

DOI: 10.13718/j.cnki.xdzk.2022.03.001

基于强化学习的增程式 新能源汽车能量管理策略研究

邹博文¹, 张炳飞¹, 凌青海², 连源²,
刘俊², 杜慎之², 马华君¹

1. 西南大学 人工智能学院, 重庆 400715; 2. 重庆小康动力有限公司, 重庆 401228

摘要: 具有油、电多能源系统的增程式插电混动是新能源汽车的重要技术途径之一。为了寻找增程汽车更优的能量消耗率, 本研究以一款增程式汽车为研究对象, 探讨了该车基于规则的最优能量管理策略与基于强化学习的最优能量管理策略。在 MATLAB 软件下搭建策略模型并与 AVL/Cruise 进行联合仿真, 仿真结果表明, 在全球统一轻型车辆测试循环(WLTC)标准下, 基于强化学习优化后的能量管理策略比原先基于规则的能量管理策略能量消耗率减少了 3.2%。

关键词: 新能源; 增程式; 强化学习; 能量管理

中图分类号: U461.8

文献标志码: A

文章编号: 1673-9868(2022)03-0002-10

开放科学(资源服务)标识码(OSID):



Optimizing Range-extended Vehicle Energy Management Strategy Based on Reinforcement Learning Method

ZOU Bowen¹, ZHANG Bingfei¹, LING Qinghai²,
LIAN Yuan², LIU Jun², DU Shenzhi², MA Huajun¹

1. School of Artificial Intelligence, Southwest University, Chongqing 400715, China;

2. Chongqing Xiaokang Power Co., Ltd., Chongqing 401228, China

Abstract: The extended-range plug-in hybrid with gasoline and electric multi-energy system is one of the important technical approaches for new energy vehicles. In order to find a better energy consumption rate for range-extended vehicles, a range-extended hybrid vehicle is taken as the research object. The rule-based optimal energy management strategy and optimal energy management strategy based on reinforcement learning were studied. The strategy model was built with MATLAB software and co-simulated with

收稿日期: 2021-11-21

基金项目: 重庆市技术创新与应用发展专项资助项目(cstc2019jscx-zdztzxX0042)。

作者简介: 邹博文, 博士, 教授, 主要从事新能源汽车与智能汽车控制研究。

AVL/Cruise. The simulation results showed that the optimized energy management strategy based on reinforcement learning under the global unified light vehicle test cycle (WLTC) standard is more energy-efficient than the original rule-based energy management strategy by 3.2%.

Key words: new energy; range-extended vehicle; reinforcement learning; energy management

增程式汽车较好解决了续驶里程和污染物排放问题,是实现“碳达峰、碳中和”的重要抓手.增程式汽车的结构相对复杂且具有多个工作模式,对动力系统进行精确控制以便提升车辆的节油率是当前的研究热点^[1].

目前,增程式电动汽车的能量管理策略(Energy Management Strategy, EMS)主要分为基于规则的 EMS 和基于优化的 EMS^[2-4].常见的基于规则的 EMS 有恒温器控制、发动机多工作点控制、功率跟随控制.如 Banvait 等人提出在车辆启动时电池给发动机提供助力使其工作点落在高效区,以改善车辆燃油经济性^[5].不足之处是该方法仅适用于特定的工况下,局限性较强.基于优化的 EMS 主要有动态规划^[6-7]、庞特里亚金最小值原理^[8]、模型预测控制^[9-10]、等效能耗最小^[11]、人工智能技术等^[12-15].如 Lin 等人利用动态规划设计了最优 EMS 策略,并在 NEDC 工况下验证了该方法的有效性^[16].但该方法往往需要事先获得行驶工况信息,一般不能实现即时控制.

鉴于基于规则的 EMS 和基于优化的 EMS 都存在不足^[17-18],因此有必要探索新的方法.人工智能是模拟、延伸和扩展人的智能的一门技术科学,它擅长于解决决策、控制、优化系统中的复杂问题.本研究将其中的强化学习方法应用于增程式汽车能量管理问题形成基于强化学习的 EMS,并与基于规则的 EMS 在相同条件下进行对比仿真,以期研究该方法的节油效果.

1 增程汽车模型及基于规则的能量管理策略

1.1 增程汽车模型

增程式电动汽车按照其增程器和动力电池组的工作状态可将该车的工作模式划分为:完全纯电(驱动电机能量供给仅由电池)、发电补电(电池提供能量给驱动电机)、充电驱动(增程器提供能量给驱动电机,且为电池充电)、发电驱动(增程器只给驱动电机提供能量)、再生制动等主要混动工作模式.其能量流向图与结构图如图 1a、b 所示.

