-转矩关系为

$$\left. \begin{aligned} n_g + \alpha n_{\text{out}} - (1 + \alpha) n_e &= 0 \\ T_g : T_{\text{out}} : T_e &= 1 : \alpha : - (1 + \alpha) \end{aligned} \right\} \quad (8)$$

式中: α 为特征参数; n_g 、 n_{out} 与 n_e 分别为发电机、齿轮与发动机的转速; T_g 、 T_{out} 与 T_e 分别对应上述部件的

转矩。

1.3 发动机模型

发动机、发电机与牵引电机均采用由台架试验获得的效率特性图进行表征。其中,发动机特性图可详细描述转速、转矩与有效燃油消耗率(broke specific fuel consumption,BSFC)之间的非线性映射关系,如图2所示。

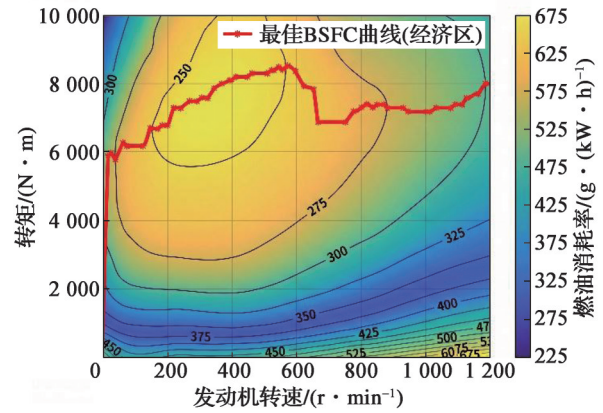


图2 发动机特性曲线图

Fig. 2 Engine characteristic curves

电池系统采用安时积分法结合等效内阻(R_{int})模型进行建模,其等效电路结构如图3所示,电池等效电路模型具体如下:

$$P_{\text{batt}}(t) = V_{\text{oc}}(t) \cdot I(t) - R_0 \cdot I^2(t) \quad (9)$$

$$I(t) = \frac{V_{\text{oc}}(t) - \sqrt{V_{\text{oc}}^2(t) - 4R_0 \cdot P_{\text{batt}}(t)}}{2R_0} \quad (10)$$

$$A_{\text{soc}}(t) = \frac{Q_0 - \int_0^t I(t) dt}{Q} \quad (11)$$

式中: $P_{\text{batt}}(t)$ 为电池输出功率; $V_{\text{oc}}(t)$ 为开路电压; $I(t)$ 为输出电流; R_0 为内阻。 $A_{\text{soc}}(t)$ 为荷电状态值; Q_0 为电池的初始容量; Q 为额定电池容量。

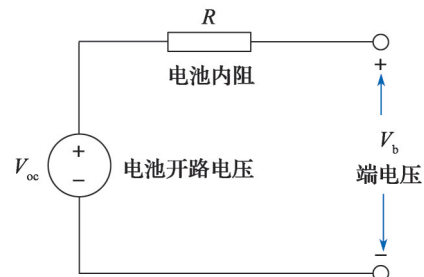


图3 电池等效电路模型

Fig. 3 Equivalent circuit model of the traction battery

为兼顾实时性,模型暂未考虑温度变化或电池老化对参数的动态影响。电池作为牵引电机与发电机的关键能量缓冲带,其内阻与开路电压随荷电状态(state of charge,SOC)值变化情况如图4所示。

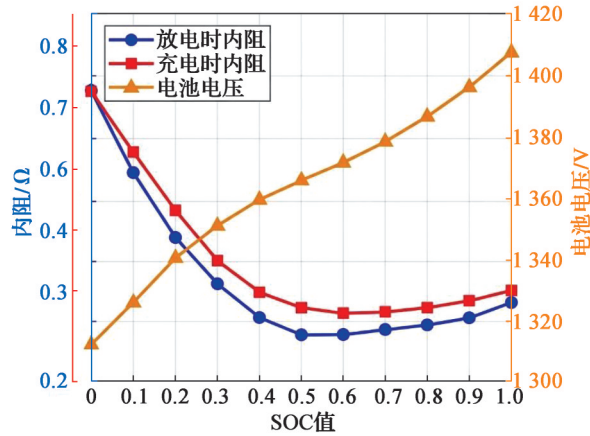


图4 电池特性

Fig. 4 Battery characteristics

2 基于规则增强的深度强化学习

本文将提出一种RDRL框架,旨在优化DEHT的能量管理策略。框架的总体架构如图5所示,集成3类关键领域知识:发动机BSFC曲线、动力电池电化学特性以及由离线动态规划生成的全局最优参考轨迹。通过将专家知识转化为DDPG算法中的硬性约束与奖励函数正则项,有效引导能量管理策略向全局最优解演进。首先,将最佳BSFC曲线、电池充放电特性及离线动态规划生成的全局最优参考轨迹映射为动作空间约束和多目标奖励函数。在这些约束引导下,DDPG智能体结合优先经验回放机制与衰减的奥恩斯坦-乌伦贝克噪声进行训练,从而提高学习效率。

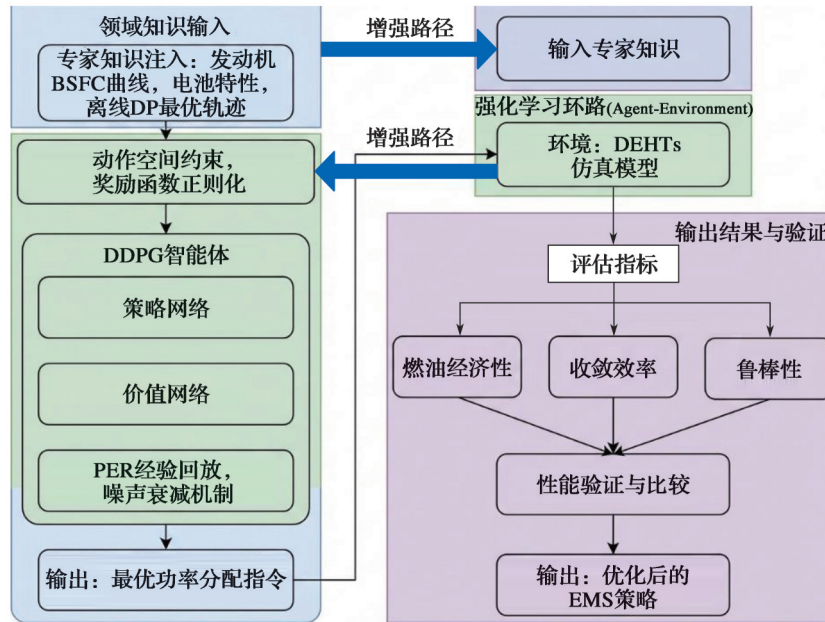


图5 规则增强深度强化学习框架

Fig. 5 Rule-augmented deep reinforcement learning framework

2.1 专家知识集成与建模

为实现更优的燃油经济性与排放性能,发动机应尽可能工作在BSFC较低的区域^[36]。图2中的最佳BSFC曲线即为本文采用的先验知识之一,通过将控制变量锚定在该曲线上,可有效简化发动机动作空间,提升策略搜索效率^[37]。电池电化学特性对EMS性能亦有显著影响。试验数据表明,当SOC值 A_{SOC} 处于0.40~0.75区间时,电池内阻较低,系统效率较高;在 $A_{\text{SOC}}=0.62$ 时内阻最低,为最优运行状态。此外,通过离线求解典型工况下的DP最优解,可获得随里程变化的参考荷电状态与参考功率,使EMS具备全局视角,预判能量需求趋势,动态调整功率分配,防止 A_{SOC} 偏离最优路径。

