

有轨电车基于工况识别的强化学习 能量管理策略研究

莫浩楠 杨中平 林飞 王珣 安星锟

(北京交通大学电气工程学院 北京 100044)

摘要 储能式混合动力有轨电车以储能系统作为唯一动力源,对能量管理策略进行优化设计,可以提高有轨电车的运行性能及经济效益。将有轨电车的需求功率看做马尔科夫过程,且为避免驾驶工况变化较大时对能量管理策略的影响,提出基于工况识别的强化学习能量管理策略。通过历史行驶数据构建有轨电车驾驶工况并得到不同工况下的马尔科夫功率状态转移矩阵。然后以混合储能系统能耗最小为目标,通过强化学习算法得到不同工况下的功率分配策略。最后,以改进的学习向量量化(LVQ)神经网络对当前的驾驶工况进行实时识别,控制系统通过当前识别的工况以及列车状态做出实时决策。以实车数据进行仿真验证,优化后的策略可以降低储能系统损耗,并且可应用于不同的驾驶工况中。以90kW混合储能平台进行实验验证,验证了该策略在实际工程应用中的可行性。

关键词: 有轨电车 混合储能系统 工况识别 强化学习 能量管理策略

中图分类号: TM 921

Research on Reinforcement Learning Energy Management Strategy of Tram Based on Condition Identification

Mo Haonan Yang Zhongping Lin Fei Wang Yu An Xingkun

(School of Electrical Engineering Beijing Jiaotong University Beijing 100044 China)

Abstract Energy storage hybrid trams use the energy storage system as the only power source, and optimize the design of energy management strategy, which can improve the running performance and economic benefits of the tram. Regarding the demand power of the tram as a Markov process, and in order to avoid the impact on the energy management strategy when the driving conditions change greatly, a reinforcement learning energy management strategy based on the recognition of the operating conditions is proposed. Based on historical driving data, the tram driving conditions are constructed and the Markov power state transition matrix under different operating conditions is obtained. Then, with the goal of minimizing the energy consumption of the hybrid energy storage system, the power allocation strategy under different working conditions is obtained through the reinforcement learning algorithm. Finally, the improved learning vectorization (LVQ) neural network is used to recognize the current driving conditions in real time, and the control system makes real-time decisions based on the current recognized conditions and train status. Real vehicle data is used for simulation verification. The optimized strategy can reduce the energy storage system loss and can be applied to different driving conditions. Experimental verification with a 90kW hybrid energy storage platform verifies the feasibility of the strategy in practical engineering applications.

Keywords: Tram, hybrid energy storage system, operating condition recognition, reinforcement learning, energy management strategy

0 引言

随着城市建设规模的不断扩大,人们对轨道交通的需求越来越高,环境保护意识也不断增强。由于有轨电车相比于电动客车其具有载客量大、运行速度快、污染少等优点。因此大力发展有轨电车将会缓解城市交通压力,提高出行效率^[1]。其中车载储能的供电方式改善了城市景观以及提高了出行的安全性^[2],然而单一的储能元件难以满足有轨电车在运行时能量和功率的全部需求,需要结合高密度和高功率密度的储能元件进行混合储能。

混合储能系统需要提供有轨电车全部牵引能量以满足其在运行过程中的速度需求,并且能够充分吸收再生制动能量,实现能量循环利用。其中能量管理策略的设计至关重要,其首要任务是在满足驾驶员需求功率前提下,实现对动力源功率优化分配,使整车性能最佳。然而,实际驾驶工况的不确定性和扰动性极大地增加了能量管理策略的设计难度。为此,开发高效、适应性强的能量管理策略是目前研究的关键问题。

近年来,针对混合储能系统的能量管理策略有多种,主要分为两大类:基于规则的能量管理策略以及基于优化的能量管理策略。其中,基于规则的能量管理策略要包括分为逻辑阈值法^[3-4]、比例法^[5]、模糊控制法^[6-7]等。主要通过系统需求功率以及储能系统的状态进行判断,从而做出实时的功率分配。该类方法由于控制逻辑简单、实用性强、实时响应快,在工程领域受到了广泛的应用。但控制逻辑的设定主要依据设计人员的工程经验,无法实现优化控制,若规则设置不合理,则会降低系统的控制效果。

基于优化的能量管理策略主要有全局优化^[8-10]以及实时优化^[15-21]的策略。其中全局优化的算法需提前获知整个行驶工况,计算量大,无法直接应用于实时控制系统。常见的做法是根据离线优化的结果提取相应的规则,转化为在线的能量管理策略。或者通过结合驾驶工况识别和模型预测控制等技术提高能量管理策略的适用性。文献[8,9]针对动态规划优化结果无法实时应用的问题,从优化结果提取相应的规则,提出新的基于规则的能量管理策略。这种以离线指导在线的方法未考虑驾驶条件变化较大时对能量管理策略产生的影响。文献[11,12]采用神经网络对驾驶工况进行实时识别,结合全局优化算法的结果来提取功率分配规则并存储于控制模块

中以供不同工况选择。这种方法虽然对不同工况下的能量管理策略进行了优化,但其优化结果受限,不能实现准确控制。为了使得能量管理策略更加适应当前驾驶工况,文献[13,14]引入了模型预测算法,将整个行驶工况内的全局最优问题转化为预测时域内的局部优化问题,通过滚动优化不断更新预测时域内未来行驶状态,获得优化结果。但这种方法很大程度上依赖对未来工况预测的精准性,并且不能得到全局最优解。

目前,机器学习、深度学习等人工智能算法应用到了混合储能实时能量管理策略中。强化学习是一种机器学习算法,在机器人控制、交通运输和运筹学等领域有着广泛的应用^[15]。文献[15-18]将强化学习应用到了求解插电式混合动力车辆的能量管理问题。通过强化学习离线优化结果对储能系统进行实时控制,并较好的优化了燃油经济性。强化学习对工况变化具有一定的鲁棒性,但它的优化性能只有在具有相似工况中才能得到保证。文献[19-21]引入了 Kullback-Leibler 偏异率,通过实时递归算法更新需求功率状态转移概率矩阵,开发基于强化学习的在线能量管理策略,使其更加适应当前驾驶工况,提高了能量管理策略的适用性和鲁棒性。

由于有轨电车启停频繁且系统功率等级高,通过在线更新强化学习策略的方法在有轨电车中受到限制。本文针对有轨电车的驾驶工况以及大功率应用场合,提出了基于工况识别的强化学习能量管理策略。

本文首先对有轨电车混合储能系统进行介绍,根据实车驾驶历史数据构建了有轨电车驾驶工况。然后通过强化学习算法得到了不同驾驶工况下的能量管理策略,并采用改进的 LVQ 神经网络对当前的驾驶工况进行识别,有轨电车根据当前识别的工况选择相应的最优控制表实时做出决策。最后采用实车运行数据进行仿真与实验,验证了该策略的有效性及可行性。

1 有轨电车储能系统建模与工况构建

1.1 混合储能系统建模

对于混合储能系统而言,不同的储能元件通过不同的储能方式连接到公共直流母线上,通过双向 DC-DC 变换器用于控制不同的电压等级之间的功率流动。由于钛酸锂电池安全性高、能量密度以及功率密度兼具,超级电容充放电速度快、功率密度高且循环寿命长。为此,本文采用钛酸锂电池和超

级电容进行混合储能，由戴维宁定理可知，储能系统串并联以后其等效电路模型依然不变，系统拓扑结构如图 1 所示。

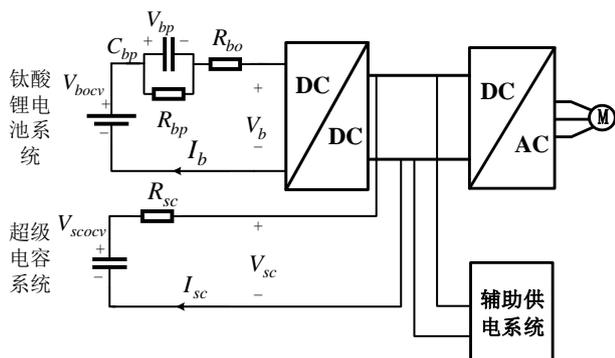


图 1 混合储能系统拓扑结构

Fig.1 Topology of hybrid energy storage system

该拓扑钛酸锂电池系统通过 DC-DC 并联到直流母线，超级电容系统直接并联到直流母线。钛酸锂电池功率可通过 DC-DC 直接控制，可以通过控制电池的充放电电流，延长其寿命周期，适用于超级电容为主要供电的场合。其中，钛酸锂电池系统使用模型为一阶模型，采用 Arbin 单体测试仪测得其单体开路电压 V_{bocv} 、内阻 R_{cbo} 、极化电阻 R_{cbp} 随荷电状态 (State of Charge, SOC) 变化曲线如图 2 所示。

