

## 基于强化学习考虑电池损耗的电动汽车充放电控制算法

卢钺, 王琼, 刘顺, 李清涛, 刘洋, 王洪彪

### 引用本文

卢钺, 王琼, 刘顺, 李清涛, 刘洋, 王洪彪. 基于强化学习考虑电池损耗的电动汽车充放电控制算法[J]. 计算机科学, 2024, 51(11A): 231200147-7.

LU Yue, WANG Qiong, LIU Shun, LI Qingtao, LIU Yang, WANG Hongbiao. [Reinforcement Learning Algorithm for Charging/Discharging Control of Electric Vehicles Considering Battery Loss](#) [J]. Computer Science, 2024, 51(11A): 231200147-7.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

#### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于数据与模型联合驱动的电动汽车充电负荷时空分布预测](#)

Prediction of Spatial and Temporal Distribution of Electric Vehicle Charging Loads Based on Joint Data and Modeling Drive

计算机科学, 2024, 51(11A): 231100110-6. <https://doi.org/10.11896/jsjcx.231100110>

#### [基于双层优化的电动车与电网实时协同定价机制](#)

Real-time Collaborative Pricing Mechanism of Between Vehicle and Power Grid Based on Bi-level Optimization

计算机科学, 2024, 51(11A): 240300013-7. <https://doi.org/10.11896/jsjcx.240300013>

#### [基于深度强化学习的云边协同任务迁移与资源再分配优化研究](#)

Cloud-Edge Collaborative Task Transfer and Resource Reallocation Optimization Based on Deep Reinforcement Learning

计算机科学, 2024, 51(11A): 231100170-10. <https://doi.org/10.11896/jsjcx.231100170>

#### [基于多任务学习的复杂城市遥感图像道路提取](#)

Road Extraction from Complex Urban Remote Sensing Images Based on Multi-task Learning

计算机科学, 2024, 51(11A): 240300095-8. <https://doi.org/10.11896/jsjcx.240300095>

#### [基于深度强化学习的无人机自主探索方法](#)

Autonomous Exploration Methods for Unmanned Aerial Vehicles Based on Deep Reinforcement Learning

计算机科学, 2024, 51(11A): 231100139-6. <https://doi.org/10.11896/jsjcx.231100139>

# 基于强化学习考虑电池损耗的电动汽车充放电控制算法

卢 钺<sup>1</sup> 王 琼<sup>2</sup> 刘 顺<sup>1</sup> 李清涛<sup>1</sup> 刘 洋<sup>1</sup> 王洪彪<sup>2</sup>

1 国网北京海淀供电公司 北京 100080

2 国网北京市电力公司 北京 100032

(339545765@qq.com)

**摘 要** 随着电动汽车数量的逐步增加,其接入对电网的负荷带来显著影响。在这一背景下,V2G/G2V技术被广泛认为能在电网管理方面发挥重要作用。以电动汽车的充放电控制算法为研究对象,引入了一种基于软演员评论家(SAC)的深度强化学习算法,从而实现了对电动汽车连续充放电行为的精细控制。研究着眼于解决电网中负荷时序动态变化的难题,通过调整不同车辆在不同电价条件下的充放电功率,最大程度地提升用户的经济效益。此外,为应对充放电过程中可能导致电池损耗加剧的问题,引入了基于物理混合神经网络(PHNN)的电池损耗预测模型。同时,通过将充放电过程建模为马尔可夫决策问题,并将PHNN模型融入电动汽车的充放电控制中,构建了一个全新的奖励函数,以精确量化电池损耗所带来的成本。基于SAC算法,该奖励函数用于学习最优的充放电策略。实验结果显示,该算法能够有效地调控车辆的充放电行为,发挥电力网络调控作用,同时在充放电过程中降低对电池寿命造成的损耗,进一步保障用户经济利益。

**关键词:** 电动汽车;电动汽车充放电控制;深度强化学习;电网调控;电池建模

**中图分类号** TP311

## Reinforcement Learning Algorithm for Charging/Discharging Control of Electric Vehicles Considering Battery Loss

LU Yue<sup>1</sup>, WANG Qiong<sup>2</sup>, LIU Shun<sup>1</sup>, LI Qingtao<sup>1</sup>, LIU Yang<sup>1</sup> and WANG Hongbiao<sup>2</sup>