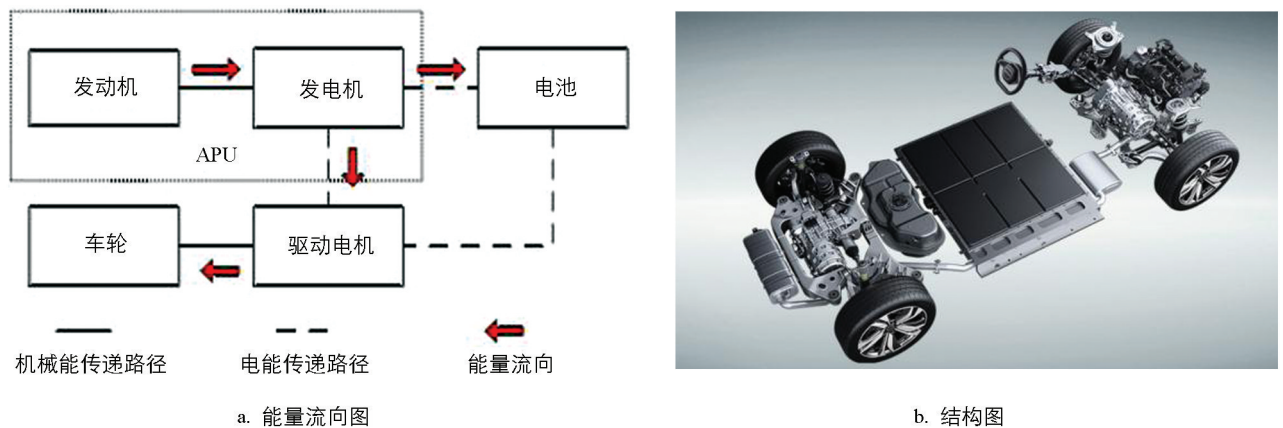


图 1 增程式插电混动汽车结构

汽车运动模型:汽车行驶时,设汽车驱动力为 F ,则汽车动力学方程可表示为:

$$F - \frac{1}{2}\rho_0 C_D A V^2 - mgk \cos\theta - mg \sin\theta = \delta ma \quad (1)$$

式中: ρ_0 是空气密度, C_D 表示空气阻力系数, A 是汽车迎风面积, V 表示车速, m 是汽车整备质量,

g 表示重力加速度, k 是汽车滑动阻力系数, a 为汽车的加速度, θ 为坡道坡度. 而汽车驱动力 F 又可表示为

$$F = \frac{T_w}{r} \quad (2)$$

式中: T_w 为增程系统输出的轮边扭矩, r 车轮半径.

发动机模型: 采用准静态模型, 定义发动机的瞬时燃油消耗 C_0 , 一段时间 T 内的燃油消耗定义为

$$C_T = \int_0^T C_0 dt \quad (3)$$

动力电池组模型: 采用内阻模型, 电池组放电功率 P_a 与电池荷电状态值 SOC 微分形式可表示如下

$$P_a = V_0 I - I^2 r \quad (4)$$

$$\dot{SOC} = -\frac{(V_0 - \sqrt{V_0^2 - 4rP_a})}{7200Qr} \quad (5)$$

式中: V_0 是电池组开路电压, r 表示电池组内阻, Q 是电池组容量, I 是电池组放电电流.

1.2 基于规则的能量管理策略

基于规则的 EMS 整体流程主要包括以下几个阶段: ① 判断是否有外接充电信号连接; ② 根据加速和制动踏板信号计算整车需求功率; ③ 若需求功率 P_{req} 大于零, 根据动力电池 SOC 值的大小, 判断整车是进入纯电动模式还是增程模式, 且增程器输出功率由当前车速与整车控制器 VCU 计算的请求功率查表; ④ 若需求功率 P_{req} 小于零, 判断是否满足再生制动条件, 从而选择再生制动或者机械制动, 具体如图 2 所示.

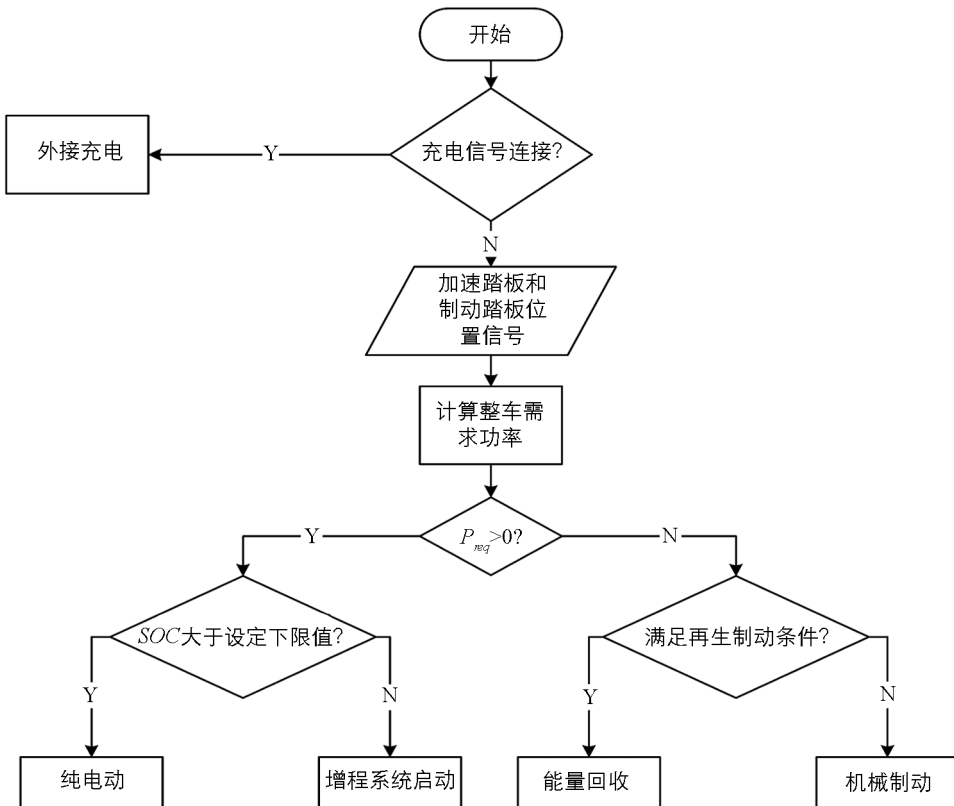


图 2 控制策略流程图

增程器若启动, VCU 根据当前车辆加速度需求、附件消耗功率、电池组当前允许充电功率等计算请求当前整车需求功率, 再结合当前车速, 查询表 1 得到请求增程系统输出功率.