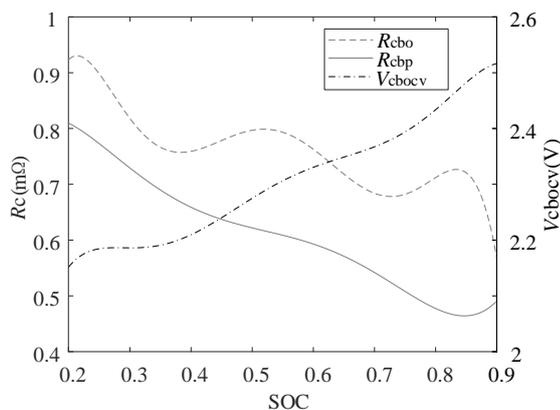


图 2 R_{cbo} 、 R_{cbp} 和 V_{bocv} 随 SOC 的变化曲线

Fig.2 R_{cbo} , R_{cbp} and V_{bocv} VS. SOC

锂电池系统离散化的电流、极化电压、外电压和时间常数表达式为

$$\begin{cases} I_b(k) = \frac{P_b(k)}{V_b(k)} \\ V_{bp}(k) = V_{bp}(k-1)e^{-\frac{\Delta t(k)}{\tau(k)}} + R_{bp}[SOC(k)]I_b(k)(1 - e^{-\frac{\Delta t(k)}{\tau(k)}}) \\ V_b(k) = V_{bocv}[SOC(k)] - I_b(k)R_{bo}[SOC(k)] - V_{bp}(k) \\ \tau(k) = R_{bp}[SOC(k)]C_{bp}[SOC(k)] \end{cases} \quad (1)$$

式中， k 为离散时间常数； I_b 表示锂电池系统电流； Δt 为时间间隔； τ 为时间常数； R_{bp} 、 C_{bp} 、 R_{bo} 表示电池系统的极化电阻、极化电容、欧姆内阻； V_{bocv} 、 V_b 、 V_{bp} 分别表示电池系统开路电压、端电压以及极化电压； P_b 表示锂电池系统端的功率。

本文使用安时积分法计算钛酸锂电池系统荷电状态 SOC，表达式如下：

$$SOC_b(k) = SOC_b(k-1) - \frac{I_b(k)\Delta t(k)}{Q_0} \quad (2)$$

式中， Q_0 为钛酸锂电池系统总电荷量。

超级电容系统采用模型为内阻阻值不变的串电阻模型。超级电容系统离散化的开路电压、外电压和电流表达式为：

$$\begin{cases} V_{scov}(k) = V_{scov}(k-1) - \frac{I_{sc}(k)\Delta t(k)}{C_{sc}} \\ V_{sc}(k) = V_{scov}(k) - I_{sc}(k)R_{sc} \\ I_{sc}(k) = \frac{V_{scov}(k) - \sqrt{V_{scov}(k)^2 - 4R_{sc}P_{sc}(k)}}{2R_{sc}} & P_{sc}(k) > 0 \\ I_{sc}(k) = -\frac{V_{scov}(k) - \sqrt{V_{scov}(k)^2 + 4R_{sc}P_{sc}(k)}}{2R_{sc}} & P_{sc}(k) < 0 \end{cases} \quad (3)$$

式中， V_{scov} 、 V_{sc} 、分别为超级电容系统的开路电压与端电压； I_{sc} 为超级电容系统的电流； R_{sc} 为超级电容系统内阻； C_{sc} 为超级电容系统容值； P_{sc} 为超级电容系统端的功率。

采用电压平方的关系表达超级电容系统的 SOC，表达式如下：

$$SOC_{sc}(k) = \left(\frac{V_{scov}(k)}{V_{scovmax}}\right)^2 \quad (4)$$

式中， $V_{scovmax}$ 为超级电容系统的额定电压。

1.2 有轨电车驾驶工况构建

本文提取武汉东湖线有轨电车连续运行 10 天的数据，包括工作日与周末。该数据从有轨电车车载监测装置中获取，记录有轨电车自出库至回库的全部运行状态，采样频率为 1Hz。有轨电车运行时间为 6:30-22:00，发车间隔约为 6min，运行时长为 15.5h。采用短行程分析法对有轨电车驾驶工况进行构建。不同于非轨道交通车辆，有轨电车的运行路径固定，且在运行时需遵循发车间隔、限速等要求，有较为明显的规则性。定义有轨电车起点和终点加速度和速度都为 0 的时刻之间的行程为短行程，如图 3 所示。将数据处理后得到 2782 个短行程。

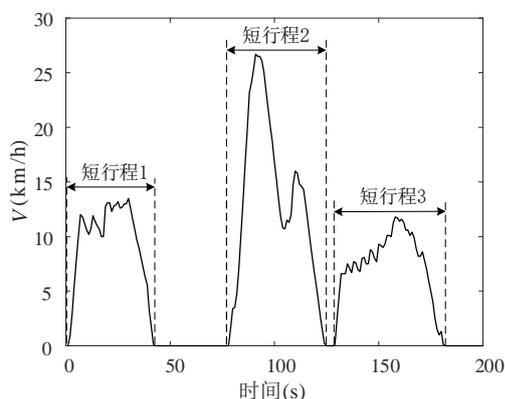


图3 短行程分割示意图

Fig.3 Schematic diagram of short stroke division

提取每段短行程的特征参数,选取 15 个重要特征参数来反应有轨电车短行程的运行信息,如表 1 所示。定义低速 $v_{low} \in (0, 20]$ km/h, 中速 $v_{mid} \in (20, 40]$ km/h, 高速 $v_{high} > 40$ km/h。

表 1 代表性特征值

Tab.1 Representative eigenvalue

特征参数	单位	特征参数	单位
平均行驶速度 v_{ave}	km/h	减速度标准差 a_{d-sd}	m/s^2
最大车速 v_{max}	km/h	加速时间比 P_a	%
平均加速度 a_{a-ave}	m/s^2	匀速时间比 P_c	%
最大加速度 a_{a-max}	m/s^2	减速时间比 P_d	%
平均减速度 a_{d-ave}	m/s^2	低速时间比值 P_{low}	%
最大减速度 a_{d-max}	m/s^2	中速时间比值 P_{mid}	%
速度标准差 v_{sd}	km/h	高速时间比值 P_{high}	%
加速度标准差 a_{a-sd}	m/s^2		

为了消除单位不同带来的影响,将原始特征参数进行标准化处理,并采用主成分分析法对标准化后的数据进行降维处理以减小数据重叠。选择累积贡献率达到 80% 以上前 4 个主成分代表原始变量。主成分贡献率如表 2 所示。

表 2 主成分贡献率

Tab.2 Principal component contribution rate

主成分	特征值	方差百分比/%	累积贡献率/%
M1	6.570	43.797	43.797
M2	2.555	17.036	60.832
M3	2.206	14.709	75.541
M4	1.198	7.985	83.527

根据东湖线有轨电车的行驶状况,可以将短行

程片段划分为 3 类,采用 K 均值聚类进行分析,其聚类结果及各工况特征参数如表 3 所示。

表 3 工况聚类结果及特征值

Tab.3 Working condition clustering results and eigenvalues

特征参数	工况 1	工况 2	工况 3
聚类数目	400	1263	1119
v_{ave} (km/h)	29.12	19.08	12.7
v_{max} (km/h)	54.14	40.36	23.75
a_{a-ave} (m/s^2)	0.47	0.40	0.32
a_{a-max} (m/s^2)	1.14	1.13	0.86
a_{d-ave} (m/s^2)	-0.34	-0.32	-0.22
a_{d-max} (m/s^2)	-0.85	-0.89	-0.61
v_{sd} (km/h)	18.69	11.82	7.44
a_{a-sd} (m/s^2)	0.32	0.31	0.25
a_{d-sd} (m/s^2)	0.21	0.22	0.16
P_a (%)	39.5	42.8	40.03
P_c (%)	7.5	3.8	4.17
P_d (%)	53	53.4	55.8
P_{low} (%)	37.28	60.78	86.18
P_{mid} (%)	28.8	34.62	13.46
P_{high} (%)	33.92	4.6	0.36