2.2 基于领域知识与DP先验的动作空间塑造

DRL在处理DEHT高维连续动作空间问题时,面

临探索效率低下的挑战。智能体在训练初期缺乏对动力系统物理特性的理解,随机探索产生的发电机功率指令极易偏离系统高效区间,导致训练收敛缓慢,且结果具有随机性。针对这个问题,本文在动作空间设计中引入了基于动态规划的基准先验。

首先,通过对列车典型运行工况(包括不同的区间距离、限速要求及坡度组合)进行离线DP全局优化求解,提取在不同荷电状态值 A_{SOC} 和实时功率需求下发电机功率的最优控制序列 $P_{\text{gen},t}^{\text{DP}}$,并基于离线最优轨迹构建动态动作约束边界 $P_{\text{gen},t}$,具体方法为

$$P_{\text{gen},t} \in [P_{\text{gen},t}^{\text{DP}} - \varepsilon, P_{\text{gen},t}^{\text{DP}} + \varepsilon] \cap [P_{\text{gen},\min}, P_{\text{gen},\max}] \quad (12)$$

式中: $P_{\text{gen},\min}$ 、 $P_{\text{gen},\max}$ 分别为发电机的物理极限功率下限、上限; ε 为根据工程经验设定的松弛因子,用于平衡DP先验的指导作用与DRL的自主寻优空间。

其次,在构建动作空间约束时,本文除考虑发电机组的静态功率边界外,还考虑了动力总成的物理动

态响应特性，引入发电机组功率爬坡率约束 $\Delta P_{\text{gen,max}} \cdot \Delta t$ ，公式表示为

$$|P_{\text{gen},t} - P_{\text{gen},t-1}| \leq \Delta P_{\text{gen,max}} \cdot \Delta t \quad (13)$$

离线全局最优经验在实时推理阶段的有效映射如图6所示。通过DP对整个运行阶段典型工况进行离线求

解，构建以列车当前速度 v 、荷电状态 SOC 值及需求功率 P_{dem} 为索引的最优控制先验数据库。在DRL在线训练与推理过程中，智能体通过多维线性内插技术，根据实时状态向量从数据库中检索对应的准优参考值 $P_{\text{gen},t}^{\text{DP}}$ ，从而兼顾全局优化的预见性与在线控制的实时性。

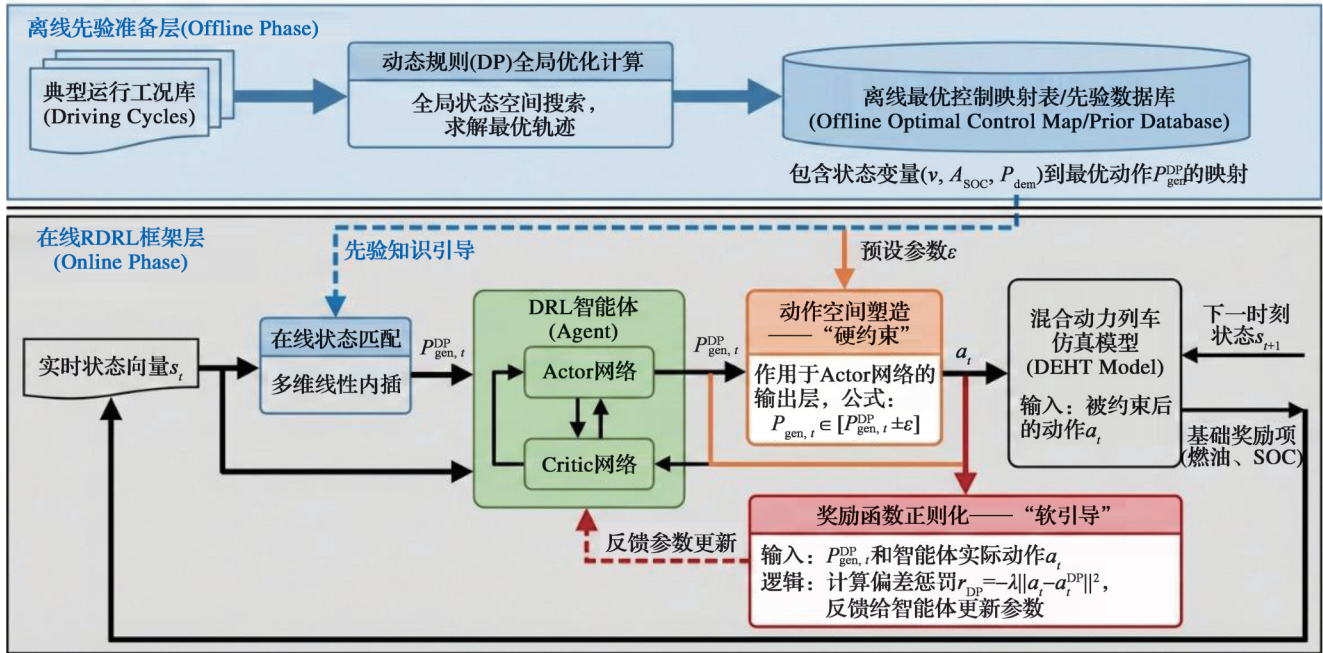


图6 DP先验引导与在线DRL结合架构图

Fig. 6 Architecture of DP prior-guided integrated with offline DRL

2.3 基于RDRL的DEHTs能量管理策略

在具有马尔可夫性质的环境中，RDRL智能体通过与环境的持续交互进行学习，根据当前状态选择动作，从环境中获取奖励，并实现向新状态的转移，EMS智能体与关键外部环境因素之间的交互关系如图7所示。

2.3.1 基础奖励函数设计

RDRL-EMS的状态与动作变量设置如下：连续动作变量来源于发动机最佳BSFC曲线。EMS的奖励函数由3部分构成：发动机瞬时燃油消耗、维持电池电量的代价以及基于DP参考 A_{SOC} 轨迹的偏差惩罚项。构建多目标奖励函数以平衡三者，公式为

$$\left. \begin{aligned} \text{state} &= \{A_{\text{SOC}}, \text{velocity}, \text{acceleration}, \text{timetable}\} \\ \text{action} &= \{\text{continuous engine power}\} \\ \text{reward} &= -\left\{ \alpha \text{fuel}(t) + \beta [A_{\text{SOC,ref}} - A_{\text{SOC}}(t)]^2 \right\} \end{aligned} \right\} \quad (14)$$

式中：state为智能体观测的状态空间，包含电池荷电状态值 A_{SOC} 、列车瞬时速度 velocity、加速度 acceleration 以及运行时刻表 timetable；action为智能体输出的连续动作变量，包含发动机输出功率 continuous engine power；reward为即时奖励函数，其中的 $\text{fuel}(t)$ 为发动机瞬时燃油消耗率， $A_{\text{SOC}}(t)$ 为当前时刻电池荷电状态值， α 与 β 分别为燃油消耗与电池荷电