表 1 请求增程系统功率(部分)

整车需求 功率/kW	车速 0 km/h 下 输出功率/kW	车速 25 km/h 下 输出功率/kW	车速 55 km/h 下 输出功率/kW	车速 75 km/h 下 输出功率/kW	车速 95 km/h 下 输出功率/kW	车速 115 km/h 下 输出功率/kW	车速 135 km/h 下 输出功率/kW
0	10	15	15	20	25	30	35
5	10	15	15	20	25	30	35
10	10	15	15	20	25	30	35
15	10	15	15	20	30	35	40
20	10	15	15	20	30	35	40
25	10	15	15	25	30	35	45
30	10	15	20	25	35	40	45
35	10	15	20	25	35	40	45
50	10	15	20	25	35	40	50
70	10	15	20	25	35	40	50
100	10	15	20	25	35	40	50

尽管该 EMS 对于特定车型直观且有效, 但其输出的增程系统功率需根据工程经验进行标定, 且不同车型需要重新设定, 会耗费许多的人力资源与时间成本. 此外, 标定的结果也存在主观性较强、精确性较差现象. 由于存在上述问题, 这里有必要尝试研究基于强化学习的能量管理策略解决这些问题.

2 基于强化学习的能量管理策略

2.1 强化学习

强化学习就是程序或智能体(agent)通过与环境不断地进行交互, 学习一个从环境到动作的映射, 学习的目标就是使累计回报最大化^[19-21], 结构如图 3 所示.

式中系统 t 时刻的状态 S_t , a_t 是智能体 t 时刻的动作, r_t 是智能体 t 时刻的奖赏值, S_{t+1} 是系统 $t+1$ 时刻的状态, r_{t+1} 是 $t+1$ 时刻的奖赏值. 强化学习方法应用场景中, 对象一般具有序惯性, 通过分析发现强化学习方法适用于增程汽车能量管理问题.

2.2 强化学习的模型对象及能量管理策略

为将强化学习方法应用于研究对象, 先简化 REEV 能量管理系统为一个非线性离散系统, 可表示为

$$S(t+1) = f(S(t), u(t)), t = 0, 1, \dots, N \quad (6)$$

式中 $S(t+1)$ 是 $t+1$ 时刻的状态值, t 表示采样时刻, N 是采样终止时刻, $S(t)$ 是 t 时刻的状态值, $u(t)$ 是 t 时刻的控制量. 状态变量 $S(t)$ 从属有限的状态空间, 控制变量 $u(t)$ 从属有限的控制空间.

2.2.1 状态

将扭矩需求 T 、当前车速 V 、动力电池荷电状态 SOC 、及行驶里程 d 选定为状态变量 S 并对所选状态值按采样时间离散, 设定扭矩需求和车速上下边界为 0 到 520 N·m、0 至 132 km/h, 行驶里程设定为行驶累计里程与总里程的比值, 其范围为 $[0, 1]$, 状态变量 S 为

$$S = [T, V, SOC, d] \quad (7)$$

2.2.2 动作

通过上面的分析, 选择增程器的输出功率 P_r 作为动作空间 A 的控制变量, 其范围为 0 到 80 kW, 动作个数按采样时间数离散.

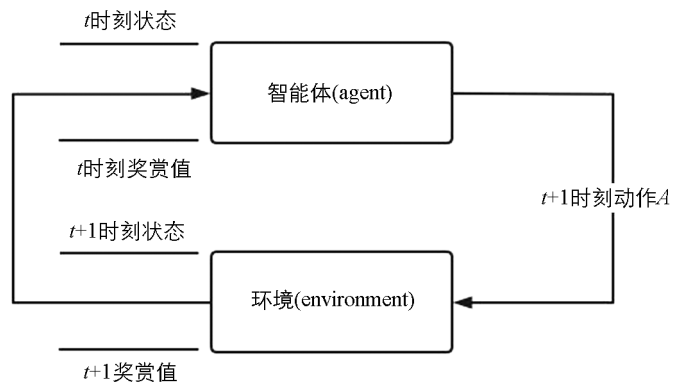


图 3 强化学习结构图

2.2.3 奖赏值

设定即时奖赏值 $r(s, a)$ 为所选时间步长内燃油消耗与电量消耗的总花费如下

$$r(s, a) = \int_{t-1}^t C_0 dt + f(\Delta soc) \quad (8)$$

式中 C_0 是前面定义的瞬时燃油消耗率, t 是采样时刻, Δsoc 是该步长内的电池组 SOC 变化量, $f(\Delta soc)$ 为电量和油耗的换算关系函数.

2.2.4 动作更新选择

本研究采用贪心策略来制定选择动作的规则, 动作更新公式如下

$$\pi(s, a) = \begin{cases} \operatorname{argmax}_a Q(s, a) & \text{以概率 } 1 - \epsilon \\ \text{随机从 } A \text{ 中选取动作} & \text{以概率 } \epsilon \end{cases} \quad (9)$$

式中 ϵ 为探索率, 本文设置的探索率会根据训练次数进行变化, A 为动作空间.