由表 3 可以看出,三种工况特征参数区分明显,对各工况特征参数进行分析。第一类工况数目最少,平均运行速度、平均加速度和减速度绝对值、高速运行时间占比最高,称为高速工况。第三类工况平均运行速度、平均加速度和减速度绝对值最低,低速运行时间占比高达 86.18%,称为低速工况。工况二特征参数介于工况一和工况三之间,称为中速工况。

2 基于工况识别的强化学习能量管理策略

强化学习 (Reinforcement Learning, RL) 是一种自适应最优控制方法,其目的是在观察和分析系统行为的基础上,通过试错学习来作出最优决策以改进系统性能,得到的策略是状态到动作的映射,与时间无关,因而可以直接应用到在线控制。对于混合动力有轨电车而言,运行过程中的牵引制动功率是一个随机状态,这样一个随机过程满足马尔可夫性质,即从一个状态转移到另一状态的概率只与

当前系统所处状态有关,与之前系统所处状态无关。因此,混合动力有轨电车的能量管理问题本质上是一个马尔可夫决策过程(MDP)。

2.1 目标函数及约束条件

有轨电车能量管理问题可以表示为在约束范围内最小化目标函数的约束优化问题,目标函数选择储能系统的能量损失。

$$\begin{aligned} J &= \sum P_{\text{loss}}(k) \Delta t(k) \\ &= \sum [P_{\text{bloss}}(k) + P_{\text{scloss}}(k) + P_{\text{dcloss}}(k)] \Delta t(k) \end{aligned} \quad (5)$$

式中,目标函数 J 包括储能系统和 DC-DC 转换器在整个时间跨度内的功率损耗。其中, P_{loss} 为储能系统总损耗, P_{bloss} 为电池内阻损耗, P_{scloss} 为超级电容内阻损耗, P_{dcloss} 为 DC-DC 损耗。各系统损耗通过公式 6 来计算。

$$\begin{cases} P_{\text{scloss}}(k) = I_{\text{sc}}(k)^2 \times R_{\text{sc}} \\ P_{\text{bloss}}(k) = I_{\text{b}}(k)^2 R_{\text{bo}} [\text{SOC}(k)] + \frac{V_{\text{bp}}(k)^2}{R_{\text{bp}} [\text{SOC}(k)]} \\ P_{\text{dcloss}}(k) = P_{\text{b}}(k) \times \left(\frac{1}{\eta_{\text{dcdc}}} - 1 \right) \quad P_{\text{b}}(k) > 0 \\ P_{\text{dcloss}}(k) = |P_{\text{b}}(k)| \times (1 - \eta_{\text{dcdc}}) \quad P_{\text{b}}(k) < 0 \end{cases} \quad (6)$$

式中, η_{dcdc} 表示 DC-DC 转换器的效率。

为了保证储能系统工作在合理范围内,能量管理问题受到以下约束条件的限制。

$$\begin{cases} \text{SOC}_{\text{bmin}} \leq \text{SOC}_{\text{b}}(k) \leq \text{SOC}_{\text{bmax}} \\ \text{SOC}_{\text{scmin}} \leq \text{SOC}_{\text{sc}}(k) \leq \text{SOC}_{\text{scmax}} \\ I_{\text{bmin}} \leq I_{\text{b}}(k) \leq I_{\text{bmax}} \\ I_{\text{scmin}} \leq I_{\text{sc}}(k) \leq I_{\text{scmax}} \end{cases} \quad (7)$$

式中, SOC_{bmax} 和 SOC_{bmin} 分别表示电池系统荷电状态的上限和下限。 $\text{SOC}_{\text{scmax}}$ 和 $\text{SOC}_{\text{scmin}}$ 分别表示超级电容系统荷电状态的上下限。 I_{bmax} 和 I_{scmax} 分别代表电池和超级电系统的最大放电电流, I_{bmin} 和 I_{scmin} 分别代表电池和超级电系统最大充电电流。

2.2 转移概率矩阵

基于强化学习能量管理策略的一个基本步骤是对列车运行过程中的需求功率进行建模。需求功率变化可以看作是一个平稳的马尔可夫过程,而需求功率状态转移概率矩阵可以通过公式 8 最邻近法和最大似然估计法来计算。

$$\begin{cases} P_{\text{m}} = \{P_1, P_2, P_3, \dots, P_n\} \\ P_{ij} = \frac{N_{ij}}{N_i} \\ N_i = \sum_{j=1}^n N_{ij} \end{cases} \quad (8)$$

式中,将需求功率分为 n 个状态, P_{ij} 表示在一定速度下功率 P_i 到 P_j 的转移概率, N_{ij} 表示 P_i 到 P_j 的次数, N_i 表示状态 P_i 产生的总次数。

计算不同驾驶工况下的状态转移概率过程如图 4 所示,根据 1.2 节构建的三种驾驶工况提取状态转移概率。

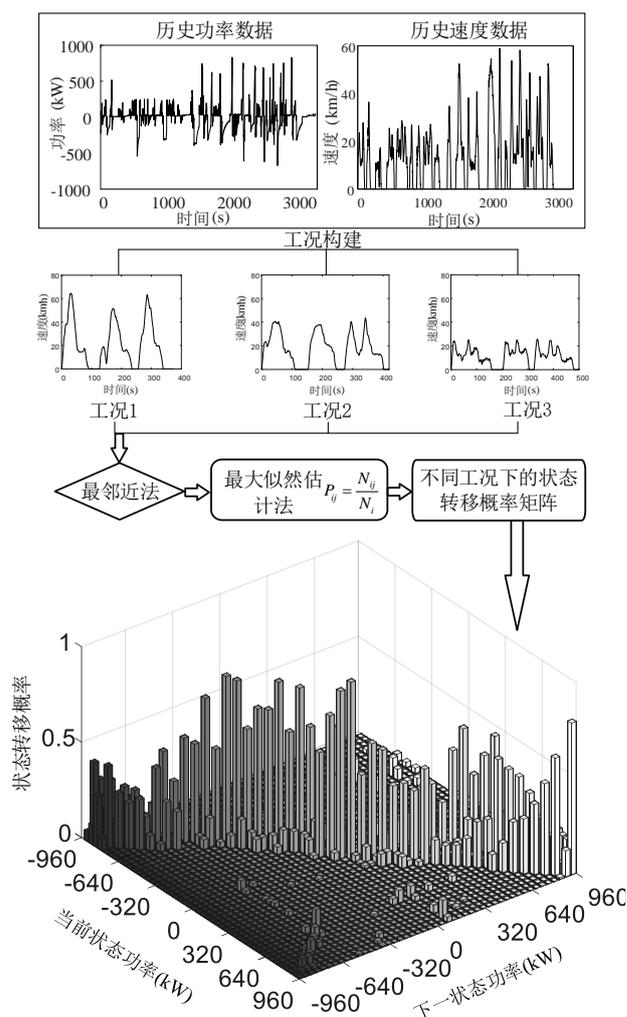


图 4 不同工况状态转移概率计算过程

Fig.4 Calculation process of state transition probability under different working conditions

2.3 基于工况识别的强化学习控制策略建立流程

对于有轨电车能量管理问题,选择有轨电车运行速度、有轨电车功率需求、电池 SOC 和超级电容 SOC 作为状态变量: $s_t \in S = \{V, P_{\text{req}}, \text{SOC}_{\text{b}}, \text{SOC}_{\text{sc}}\}$ 。

电池输出功率为决策变量： $a_t \in A = \{P_{bat}\}$ 。即时奖励定义为功率损耗的倒数： $r_t \in R = \{1/P_{loss}(s_t, a_t)\}$ 。

基于强化学习的控制策略是一个从状态到动作的映射函数 $\pi: S \rightarrow A$ ，当 π 被用作一系列完整的决策策略时，状态 s 的最优值被定义为折扣回报函数的期望之和，如公式 9 所示。

$$V^*(s) = \max_{a \in \pi} E\left(\sum_{t=0}^{+\infty} \gamma^t r_{t+1}\right) \quad (9)$$