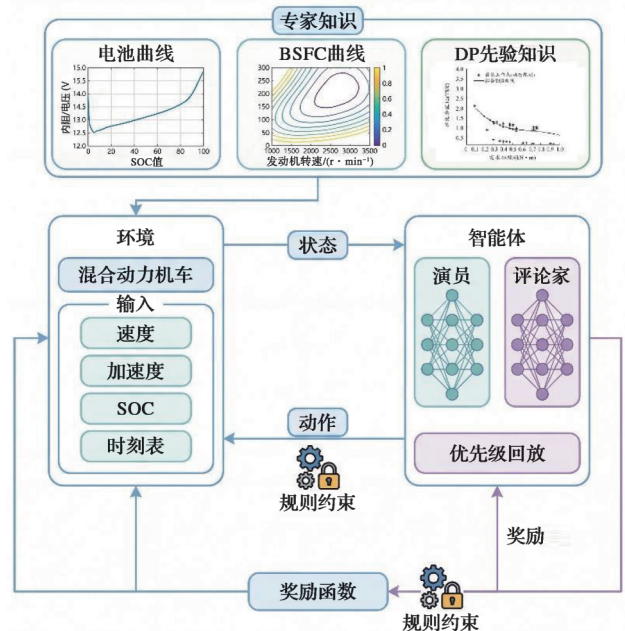


图7 DEHTs能量管理的智能体-环境交互

Fig. 7 Agent-environment interaction for DEHT energy management

维持的权重系数， $A_{\text{SOC,ref}}$ 为 A_{SOC} 的基准值。

多目标奖励函数优化的关键难点在于调节权重参数 α 与 β ，为将 A_{SOC} 参考值保持在安全工作区间，设定 $A_{\text{SOC,ref}}=0.62$ ，并将其约束在 $[0.58, 0.64]$ 范围之内。采用

两阶段的权重调节流程: 阶段1, 固定 $\alpha=1$ 并在窄范围内扫描 β , 使燃油消耗项与 A_{SOC} 偏差二次项在代表性工作点处量级相当; 阶段2, 针对每组权重组合进行训练, 并评估4项核心指标——燃油经济性、 A_{SOC} 稳定性、平均回合回报以及收敛速度。验收标准设定为: 终端 A_{SOC} 位于 $[0.58, 0.64]$ 内; A_{SOC} 波动不大于 0.05。

2.3.2 引入DP的奖励函数设计

为解决 RDRL 在大规模连续动作搜索中收敛缓慢且容易陷入局部最优的问题, 在上述奖励函数的基础上, 引入基于 DP 最优轨迹的正则化项 $r_{\text{best},t}$, 利用离线全局最优经验作为专家知识引导, 修正智能体在复杂工况下的策略偏差, 融合后的完整奖励函数 r_t 定义为

$$r_t = r_{\text{best},t} + r_{\text{DP},t} \quad (15)$$

式中: $r_{\text{DP},t}$ 为对应 DP 的正则项, 具体见式(16)。

$$r_{\text{DP},t} = -\lambda_{\text{DP}} \|P_{\text{gen},t} - P_{\text{gen},t}^{\text{DP}}\|^2 \quad (16)$$

式中: λ_{DP} 为拉格朗日乘子; $P_{\text{gen},t}^{\text{DP}}$ 为将 DP 求解的全局最优控制序列。

超参数 ε 与 λ_{DP} 的设置直接决定先验知识引导的强度。其中, 松弛因子 ε 设定为发电机额定功率的 5%~

10%, 以确保在压缩搜索空间的同时, 为 DRL 留出足够的随机探索裕量。为了平衡训练初期的模仿学习与后期的自主策略演化, 本文为正则项权重 λ_{DP} 引入分阶段线性衰减机制, 具体为

$$\lambda_{\text{DP}}(k) = \max(\lambda_{\text{end}}, \lambda_{\text{start}} - (k/K)(\lambda_{\text{start}} - \lambda_{\text{end}})) \quad (17)$$

式中: $\lambda_{\text{DP}}(k)$ 为第 k 次迭代时 DP 正则项的实际权重; λ_{start} 和 λ_{end} 分别为正则项权重的初始值和终止值; k 为当前迭代次数; K 为预设的退火步数, 即权重衰减过程的总迭代步数。

该策略使智能体在训练初期 ($\lambda_{\text{DP}}(k)$ 较大) 能够迅速向 DP 全局最优分布靠拢以越过收敛瓶颈, 而在训练后期 ($\lambda_{\text{DP}}(k)$ 逐渐减小至 λ_{end}) 逐渐放开软约束, 允许策略根据高保真环境反馈进行微调。

RDRL 算法采用独立的动作网络将状态映射为确定的连续动作序列, 适用于连续状态与动作空间问题。算法结合优先经验回放与探索噪声以优化决策流程, 详细步骤如图8所示。Critic 与 Actor 网络均采用多层感知机结构的深度神经网络以适配大状态和动作空间。

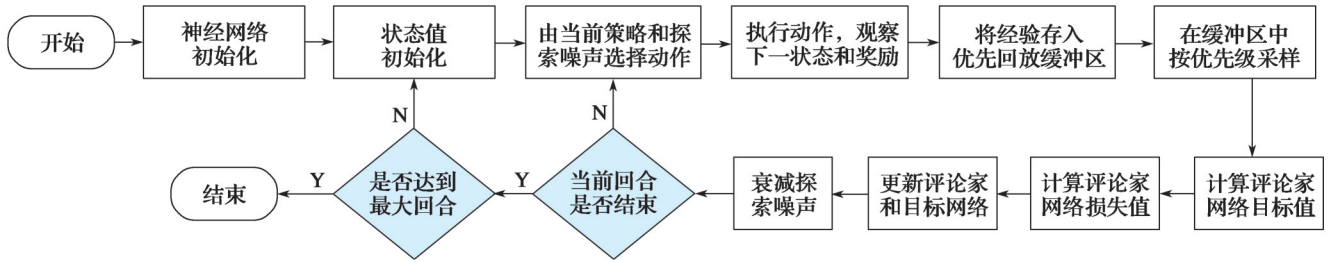


图8 RDRL算法流程

Fig. 8 Workflow of the RDRL algorithm

Actor-Critic 网络结构针对混合动力列车 EMS 任务设计的结构如图9所示。Critic 网络通过贝尔曼方程学习, Actor 网络通过采样策略梯度更新^[14]。

为进一步提升学习效率并加速 RDRL 收敛, 本文采用优先经验回放机制, 按时间差分误差为转移样本分配优先级^[38]。该方法通过更频繁地回放更重要的观测区域提高深度强化学习效果, 相比标准随机回放更高效。深度 Q 学习(DQL)适用于离散动作空间^[39], 但在高维连续空间中依赖最大化动作价值函数, 因此在本文不能直接采用。本文中动作被定义为发动机功率的增减量, 并基于发动机 BSFC 曲线将其离散为以下 10 种选项: 保持不变, 增加 20 kW, 减少 20 kW, 增加 40 kW, 减少 40 kW, 增加 100 kW, 减少 100 kW, 增加 200 kW, 减少 200 kW, 关闭发动机 (功率设为 0 kW)。

2.4 RDRL-EMS 评估指标

为评估 RDRL 在 DEHT 能量管理任务中的表现, 建立了以下 3 项与 RDRL 和 EMS 特性相匹配的评价指标:

