

基于 TD3 的电动汽车复合电源能量管理策略研究

刘家成¹, 张向文^{1,2}

(1. 桂林电子科技大学电子工程与自动化学院, 广西 桂林 541004;
2. 广西自动检测技术与仪器重点实验室, 广西 桂林 541004)

摘要: 将蓄电池与超级电容组成复合电源系统并结合有效的能量管理策略, 能显著提高能量利用率, 延长储能系统的使用寿命。为了实现复合电源系统能耗损失的最小化, 设计了一种基于双延迟深度确定性策略梯度 (TD3) 算法的能量管理策略。与深度确定性策略梯度 (DDPG) 算法相比, 该算法解决了 Q 值过高估计问题, 能耗损失更小。利用电动汽车行驶方程式和复合电源系统等效电路模型, 搭建了基于 TD3 算法的 MATLAB/Simulink 仿真模型, 并进行测试。仿真结果显示, 所提出的能量管理策略能降低大电流对蓄电池的冲击, 与 DDPG 算法相比, 能量利用率提高了 1.36%, 蓄电池峰值电流输出降低了 14.68%, 蓄电池温升降低了 3.52%, 系统总能耗降低了 2.17%。

关键词: 电动汽车; 复合电源系统; 能量管理; 深度强化学习

中图分类号: U469.72

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202230

TD3-based energy management strategy for hybrid energy storage system of electric vehicle

LIU Jiacheng¹, ZHANG Xiangwen^{1,2}

1. School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China
2. Guangxi Key Laboratory of Automatic Detection Technology and Instruments, Guilin 541004, China

Abstract: Combining batteries and super capacitors into a composite power system (CPS) with an effective energy management strategy can significantly improve the energy utilization, and increase the service life of the energy storage system. To minimize the energy loss of the system, an energy management strategy based on the twin delayed deep deterministic policy gradient (TD3) algorithm was designed. Compared with the deep deterministic policy gradient (DDPG) algorithm, TD3 algorithm solved the problem of overestimation of Q value and less energy loss. A MATLAB/Simulink simulation model based on the TD3 algorithm was built, and tested with the electric vehicle driving equation and the equivalent circuit model of CPS. The outcomes indicate that the proposed energy management strategy can effectively reduce the impact of high current on the battery, and compared with DDPG algorithm, the energy utilization efficiency is improved by 1.36%, the peak of output current of the battery is reduced by 14.68%, the temperature rise of the battery is reduced by 3.52%, the total energy consumption of the system is reduced by 2.17%.

Key words: electric vehicle, composite power system, energy management, deep reinforcement learning

收稿日期: 2022-02-25; 修回日期: 2022-04-15

通信作者: 张向文, zxw@guet.edu.cn

基金项目: 国家自然科学基金资助项目 (No.51465100); 广西自然科学基金资助项目 (No.2018GXNSFAA281282); 广西自动检测技术与仪器重点实验室基金资助项目 (No.YQ17110); 桂林电子科技大学研究生教育创新计划资助项目 (No.2021YCXS120)

Foundation Items: The National Natural Science Foundation of China (No.51465100), The Natural Science Foundation of Guangxi (No.2018GXNSFAA281282), The Key Laboratory of Automatic Detection Technology and Instrumentation Foundation of Guangxi (No.YQ17110), Graduate Education Innovation Program of Guilin University of Electronic Technology (No.2021YCXS120)

0 引言

随着石油资源的短缺和气候的异常变化,在过去几十年里,纯电动汽车的发展受到越来越多人的关注。与燃油车相比,纯电动汽车具有零污染、噪声低等优点^[1]。蓄电池作为纯电动汽车的主要能量来源,具有能量密度高、自放电小、无记忆效应等优点^[2],但在大倍率电流充放电情况下,蓄电池容量衰减速度增大,导致蓄电池循环寿命缩短^[3-4],而超级电容具有高比功率、可以进行快速充放电等特点^[5-6]。因此,将二者组成复合电源系统,并结合有效的能量管理策略,可以充分发挥蓄电池高能量密度与超级电容高功率密度的优势。能量管理策略是其中的关键技术,目前,主要的管理策略包括基于规则的策略、基于优化的策略和基于学习的策略。

基于规则的策略包括确定性规则和模糊规则。周美兰等人^[7]研究了基于确定性规则的逻辑门限控制,该策略的主要思想是蓄电池提供汽车行驶时的平均功率,而超级电容提供剩余部分功率,从而使蓄电池的输出电流维持在一个比较安全的范围内,进而更好地保护蓄电池。这种方法简单且易于实现,具有广泛的实际应用前景,但这种方法只能用于特定的驾驶条件。胡杰等人^[8]采用基于模糊规则的模糊控制,结果表明,基于模糊控制的能量管理策略能充分发挥超级电容的优势,有效改善蓄电池的放电状态。模糊控制结构简单,对于非线性不确定系统控制具有一定的优势,但模糊规则和隶属度函数完全来自设计者的经验和专家知识,因此对未知的驾驶环境模糊控制方法缺乏适应性。

基于优化的策略主要包括动态规划、粒子群优化和模型预测控制。孙永健^[9]首先对蓄电池和超级电容进行特性分析及参数设计,其次对参数进行优化,最后利用动态规划算法对蓄电池和超级电容的功率进行分配。结果显示,超级电容的加入能减小蓄电池的输出倍率,从而可以延长蓄电池的使用寿命。虽然动态规划在理论上可以找到最优解,但其计算量大,并且需要未来驾驶循环的先验知识,不能在线实施。Chen Z Y 等人^[10]利用粒子群优化算法对其设定的规则阈值参数进行优化,结果表明,优化后的各项指标都有不同程度的提升。粒子群优化收敛速度快,但最终解可能不是最优解。Zhang S 等人^[11]利用马尔可夫模型对车辆未来一段时间范

围内的状态进行预测,并用动态规划算法对结果进行最优化求解,结果显示,模型预测控制在电动汽车的能量管理方面能起到较好的效果。虽然模型预测控制方法能实现实时优化,但是预测控制的性能在很大程度上依赖于行驶周期预测算法的准确性。