2.2.5 状态动作值更新选择

该部分状态动作值更新方式采用 Q 学习算法, 该算法的状态动作值函数 $Q(s, a)$ 直接逼近最优状态动作值函数. 状态动作值函数 $Q(s, a)$ 更新公式如下

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_t + \gamma * \max Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad (10)$$

式中 r_t 为在 t 时刻的状态选择该动作的奖赏值, $Q(s_t, a_t)$ 为在 t 时刻的状态动作值, $Q(s_{t+1}, a_{t+1})$ 为在 $t+1$ 时刻的状态动作值, γ 为学习率. 状态动作值函数的具体更新流程如图 4 所示.

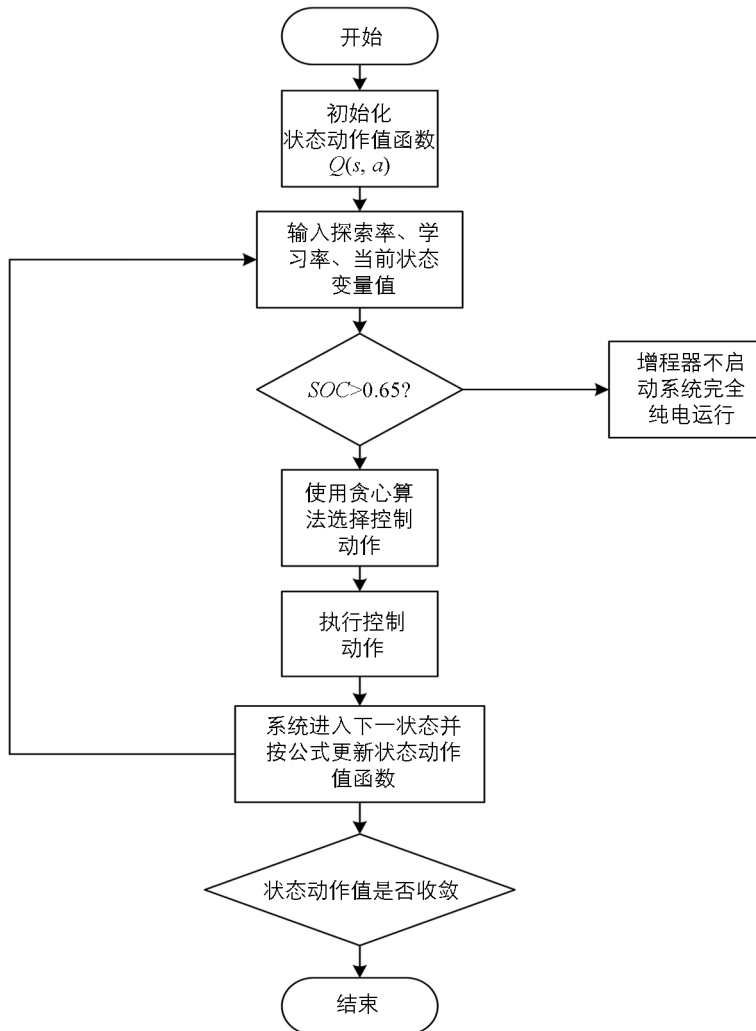


图 4 状态动作值函数更新流程图

基于强化学习的 EMS 中, 选取车速、扭矩等参数作为状态变量, 在每一个时间节点 t , 车辆产生一组控制动作 a_t 并观察即时的奖赏值 r_t , 并将一串控制动作映射到状态的函数. 考虑环境的随机输入, 如驾驶员的需求转矩及需求车速, 即时花费和下一状态都难以预测, 所以采用免模型的 Q 学习来实现能量管理控制. 对比前种方法可知于强化学习方法的 EMS 易于在仿真环境中得到控制策略, 面对不同车型的不同模型参数时也具有可扩展性.

3 仿真与结果分析

3.1 仿真车型参数与条件

为验证上述两种控制策略燃油节省的效果, 在 MATLAB/Simulink 与 AVL/Cruise 进行仿真, 选取某款增程式插混汽车为研究对象, 其部分参数如表 2 所示.

表 2 AVL/Cruise 仿真参数表

参 数	值	参 数	值
整备质量/kg	2005	滚动阻力系数/%	8.5
迎风面积/m ²	2.56	增程器功率/kW	80
风阻系数	0.32	驱动电机功率/kW	250
轮胎滚动半径/mm	366	电池组容量/kWh	35

在 AVL/Cruise 主界面搭建整车物理模型, 包括发动机、发电机、动力电池组、主减速器、车轮等组件, 并对组件按表 2 进行参数设置, 如图 5 所示.