①燃油经济性。在指定工况循环下统计 DEHT 总

燃油消耗, 并以离线 DP 作为对比基准, 同时对终点 SOC 差异进行修正。

②收敛效率。为衡量 RDRL 策略收敛到最优解的速度, 采用相邻迭代奖励的欧氏距离度量, 如式(18)所示。

$$|\bar{R}_{i+1} - \bar{R}_i| \leq \varepsilon, \text{ subject to } A_{\text{SOClo}} \leq A_{\text{SOC}} \leq A_{\text{SOCup}} \quad (18)$$

式中: $\bar{R}_i$ 为平均奖励; A_{SOClo} 、 A_{SOCup} 分别为 A_{SOC} 的预设下界与上界。

③鲁棒性。为评估 RDRL 在不同工况循环下的稳定性, 采用燃油经济性的均值与标准差构建鲁棒性指标, 如式(19)所示。

$$\left. \begin{aligned} \mu &= \sum_{i=1}^N (f_{\text{eco}})_i / N \\ \sigma &= \sqrt{\frac{1}{N} \sum_{i=1}^N ((f_{\text{eco}})_i - \mu)^2} \end{aligned} \right\} \quad (19)$$

式中: N 为测试工况循环的总次数; $(f_{\text{eco}})_i$ 为第 i 次测试的燃油经济性指标; μ 与 σ 分别为均值与标准差, 标准差越小, 表明算法在不同工况下的稳定性越好, 鲁棒性越强。

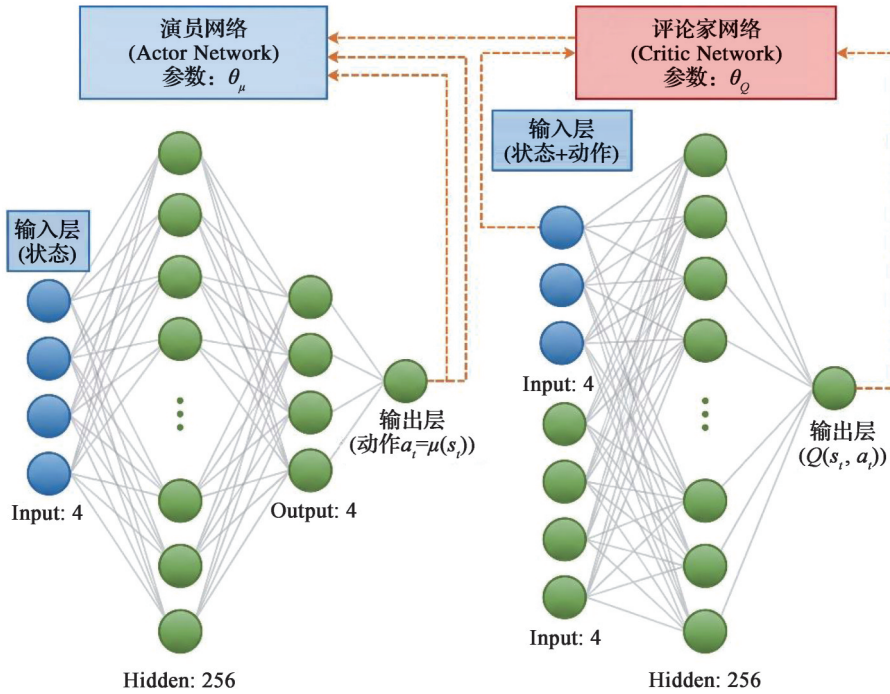


图9 Actor-Critic网络结构

Fig. 9 Actor-Critic network architecture

3 试验验证

本章将对基于RDRL的EMS进行试验分析, 试验采用DEHT的真实运行数据进行测试, 并将电池的SOC值 A_{SOC} 统一设定为0.62。

3.1 多目标函数分析

多目标奖励函数由发动机瞬时燃油消耗和基于先验知识构建的电池荷电维持成本两项组成。在式(14)中, 较大的权重系数 α 会驱动智能体在各类工况下更优先降低燃油消耗, 而较大的系数 β 会促使智

能体更关注电池健康, 需要通过参数调优来平衡整体性能。

试验中, 首先设定 $\alpha=1$, 使 β 在0.000 3~0.001 0区间动态变化。在该设定下, 电池SOC维持成本项 $\frac{1}{2}(A_{\text{SOC ref}} - A_{\text{SOC}}(t))^2$ 通常在0.01至1区间内变化, 用于引导燃油消耗与电池维持成本之间的平衡。图10展示了在8组不同权重下, 电池荷电维持奖励随 A_{SOC} 变化的关系, 当 A_{SOC} 位于0.58~0.62之间时, 该项奖励趋近于0。图11进一步展示了惩罚参数 α 逐步增大对 A_{SOC} 动态特性的影响, 随着 α 的增大, A_{SOC} 波动幅度明显减小, 这表明系统稳定性得到了增强。

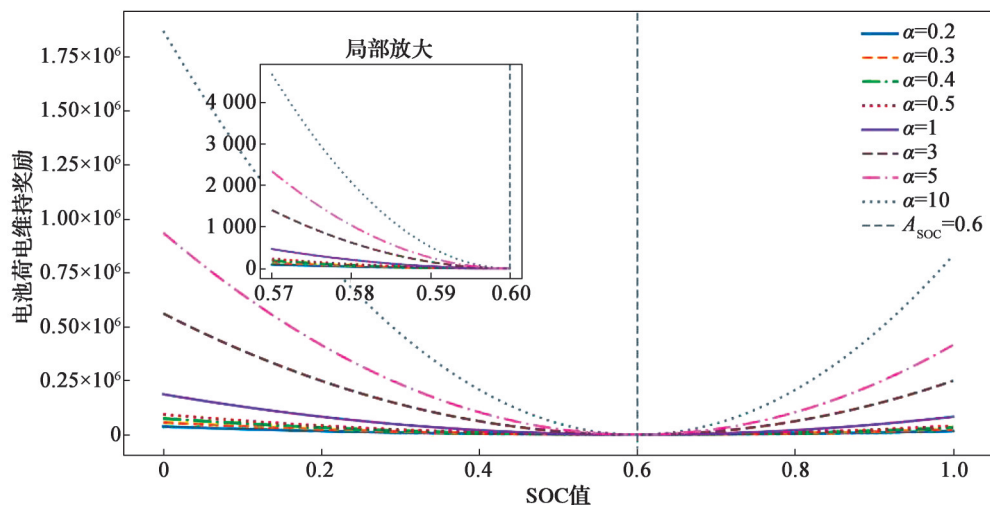


图10 不同权重设置下电池SOC值维持情况

Fig. 10 Different weight settings for battery SOC maintenance

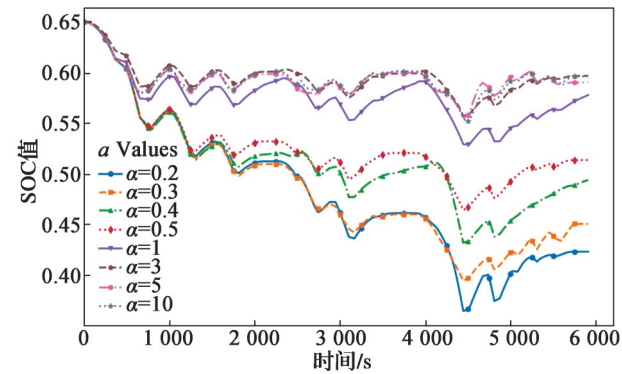


图 11 不同权重设置下的SOC值轨迹
Fig. 11 SOC trajectories under various weight settings

3.2 对比试验

为评估动作空间简化策略的有效性, 试验对比了引入和未引入最佳BSFC曲线的RDRL模型。图12展示了发动机工作点分布特征, 引入BSFC先验知识后, 工作点显著聚集在BSFC曲线附近, 从而降低了燃油消耗; 相比之下, 未引入先验知识的模型工作点更分散, 燃油消耗控制效果较差。