随着人工智能技术日益趋于成熟,强化学习在众多方面都扮演着重要角色^[12]。在电动汽车能量管理方面也不例外,基于强化学习的电动汽车能量管理策略受到越来越多人的关注。Xiong R 等人^[13]研究了一种基于马尔可夫概率矩阵的强化学习离线系统,根据 KL (Kullback-Leibler) 因子来判断何时触发能量管理策略的更新,并更新功率转移概率矩阵。结果表明,这种方法的结果要优于基于规则的方法,但它的性能取决于状态空间和动作空间的离散程度,并且随着状态空间和动作空间维度的增加会导致“维数灾难”。Wu J D 等人^[14]采用深度 Q 学习的能量管理策略,利用神经网络来拟合最优值函数,不仅解决了 Q 学习导致的“维数灾难”,而且可以获得比 Q 学习更好的燃油经济性。Kong H F 等人^[15]以功耗最小和蓄电池寿命最长为目标,针对超参数直接影响深度强化学习的性能,提出了一种基于贝叶斯的优化方法来优化超参数,结果表明,最优性和收敛性都得到了提高。Li W H 等人^[16]考虑了储能部件的电流和温度特性,提出了一种基于深度 Q 学习的能量管理策略。实验结果表明,该策略能最大限度地降低系统能量损失,并延长蓄电池的使用寿命。与基于 Q 学习的策略相比,基于深度 Q 学习的策略具有更好的能源经济性和更低的能量损耗,但性能会受到动作空间离散化的影响,不能保证策略的最优化。Tan H C 等人^[17]和 Wu Y K 等人^[18]提出深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 算法,并且使用优先经验回放算法提高能量管理策略的学习效率,其效果与动态规划接近。DDPG 算法虽然解决了动作空间离散问题,但该算法对 Q 值的过高估计可能会导致偏差增大和寻找到次优策略。

通过以上分析可以看出,目前 DDPG 算法具有较好的能量管理效果,但是存在 Q 值过高估计的问题。针对此问题,本文设计了一种基于双延迟深度确定性策略梯度 (twin delayed deep deterministic policy gradient, TD3) 算法的能量管理策略。TD3 算法建立了两个独立的评价网络来计算 Q 值,并将较小值作为目标 Q 值,从而避免 Q 值过高估计问

题。为了减小 Q 值函数的方差，降低了策略网络参数的更新频率，使策略网络参数比值网络参数更新慢。为了提高算法鲁棒性，在计算目标动作值函数的目标动作中加入服从截断正态分布的噪声。在能量管理策略中，以需求功率、速度、蓄电池荷电状态 (state of charge, SOC) 和超级电容 SOC 为输入，蓄电池功率为输出，能耗损失最小为奖励函数，通过学习和探索的方式获得最优的控制策略。最后，在城市道路循环工况 (urban dynamometer driving schedule, UDDS) 进行仿真，并将仿真结果与 DDPG 算法和近端策略优化 (proximal policy optimization, PPO) 算法^[19]进行对比。仿真结果表明，基于 TD3 算法的复合电源能量管理策略充分发挥了蓄电池和超级电容的优势，降低了大电流和大功率对蓄电池的冲击，延长了蓄电池的使用寿命。同时将制动回收能量优先回收给超级电容，充分发挥了超级电容充放电速度快、充放电效率高的优势，提高了能源利用效率，从而降低了出行成本，延长了续航里程，更有利于纯电动汽车的推广与应用。

1 复合电源系统建模

复合电源系统建模是进行能量管理策略设计的基础。首先介绍了复合电源拓扑结构，其次根据整车纵向动力学模型得到整车需求功率，建立蓄电池模型和超级电容模型，从而确定各自的输出功率，最后对直流-直流 (DC/DC) 变换器效率模型进行介绍。

1.1 复合电源拓扑结构

复合电源拓扑结构一般分为 3 种类型：被动并联拓扑结构、半主动拓扑结构和全主动拓扑结构^[20]。被动并联拓扑结构成本低、结构简单，但不能充分发挥超级电容的优势。半主动拓扑结构易于控制，具有较高的能量传递效率。全主动拓扑结构包含蓄电池和超级电容，成本会显著增加，控制策略会更复杂。因此，考虑成本和效率，本文选择如图 1 所示的半主动拓扑结构。超级电容先与 DC/DC 变换器串联，再与蓄电池并联，共同为负载提供需求功率 P_{req} 。蓄电池为主电源，其输出功率 P_b 直接施加到负载上，保持负载侧电压稳定，而超级电容为辅助电源，其输出功率 p_{sc} 通过 DC/DC 变换器连接到负载，向负载提供峰值功率和回收制动能量。这种拓扑结构可以使超级电容工作在更宽的电压范围内，防止峰值电流对蓄电池造成损害，从而保护蓄电池。

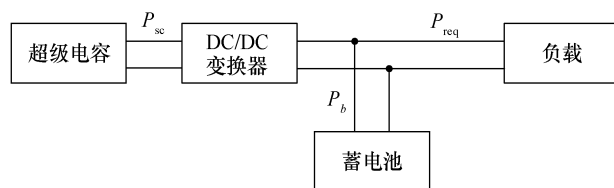


图 1 复合电源系统结构

1.2 整车纵向动力学模型

车辆行驶状态下，沿行驶方向作用于车辆的各种外力的平衡关系式就是车辆行驶方程式。外力可分为驱动力 F_t 和行驶阻力 ΣF 两大类，其中行驶阻力 ΣF 包括滚动阻力 F_f 、空气阻力 F_w 、坡度阻力 F_i 和加速阻力 F_j 。对车辆进行受力分析，可以得到车辆行驶方程式为：

$$F_t = F_f + F_w + F_i + F_j = mgf \cos \alpha + \frac{C_D A v^2}{21.15} + mg \sin \alpha + \delta m a \quad (1)$$

其中， m 为质量 (kg)， g 为重力加速度 (m/s^2)， f 为滚动阻力系数， α 为道路坡度 (rad)， C_D 为空气阻力系数， A 为迎风面积 (m^2)， v 为速度 (km/h)， δ 为旋转质量换算系数， a 为加速度 (m/s^2)。

在给定速度 v 下，由车辆行驶方程即可计算出车辆需求功率：

$$P_{\text{req}} = \frac{v}{\eta_T} \left(mgf \cos \alpha + \frac{C_D A v^2}{21.15} + mg \sin \alpha + \delta m a \right) \quad (2)$$

其中， P_{req} 为整车需求功率 (kW)， η_T 为传动系统效率。

1.3 蓄电池模型

鉴于单体模型在后期串联中使用，计算量大，以二阶电阻-电容 (resistance-capacitance, RC) 模型为蓄电池模型，该模型具有两个 RC 并联回路环节，如图 2 所示，其中， U_{oc} 为开路电压， R_0 为欧姆内阻， R_1 与 C_1 分别为电化学极化内阻与电容， R_2 与 C_2 分别为浓差极化内阻与电容， I_b 为电流， U_b 为端电压。

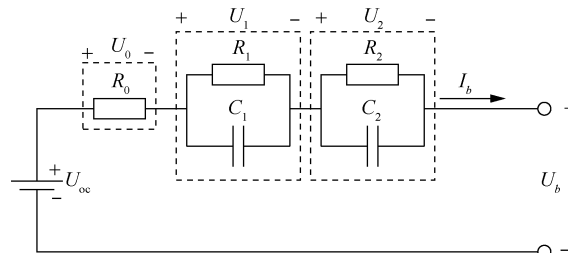


图 2 蓄电池等效电路模型

由基尔霍夫电压定律，端电压为：

$$U_b = U_{oc} - U_0 - U_1 - U_2 \quad (3)$$

则蓄电池的输出功率为:

$$P_b = (U_{oc} - U_0 - U_1 - U_2) I_b \quad (4)$$

蓄电池的输出电流为:

$$I_b = C_1 \frac{dU_1}{dt} + \frac{U_1}{R_1} = C_2 \frac{dU_2}{dt} + \frac{U_2}{R_2} \quad (5)$$