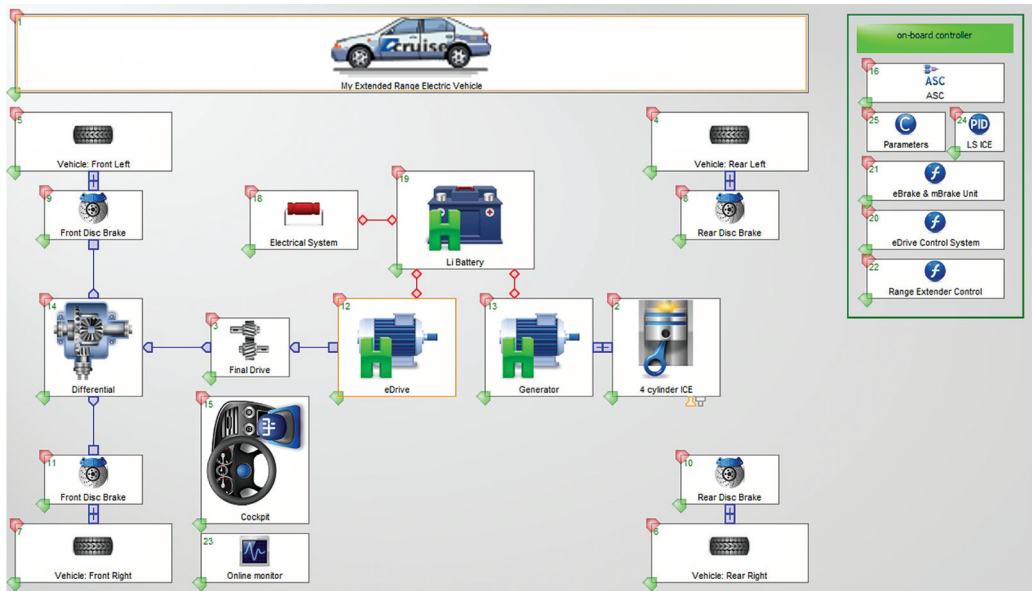


图 5 AVL/Cruise 的整车模型图

仿真选择在 WLTC(World Light Vehicle Test Cycle)工况下进行, WLTC 测试循环总共持续 1 800 s, 累计行驶里程 23.3 km, 测试当中车辆的最高时速提升至 131 km/h, 平均速度约为 47 km/h, 整个过程如图 6 所示.

3.2 能量管理策略仿真

3.2.1 基于规则的能量管理策略仿真

在 MATLAB/Simulink 中建立基于规则的整车的控制策略模型后, 将该模型嵌入到 AVL/Cruise 的整车模型中, 进行 AVL/Cruise 和 MATLAB/Simulink 的联合仿真(图 7). 即利用 Simulink 中的 Real Time Workshop 工具, 将 Simulink 模型转化成 DLL 文件, 在 AVL/Cruise 中选择 MATLAB DLL 接口模块并将生成的 DLL 文件导入. 此外, 需要在 CYCLE RUN 任务中选择 WLTC 工况作为测试工况.