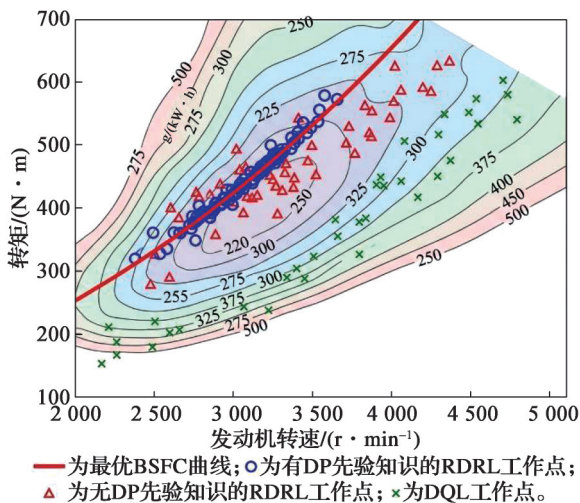


图 12 引入与未引入先验知识时的发动机工作点
Fig. 12 Engine operating points with and without prior knowledge

表2详细对比了RDRL与标准DDPG算法的性能表现。RDRL的燃油消耗量为320.667 L/100 km, 优于基准DDPG的332.296 L/100 km, 燃油经济性从91.3%提升至94.66%, 结果表明引入BSFC曲线对学习过程具有显著的促进作用。两种模型的终端SOC值基本保持一致 (RDRL为0.597, DDPG为0.598), 确保了电池的可持续利用。

图13展示了引入与未引入BSFC曲线时算法的收敛速度对比。结果表明, BSFC先验信息显著加快了算法的收敛过程。通过集成BSFC曲线, 算法能够更快地定位最优工作区域, 减少不必要的探索过程, 与标准DRL基线相比, 收敛速度提升了38.7%。

表2 RDRL与标准DDPG算法对比

Table 2 Comparison between RDRL and standard DDPG algorithms

参数项目	RDRL	DDPG
燃油消耗/(L·(100 km) ⁻¹)	320.667	332.296
终端SOC值	0.597	0.598
DP燃油消耗基准值/(L·(100 km) ⁻¹)	303.538	303.538
燃油经济性/%	94.66%	91.30%

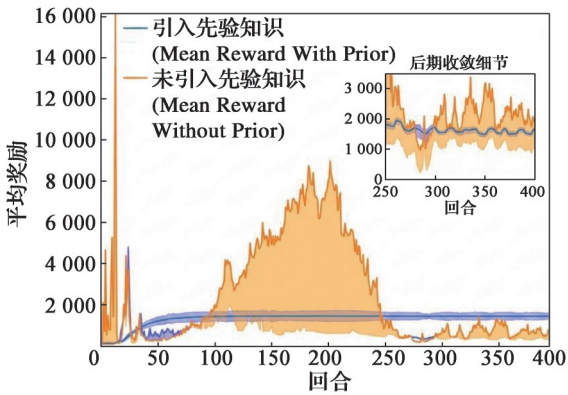


图 13 引入与未引入先验知识时的收敛速度
Fig. 13 Convergence speed with and without prior knowledge

为评估融合离线DP最优轨迹先验的影响, 通过试验对比含有与不含DP轨迹指引的RDRL模型。如图14所示, 将DP轨迹作为参考项引入EMS能够显著提升系统的稳定性与运行效率。在初始 A_{soc} 为0.65的条件下, 引入DP轨迹后的最小 A_{soc} 从0.56提升至0.60, A_{soc} 储备更为健康; 同时, A_{soc} 的波动幅值减少了50% (从0.08降至0.04), 系统稳定性明显增强。

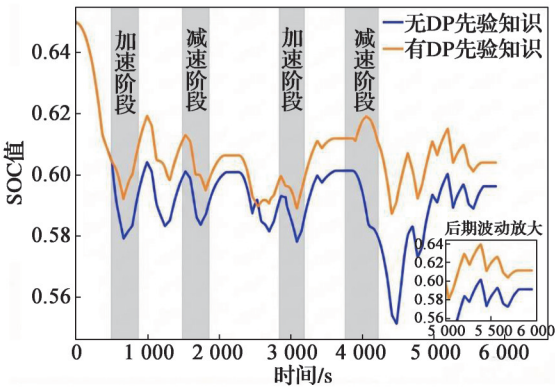


图 14 融合DP先验知识的影响
Fig. 14 Impact of incorporating DP prior knowledge

3.3 DP全局轨迹指引的消融试验

为进一步验证DP全局参考轨迹对学习过程的促进作用, 在对比试验基础上补充消融试验: 组A为标准DDPG, 组B为现有RDRL (BSFC+电池特性), 组C为本文采用的RDRL (在组B基础上增加DP轨迹指引)。3组均在相同工况与初始SOC值条件下训练, 统计燃油消耗、终端SOC值偏差等指标。试验结果如表3

所示，结果表明组 C 的燃油经济性最高，达到了 DP 基准的 94.9%。

表 3 3 种策略对比

Table 3 Comparison among three strategies

方法	燃油消耗/ (L·(100 km ⁻¹))	燃油经济性 (相对 DP)
A: 标准 DDPG	332.30	91.3%
B: 现有 RDRL	323.44	93.8%
C: RDRL + DP 先验	320.71	94.9%

3.4 可迁移性验证

表 4 对比了在不同权重系数 α 取值下 RDRL 与 DP 算法的表现，反映了该方法对不同惩罚强度的适应性，而惩罚强度直接决定了燃油消耗与终端 SOC 值的平衡关系。结果显示，随着 α 的增大，两种算法的燃油消耗均上升，且 RDRL 增幅更明显。尽管如此，RDRL 仍保持较优的燃油经济性，并在 $\alpha=0.5$ 时达到 94.9%。当 α 取 1、3、5、10 等较大值时，RDRL 燃油消耗虽高于 DP，但其相对效率仍保持稳定。

为验证所提方法的可迁移性，在半实物仿真平台直接将其集成至采用氢燃料电池与动力电池混合构型的 ART 系统中，如图 15 所示。ART 系统与 DEHT 系统

表 4 不同权重设置对比

Table 4 Comparison under various weight settings

α	燃油消耗/ (L·(100 km ⁻¹))	终端 SOC 值	DP 燃油消耗/ (L·(100 km ⁻¹))	燃油经济性/%
0.2	154.595	0.423	141.299	91.4%
0.3	175.611	0.451	161.035	91.7%
0.4	222.220	0.494	204.887	92.2%
0.5	233.449	0.514	221.544	94.9%
1	295.872	0.590	276.344	93.4%
3	304.401	0.596	285.528	93.8%
5	333.147	0.590	307.494	92.3%
10	331.419	0.597	308.551	93.1%