蓄电池 SOC 为:

$$SOC = SOC_0 - \frac{\int \eta_b I_b dt}{3600 Q_t} \quad (6)$$

其中, SOC_0 为蓄电池的初始 SOC, η_b 为蓄电池充放电效率, Q_t 为蓄电池的总容量 (Ah)。

1.4 超级电容模型

本文选用的超级电容等效电路模型如图 3 所示。图 3 中, V_{oc} 为开路电压, C_{sc} 为理想电容, R_{sc} 为等效串联内阻, I_{sc} 为电流, U_{sc} 为负载电压。

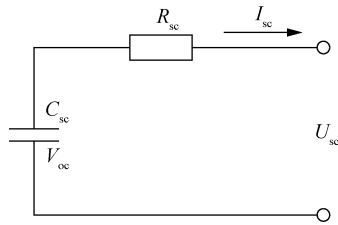


图3 超级电容等效电路模型

根据基尔霍夫电压定律, 超级电容的负载电压与开路电压的关系为:

$$U_{sc} = V_{oc} - I_{sc} R_{sc} \quad (7)$$

其中, 开路电压的计算式为:

$$V_{oc} = \frac{Q_{sc}}{C_{sc}} \quad (8)$$

其中, Q_{sc} 为超级电容存储的能量, 其计算式为:

$$Q_{sc} = \int_0^t I_{sc} dt \quad (9)$$

则超级电容输出功率为:

$$P_{sc} = (V_{oc} - I_{sc} R_{sc}) I_{sc} \quad (10)$$

由式 (10) 可得超级电容输出电流为:

$$I_{sc} = \frac{V_{oc} - \sqrt{V_{oc}^2 - 4P_{sc}R_{sc}}}{2R_{sc}} \quad (11)$$

超级电容的 SOC 定义为:

$$SOC_{sc} = \frac{Q_{remain}}{Q_{total}} = \frac{V_{oc} - U_{min}}{U_{max} - U_{min}} \quad (12)$$

其中, Q_{remain} 为超级电容剩余电量, Q_{total} 为超级电容存储的总电量, U_{max} 、 U_{min} 分别为超级电容允许的最大、最小工作电压。

1.5 DC/DC 变换器模型

DC/DC 变换器结构复杂, 构建其电路模型有一定的难度, 同时计算量大, 仿真难以实现。而本文关注的是复合电源的能量和功率特性, 因此选用 DC/DC 变换器的效率模型, 其效率如图 4 所示。效率表达式如下:

$$\eta_{dcdc} = f\left(\frac{U_b}{U_{sc}}, P_{sc}\right) \quad (13)$$

其中, $f\left(\frac{U_b}{U_{sc}}, P_{sc}\right)$ 是蓄电池与超级电容电压比和输出功率的查表函数。

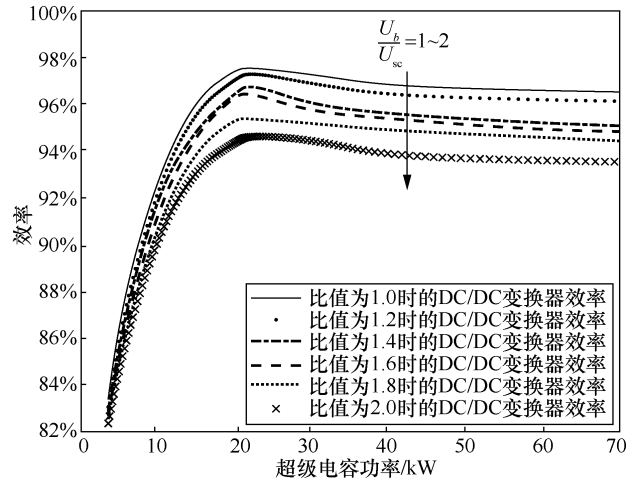


图4 蓄电池与超级电容不同电压比值下的 DC/DC 变换器效率

2 基于 TD3 的能量管理策略

强化学习算法是能量管理策略的重要组成部分, 下面对其工作原理进行阐述。首先介绍强化学习基本原理, 其次对 Q 学习算法、DDPG 算法和 TD3 算法的原理进行介绍, 最后建立基于 TD3 算法的能量管理策略。

2.1 强化学习基本原理

强化学习通过智能体与环境的交互, 利用不断的试错学习, 以累计奖励最大化为目标, 获得最优策略^[21]。强化学习算法主要包括环境和智能体两部分, 在每一个采样时间点 $t=0,1,2,3,\dots,n$, 智能体从环境获取状态 $s_t (s_t \in S)$, 结合其当前策略选择当前最优的动作 $a_t (a_t \in A)$, 并作用于环境, 从而迫使环境状态从当前状态转换为下一状态 s_{t+1} , 并返回奖励值 r_{t+1} 给智能体。智能体根据反馈的奖励值, 判断其策略的好坏, 从而更好地指导其策略改进的方向, 周而复始, 最终能学习到最优控制策略。

2.2 Q 学习算法

Q 学习算法作为一种典型的强化学习算法，近年来已被应用于复合电源系统的能量管理。例如，参考文献[13]中提出了基于 Q 学习算法的能量管理策略，实验表明，此策略比基于规则的策略消耗的能量更少。在 Q 学习中，采用离散化查询表来表示所有可能的状态-动作组合的动作值，此动作值表示智能体所采取动作的良好程度，并将学习进程引导到期望的方向。然而，随着状态空间和动作空间的增加，Q 学习算法容易陷进“维数灾难”。

2.3 DDPG 算法

DDPG 算法创建了策略网络 Actor 和评价网络 Critic 双网络结构，并结合了深度 Q 学习的经验回放^[22]，可用于解决连续状态和动作控制问题。Critic 网络损失函数 L 被定义为：

$$L(\theta_Q) = \frac{1}{M} \sum_{i=1}^M [y_i - Q(s_i, a_i | \theta_Q)]^2 \quad (14)$$

其中， θ_Q 为 Critic 当前网络的参数， M 为从经验回放缓冲区中选择的学习样本个数， y_i 为 Critic 目标网络的 Q 值， Q 为 Critic 当前网络。

y_i 的计算如下：

$$y_i = r_i + \gamma Q'[s_{i+1}, \mu'(s_{i+1} | \theta_{\mu'}) | \theta_{Q'}] \quad (15)$$