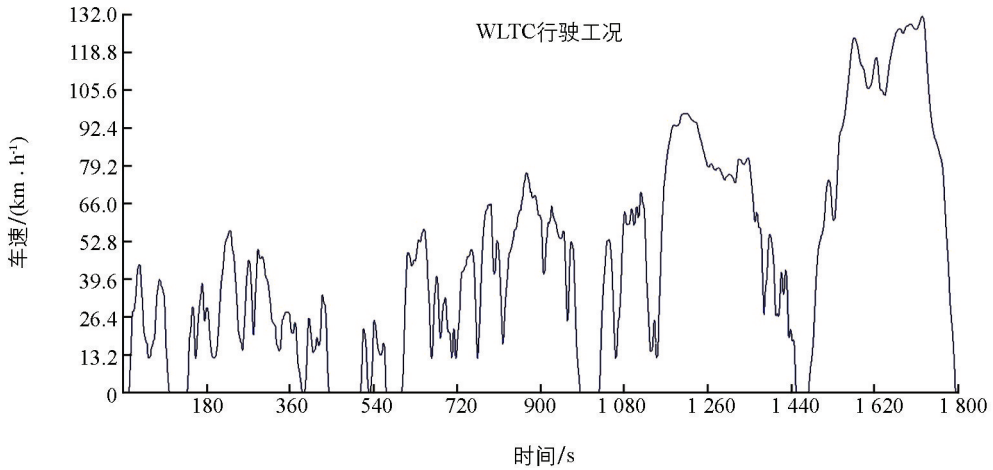


图 6 WLTC 循环工况

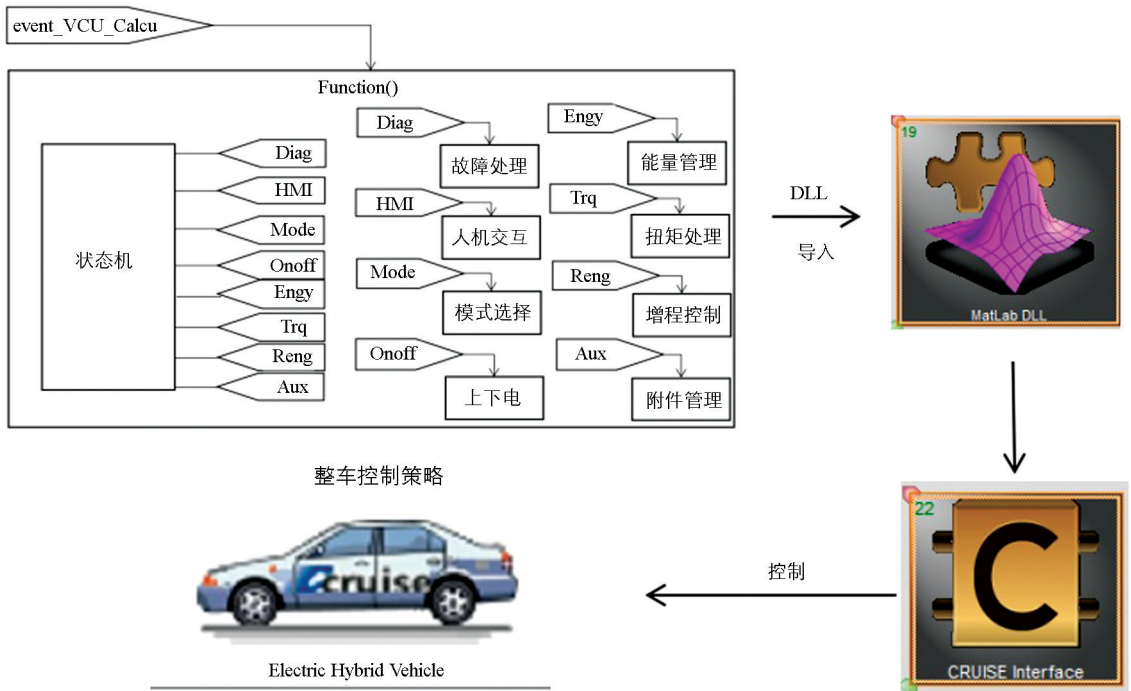


图 7 联合仿真实现

3.2.2 基于强化学习的能量管理策略仿真

如图 8 整个仿真的过程是首先将 WLTC 测试工况输入的目标车速和加速度信息传给 VCU 控制器, VCU 控制器一方面接收输入的信息和整车模型反馈的实际车速、累计里程, 另一方面将计算后得到的需求扭矩 T 、电池 SOC、车速 V 、剩余里程 d 作为状态变量输入给强化学习控制器, 强化学习控制器负责迭代训练并产生控制动作, 且会根据燃油电量消耗更新输出的控制动作, 训练完成后选择最优结果得到强化学习控制策略。

建立强化学习控制策略时, 选择 Matlab 中强化学习工具箱 (Reinforcement Learning Toolbox) 来完成, 如图 9 所示. 首先初始化 Q 值, 再通过 `rlNumericSpec` 函数对状态进行离散, 并将状态 (State)、动作 (Action) 和 Q 值存储在三维数据表格 `rlTable` 中, 并根据表格选取最优动作, 与整车 Cruise 模型交互, 并基于即时奖励函数 r , 按式 (10) 更新 Q 值. 重复上述步骤完成 Q 学习过程。

基于强化学习的控制策略需要经过一定次数的训练才能获得稳定收敛状态动作值, 在 WLTC 工况下每一次行程便是一次完整的训练. 为了加快收敛的速度, 在训练过程中并未采用固定的探索率 ϵ_t 和学习率 γ , 而是在训练的前期选择较大的探索率和学习率, 随着训练次数的增加再逐步减小, 最终获得良好的收

敛. 探索率及学习率的设定为

$$\epsilon_i = \gamma = [0.2; -0.0005; 0.005] \quad (11)$$

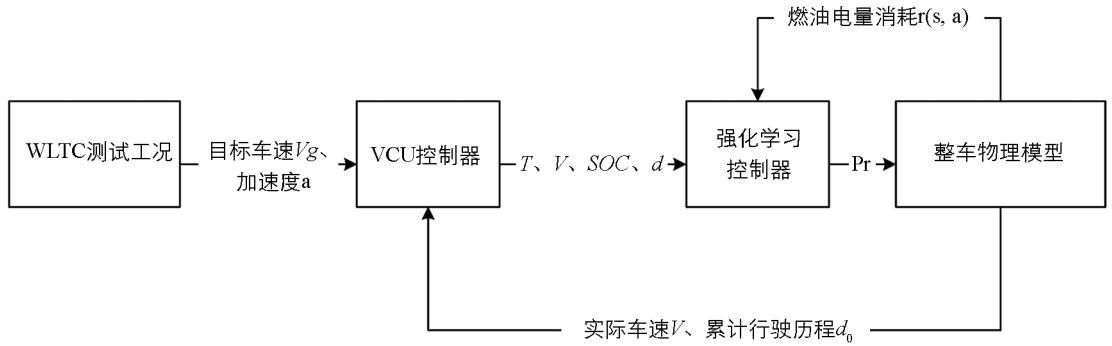


图 8 强化学习策略结构图

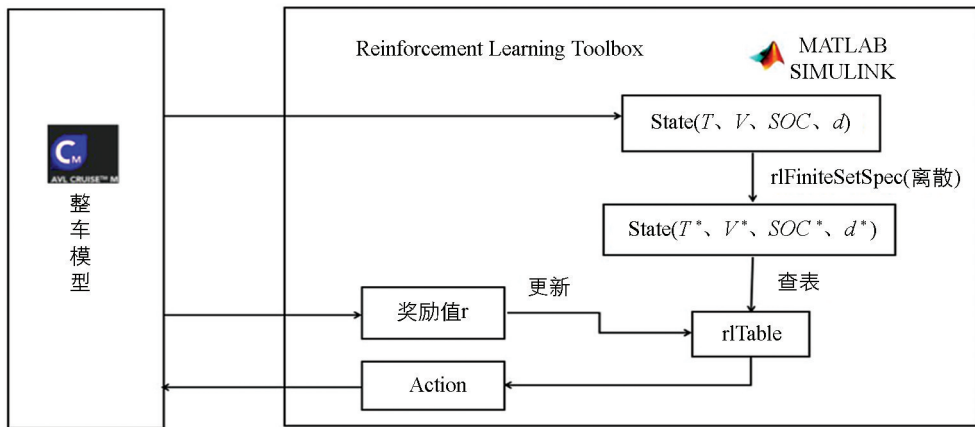


图 9 Q 学习迭代过程

由图 10 可以看出, 随着训练次数的增加, 每次行程的花费值渐渐趋于收敛, 并在第 513 次训练时有最小耗油量 0.675 L.