的动力架构存在本质差异，ART 采用氢燃料电池与锂电池的动力组合，先验数据库无法直接复用。本文针对 ART 系统重新采集典型运行工况，并离线求解 DP 全局最优轨迹，构建适用于 ART 系统的先验数据库后，对 RDRL 框架进行训练与验证。迁移性试验验证了 RDRL 整体架构的适用性，包括动作空间塑造方法、奖励函数设计、正则化项引入策略，仅需替换特性模型与先验数据即可实现跨车型部署。试验结果如图 16 所示，RDRL 框架在 ART 系统上实现了稳定的 SOC 值维持与氢气消耗降低，验证了该框架作为通用能量管理架构的工程适用性。

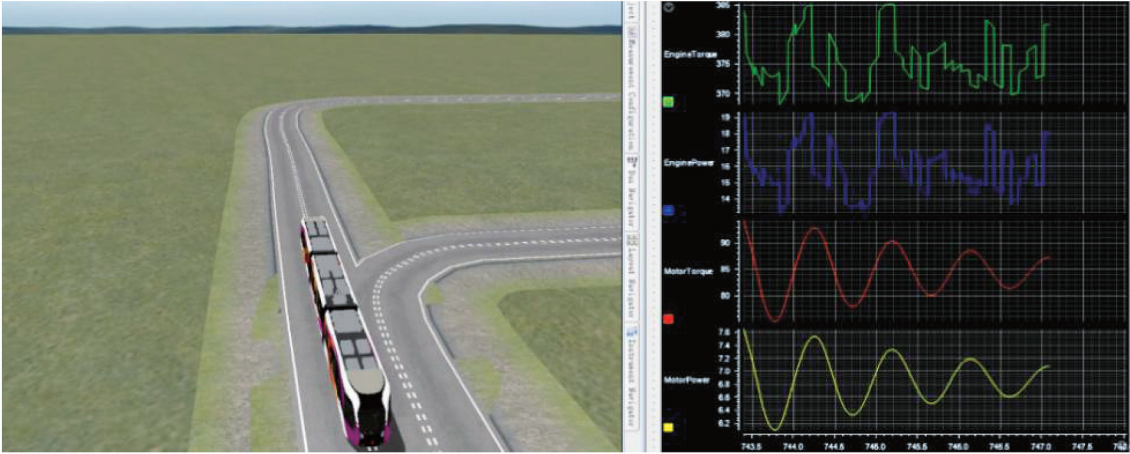


图 15 ART 半实物仿真平台

Fig. 15 Hardware-in-the-loop simulation platform for ART

3.5 多算法对比试验

图 17 对比了 RDRL、深度 Q 学习 (DQL) 与 DP 3 种算法的 SOC 值轨迹，将其中的 DP 作为最优基准。结果显示，RDRL 的 SOC 值轨迹更接近 DP 的最优轨迹，且相较于 DQL 能够更有效地抑制波动，说明其在变工况下更擅长稳定 SOC 值。相比之下，DQL 与 DP 基准的偏差较大，反映了其在动态负载变化下的适应能力不足。

图 18 进一步分析了 RDRL、DQL 与 DP 3 种算法的发动机工作点分布，以评估其燃油消耗效率。结果表

明，RDRL 能够将工作点稳定保持在最佳燃油消耗区域，聚集度高于 DQL 与 DP。这种对最优工作点的精准维持不仅降低了燃油消耗，而且提升了整体能量管理的效果。表 5 详细对比了 RDRL 与 DQL 在燃油消耗、终点 SOC 值以及相对 DP 基准的燃油经济性等方面的表现。图 18 中 RDRL 的发动机工作点聚集密度高于 DP，但表 5 中 DP 的燃油消耗仍低于 RDRL，这一现象并不矛盾，原因在于 DP 作为全局优化算法，其优化目标是全工况累积燃油消耗最小化，允许发动机在部分时段偏离最佳 BSFC 曲线，以换取更优的电池 SOC 值轨迹，从而实现全局能量分配的最优解；而 RDRL 在训练过

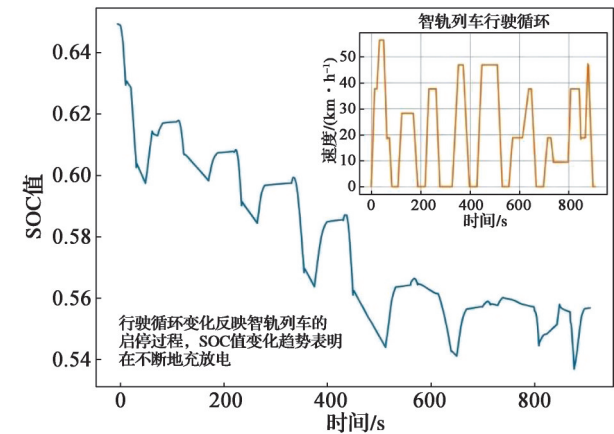


图16 ART的SOC值轨迹

Fig. 16 SOC trajectory of the ART system

程中同时受燃油消耗、SOC值维持及DP轨迹正则化3项约束,其工作点更紧密地聚集于最佳BSFC曲线附近,体现了对瞬时燃油效率的更高追求。DP以牺牲部分时刻的瞬时燃油效率换取全局最优,而RDRL在逼