其中， r_i 为即时奖励值， γ 为折扣因子。 Q' 与 $\theta_{Q'}$ 分别为 Critic 目标网络及其参数， μ' 与 $\theta_{\mu'}$ 分别为 Actor 目标网络及其参数。

通过动作值函数将当前状态映射到指定的动作，并通过梯度反向传播算法更新 Actor 网络参数。其损失梯度为^[21]：

$$\nabla_{\theta_{\mu}} J \approx \frac{1}{M} \sum_{i=1}^M [\nabla_a Q(s_i, a | \theta_Q) |_{a=\mu(s_i | \theta_{\mu})} \cdot \nabla_{\theta_{\mu}} \mu(s_i | \theta_{\mu})] \quad (16)$$

其中， ∇ 为梯度， θ_{μ} 为 Actor 当前网络的参数。 μ 为 Actor 当前网络。

最后，Critic 和 Actor 目标网络参数 $\theta_{Q'}$ 和 $\theta_{\mu'}$ 可以通过指数平滑的方法来更新，即：

$$\theta_{Q'} = \tau \theta_Q + (1 - \tau) \theta_{Q'} \quad (17)$$

$$\theta_{\mu'} = \tau \theta_{\mu} + (1 - \tau) \theta_{\mu'} \quad (18)$$

其中， τ 为软更新因子，由于参数 $\tau \ll 1$ ，目标网络变化平稳且缓慢，从而提高学习的稳定性。

由于 DDPG 算法遵循与深度 Q 网络相同的更新 Q 值的方法，即按 $y = r + \gamma \max Q(s_{i+1}, a_{i+1})$ 更新，

容易给某些行动带来 Q 值过高的问题，导致偏差增大和寻找到次优策略。此外，DDPG 对超参数的设置非常敏感，这会直接影响其收敛稳定性。

2.4 TD3 算法

针对上述缺陷，本文采用最新的 TD3^[23] 算法，构建了电动汽车复合电源能量管理策略。

首先，TD3 模仿了双 Q 学习的思想，建立了两个独立的 Critic 网络来计算目标值^[23]，即：

$$\begin{cases} y_1 = r + \gamma Q'_1(s_{i+1}, \tilde{a} | \theta_{Q'_1}) \\ y_2 = r + \gamma Q'_2(s_{i+1}, \tilde{a} | \theta_{Q'_2}) \end{cases} \quad (19)$$

其中， y_1 、 y_2 分别为 Critic 目标网络 1 和 Critic 目标网络 2 的目标值， Q'_1 和 $\theta_{Q'_1}$ 分别为 Critic 目标网络 1 及其参数， Q'_2 和 $\theta_{Q'_2}$ 分别为 Critic 目标网络 2 及其参数， $\tilde{a}$ 为加了正态分布噪声的动作。

选取两个目标值中较小的值为目标 Q 值，从而可以避免 Q 值估计过高的问题，并将其代入贝尔曼方程，计算损失函数，即：

$$y_i = r_i + \gamma \min_{k=1,2} Q'_k(s_{i+1}, \tilde{a} | \theta_{Q'_k}) \quad (20)$$

$$L(\theta_{Q_k}) = \frac{1}{M} \sum_{i=1}^M [y_i - Q_k(s_i, a_i | \theta_{Q_k})]^2 \quad (21)$$

其中， Q_k 和 θ_{Q_k} 分别为 Critic 当前网络及其参数。

其次，目标网络为学习过程提供一个稳定的目标，有助于提高收敛性能。然而，当 Critic 网络误差较大时，可能会导致策略出现离散状况。因此，为了防止算法出现不收敛的情况，Actor 网络参数更新速度比 Critic 网络慢，从而导致值函数的更新方差越小，得到的策略越好。

最后，为了避免过拟合，必须对 Q 值的计算进行平滑处理，因此，在每个动作中加入服从截断正态分布的噪声，对目标更新修改如下^[23]：

$$\tilde{a} \leftarrow \mu'(s_{i+1} | \theta_{\mu'}) + \varepsilon, \varepsilon \sim \text{clip}(N(0, \sigma), -c, c), c > 0 \quad (22)$$

其中， ε 为服从截断正态分布的噪声， σ 为方差， c 为裁剪幅度。

2.5 基于 TD3 算法的能量管理策略

由于智能体需要从状态中获取环境知识，状态的选择对于 TD3 能量管理策略的成功至关重要。在本文中，选取蓄电池和超级电容的 SOC 为状态值。此外，还将整车的总需求功率和车速作为另外两种状态来表示车辆的动力学特性。基于 TD3 的能量管理策略的整个状态空间可以表示为：

$$S = \{P_{\text{req}}, v, \text{SOC}_b, \text{SOC}_{\text{sc}}\} \quad (23)$$

其中, SOC_b 、 SOC_{sc} 分别是蓄电池和超级电容的 SOC。

由于能量管理策略的主要任务是在满足需求功率前提下, 充分发挥蓄电池与超级电容二者各自的优势, 因此将蓄电池提供的功率占总需求功率的比值作为控制动作, 即 $A = \{P_b/P_{req}\}$ 。

基于 TD3 的能量管理策略的目标是在系统能量损失最小的情况下, 获得一系列最优策略。为了总奖励累计的最大化, 本文将即时奖励定义为:

$$r = -\left\{\sum_{k=0}^{T-1} [L_b(k) + L_{sc}(k) + L_{dcdc}(k)]\right\} \quad (24)$$

其中, T 为典型循环工况长度, 瞬时损耗包括蓄电池能量损耗 $L_b(k)$ 、超级电容能量损耗 $L_{sc}(k)$ 和 DC/DC 变换器损耗 $L_{dcdc}(k)$ 。各部分损耗可通过下式计算:

$$\begin{cases} L_b(k) = I_b^2(k)R_b(k) \\ L_{sc}(k) = I_{sc}^2(k)R_{sc}(k) \\ L_{dcdc}(k) = P_{sc}(k)[1 - \eta_{dcdc}(k)]\eta_{dcdc}^{-z}(k) \end{cases} \quad (25)$$

其中, $\eta_{dcdc}(k)$ 为 DC/DC 变换器的效率; $P_{sc}(k)$ 为超级电容输出功率; z 为逻辑值, 充电时设为 1, 放电时设为 0。