其中 $\gamma \in [0,1]$ 为折扣因子。

根据状态转移概率以及贝尔曼最优方程的定义，可以将公式 9 的方程改写为公式 10。

$$V^*(s) = \max_{a \in A} \sum_{s' \in S} P_{s \rightarrow s'}^a (r_{s \rightarrow s'}^a + \gamma V^*(s')) \quad (10)$$

基于给定的最优值函数，可通过公式 11 计算最优策略。

$$\pi^*(s) = \arg \max_{a \in A} \sum_{s' \in S} P_{s \rightarrow s'}^a (r_{s \rightarrow s'}^a + \gamma V^*(s')) \quad (11)$$

对于状态 s 和动作 a 所对应的值函数可用 Q 来表示。

$$Q^*(s, a) = \sum_{s' \in S} P_{s \rightarrow s'}^a (r_{s \rightarrow s'}^a + \gamma \max_{a' \in A} Q^*(s', a')) \quad (12)$$

式中， $P_{s \rightarrow s'}^a$ 为采取动作 a 从状态 s 转移到 s' 的转移概率， $r_{s \rightarrow s'}^a$ 为采取动作 a 从状态 s 转移到 s' 的即时回报。在 Q-learning 算法中， Q 值可以按照公式 13 进行更新：

$$Q(s, a) = Q(s, a) + \alpha (r(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)) \quad (13)$$

式中， $\alpha \in [0,1]$ 为算法中的学习率，学习率越大收敛速度越快，但会导致过拟合问题。本文选择的学习率为 $(1/\sqrt{(n/100+3)})$ ， n 为迭代次数。

由于不同工况下的功率状态转移概率有较大差异，若直接依据历史行驶数据构建状态转移矩阵，会使系统控制性能下降，降低鲁棒性。为使得能量管理策略能更加适应有轨电车复杂驾驶工况，将具有相似的马尔科夫链模型的驾驶工况进行归类，并采用改进的 LVQ 神经网络进行工况识别，以规避驾驶工况变化较大时对能量管理策略的影响。如图 5 所示，LVQ 神经网络主要由输入层、竞争层和线性输出层组成，通过不断训练输入层和隐含层之间的权值进而获得更好的分类结果。输入层的 15 个输入节点分别对应行驶工况的 15 个特征参数，竞争层选择 80 个神经元，线性输出层神经元个数为 3，对应着期望识别的三种工况。

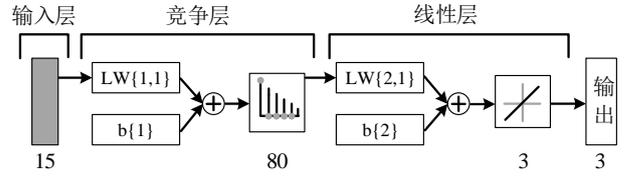


图 5 LVQ 神经网络训练结构

Fig.5 LVQ neural network training structure

由于 LVQ 网络的初始权值向量对网络训练的影响很大，为了获得良好的初始权值，提高网络分类性能，本文采用粒子群算法对网络初始权值进行优化，目标函数为训练数据预测误差。

$$fitness = \sum_{i=1}^n abs(T_i - Y_i) \quad (14)$$

式中， T_i 为神经网络训练输出数据， Y_i 为神经网络预测数据。

采用移动时间窗口的形式进行工况数据的更新，识别窗口 $T=120s$ ，预测窗口 $T_f=20s$ 。通过实时计算时间窗口 T 内的工况特征参数进行工况识别，作为未来 T_f 窗口的驾驶工况。对东湖线某天一段时间内的实车数据进行实时工况识别，识别结果如图 6 所示。

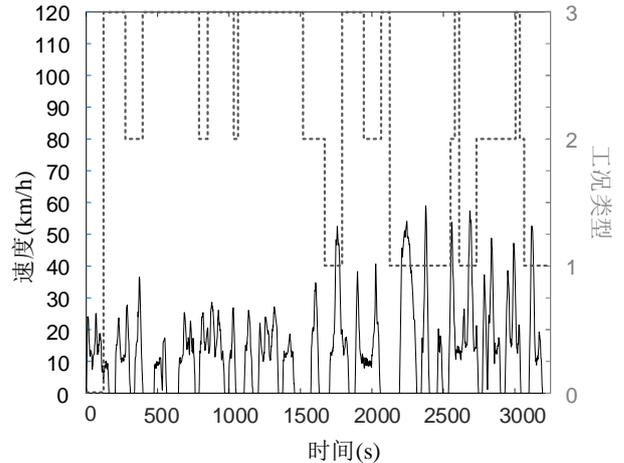


图 6 工况在线识别结果

Fig.6 Working condition online recognition result

由于采用移动窗口的形式进行工况识别，窗口内的特征参数可能和当前短行程的特征参数有一定的差异，使得识别结果出现一定的误差，但由于不同工况特征参数差异较大，LVQ 神经网络能够大致识别工况类型。

基于工况识别的强化学习能量管理策略流程如图 7 所示，主要包括离线优化求解和在线实时控制两个流程。在线实时控制时，通过截取时间窗内的

实际工况信息进行在线工况识别，有轨电车根据当前识别工况、运行速度、需求功率、及储能系统 SOC 并结合该类工况下的最优控制表实时输出电池需要承担的功率。

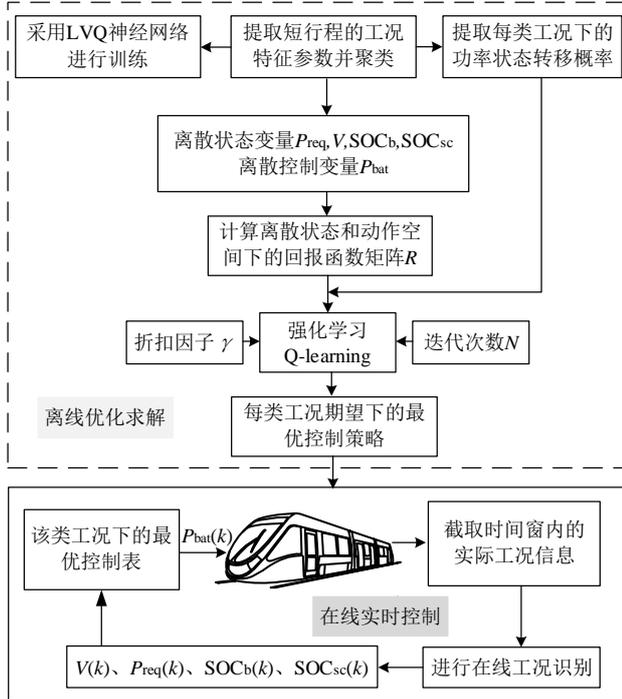


图 7 基于工况识别的强化学习能量管理策略建立流程
Fig.7 Process of establishing RL energy management strategy based on working condition recognition

3 案例分析

3.1 仿真条件及优化结果分析

本文以东湖线实车驾驶工况，采用遍历方法对混合储能系统进行容量配置，得到满足边界条件的储能系统参数如表 4 所示。

表 4 储能系统参数

Tab.4 Energy storage device parameters

参数	钛酸锂电池	超级电容
额定容量	140Ah	63.06F
额定电压/V	384.1	899.1
内阻/mΩ	28.63	13.8
可用能量/kWh	21.683	4.9
最大充放电电流/A	840	2100
SOC 使用区间	0.4-0.8	0.4-1

以工况一为例，图 8 给出在速度为 40km/h 时强化学习算法中每 500 次迭代后的 Q 值平均误差变化曲线。选择折扣因子为 0.9，并采用贪婪概率 $(1 -$

$1/\log(n/100+2.8))$ 进行动作选择。在迭代初期，随机选择动作概率较大，主要选择探索环境扩充样本，随着训练次数的上升，随机选择动作概率逐渐减小，经过充分“试错”学习之后，由探索环境状态逐渐转为利用知识状态。当迭代次数为 5000 万次时， Q 值平均误差逐渐趋近于 0，算法达到收敛。

图 9 为在工况一条件下，速度为 40km/h，不同需求功率等级下的动作变量变化情况。从图中可以看到，钛酸锂电池输出功率随着钛酸锂电池 SOC_b、超级电容 SOC_{sc} 和需求功率变化而变化。当系统牵引功率较小时主要由超级电容提供功率，且超级电容 SOC_{sc} 较低时，电池提供较高功率。当系统制动功率较小且超级电容 SOC_{sc} 较高时，主要由电池吸收制动功率。当系统牵引、制动功率较高时，由于电池额定功率限制，电池和超级电容分别承担相应的功率。