1 State Grid Beijing Haidian Power Supply Company, Beijing 100080, China

2 State Grid Beijing Electric Power Company, Beijing 100032, China

**Abstract** With the gradual increase in the number of electric vehicles, their integration has a significant impact on the load of the power grid. In this context, V2G/G2V technology is widely believed to play an important role in power grid management. Taking the charging and discharging control algorithm of electric vehicles as the research object, a deep reinforcement learning algorithm based on Soft Actor-Critic(SAC) is introduced. In terms of the dynamic change of load sequence in the power grid, the charging/discharging rate of different vehicles is controlled to maximize the benefits for users under different electricity prices. In addition, in order to address the issue of increased battery loss during the charging and discharging process, a battery loss prediction model based on physical hybrid neural network(PHNN) is introduced in the research. Meanwhile, the charging/discharging process is modeled as a Markov decision process. By integrating the PHNN model into the charging and discharging control of electric vehicles, a new reward function is constructed to accurately quantify the cost of battery loss. Based on the SAC algorithm, this reward function is used to learn the optimal charging and discharging strategy. Experimental results show that this algorithm can effectively regulate the charging and discharging behavior of vehicles, play a regulatory role in the power network, and reduce the loss of battery life during the charging and discharging process, further ensuring the economic interests of users.

**Keywords** Electric vehicle, Electric vehicle charging/discharging control, Deep reinforcement learning, Power grid regulation, Battery modeling

### 1 概述

电动汽车(Electric Vehicle, EV)作为传统化石燃料驱动汽车的有力替代方案,具备减少温室气体排放和空气污染的潜力<sup>[1]</sup>。然而,电动汽车的迅猛增长也为电力网络带来了诸多新的挑战,包括电力负荷的增加以及峰值电力需求的扩大。

相关研究指出,即便电动汽车仅增加10%,也可能导致电力网络电压曲线出现显著波动<sup>[2]</sup>。

车辆到电网(Vehicle-to-Grid, V2G)和电网到车辆(Grid-to-Vehicle, G2V)被认为是解决这一问题的有效技术<sup>[3-4]</sup>。车辆可以在电力低谷时充电(G2V),在电力高峰时放电(V2G),参与到电网的电力调控。然而,动态电价、电池健康度以及

基金项目:国网北京市电力公司科技项目:电动汽车充电站 V2G/S2G 车网互动及智慧集群调控技术研究及示范(520204220008)

This work was supported by the State Grid Beijing Electric Power Company Technology Project: Research and Demonstration of V2G/S2G Vehicle Network Interaction and Intelligent Cluster Control Technology for Electric Vehicle Charging and Discharging Stations(520204220008).

通信作者:王琼(coycathy@163.com)

人类行为等不确定性因素,增加了制定充电策略的复杂性<sup>[5]</sup>。因此,如何在动态环境中有效地协调电动汽车的充放电行为,使其能够积极参与输电系统,是目前亟须解决的问题。

传统研究通常将充电控制视为一个优化问题,其目标在于最大化驾驶员的利益<sup>[6]</sup>。近年来,国内外学者针对电动汽车充放电控制提出了大量算法。文献[7]构建了双阶段成本优化模型,以减少工作场所运营成本。在另一研究中<sup>[8]</sup>,作者对配电变压器容量限制进行考虑,并将电动汽车的连续充电过程转化为线性规划问题。文献[9]设计了一种带有电池能量约束的成本函数,并通过动态规划方法计算了最优的电动汽车充电控制策略。文献[10]将风能供应的不确定性纳入考虑范畴,并建模为马尔可夫决策过程(Markov Decision Process, MDP)来解决电动汽车充放电控制问题。文献[11]考虑了用户对充电参数的动态偏好,并采用随机博弈方法来解决不确定性问题。文献[12]引入了时间方面的不确定性,并使用广义纳什均衡(NE)博弈来解决电动汽车充电问题。总体而言,上述各种算法均在基于模型的完全可观察环境假设下,保持明确的优化模型。最佳充电控制策略是基于精确的系统模型生成的。然而这些方法都依赖于完全可观测的环境,这在实际情况往往是不可满足的,并且环境中的不确定因素会进一步影响算法的准确性。