最后, 设计基于 TD3 算法的能量管理策略流程如图 5 所示, TD3 算法的伪代码如算法 1 所示。

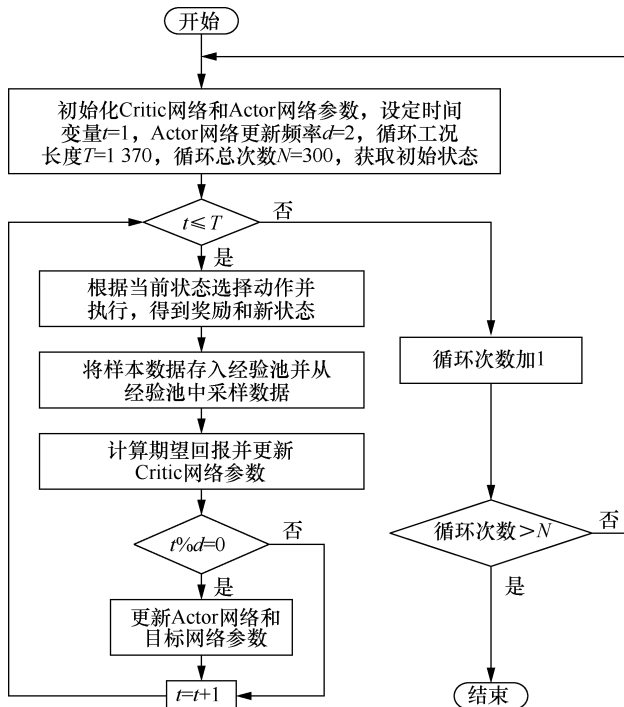


图5 基于 TD3 算法的能量管理策略流程

算法1 TD3 算法的伪代码

for episode = 1:N do

初始化参数 θ_{Q_k} 、 θ_{μ} , 目标网络参数 $\theta_{Q_k'} \leftarrow \theta_{Q_k}$, $\theta_{\mu'} \leftarrow \theta_{\mu}$, 经验池 B , 样本学习个数 M , 时间变量 t , Actor 网络更新频率 d , 循环工况长度 T , 循环次数 N , 获取初始状态 $s_1: P_{req1}$, v_1 , SOC_{b1} , SOC_{sc1} 。

for $t = 1:T$ do

根据当前状态 s_t , 选择动作:

$$a_t \sim \mu(s_t | \theta_{\mu}) + \varepsilon, \varepsilon \sim N(0, \sigma)$$

执行动作 a_t , 得到奖励 r_t 和新状态 s_{t+1} 。

将四元组 (s_t, a_t, r_t, s_{t+1}) 存至经验池 B 并从中采样 M 个样本数据 (s_i, a_i, r_i, s_{i+1}) ,

$i=1, 2, 3, \dots, M$ 。

计算动作的期望回报:

$$\tilde{a} \leftarrow \mu'(s_{i+1} | \theta_{\mu'}) + \varepsilon, \varepsilon \sim \text{clip}(N(0, \sigma), -c, c), c > 0$$

$$y_i = r_i + \gamma \min_{k=1,2} Q'_k(s_{i+1}, \tilde{a} | \theta_{Q_k'})$$

更新 Critic 网络参数:

$$L(\theta_{Q_k}) = \frac{1}{M} \sum_{i=1}^M [y_i - Q_k(s_i, a_i | \theta_{Q_k})]^2$$

if $t \% d == 0$ then

更新 Actor 网络参数:

$$\nabla_{\theta_{\mu}} J \approx \frac{1}{M} \sum_{i=1}^M [\nabla_a Q(s_i, a | \theta_Q)_{a=\mu(s_i | \theta_{\mu})}]$$

$$\nabla_{\theta_{\mu}} \mu(s_i | \theta_{\mu})$$

更新目标网络参数:

$$\theta_{Q_k'} = \tau \theta_{Q_k} + (1 - \tau) \theta_{Q_k'}$$

$$\theta_{\mu'} = \tau \theta_{\mu} + (1 - \tau) \theta_{\mu'}$$

end if

end for

end

3 仿真结果和讨论

为了验证所提基于 TD3 算法的能量管理策略 (以下简称 TD3 控制策略) 的有效性, 搭建 MATLAB/Simulink 模型进行仿真分析, 并将其与单一电源、基于 DDPG 算法的能量管理策略 (以下简称 DDPG 控制策略) 和基于 PPO 算法的能量管理策略 (以下简称 PPO 控制策略) 做对比。TD3 算法主要超参数见表 1。仿真所用循环工况选用比较常用的 UDDS, UDDS 是美国国家环境保护局为轻型

表1 TD3 算法主要超参数

参数名称	描述	参数值
Actor 网络层参数	各连接层神经元数量	(4,100,100,1)
Critic 网络层参数	各连接层神经元数量	(5,100,100,1)
折扣因子	用于计算长期折扣奖励	0.95
Actor 网络学习率	策略梯度更新步长	0.001
Critic 网络学习率	Q 函数网络梯度下降更新步长	0.001
优化算法类型	用于梯度下降	Adam
经验池尺寸	用于样本存储	10^4
样本学习个数	一次批量梯度下降中的样本数量	64
软更新因子	目标网络参数更新速度	0.001
策略频率	用于延迟策略更新	2
策略噪声	添加在目标动作中的正态分布方差	0.2
裁剪幅度	指定裁剪的高斯噪声范围	(-0.5,0.5)
循环工况长度	所用 UDDS 长度	1 370
循环次数	训练次数	300

汽车设计的城市循环工况^[24]。一个完整的 UDDS 的循环时间为 1 370 s, 行驶里程为 11.99 km, 车辆行驶最高车速为 91.25 km/h, 平均车速为 31.51 km/h。UDDS 时间与车速的关系如图 6 所示, 对应的 UDDS 下功率需求曲线如图 7 所示。

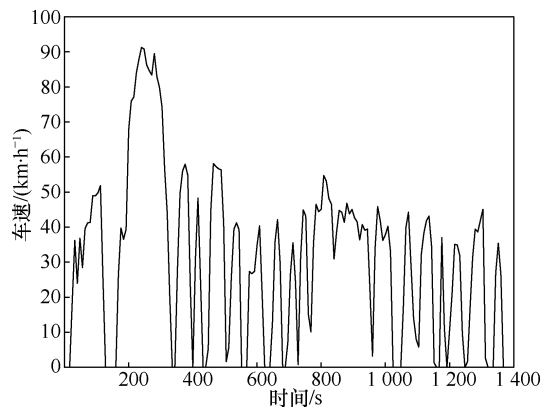


图6 UDDS 时间与车速的关系

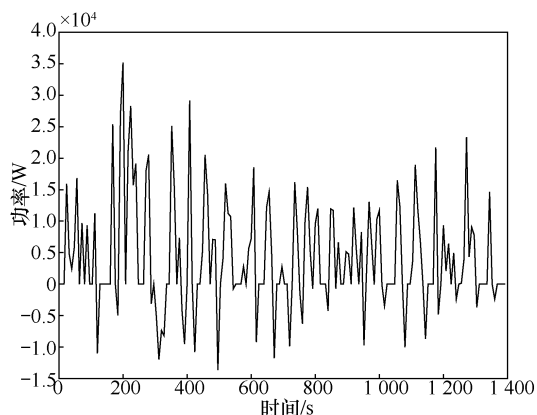


图7 UDDS 下功率需求曲线

仿真得到蓄电池 SOC 变化曲线如图 8 所示。从图 8 可以看出, 与单一电源相比, 复合电源中蓄电池 SOC 曲线下降速率明显降低, 整个循环工况结束后, 复合电源系统中蓄电池 SOC 在 DDPG 控制策略、PPO 控制策略和 TD3 控制策略下分别提高了 3.59%、4.29%和 5.00%。相比于 DDPG 控制策略和 PPO 控制策略, 蓄电池 SOC 在 TD3 控制策略下分别提高了 1.36%和 0.68%。这说明超级电容的利用率提高, 蓄电池的放电状态得到改善, 蓄电池的能量消耗降低, 从而可以延长蓄电池的使用寿命。