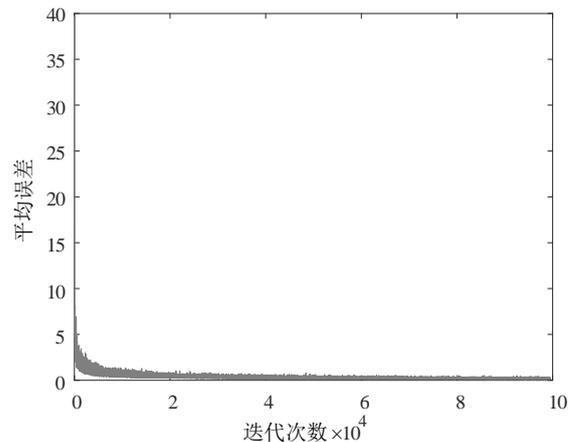
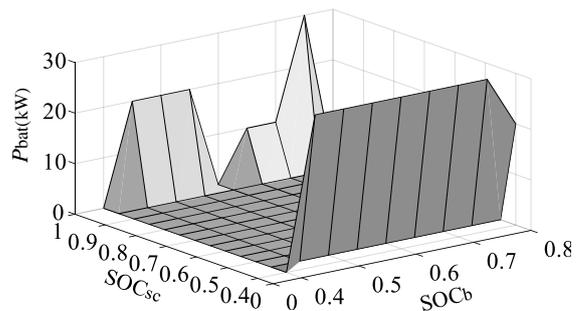


图 8 每 500 次迭代 Q 值平均误差

Fig.8 Mean discrepancy of Q value per 500 iterations.

需要注意的是，由于有轨电车到站充电的过程不具有马尔科夫性，由强化学习得到的能量管理策略只在有轨电车运行过程中使用，到站进行恒流充电，采用优先超级电容充电策略。



(a) 40kW 牵引

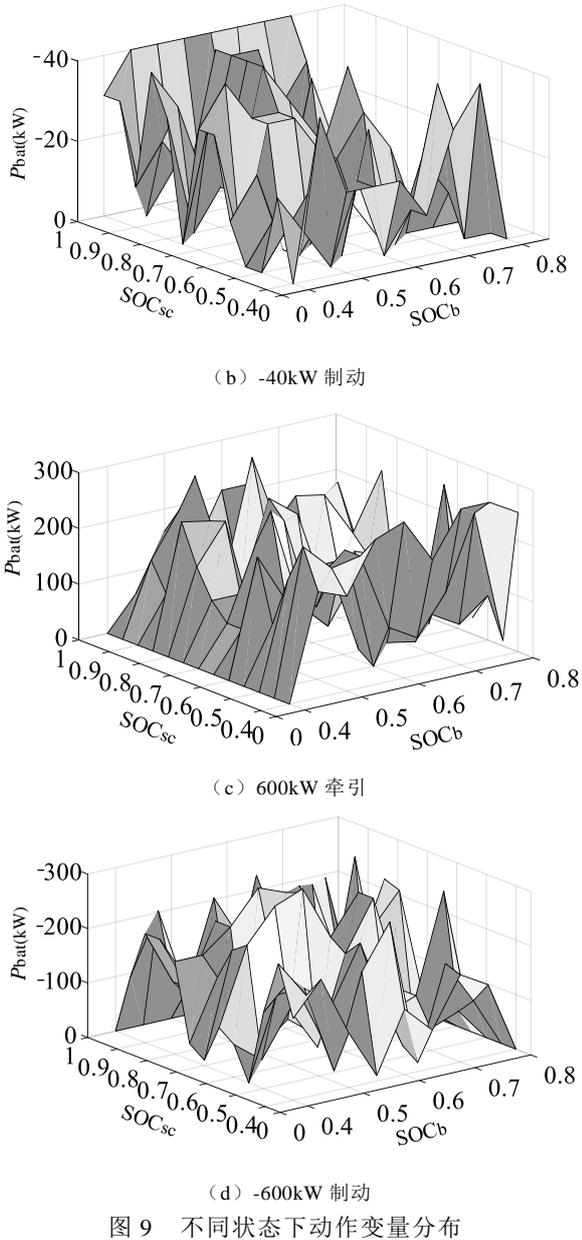
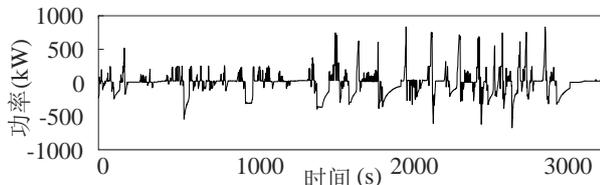


图 9 不同状态下动作变量分布

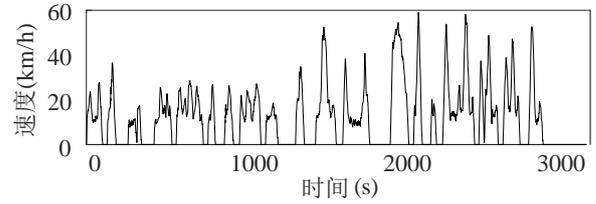
Fig.9 Distribution of action variables in different states

3.2 能量管理策略验证及对比分析

以一条实车驾驶数据作为输入条件，如图 10 所示，来验证工况识别的强化学习能量管理策略的优化效果。其中充电站以 500A 恒流充电。



(a) 功率需求曲线 1

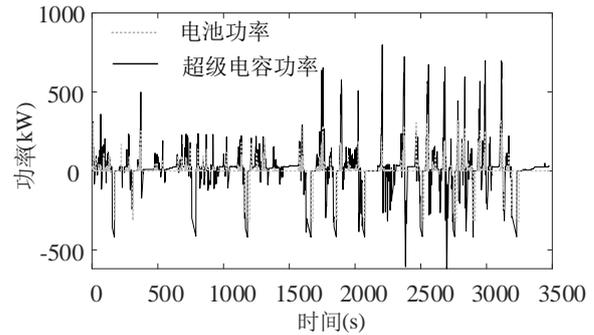


(b) 速度需求曲线 1

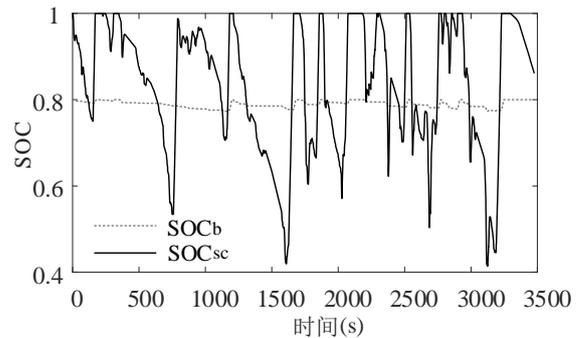
图 10 实车驾驶曲线 1

Fig.10 Real vehicle driving curve 1

图 11 为储能系统功率分配曲线以及 SOC 曲线，从中可以看出，由于钛酸锂电池系统 SOC 充足，制动功率基本都由超级电容系统吸收，且在牵引工况时，充分利用超级电容供电，减少了电池的使用，在一定程度上可以增加电池的使用寿命并降低储能系统损耗。



(a) 功率分配曲线 1



(b) SOC 变化曲线 1

图 11 储能系统功率分配曲线及 SOC 变化曲线 1

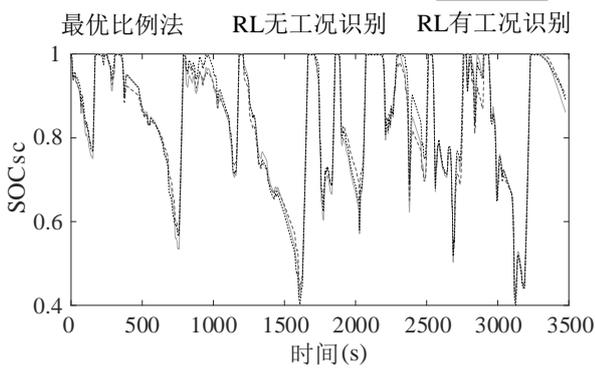
Fig.11 Energy storage system power distribution curve and SOC change curve 1