深度强化学习(Deep Reinforcement Learning, DRL)在缺乏环境先验知识的情况下,在时序决策问题中展现了巨大的潜力。部分学者将深度强化学习算法用在了电动汽车充放电控制中。文献[13]利用表示网络提取电价的变化特征,采用Q学习方法(Q-learning)获得了最优的离散电动汽车充电控制策略。文献[14]采用了长短期记忆(LSTM)技术来提取电价趋势,并将预测的电价应用于深度确定性策略梯度算法(DDPG)中。文献[15]将电动汽车充电控制问题重新构建为约束型马尔可夫决策过程(MDP),并开发了一种安全的DRL方法来学习最优调度策略。在文献[16]中,作者使用统计原理对驾驶员焦虑进行了数学建模,并提出使用软演员评论家算法(Soft Actor-Critic, SAC)来学习最优策略。

以上算法都站在用户侧考虑充放电的收益问题,忽略了电动汽车电池的寿命限制。文献[17]表明,过多的电池循环次数会显著降低电池寿命,提升用户的使用隐性开销。因此,电动汽车的充放电控制应当同时考虑充放电收益和电池损耗的影响。然而,电池健康度状态(State of Health, SoH)受到电压、电流和温度等因素的影响,其预测是一项难题。目前,SoH的预测方法可以大致分为两种不同的类型:模型驱动的方法和数据驱动的方法。模型驱动的方法主要依赖于基于物理先验知识建立的模型,如等效电路模型<sup>[18]</sup>和电化学模型<sup>[19]</sup>。文献[20]中提出了一种具有退化意识的高保真电化学热模型,该模型利用进化插值卡尔曼滤波器来估计电池窗口。文献[21]提出了一种改进的快速傅里叶变换方法来筛选基于电化学阻抗谱的特征,并利用极限学习机建立电池预测模型。在文献[22]中,作者将电池容量增量分析与支持向量机相结合来估计电池SoH。尽管基于模型的方法已经取得了重大进展,但它们往往具有较大的局限性:通常基于简化的假设或狭义的物理定律构建,这可能会导致大量的计算误差,导致不准确和不确定的估计。

数据驱动的方法依赖于电池的循环数据,而没有明确地对潜在的物理过程进行建模。这些方法旨在捕捉电池性能与

从电池运行数据中提取的各种特征之间的复杂关系。其中,LSTM网络由于其有效预测长序列的卓越能力,在基于数据驱动的电池SoH预测领域得到了广泛应用。在文献[23]中,作者提出了一种基于改进的卷积神经网络和LSTM的方法来估计电池剩余使用寿命。文献[24]提出了一个可迁移的多级SoH估计模型,该模型使用LSTM在跨电池型号间进行迁移学习。文献[25]通过引入注意力机制提升了双向LSTM的预测精度,并提出了多维温度补偿以增强温度变化下的预测。

综上所述,本文考虑到以往电池SoH预测模型的局限性,结合了基于模型和数据驱动方法的优势,提出了一个基于物理混合神经网络(Physical Hybrid Neural Network, PHNN)的电池健康度预测模型,并将其应用到基于SAC的电动汽车充放电控制中。相比于已有的工作,本文的创新点如下:

(1)将物理先验知识与神经网络相结合,设计了一种新颖的神经网络结构,显著提高了SoH预测的准确性。

(2)基于提出的PHNN电池SoH预测模型,设计了一个结合电池损耗的奖励函数,精确量化了V2G中电池损耗的成本开销。结合SAC方法优化电动汽车充放电策略,保障用户经济效益。