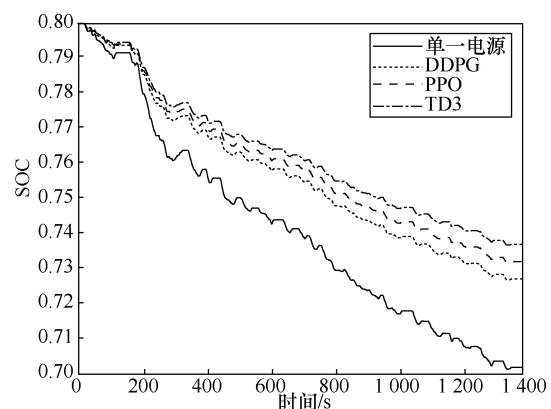


图8 蓄电池 SOC 变化曲线

蓄电池功率对比曲线如图 9 所示。从图 9 可以看出, 蓄电池作为单一电源时, 其充放电功率有很大一部分超过了 15 kW, 甚至在某些时刻放电功率达到了 35 kW 左右, 此时, 蓄电池的放电功率远远超出其放电范围, 会对蓄电池造成一定程度的损害, 而复合电源系统中的蓄电池的放电功率大部分都在 15 kW 以

内。相比于 DDPG 控制策略和 PPO 控制策略, 基于 TD3 控制策略的蓄电池输出功率较小, 使得蓄电池功率在较小范围内波动, 从而可以更好地保护蓄电池。当需求功率较大的时候, 大功率由超级电容提供, 而剩余的较低功率则由蓄电池提供, 从而降低对蓄电池的损害。当汽车制动时, 由于超级电容比能量较低, 回收的大部分能量优先由超级电容吸收, 从而提高能源的使用效率。

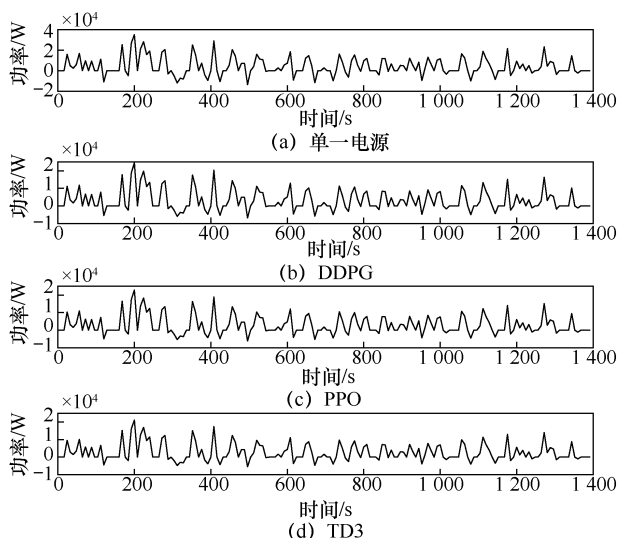


图9 蓄电池功率对比曲线

单一电源和复合电源蓄电池电流对比曲线如图 10 所示。从图 10 可以明显看到, 相比于单一电源, 复合电源系统中的蓄电池在 DDPG 控制策略、PPO 控制策略和 TD3 控制策略下, 其电流大小明显降低, 峰值电流降低显著。单一电源的蓄电池放电电流最高可达到 107.61 A, 不利于蓄电池健康。而复合电源系统中采用 DDPG 控制策略、PPO 控制策略和 TD3 控制策略的蓄电池峰值电流分别为 74.23 A、68.77 A 和 63.33 A, 分别降低了 31.02%、36.09% 和 41.15%。与 DDPG 控制策略和 PPO 控制策略相比, 蓄电池峰值电流在 TD3 控制策略下分别降低了 14.68% 和 7.91%。超级电容输出功率曲线如图 11 所示。从图 11 可以明显看到, 在大部分时间里, 超级电容的输出功率在 TD3 控制策略下要高于 DDPG 控制策略和 PPO 控制策略, 从而使得蓄电池功率减小, 更有利于蓄电池的健康状况, 这表明 TD3 控制策略的控制效果优于 DDPG 控制策略和 PPO 控制策略。

蓄电池温度对比曲线如图 12 所示, 可以看到, 蓄电池为单一电源时, 温度上升速度最快, 而复合

电源在控制策略下温度上升速度都有所下降, 相比于 DDPG 控制策略和 PPO 控制策略, 蓄电池温升在 TD3 控制策略下分别降低了 3.52% 和 2.04%。由此可见, 加入超级电容作为辅助电源并加以有效的能量管理策略, 可以延长蓄电池的使用寿命。