为进一步验证基于工况识别的 RL 策略的有效性，将其与无工况识别的 RL 策略和基于规则的最优比例法进行对比。

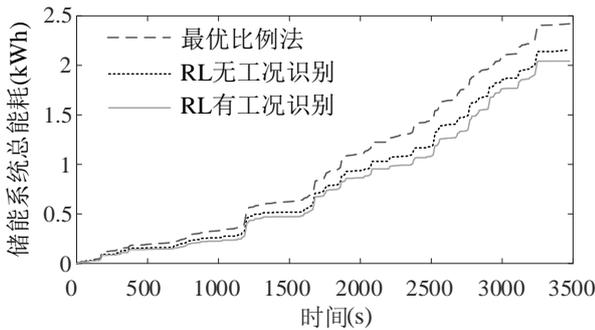
图 12(a)对三种策略超级电容的 SOC 进行了对比。三种方法 SOC 轨迹均有差异，由于超级电容 SOC 使用范围限制，其 SOC 都在 0.4-1 之间。可以

看出基于RL的策略对超级电容进行了充分的使用。

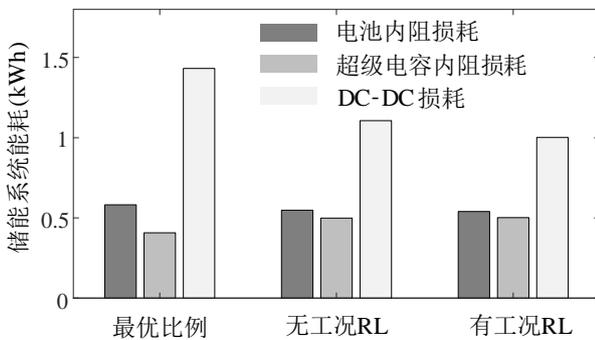
图 12(b)和图 12(c)对三种策略储能系统能耗进行了对比。从储能系统能耗分布来看,基于 RL 的策略电池内阻损耗和 DC-DC 损耗都会减小,超级电容内阻损耗会增加,但储能系统总能耗都会减小。其中无工况识别的 RL 策略总能耗降低 11.2%,有工况识别的 RL 策略总能耗降低了 15.7%,均能减少储能系统能量损失,提高系统效率。可以看出,在进行强化学习离线优化时,对有轨电车驾驶工况进行聚类分析,将具有相似马尔科夫状态转移概率的工况归为一类,可以得到更好的节能效果。



(a) 超级电容 SOC 对比曲线 1



(b) 储能系统总能耗对比曲线 1



(c) 储能系统能耗分布对比 1

图 12 三种策略下超级电容 SOC 及总能耗对比 1

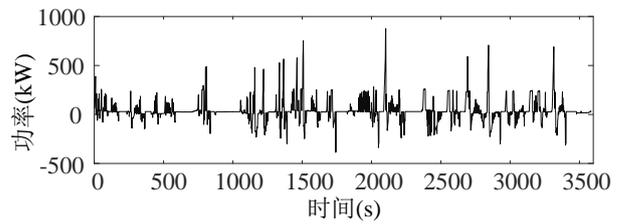
Fig.12 Comparison of super capacitor SOC and total energy consumption under three strategies 1

表 5 三种能量管理策略下的储能系统能耗对比 1

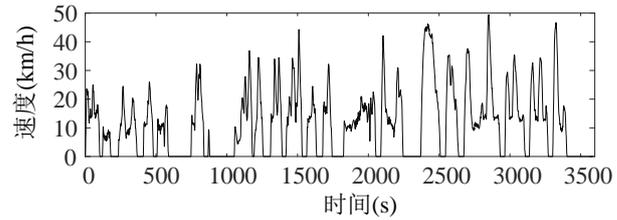
Tab.5 Comparison of energy consumption of EMS under three energy management strategies 1

能量管理策略	总能耗/kWh	相对降低率/%
最优比例法	2.4250	—
无工况识别 RL	2.1533	11.2%
有工况识别 RL	2.0436	15.7%

为充分验证基于工况识别强化学习策略的适应性,将求得的策略应用于东湖线另外一条实际驾驶工况,其中充电站以 1000A 恒流充电。如图 13 所示。



(a) 功率需求曲线 2

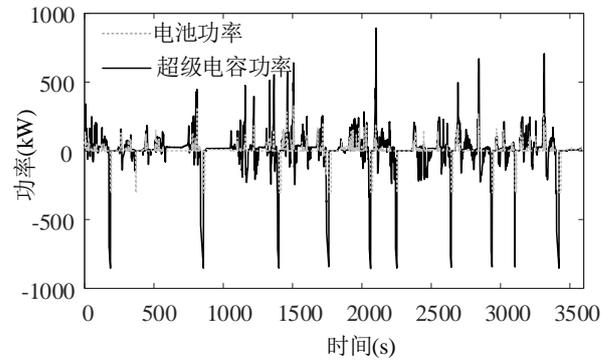


(b) 速度需求曲线 2

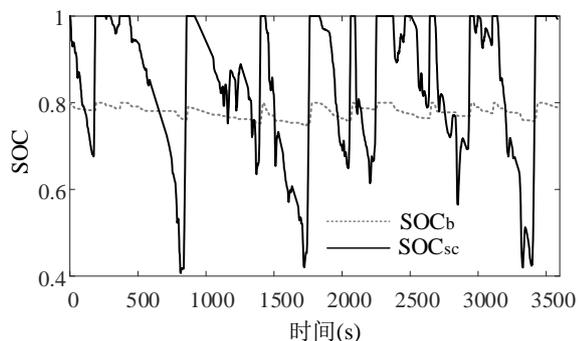
图 13 实车驾驶曲线 2

Fig.13 Real vehicle driving curve 2

得到的基于工况识别的强化学习策略适应性验证工况的功率分配曲线以及储能系统 SOC 曲线如图 14 所示。



(a) 功率分配曲线 2



(b) SOC变化曲线 2

图 14 储能系统功率分配曲线及 SOC 变化曲线 2

Fig.14 Energy storage system power distribution curve and SOC change curve 2

将适应性验证工况的三种能量管理策略进行对比, 分别得到超级电容 SOC 曲线以及储能系统能耗, 如图 15 和表 6 所示。

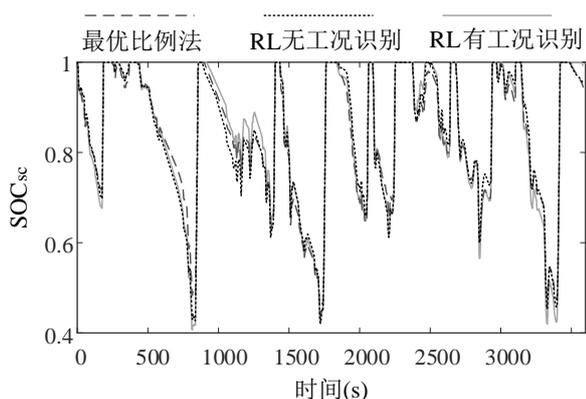


图 15 三种策略下超级电容 SOC 对比 2

Fig.15 State of charge of super capacitors comparison under three strategies 2

表 6 三种能量管理策略下的储能系统能耗对比 2

Tab.6 Comparison of energy consumption of EMS under three energy management strategies 2

能量管理策略	总能耗/kWh	相对降低率/%
最优比例法	2.6175	—
无工况识别 RL	2.3290	11.0%
有工况识别 RL	2.2223	15.1%

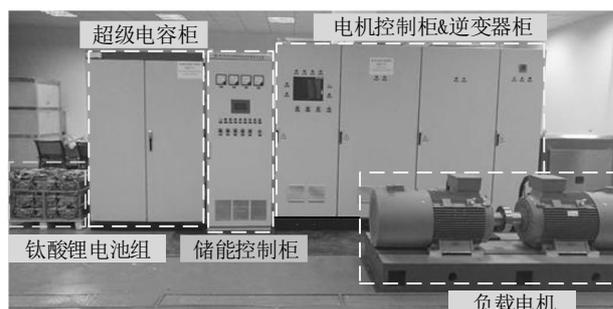
可以看到适应性验证工况与原工况的仿真结果趋于一致, 通过改变三种能量管理策略的验证工况, 得到基于工况识别的强化学习策略在不同工况下均优于无工况识别强化学习策略以及最优比例法, 从而说明强化学习算法对不同工况的适应性。