(3)基于公开数据集,设计了几个模拟实验来验证算的有效性。实验结果表明,提出的电动汽车充放电控制算法能显著减缓V2G过程中电池健康度的下降,具有较强的实际意义。

本文第2章介绍了电池损耗模型和电动汽车充放电建模;第3章介绍了电池损耗的预测算法以及电动汽车的充放电控制算法;第4章设计了几个仿真实验验证了算法的有效性。

## 2 基于电池损耗的电动汽车充放电成本优化模型

### 2.1 考虑电池损耗的电动汽车电池的充放电建模

本文从用户角度研究电动汽车的充放电调度中的成本优化问题。将电价波动和电池损耗纳入电动汽车充放电建模,实现全面鲁棒的充放电管理。

对于电动汽车的电池,通常有两个重要的参数来描述其当前的状态:电池荷电状态(State of Charge, SoC)和电池健康状态(State of Health, SoH)。其中,SoC通常以百分比的形式表示,反映了电池当前存储的电量相对于其最大充电容量的比例;SoH也通常以百分比的形式表示,衡量了电池的实际容量与其初始容量之间的比例,用于表示电池的老化程度和性能衰减情况。为了在电动汽车充放电中考虑电池损耗带来的成本开销问题,本文提出了一个全新的参数:电池初始荷电状态(Initial State of Charge, ISoC)。ISoC表示了电池当前存储的电量与初始容量之间的比例,其同样由百分比表示。具体来说, $t$ 时刻的ISoC为:

$$ISoC_t = SoC_t \cdot SoH_t \quad (1)$$

当连接上充电桩后,电动汽车为了满足充电需求,可以在单位时间内充电,也可以在高电价时放电来获取电价波动带来的收益。假设电动汽车在 $t_a$ 时刻连接上充电桩,用户计划在 $t_d$ 时刻出行。在单位时间内,电池电量的变化为:

$$ISoC_{t+1} = \begin{cases} ISoC_t, & t < t_a, t \geq t_d \\ ISoC_t + \eta \cdot a_t, & t_a \leq t < t_d \end{cases} \quad (2)$$

其中, $a_t$ 表示电动汽车在单位时间内的充/放电的速率,若 $a_t > 0$ 则电动汽车处于充电状态。反之则处于放电状态, $a_t$

同样相对于电池初始容量表示为百分比的形式。 $\eta$  表示充放电损耗,因此  $0 < \eta < 1$ 。

## 2.2 基于马尔可夫决策的电动汽车充放电建模

电动汽车的充放电行为可以被建模为一个马尔可夫决策过程,其作为一种数学框架,提供了一种有效的方式来描述电动汽车充放电行为的随机性和不确定性。该模型允许根据不同时间段的电价和电池健康状态,做出相应的充电和放电决策,以优化用户的经济效益。为了描述电动汽车的连续充放电过程,将连续的状态变化拆分成足够小的时隙变化。如下文中的  $s_t$  即表示  $t$  时隙内车辆的状态。

### 1) 状态

本文将从环境中获得的状态用作充放电控制策略的输入,以生成充放电速率。具体地,在  $t$  时段内,状态  $s_t$  定义为:

$$s_t = (\xi_{t-47}, \xi_{t-46}, \dots, \xi_t, t_d, I\text{SoC}_t, I\text{SoC}_d) \quad (3)$$

其中,  $\xi_{t-47}, \xi_{t-46}, \dots, \xi_t$  表示过去 48 个时隙内电力公司的定价。 $I\text{SoC}_t$  表示时刻  $t$  时电池中的电量。本文还引入了  $t_d$  和  $I\text{SoC}_d$  来表示用户对于电动汽车充放电过程的控制效果,  $t_d$  表示用户计划结束充电的时间,  $I\text{SoC}_d$  表示充电结束时希望电池中拥有的电量,该值为用户人为设定。

### 2) 动作

在车辆的充放电过程中,在时间段  $[t, t+1)$  内,车辆的充放电行为由动作变量  $a_t$  所表示。该变量反映了车辆在该时间段内电池电量的百分比变化(相对于电池初始容量)。动作变量  $a_t$  的取值范围限制在  $[-0.2, 0.2]$  之间,其中正数表示车辆处于充电模式,负数则表示车辆处于放电模式。

### 3) 状态转移

用户的行为以及各种环境随机因素会影响车辆状态之间的转移。我们使用  $P(s_{t+1} | s_t, a_t)$  来表达这种随机性,其表示当采取动作  $a_t$  时车辆状态从  $s_t$  转换为  $s_{t+1}$  的概率。