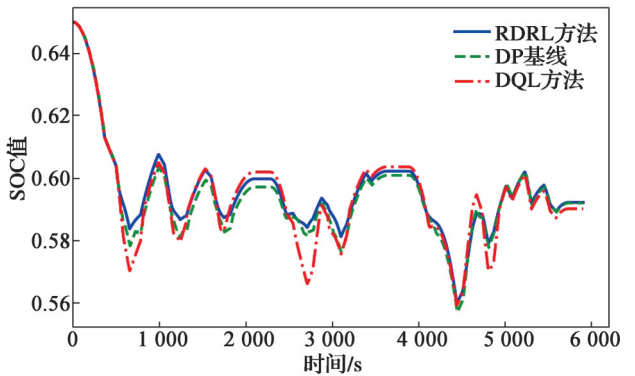


图17 3种EMS模型的SOC值轨迹对比

Fig. 17 Comparison of SOC trajectories among three EMS models

近DP全局最优的同时,倾向于更严格地遵循最佳BSFC曲线。

表5数据显示,RDRL的燃油消耗更低、燃油经济性更高(达到93.5%),在能量管理效率上显著优于DQL(84.7%)。

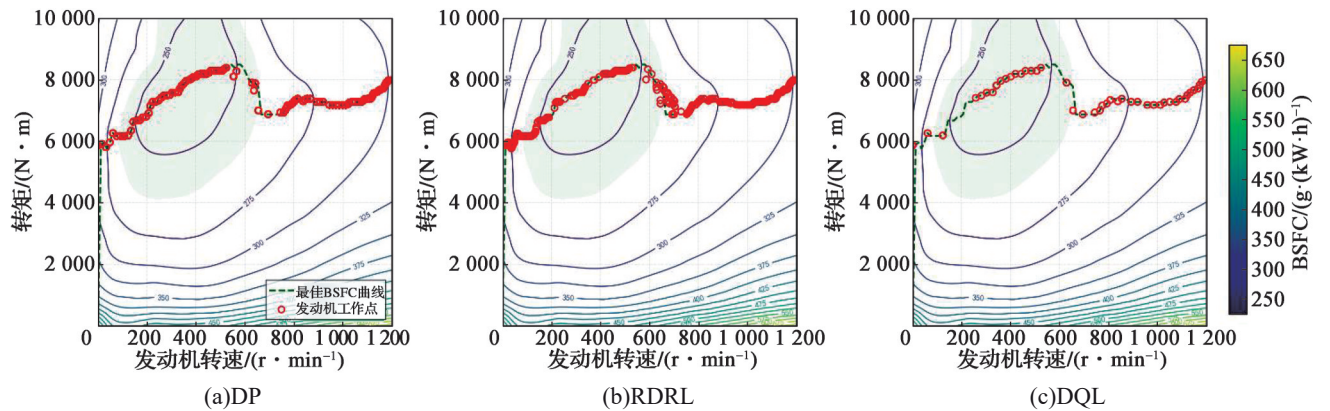


图18 发动机工作点分布

Fig. 18 Distribution of engine operating points

表5 不同工况下RDRL与DQL对比

Table 5 Comparison between RDRL and DQL algorithms under various conditions

算法	燃油消耗/ (L·(100 km) ⁻¹)	终端 SOC值	DP燃油消耗/ (L·(100 km) ⁻¹)	燃油经济 性/%
RDRL	323.599	0.593	302.565	93.5%
DQL	353.632	0.594	299.527	84.7%

真实机车测试的试验展示了本文框架的工程应用效果显著。在此基础上,图19进一步分析了RDRL得到的运行点分布,并重点关注其燃油消耗表现。图19运行点在低燃油消耗区域的明显聚集再次证明了RDRL在优化燃油使用方面的能力,同时成功将电池SOC值保持在0.45~0.70的合理区间。此外,在突发负载变化的运行模式下,本框架相对全局最优解的偏差始终保持在6.5%以内。这些试验结果共同验证了所提方法在真实场景中的鲁棒性与实用性。

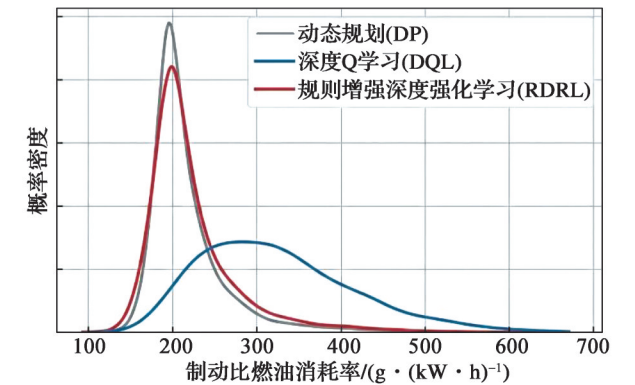


图19 燃油消耗率分布

Fig. 19 Distribution of fuel consumption rates

4 结束语

本文提出并验证了一种面向柴油-电力混合动力列车实时能量管理的规则增强深度强化学习框架。通过

将发动机最佳BSFC曲线、电池电化学特性及离线DP全局最优参考轨迹等先验知识显式注入DDPG智能体,有效解决了深度强化学习在重载铁路应用中面临的高维动作空间过度探索与未知工况鲁棒性不足两大瓶颈。研究表明,该机制使发动机工作点紧密收敛于BSFC曲线附近,燃油消耗较传统规则型控制器降低了12%~15%,较标准DDPG算法燃油消耗降低了2.3%~3.5%;优先经验回放与动作空间简化相结合使收敛速度提升了38.7%,在1500个训练回合内即可实现可靠部署;多目标奖励函数使电池SOC值在多种循环工况下稳定保持在0.58~0.64内,满足荷电维持要求;在氢燃料电池ART半实物仿真平台上的测试进一步验证了该框架在不同动力架构下的适用性,氢气消耗降低9%。

未来研究工作将聚焦于多物理场模型融合、系统安全韧性增强、多车协同控制、工程化落地验证以及多工况泛化能力研究等方向,具体包括集成热-老化耦合模型与在线健康状态估计、引入失效安全回退机制与硬件在环验证、扩展至多智能体协同控制场景、开展分阶段现场试验,以及探索基于元学习或域自适应技术的跨工况快速适配机制,以进一步提升算法在复杂运行环境下的鲁棒性与部署效率。

参考文献:

- [1] SINGH S. Energy crisis and climate change: global concerns and their solutions[J]. *Energy: crises, challenges and solutions*, 2021, 8(1): 1-17.
- [2] KLJAIĆ Z, PAVKOVIĆ D, CIPEK M, et al. An overview of current challenges and emerging technologies to facilitate increased energy efficiency, safety, and sustainability of railway transport[J]. *Future internet*, 2023, 15(1): 347.
- [3] DOLINAYOVÁ A, MORIHLADKO P. Rail transport performances on electrified and non-electrified railway lines: sustainable rail transport point of view[C]//*Proceedings of the 8th International Conference on Road and Rail Infrastructure*. Cavtat, Croatia: University of Zagreb, 2024: 1-8.
- [4] ALARROUQI R A, AHMAD F, BAYHAN S, et al. Electric buses in hot climates: challenges, technologies, and the road ahead[J]. *IEEE access*, 2024, 12: 55531-55550.
- [5] ZHU C Y, ZHANG Y X, WANG M X, et al. Optimization, validation and analyses of a hybrid PV-battery-diesel power system using enhanced electromagnetic field optimization algorithm and ε -constraint[J]. *Energy reports*, 2024, 11: 5335-5349.
- [6] FARAJ J, ALSAYAH A M, ALSHUKRI M J, et al. New co-generation concept applied to diesel power generators-thermal modeling, parametric analysis and first feasibility study[J]. *International journal of thermofluids*, 2025, 26: 101096.
- [7] AREDAH A, DU J H, HEGAZI M, et al. Comparative analysis of alternative powertrain technologies in freight trains: a numerical examination towards sustainable rail transport[J]. *Applied energy*, 2024, 356: 122411.
- [8] FONT A, HEDGES M, HAN Y Q, et al. Air quality on UK diesel and hybrid trains[J]. *Environment international*, 2024, 187: 108682.
- [9] ZAFAR S, KHAN A. Integrated hydrogen fuel cell power system as an alternative to diesel-electric power system for conventional submarines[J]. *International journal of hydrogen energy*, 2024, 51: 1560-1572.
- [10] BILAL M, ALGETHAMI A A, HAMEED S. Review of computational intelligence approaches for microgrid energy management[J]. *IEEE access*, 2024, 12: 123294-123306.
- [11] JUI J J, AHMAD M A, MOLLA M M I, et al. Optimal energy management strategies for hybrid electric vehicles: a recent survey of machine learning approaches[J]. *Journal of engineering research*, 2024, 12(3): 454-467.
- [12] YUAN H B, ZOU W J, JUNG S, et al. Optimized rule-based energy management for a polymer electrolyte membrane fuel cell/battery hybrid power system using a genetic algorithm[J]. *International journal of hydrogen energy*, 2022, 47(12): 7932-7948.
- [13] YUAN H B, ZOU W J, JUNG S, et al. A real-time rule-based energy management strategy with multi-objective optimization for a fuel cell hybrid electric vehicle[J]. *IEEE access*, 2022, 10: 102618-102628.
- [14] GHARIBEH H F, FARROKHIFAR M. Online multi-level energy management strategy based on rule-based and optimization-based approaches for fuel cell hybrid electric vehicles[J]. *Applied sciences*, 2021, 11(9): 3849.
- [15] DU C Q, HUANG S Y, JIANG Y Y, et al. Optimization of energy management strategy for fuel cell hybrid electric vehicles based on dynamic programming[J]. *Energies*, 2022, 15(12): 4325.
- [16] SHI D P, LIU K J, WANG Y, et al. Adaptive energy control strategy of PHEV based on the Pontryagin's minimum principle algorithm[J]. *Advances in mechanical engineering*, 2021, 13(8): 1-14.
- [17] HUANGFU Y G, LI P, PANG S Z, et al. An improved energy management strategy for fuel cell hybrid vehicles based on Pontryagin's minimum principle[J]. *IEEE transactions on industry applications*, 2022, 58(3): 4086-4097.
- [18] SHAFIKHANI I, ÅSLUND J. Analytical solution to equivalent consumption minimization strategy for series hybrid electric vehicles[J]. *IEEE transactions on vehicular technology*, 2021, 70(3): 2124-2137.
- [19] CHEN Z, LIU Y G, YE M, et al. A survey on key techniques and development perspectives of equivalent consumption minimisation strategy for hybrid electric vehicles[J]. *Renewable and sustainable energy reviews*, 2021, 151: 111607.
- [20] CHOI K, BYUN J, LEE S M, et al. Adaptive equivalent consumption minimization strategy (A-ECMS) for the HEVs