3.3 仿真结果分析

图 11 表现了训练完成后, WLTC 工况下使用两种能量管理策略仿真后 SOC 的轨迹. 测试的初始阶段, 因为增程器未启动, 电池组提供驱动所需功率, 而相同测试对功率需求一致, 故 SOC 轨迹在这段时间重合. 当 SOC 下降到 0.65 时, 增程器启动. 此阶段基于规则的 EMS 电池 SOC 轨迹上升的总体来说更快, 表明基于强化学习 EMS 策略在增程器启动后选择输出的功率平均值更小.

为更直观地体现燃油节省率, 下面将 WLTC 循环的电耗统一折算为油耗进行对比, 电池 SOC 值与油耗的换算公式(根据 GB/T 37340-2019《电动汽车能耗折算方法》)如下式所示:

$$F_c = E \times F_E \quad (12)$$

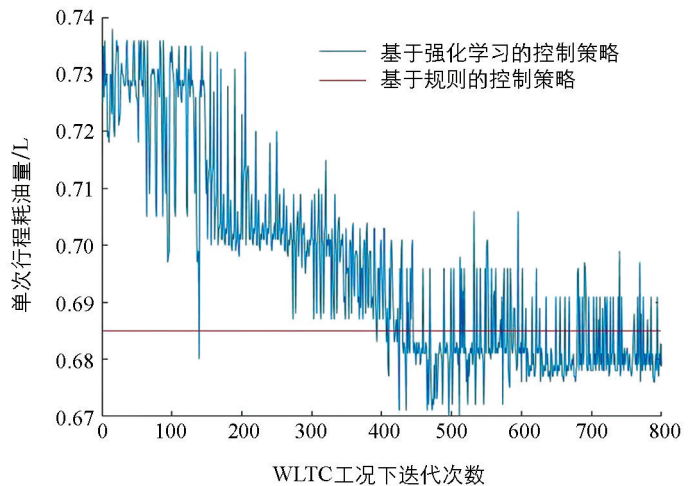


图 10 WLTC 工况下仿真迭代次数

式中： F_c 表示燃油的消耗量； E 为电能消耗； F_E 是能量折算因子，92 号汽油为 0.1161。

经训练后，基于强化学习的最优能量管理策略与基于设定规则的能量管理策略在 WLTC 工况的运行燃油消耗（电量消耗折算后）对比如图 12 所示。

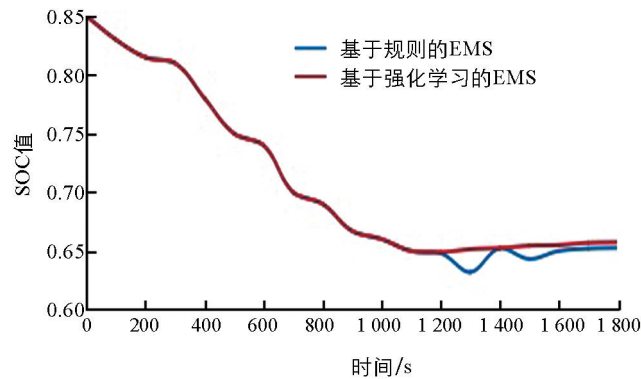


图 11 WLTC 工况下两种策略 SOC 轨迹对比

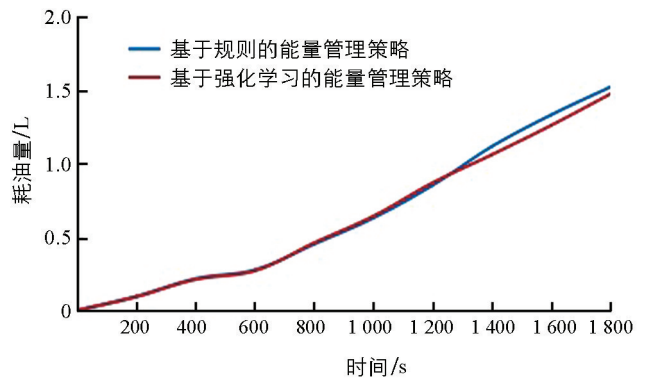


图 12 WLTC 循环工况下两种策略综合能耗对比

WLTC 工况下基于强化学习的能量管理策略与基于规则的能量管理策略能耗结果对比如表 3 所示。

表 3 WLTC 工况下能量管理策略结果对比

能量管理策略	电池 SOC 值变化	油耗/L	折算后总油耗/L	能耗节省率/%
基于规则	0.197 4	0.719 7	1.522	0
基于强化学习	0.192 5	0.696 2	1.474 4	3.2

上表中将电池组 SOC 值变化折算成油耗得到 WLTC 工况下等效油耗，可以看到：跑完整个 WLTC 循环，基于规则 EMS 油耗为 0.719 7 L，而基于强化学习 EMS 油耗为 0.696 2 L。折算后，若设基于规则的能量管理策略能耗节省率 0% 作为对比基准，基于强化学习的能量管理策略比基于规则的能量管理策略能耗节省率提高了 3.2%。

图 11 中 SOC 轨迹的变化表征了电池组充电功率的变化，也间接反映了增程器启动后两种策略下的增程器输出功率变化，计算后可知整个 WLTC 循环过程中，基于强化学习 EMS 增程器启动时输出功率平均值小于基于强化学习 EMS 增程器启动时输出功率的平均值。因油耗的来源是增程器中发动机的运行，更小的增程器输出功率意味着更低的燃油消耗。