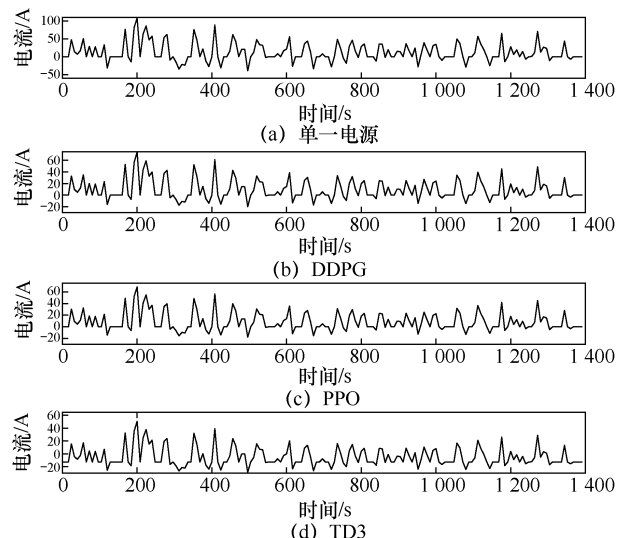


图 10 单一电源和复合电源蓄电池电流对比曲线

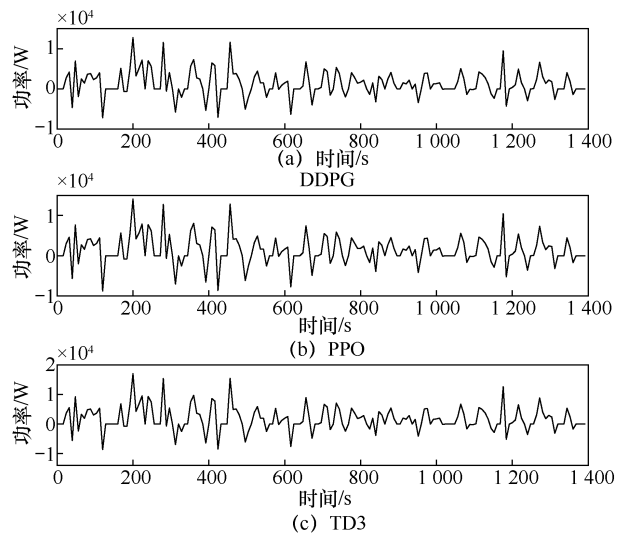


图 11 超级电容输出功率曲线

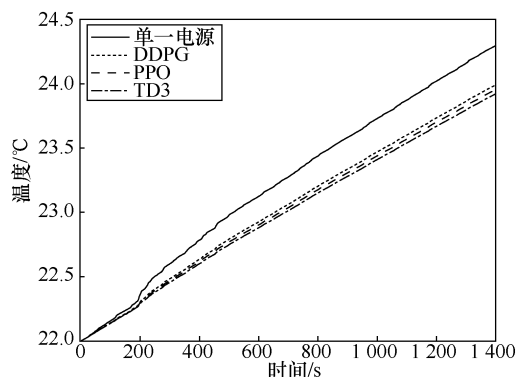


图 12 蓄电池温度对比曲线

超级电容 SOC 变化曲线如图 13 所示。从图 13 可以看出,相比于 DDPG 控制策略和 PPO 控制策略,基于 TD3 控制策略的电容 SOC 曲线下落速率更快。在仿真结束后,DDPG 控制策略的电容 SOC 剩余量为 0.6606, PPO 控制策略的电容 SOC 剩余量为 0.6361,而 TD3 控制策略的电容 SOC 剩余量为 0.5418,说明 TD3 控制策略下的超级电容为蓄电池承担了更多的功率分配,充分发挥了其“削峰填谷”作用。

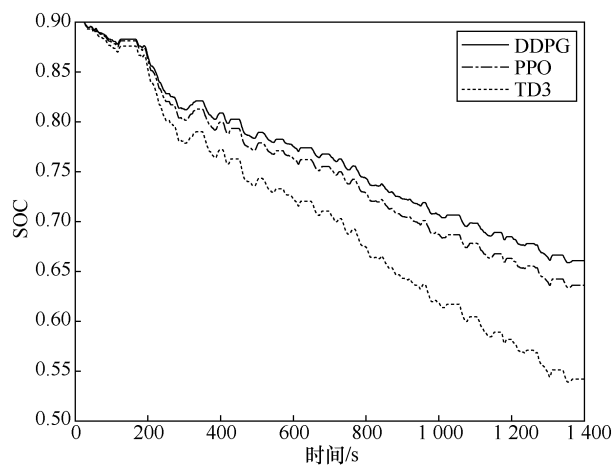


图 13 超级电容 SOC 变化曲线

系统总能耗仿真结果如图 14 所示。从图 14 可以看出,相比于单一电源,复合电源的系统总能耗在 DDPG 控制策略、PPO 控制策略和 TD3 控制策略下分别降低了 12.71%、13.57%和 14.60%。相比于 DDPG 控制策略和 PPO 控制策略,TD3 控制策略下系统能耗分别降低了 2.17%和 1.19%。

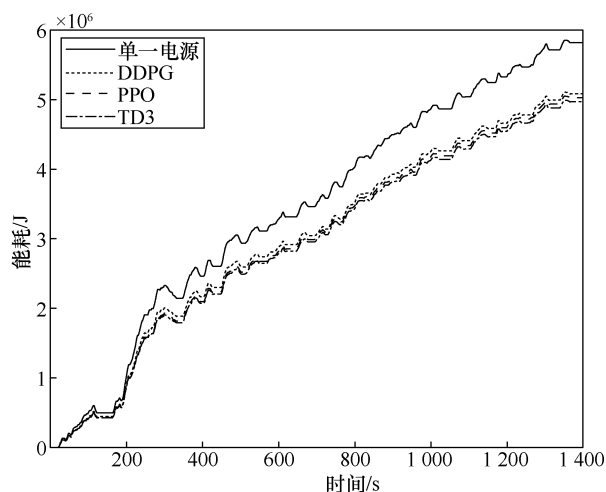


图 14 系统总能耗仿真结果

计算整个循环工况结束后不同控制策略的结果,具体见表 2。

表 2 不同控制策略结果

性能指标	单一电源	DDPG	PPO	TD3
蓄电池 SOC 剩余量	0.7019	0.7271	0.7320	0.7370
蓄电池峰值电流/A	107.61	74.23	68.77	63.33
蓄电池温升/℃	2.30	1.99	1.96	1.92
超级电容 SOC 剩余量	—	0.6606	0.6361	0.5418
系统总能耗/J	5.82×10^6	5.08×10^6	5.03×10^6	4.97×10^6

从表 2 可以看出,相比于单一电源,复合电源的蓄电池各项性能指标和系统总能耗在 DDPG 控制策略、PPO 控制策略和 TD3 控制策略下都有不同程度的改善。与单一电源相比,蓄电池 SOC 剩余量在 DDPG 控制策略和 PPO 控制策略下分别提高了 3.59%和 4.29%,峰值电流分别降低了 31.02%和 36.09%,温升分别降低了 13.48%和 14.78%,系统总能耗分别降低了 12.71%和 13.57%。而相比于 DDPG 控制策略和 PPO 控制策略,TD3 控制策略的蓄电池 SOC 分别提高了 1.36%和 0.68%,峰值电流分别降低了 14.68%和 7.91%,温升分别降低了 3.52%和 2.04%,超级电容 SOC 剩余量分别降低了 17.98%和 14.82%,系统总能耗分别降低了 2.17%和 1.19%。这说明 TD3 控制策略可以更充分地发挥超级电容功率密度高、充放电效率高的优势,从而更好地保护蓄电池,降低电动汽车的行驶成本。

4 结束语

本文针对基于 DDPG 算法的能量管理策略存在对 Q 值过高估计的问题,设计了基于 TD3 算法的能量管理策略。TD3 算法建立了两个独立的评价网络计算 Q 值,并将较小值作为目标 Q 值,从而解决对 Q 值过高估计问题。通过仿真对比实验得到以下结论。

① 在满足车辆动力性的前提下,相比于 DDPG 控制策略和 PPO 控制策略,基于 TD3 算法的能量管理策略可以充分发挥超级电容充放电速度快并输出较大功率的优势,将蓄电池峰值电流分别降低了 14.68%和 7.91%,温升分别降低了 3.52%和 2.04%,能更好地保护蓄电池。