3.3 实验验证

为验证基于工况识别的强化学习策略的在线决

策的有效性及可行性, 在如图 16(a)所示的 90kW 电池-超级电容混合储能系统实验平台上进行实验。实验平台的储能系统基本参数见表 7。

本实验主要验证混合储能系统在不同工况下的强化学习能量管理策略的控制效果。设计验证实验接线如图 16(b)所示。上位机通过 CAN 通讯接收由电机通过 DSP 传输的数据, 输入至 Matlab 软件中进行工况在线识别并通过强化学习已优化出的最优控制表进行在线决策, 采用 Python 读取功率分配结果, 并通过 CAN 通讯把指令传回 DSP, 使混合储能系统执行充放电指令。上位机通过以太网-CAN 转换模块输出信号, 与电池的 BMS 模块所输出的信息并联在 CAN 总线上。



(a)混合储能实验平台结构



(b)实验平台通讯连接示意图

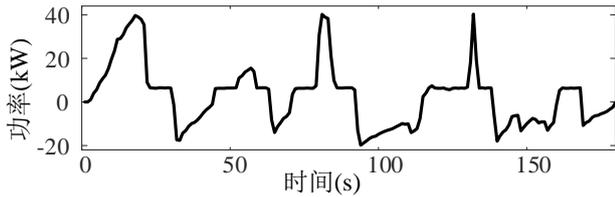
图 16 混合储能实验验证平台

Fig.16 Hybrid energy storage experimental verification platform

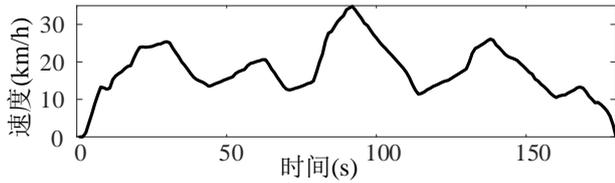
表 7 混合储能平台参数

参数	电池系统	超级电容系统
品牌	微宏	Maxwell
配置方式	216S2P	48S4P
额定电压 (V)	496.8	500
容量	20Ah	15.75F
能量(kWh)	9.94	0.56
最大持续电流 (A)	60	140
最大持续功率/kW	29.8	70

以有轨电车实际运行数据作为输入，由于实验平台功率等级的限制，对列车负载曲线进行等比例放缩处理，放缩后最大功率约为 40kW，如图 17 所示。



(a) 实验输入功率曲线 1

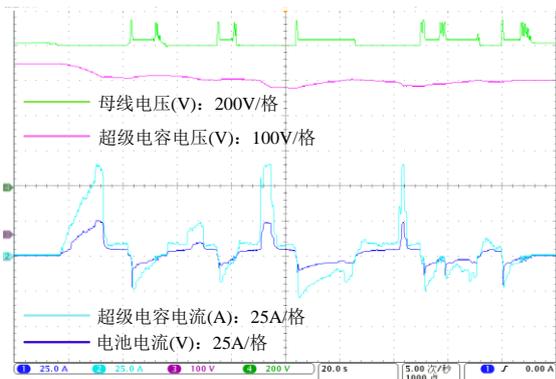


(b) 实验输入速度曲线 1

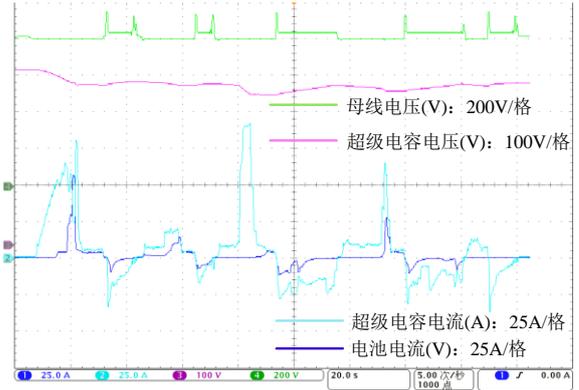
图 17 实验输入曲线 1

Fig.17 Experimental input curve 1

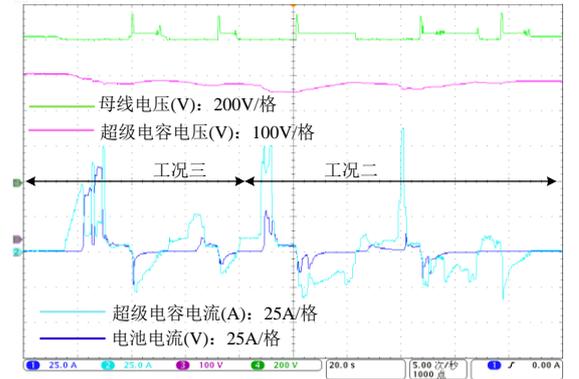
以最优比例法、无工况识别的强化学习策略以及基于工况识别的强化学习策略依次进行实验。所得电池、超级电容的电流曲线、超级电容的电压曲线及母线电压曲线如图 18 所示。



(a) 最优比例法实验波形 1



(b) 无工况识别强化学习策略实验波形 1

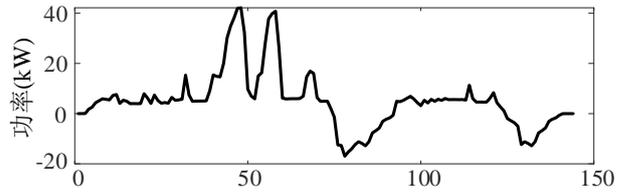


(c) 有工况识别强化学习策略实验波形 1

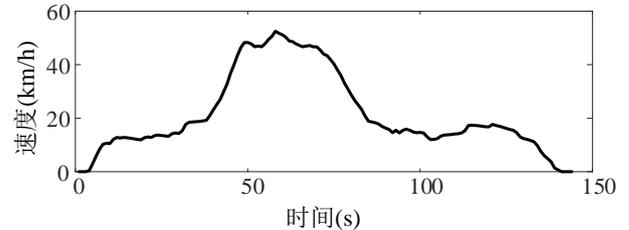
图 18 三种策略下的实验波形 1

Fig.18 Three strategy experimental waveforms 1

采用另一条实际运行数据进行实验验证，如图 19，验证强化学习策略对不同工况的适应性。



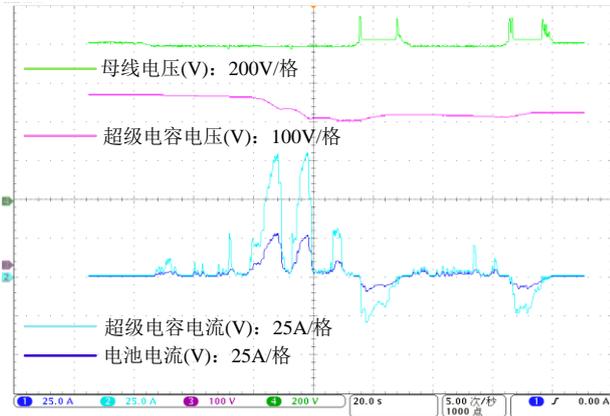
(a) 实验输入功率曲线 2



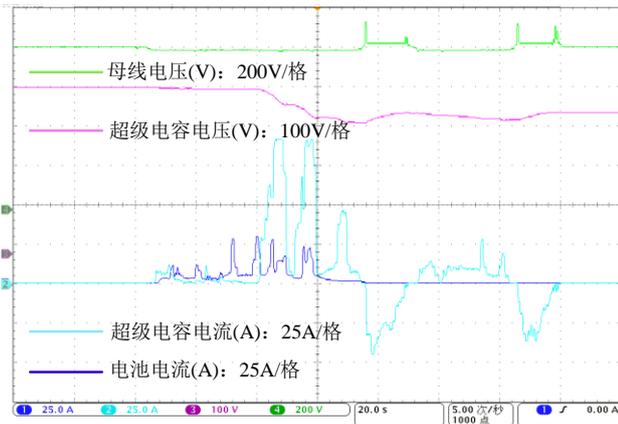
(b) 实验输入速度曲线 2

图 19 实验输入曲线 2

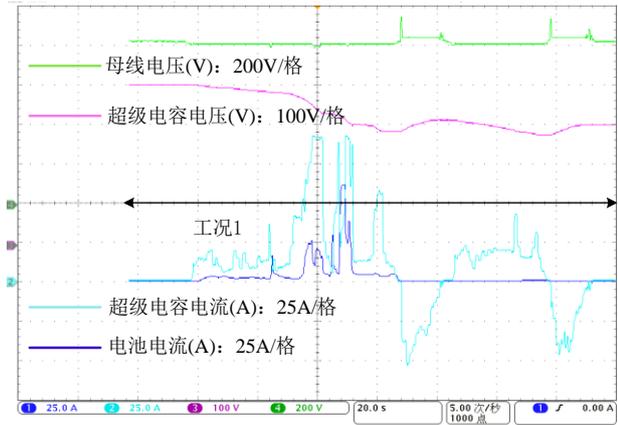
Fig.19 Experimental input curve 2



(a) 最优比例法实验波形 2



(b) 无工况识别强化学习策略实验波形 2



(c) 有工况识别强化学习策略实验波形 2

图 20 三种策略下的实验波形 2

Fig.20 Three strategy experimental waveforms 2

通过 90kW 样机实验可得，在混合储能系统实际运行时，通过上位机向控制器实时发送指令，控制混合储能系统动作，储能系统能够跟随功率分配结果进行相应充放电。由实验波形可以看出基于强化学习的能量管理策略能够根据系统状态做出实时决策，充分利用超级电容进行供电以减少系统损耗。