基于规则的增程器功率输出来源于人工标定，其值设置的偏大，会导致电气传输过程中，线路电流会偏大，内阻一定时能量传递损耗的越多。而基于强化学习的 EMS 在离线多次迭代后找到了当前采样时刻下更合适的增程器功率输出值，从而有更少的能量消耗。

4 总结

根据研究对象的结构建立模型并进行仿真，结果表明基于强化学习的能量管理策略比基于规则的能量管理策略能量消耗率减少了 3.2%。因此，在 WLTC 循环工况下基于强化学习的能量管理策略是一种相对更优的能量管理策略。考虑到车辆实际运行的工况比模型要复杂，故还需进行实车测试以便检验能量分配策略的可靠性。但基于强化学习方法训练得出的 EMS 可为工程上最优增程器输出功率标定提供范围，该范围内将显著减少工程人员标定工作量。

参考文献：

- [1] SHABBIR W, EVANGELOU S A. Exclusive Operation Strategy for the Supervisory Control of Series Hybrid e Lectric Vehicles [J]. IEEE Transactions on Control Systems Technology, 2016, 24(6): 1-9.
- [2] ZHUANG W, ZHANG X, DING Y, et al. Comparison of Multi-mode Hybrid Powertrains with Multiple Planetary Gears [J]. Applied Energy, 2016, 178: 624-632.
- [3] 胡平, 张浩. 基于用户接受度的增程式混合动力汽车控制策略研究 [J]. 汽车工程学报, 2011, 1(5): 455-463.

- [4] 席利贺,张欣,孙传扬,等. 增程式电动汽车自适应能量管理策略[J]. 吉林大学学报(工学版), 2018, 38(6): 1636-1644.
- [5] BANVAIT H, ANWAR S, CHEN Y. A Rule-based Energy Management Strategy for Plug-in Hybrid Electric Vehicle (PHEV) [C]. St. Louis: American Control Conference, 2009.
- [6] PENG J, HE H, XIONG R. Rule Based Energy Management Strategy for a Series-parallel Plug-in hybrid Electric Bus Optimized by Dynamic Programming [J]. Applied Energy, 2016, 185: 1633-1643.
- [7] WANG Y, SHEN Y, YUAN X, et al. Operating Point Optimization of Auxiliary Power Unit Based on Dynamic Combined Cost Map and Particle Swarm Optimization [J]. IEEE Transactions on Power Electronics, 2015, 30(12): 7038-7050.
- [8] ONORI S, TRIBIOLI L. Adaptive Pontryagin's Minimum Principle Supervisory Controller Design for the 12plug-in Hybrid GM Chevrolet Volt [J]. Applied Energy, 2015, 147: 224-234.
- [9] 秦大同,秦岭. 基于显式随机模型预测控制的功率分流式混合动力车辆能量管理策略[J]. 华南理工大学学报(自然科学版), 2019, 47(7): 112-120.
- [10] 赵治国,王茂垚,郭元强,等. 新型功率分流混合动力系统能量管理预测优化[J]. 西安交通大学学报, 2019, 53(1): 52-61.
- [11] 林歆悠,冯其高,张少博. 等效因子离散全局优化的等效燃油瞬时消耗最小策略能量管理策略[J]. 机械工程学报, 2016, 52(20): 102-110.
- [12] LIU J, CHEN W, LI W, et al. Hybrid-trip-model-based Energy Management of a PHEV with Computation-optimized Dynamic Programming [J]. IEEE Transactions on Vehicular Technology, 2018, 67: 338-353.
- [13] 刘辉,李训明,王伟达. 基于最优功率分配因子的插电式混合动力汽车实时能量管理策略研究[J]. 机械工程学报, 2019, 55(4): 91-101.
- [14] 胡晓松,陈科坪,唐小林,等. 基于机器学习速度预测的并联混合动力车辆能量管理研究[J]. 机械工程学报, 2020, 56(16): 181-192.
- [15] FRIES M, KRUTTSCHNITT M, LIENKAMP M. Operational Strategy of Hybrid Heavy-duty Trucks by Utilizing a Genetic Algorithm to Optimize the Fuel Economy Multiobjective Criteria [J]. IEEE Transactions on Industry Applications, 2018, 54(4): 3668-3675.
- [16] LIN C, PENG H, GRIZZLE J W, et al. Power Management Strategy for a Parallel Hybrid Electric Truck [J]. IEEE Transactions on Control Systems Technology, 2004, 11: 839-849.
- [17] 杨超,杜雪龙,王伟达,等. 智能网联环境下的 PHEV 实时优化能量管理策略研究[J]. 汽车安全与节能学报, 2021, 12(2): 210-218.
- [18] 陈征,刘亚辉,杨芳. 基于进化-增强学习方法的插电式混合动力公交车能量管理策略[J]. 机械工程学报, 2017, 53(16): 86-93.
- [19] QI X, LUO Y, WU G, et al. Deep Reinforcement Learning Enabled Self-learning Control for Energy Efficient Driving [J]. Transportation Research Part C: Emerging Technologies, 2019, 99: 67-81.
- [20] 杨宁康,韩立金,刘辉,等. 基于效率优化的混合动力车辆强化学习能量管理策略研究[J]. 汽车工程, 2021, 43(7): 1046-1056.
- [21] 高阳,陈世福,陆鑫. 强化学习研究综述[J]. 自动化学报, 2004, 30(1): 86-100.

责任编辑 欧宾