② 相比于 DDPG 控制策略和 PPO 控制策略,

基于 TD3 算法的能量管理策略将蓄电池 SOC 剩余量分别提高了 1.36% 和 0.68%，有效地改善了蓄电池的放电状态，从而保护蓄电池，进而延长储能系统的工作寿命，提高系统的能量利用率。

③ 采用的复合电源系统及提出的基于 TD3 算法的能量管理策略充分利用了超级电容快速充放电及充放电效率高的特点，回收了更多的制动能量，相比于 DDPG 控制策略和 PPO 控制策略，TD3 控制策略使系统总能耗分别降低了 2.17% 和 1.19%，降低了整车行驶成本。

参考文献:

- [1] 崔淑梅, 宋贝贝, 王志远. 电动汽车动态无线供电电磁耦合机构研究综述[J]. 电工技术学报, 2022, 37(3): 537-554.
CUI S M, SONG B B, WANG Z Y. Overview of magnetic coupler for electric vehicles dynamic wireless charging[J]. Transactions of China Electrotechnical Society, 2022, 37(3): 537-554.
- [2] 张文爽, 李健, 余文涛, 等. 基于新一代电源控制器的锂电池在轨管理分析[J]. 电源技术, 2022, 46(1): 109-112.
ZHANG W S, LI J, YU W T, et al. Analysis and verification of in-orbit management of lithiumion battery based on the new generation of power control unit[J]. Chinese Journal of Power Sources, 2022, 46(1): 109-112.
- [3] 靳聪, 高申, 孙世光, 等. 基于无损检测的生命周期锂离子动力电池安全性分析[J]. 储能科学与技术, 2019, 8(2): 259-263.
JIN C, GAO S, SUN S G, et al. Analysis of safety performance of lithium-ion power battery during life cycle based on non-destructive testing[J]. Energy Storage Science and Technology, 2019, 8(2): 259-263.
- [4] 魏东涛, 黄之杰, 孔华, 等. 蓄电池 SOC 的研究及预测方法[J]. 电源技术, 2016, 40(6): 1321-1323.
WEI D T, HUANG Z J, KONG H, et al. Battery SOC research and prediction methods[J]. Chinese Journal of Power Sources, 2016, 40(6): 1321-1323.
- [5] ZHAI C J, LUO F, LIU Y G. A novel predictive energy management strategy for electric vehicles based on velocity prediction[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 12559-12569.
- [6] ZHANG L J, YE X M, XIA X H, et al. A real-time energy management and speed controller for an electric vehicle powered by a hybrid energy storage system[J]. IEEE Transactions on Industrial Informatics, 2020, 16(10): 6272-6280.
- [7] 周美兰, 冯继峰, 张宇. 纯电动汽车复合储能系统及其能量控制策略[J]. 电机与控制学报, 2019, 23(5): 51-59.
ZHOU M L, FENG J F, ZHANG Y. Composite energy storage system and its energy control strategy for electric vehicles[J]. Electric Machines and Control, 2019, 23(5): 51-59.
- [8] 胡杰, 刘迪, 杜常清, 等. 电动汽车复合能源系统能量管理策略研究[J]. 机械科学与技术, 2020, 39(10): 1606-1614.
HU J, LIU D, DU C Q, et al. Study on energy management strategy of hybrid energy storage system for electric vehicles[J]. Mechanical Science and Technology for Aerospace Engineering, 2020, 39(10): 1606-1614.
- [9] 孙永健. 某型纯电动客车复合电源系统优化设计与控制[D]. 长春: 吉林大学, 2017.
SUN Y J. The optimization design and control of hybrid energy storage system for A certain type electric bus[D]. Changchun: Jilin University, 2017.
- [10] CHEN Z Y, XIONG R, CAO J Y. Particle swarm optimization-based optimal power management of plug-in hybrid electric vehicles considering uncertain driving conditions[J]. Energy, 2016, 96: 197-208.
- [11] ZHANG S, XIONG R, SUN F C. Model predictive control for power management in a plug-in hybrid electric vehicle with a hybrid energy storage system[J]. Applied Energy, 2017, 185: 1654-1662.
- [12] VÁZQUEZ-CANTELI J R, NAGY Z. Reinforcement learning for demand response: a review of algorithms and modeling techniques[J]. Applied Energy, 2019, 235: 1072-1089.
- [13] XIONG R, CAO J Y, YU Q Q. Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle[J]. Applied Energy, 2018, 211: 538-548.
- [14] WU J D, HE H W, PENG J K, 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.
- [15] KONG H F, YAN J P, 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.
- [16] LI W H, CUI H, NEMETH T, et al. Deep reinforcement learning-based energy management of hybrid battery systems in electric vehicles[J]. Journal of Energy Storage, 2021, 36: 102355.
- [17] TAN H C, ZHANG H L, PENG J K, et al. Energy management of hybrid electric bus based on deep reinforcement learning in continuous state and action space[J]. Energy Conversion and Management, 2019, 195: 548-560.
- [18] WU Y K, TAN H C, PENG J K, 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.
- [19] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[J]. arXiv preprint, 2017, arXiv:1707.06347.
- [20] SONG Z Y, HOU J, HOFMANN H, et al. Sliding-mode and Lyapunov function-based control for battery/supercapacitor hybrid energy storage system used in electric vehicles[J]. Energy, 2017, 122: 601-612.

- [21] QI X W, WU G Y, BORIBOONSOMSIN K, et al. Data-driven reinforcement learning-based real-time energy management system for plug-in hybrid electric vehicles[J]. Transportation Research Record: Journal of the Transportation Research Board, 2016, 2572(1): 1-8.
- [22] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[J]. arXiv preprint, 2015, arXiv:1509.02971.
- [23] FUJIMOTO S, VAN HOOF H, MEGER D. Addressing function approximation error in actor-critic methods[J]. arXiv preprint, 2018, arXiv:1802.09477.
- [24] 丁镇涛, 邓涛, 李志飞, 等. 基于安时积分和无迹卡尔曼滤波的锂离子电池 SOC 估算方法研究[J]. 中国机械工程, 2020, 31(15): 1823-1830.
- DING Z T, DENG T, LI Z F, et al. SOC estimation of lithium-ion battery based on ampere hour integral and unscented Kalman filter[J]. China Mechanical Engineering, 2020, 31(15): 1823-1830.

[作者简介]



刘家成(1995-), 男, 桂林电子科技大学电子工程与自动化学院硕士生, 主要研究方向为汽车电子、电动汽车能量管理。



张向文(1976-), 男, 博士, 桂林电子科技大学电子工程与自动化学院研究员, 主要研究方向为电动汽车能量管理、动力电池管理系统、电动汽车再生制动控制。