- with a near-optimal equivalent factor considering driving conditions[J]. IEEE transactions on vehicular technology, 2022, 71(3): 2538-2549.
- [21] XU F G, FU Q B, SHEN T L. PMP-based numerical solution for mean field game problem of general nonlinear system[J]. Automatica, 2022, 146: 110655.
- [22] LI Y P, WANG F, TANG X L, et al. Real-time multiobjective energy management for electrified powertrains: a convex optimization-driven predictive approach[J]. IEEE transactions on transportation electrification, 2022, 8(3): 3139-3150.
- [23] SCHMID R, BUERGER J, BAJCINCA N. Energy management strategy for plug-in-hybrid electric vehicles based on predictive PMP[J]. IEEE transactions on control systems technology, 2021, 29(6): 2548-2560.
- [24] CAO K. A machine learning based approach to railway logistics transport path optimization[J]. Mathematical problems in engineering, 2022, 2022(1): 691215.
- [25] DENG K, LIU Y X, HAI D, et al. Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging[J]. Energy conversion and management, 2022, 251: 115030.
- [26] LIU W Y, FENG Q S, LI H. Optimizing passengers' experience: a goal-oriented reinforcement learning speed control approach for urban railway trains[J]. Journal of rail and rapid transit, 2024, 238(10): 1283-1294.
- [27] SHANG M Y, ZHOU Y H, FUJITA H. Deep reinforcement learning with reference system to handle constraints for energy-efficient train control[J]. Information sciences, 2021, 570: 708-721.
- [28] LI G N, OR S W, CHAN K W. Intelligent energy-efficient train trajectory optimization approach based on supervised reinforcement learning for urban rail transits[J]. IEEE access, 2023, 11: 31508-31521.
- [29] WANG J W, PINSON P, CHATZIVASILEIADIS S, et al. On machine learning-based techniques for future sustainable and resilient energy systems[J]. IEEE transactions on sustainable energy, 2023, 14(2): 1230-1243.
- [30] ZHANG J Y, TAO J L, HU Y M, et al. An energy management strategy based on DDPG with improved exploration for battery/supercapacitor hybrid electric vehicle[J]. IEEE transactions on intelligent transportation systems, 2024, 25(5): 3999-4008.
- [31] TANG X W, SHI L, ZHANG Y J, et al. Degradation adaptive energy management strategy for FCHEV based on the rule-DDPG method: tailored to the current SOH of the powertrain[J]. IEEE transactions on transportation electrification, 2025, 11(1): 993-1003.
- [32] CHEN Y, NIU Z G, HE H W, et al. Hybrid energy management optimization combining offline and online reinforcement learning for connected hybrid electric buses[J]. IEEE transactions on transportation electrification, 2025, 11(2): 6344-6354.
- [33] LIU Z E, ZHOU Q, LI Y F, et al. Safe deep reinforcement learning-based constrained optimal control scheme for HEV energy management[J]. IEEE transactions on transportation electrification, 2023, 9(3): 4278-4293.
- [34] WANG X, WANG S, LIANG X X, et al. Deep reinforcement learning: a survey[J]. IEEE transactions on neural networks and learning systems, 2024, 35(4): 5064-5078.
- [35] DJORDJEVIĆ B, FRÖIDH O, KRMAC E. Determinants of autonomous train operation adoption in rail freight: knowledge-based assessment with Delphi-ANP approach[J]. Soft computing, 2023, 27: 7051-7069.
- [36] ZHU L, CHEN C, WANG H W, et al. Machine learning in urban rail transit systems: a survey[J]. IEEE transactions on intelligent transportation systems, 2024, 25(3): 2182-2207.
- [37] LUO H L, ZHOU B N, LIU Y, et al. Characteristics of hydrogen enrichment on RNG combustion under various engine speeds in a retrofitted gas engine[J]. Process safety and environmental protection, 2024, 188: 629-642.
- [38] SCHAUL T, QUAN J, ANTONOGLOU I, et al. Prioritized experience replay[C]//International Conference on Learning Representations. San Juan, Puerto Rico: OpenReview. net, 2016: 322-355.
- [39] IQBAL N, WANG H, ZHENG Z Q, et al. Reinforcement learning-based heuristic planning for optimized energy management in power-split hybrid electric heavy duty vehicles [J]. Energy, 2024, 302: 131773.

Research on DRL-based energy management for trains

HU Enze, CHEN Lin, XU Zuming

(Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou, Hunan 412001, China)

Abstract: Diesel-electric hybrid trains represent an indispensable transitional solution for the decarbonization of heavy-haul railways. To address the challenge of achieving minimal diesel consumption while maintaining battery health under highly dynamic operating conditions, a rule-augmented deep reinforcement learning (RDRL) method was proposed. First, a high-fidelity backward simulation model of a 1.5 MW diesel-electric hybrid train was developed, in which the efficiency maps of the engine, generator, and traction motor were derived from bench tests, and the battery was characterized by an equivalent circuit model. Second, three types of prior knowledge—the engine’s optimal brake-specific fuel consumption curves, the charge – discharge characteristics of the traction battery, and the optimal action trajectories generated by offline dynamic programming (DP)—were encoded as action-space shaping and reward regularization terms. These elements were subsequently integrated into a deep deterministic policy gradient (DDPG) agent. Third, the RDRL agent was trained using prioritized experience replay and decaying Ornstein – Uhlenbeck noise to enhance learning efficiency and policy stability. Finally, the effectiveness of the proposed method was validated through simulations and semi-physical experiments. Experimental results demonstrate that, compared with the baseline DDPG algorithm and deterministic rule-based strategies, the proposed method reduces fuel consumption by 3.5% and accelerates convergence by 38.7%. Furthermore, the fuel economy achieves 94.9% of the global optimum obtained by offline DP. The framework is further validated on an autonomous rail rapid transit platform, demonstrating its strong transferability to hydrogen fuel cell hybrid systems. Overall, the proposed method offers a practical and effective solution for dynamic energy management in hybrid railway systems.

Keywords: diesel-electric hybrid trains; energy management strategy; rule-augmented deep reinforcement learning; fuel economy; prior knowledge; simulation