### 4) 奖励函数

在电动汽车的充放电过程中,以用户的经济收益为主要目标,即通过最优化充放电策略,尽可能提高用户在参与电力调控过程中获得的经济收益。考虑到这一因素的影响,本文设计了相应的奖励函数:

$$r_t(s_t, a_t) = \begin{cases} -\lambda_p \cdot \xi_t \cdot a_t - \lambda_h \cdot c \cdot (SoH_t - SoH_{t+1}), & t_a \leq t < t_d \\ -\lambda_d \cdot \max(I\text{SoC}_d - I\text{SoC}_t, 0), & t = t_d \end{cases}$$

其中,  $\xi_t \cdot a_t$  表示  $[t, t+1)$  时隙内电动汽车充放电的开销/收益;  $SoH_t - SoH_{t+1}$  表示当前时隙下的充放电动作对电池容量造成的损耗,  $c$  则代表用户更换电池的成本,因此  $c \cdot (SoH_t - SoH_{t+1})$  的含义为单位时间内电池损耗造成的经济损失。电动汽车并非等同于储能设备,其无法无限制地进行放电,过度放电可能影响用户正常的出行体验。因此,本文在奖励函数中引入了一个惩罚项  $\max(I\text{SoC}_d - I\text{SoC}_t, 0)$ ,即汽车电量与预期电量的差值。通过引入这一惩罚项,用户能够间接介入汽车充放电控制,使其尽可能满足用户对电量的预期需求。上式中的  $\lambda_p, \lambda_h$  和  $\lambda_d$  为各项的权重系数,用于根据实际情况调节对各种优化目标的偏好程度。

## 3 基于强化学习和电池损耗模型的电动汽车充放电控制算法

### 3.1 基于物理混合神经网络模型的电池损耗预测模型

模型驱动法和数据驱动法是两种常用的 SOH 估算方

法。模型驱动法依赖于物理、电化学或经验模型,而数据驱动法则利用机器学习技术。然而,选择一个合适的模型,在准确性和计算复杂性之间取得平衡,仍然是一项具有挑战性的任务<sup>[26]</sup>。为了突破这一局限性,本文提出了一种物理混合神经网络模型(Physical Hybrid Neural Network, PHNN)。通过将基于模型的方法与神经网络相结合,并结合物理化学的先验知识,提高了 SoH 损耗预测的准确性和鲁棒性。

在我们的方法中,锂离子电池(LIB)的 SoH 由两个变量表征:放电周期  $t$  之前的非强制动态  $D_t$  (代表系统在没有外部干扰的情况下发生变化的固有趋势)和放电周期期间发生的估计容量损失  $L_t$ 。放电周期  $t$  结束时,SoH 的估计值可计算为:

$$SoH_t = D_t - L_t \quad (4)$$

图 1 给出了所提出的模型结构,它由两个不同的部分组成。第一部分被浅色阴影覆盖,代表负责计算  $L_t$  的基于模型驱动模块;第二部分被深色阴影覆盖,代表负责计算  $D_t$  的机器学习模块。

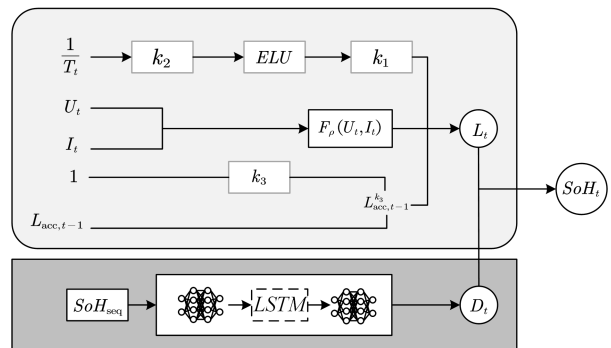


图 1 PHNN 网络结构

Fig.1 Network structure of PHNN

根据现有研究,容量衰减主要源于负极颗粒表面的固态电解质相(SEI)膜逐渐增厚这一化学过程<sup>[27]</sup>。由于阿伦尼乌斯方程(Arrhenius Equation)描述了反应速率与温度之间的关系,我们引入了温度因素的影响,这一方法在涉及电池容量的研究中已经得到广泛应用和验证<sup>[28]</sup>。此外,容量损失被认为与循环次数呈幂律关系<sup>[29]</sup>,因此可以通过以下公式估算累积容量损耗:

$$\omega = A \cdot e^{-\frac{E_a}{RT}} \cdot t^z \quad (5)$$

其中,  $\omega$  表示累积容量损耗;  $A$  是一个常量;  $E_a$  表示反应活化能;  $R$  表示通用气体常数;  $T$  表示锂电池的工作温度;  $z$  表示幂指系数。

因为模型的目标是以放电周期为单位估算电池容量损耗,需要将式(16)转化为离散版本。首先获得其一阶导为:

$$\omega' = A \cdot e^{-\frac{E_a}{RT}} \cdot z \cdot t^{z-1} \quad (6)$$

通过结合式(5)和式(6),可以得到电池容量逐周期的损耗为:

$$\omega' = A^{\frac{1}{z}} \cdot e^{-\frac{E_a}{RTz}} \cdot z \cdot \omega_t^{\frac{z-1}{z}} \quad (7)$$

在单步预测中,循环次数之差可视为 1,因此  $t$  次放电期间产生的容量损耗  $L_t$  可计算为:

$$L_t = \omega_t - \omega_{t-1} \approx A^{\frac{1}{z}} \cdot e^{-\frac{E_a}{RTz}} \cdot z \cdot \omega_t^{\frac{z-1}{z}} \quad (8)$$

其中,损耗的计算仅考虑了温度  $T$  和循环次数  $t$  对电池容量的影响。然而,在高放电率下,电池容量衰减比标准放电率下

更为迅速<sup>[30]</sup>,这可能会进一步影响容量损失估计的准确性。

因此,我们额外引入了一个由全连接层构成的神经网络  $F_\rho(U_t, I_t)$ ,用于估计放电率效应的损失。其中,  $\rho$  代表神经网络参数,  $U_t$  代表整个周期的平均电压,  $I_t$  代表平均电流。由于网络是端到端训练的,因此参数  $\rho$  可以与网络的其他部分同时学习。然后,式(8)可以进一步扩展为:

$$L_t = A^{\frac{1}{z}} \cdot e^{-\frac{E_a}{RTz}} \cdot z \cdot \omega_{t-1}^{\frac{z-1}{z}} \cdot F_\rho(U_t, I_t) \quad (9)$$

$$\text{最后,通过引入 3 个因子 } k_1 = zA^{\frac{1}{z}}, k_2 = -\frac{E_a}{Rz}, k_3 =$$

$\frac{z-1}{z}$ ,式(9)可以简化为:

$$L_t = k_1 \cdot e^{\frac{k_2}{T}} \cdot \omega_{t-1}^{k_3} \cdot F_\rho(U_t, I_t) \quad (10)$$

根据式(10),我们构建了一个具有特殊结构的神经网络,如图 1 深色阴影部分所示。该部分中的所有组件都由一个神经节点组成。ELU(Exponential Linear Unit)节点用于指数计算,而其余节点则表示提出的 3 个因子,代表相应系数。通过利用神经网络固有的学习和参数更新能力,该模型可以根据当前电池的特性自主学习适当的参数。因此,在先验知识的指导下,神经网络可以根据式(10)计算得到物理先验知识的  $L_t$ 。

机器学习模块采用一个 LSTM 网络来获取电池健康度的时序信息。其输入为一个 SoH 序列  $SoH_{seq}$ ,代表过去若干周期的电池容量值,因此  $D_t = LSTM(SoH_{seq})$ 。

在电动汽车充电中,电池的实际功率可以表示为:

$$P_t = C \cdot a_t \quad (11)$$

其中,  $C$  为电池的实际容量大小。最后,电动汽车的充放电电压电流可以由式(12)得到。

$$P_t = U_t \cdot I_t \quad (12)$$

### 3.2 基于 PHNN 和 SAC 的电动汽车充放电控制算法

在电动汽车的充放电过程中,涉及到电价以及用户行为等多个动态变量,这些变量都具有一定的随机性。深度强化学习算法在处理这种问题时无需对动态变量进行精确的预测,因此能够为电动汽车的充放电控制生成更为鲁棒的策略。由于电动汽车的充放电速率是连续的,本文中采用 SAC 算法作为充放电控制的主要算法。