通过加入对工况的在线识别过程，可提升在线决策过程中的控制效果，并满足实时控制需求，可实现工程应用。

4 结论

本文以有轨电车车载混合储能系统为研究对象，由于需求功率的随机性，将有轨电车的功率需求看做马尔科夫过程，并考虑到驾驶工况变化较大时对能量管理策略的影响，提出了基于工况识别的强化学习能量管理策略。通过主成分分析以及 K 均值聚类得到了高速、中速、低速三种驾驶工况，并得到了不同工况下的功率状态转移概率。采用强化学习算法得到了不同工况以及速度下的电池功率的动作值，并通过改进的 LVQ 神经网络实时识别当前的驾驶工况进而做出相应的决策，使其能够在相似工况中获得更好的控制效果。与最优比例法相比，该方法储能系统总能耗降低了约 15.7% 且比无工况识别的 RL 能量管理策略具有更好的节能效果。并将求得的策略应用于东湖线另外一条实际驾驶工况，验证了基于工况识别的强化学习策略对不同工况的适应性。通过 90kW 的实验平台进行模拟运行，验证了本文提出的基于工况识别的强化学习策略的有效性，证明了该策略在工程应用中的可行性。

参考文献

- [1] 姜军.有轨电车的复兴与思考[J]. 江苏城市规划,2014(05): 44-47.
Jiang Jun. The revival and reflection of trams[J]. Jiangsu City Planning, 2014(05): 44-47.
- [2] 王健全,袁富卫.城市有轨电车供电方式探讨[J].电力机车与城轨车辆,2015,38(01):47-51+58.
Wang Jianquan, Yuan Fuwei.Discussion on power supply method of urban tram[J].Electric Locomotive and Urban Rail Vehicle,2015,38(01):47-51+58.
- [3] 诸斐琴,杨中平,林飞,等. 基于加速时间预测的现代有轨电车储能系统能量管理与容量配置优化研究[J]. 电工技术学报,2017,32(23):158-166.
Zhu Feiqin, Yang Zhongping, Lin Fei, et al. Research on energy management and capacity allocation optimization of modern tram energy storage system based on acceleration time prediction[J]. Transactions of China Electrotechnical Society, 2017, 32(23): 158-166.
- [4] 王珂,杨中平,李峰,等.有轨电车混合动力系统能量

- 交互型管理策略与容量配置协同优化研究[J]. 电工技术学报, 2019, 34(08): 1780-1788.
- Wang Yu, Yang Zhongping, Li Fen, et al. Study on energy interactive management strategy and capacity allocation collaborative optimization of tram hybrid power system[J]. Transactions of China Electrotechnical Society, 2019, 34(08): 1780-1788.
- [5] 张弛. 现代有轨电车车载混合储能系统能量管理优化研究[D]. 北京交通大学, 2018.
- [6] 韦绍远, 姜久春, 程龙, 等. 储能式有轨电车车地一体化配置模型[J]. 电工技术学报, 2019, 34(02): 427-436.
- Wei Shaoyuan, Jiang Jiuchun, Cheng Long, et al. Ground-storage model of energy-storage trams[J]. Transactions of China Electrotechnical Society, 2019, 34(02): 427-436.
- [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(05): 52-59.
- Zhou Meilan, Feng Jifeng, Zhang Yu. Compound energy storage system of pure electric vehicle and its energy control strategy[J]. Electric Machines and Control, 2019, 23(05): 51-59.
- [8] 安星锴, 杨中平, 王珂, 等. 基于改进型凸优化算法的有轨电车混合储能系统容量配置帕累托解集[J]. 电工技术学报, 2020, 35(14): 3116-3125.
- An Xingkun, Yang Zhongping, Wang Yu, et al. Pareto solution set of capacity configuration of tram hybrid energy storage system based on improved convex optimization algorithm[J]. Transactions of China Electrotechnical Society, 2020, 35(14): 3116-3125.
- [9] 王珂, 杨中平, 林飞, 等. 有轨电车车载混合储能系统动态比例分配策略[J]. 电工技术学报, 2019, 34(S1): 405-413.
- Wang Yu, Yang Zhongping, Lin Fei, et al. Dynamic proportional distribution strategy for on-board hybrid energy storage system of trams[J]. Transactions of China Electrotechnical Society, 2019, 34(S1): 405-413.
- [10] 李峰, 杨中平, 王珂, 等. 基于庞特里亚金极小值原理的混合储能有轨电车能量管理策略[J]. 电工技术学报, 2019, 34(S2): 752-759.
- Li Feng, Yang Zhongping, Wang Yu, et al. Energy management strategy of hybrid energy storage tram based on Pontrya's gold minimum principle[J]. Transactions of China Electrotechnical Society, 2019, 34(S2): 752-759.
- [11] 林歆悠, 孙冬野. 基于工况识别的混联式混合动力客车控制策略研究[J]. 中国机械工程, 2012, 23(07): 869-874.
- Lin Xinyou, Sun Dongye. Research on Control Strategy of Hybrid Hybrid Electric Bus Based on Condition Identification[J]. China Mechanical Engineering, 2012, 23(07): 869-874.
- [12] 罗少华. 基于工况识别的混联式混合动力汽车能量管理策略研究[D]. 重庆大学, 2016.
- [13] Changle Xiang, Feng Ding, Weida Wang, Wei He. 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.
- [14] 张洁丽. 基于模型预测控制的插电式混合动力客车能量管理策略研究[D]. 北京: 北京理工大学, 2016.
- [15] Liu T, Zou Y, Liu D, et al. Reinforcement Learning of Adaptive Energy Management With Transition Probability for a Hybrid Electric Tracked Vehicle[J]. Industrial Electronics, IEEE Transactions on, 2015, 62(12): 7837-7846.
- [16] Teng L, Yuan Z, Dexing L, et al. Reinforcement Learning - Based Energy Management Strategy for a Hybrid Electric Tracked Vehicle[J]. Energies, 2015, 8(7): 7243-7260.
- [17] LIU T, HU X, HU W, et al. A Heuristic Planning Reinforcement Learning-Based Energy Management for Power-Split Plug-in Hybrid Electric Vehicles[J]. IEEE Transactions on Industrial Informatics, 2019, 15: 15(12): 6436-6445.
- [18] Liu T, Zou Y, Liu D, et al. Reinforcement Learning of Adaptive Energy Management With Transition Probability for a Hybrid Electric Tracked Vehicle[J]. Industrial Electronics, IEEE Transactions on, 2015, 62(12): 7837-7846.
- [19] R. Xiong, J. Cao, and Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," Appl. Energy, vol. 211, pp. 538-548, Nov. 2017. doi: 10.1016/j.apenergy.2017.11.072.
- [20] Rui Xiong, Jiayi Cao, Quanqing Yu. 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

- [21] Teng Liu,Bo Wang,Chenglang Yang. Online Markov Chain-based energy management for a hybrid tracked vehicle with speedy Q-learning [J]. Energy, 2018,160:544-555

作者简介

莫浩楠 男, 1997 年生, 硕士, 研究方向为车载储能系统容量配置与能量管理策略。

E-mail: 19126139@bjtu.edu.cn (通信作者)

杨中平 男, 1970 年生, 教授, 博士生导师, 研究方向为轨道交通电力牵引传动、节能、高速列车系统优化设计等。

E-mail: zhpyang@bjtu.edu.cn