SAC 是基于最大熵的离线强化学习框架之一,其主要目标是解决无模型深度强化学习方法稳定性差的问题。SAC 通过平衡探索和利用的关系,并引入熵正则化来促使学习到更多样化的策略,以避免智能体陷入局部最优解。

在电动汽车充放电控制中,SAC 的优化目标为学习一个充放电策略  $\pi$ ,使得累计奖励和策略熵最大化:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{a_t \sim \pi(\cdot | s_t)} \left[ \sum_{t=0}^{\infty} \gamma^t (r_t(s_t, a_t) + \alpha \mathcal{H}(\pi(\cdot | s_t))) \right]$$

其中,  $\mathcal{H}(\cdot)$  表示熵正则化;  $\alpha$  是温度系数,用于控制累计奖励和策略熵之间的权重。

SAC 引入了一种新的软 Q 函数,其被定义为:

$$Q(s_t, a_t) = r_t(s_t, a_t) + \gamma \mathbb{E}_{s_{t+1} \sim P} [V(s_{t+1})]$$

其中,  $P$  表示状态转移的概率;  $V$  是一个软状态-价值函数,其通常可以由 Q 函数来近似:

$$V(s_t) = \mathbb{E}_{a_t \sim \pi} [Q(s_t, a_t) - \alpha \log(\pi(a_t | s_t))]$$

在 SAC 中,软 Q 函数使用一个神经网络  $\theta$  来表示。其作为一个回归任务,通过最小化贝尔曼残差来实现:

$$J_Q(\theta) = \mathbb{E}_{(s_t, a_t) \sim \mathcal{D}} \left[ \frac{1}{2} (Q_\theta(s_t, a_t) - Q_\theta^\delta(s_t, a_t))^2 \right] \quad (13)$$

其中,  $Q_\theta^\delta(s_t, a_t) = r_t(s_t, a_t) + \gamma \cdot \mathbb{E}_{s_{t+1} \sim P} V_\theta^\delta(s_{t+1})$ ,状态与动作信息从一个经验回放缓存  $\mathcal{D}$  获取。  $V_\theta^\delta$  是一个目标软 V 函数,其同样由一个神经网络估计。该网络通过移动平均算法来更新  $\hat{\theta} = \tau\theta + (1-\tau)\hat{\theta}$ 。

策略函数也通过神经网络(参数为  $\phi$ )来近似,其通过最小化 Kullback-Leibler(KL)散度来优化:

$$J_\pi(\phi) = \mathbb{E}_{s_t \sim \mathcal{D}, a_t \sim \pi_\phi} [\alpha \log \pi_\phi(a_t | s_t) - Q_\theta(s_t, a_t)] \quad (14)$$

由于电动汽车的充电率在充电控制过程中是连续的,我们将策略  $\pi$  设置为高斯分布,即:

$$\pi_\phi(a_t | s_t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(a_t - \mu)^2}{2\sigma^2}\right) \quad (15)$$

其中,  $\mu$  和  $\sigma$  分别表示均值和方差,也是策略网络的两个输出参数。为了支持策略网络的反向传播,SAC 引入了一种重参数化的技巧,策略函数可以写作  $f_\phi(\epsilon; s_t) = \mu_\phi(s_t) + \epsilon \sigma_\phi(s_t)$ ,  $\epsilon$  从标准正态分布中随机采样。

温度系数  $\alpha$  的选择影响策略网络的最终效果,而手动确认需要大量的经验与尝试,因此,SAC 通过最小化来自动调节,此处的为  $\mathcal{H}$  最小期望目标熵。

$$J(\alpha) = \mathbb{E}_{s_t \sim \mathcal{D}, a_t \sim \pi} [-\alpha \log \pi_\phi(a_t | s_t) - \alpha \mathcal{H}] \quad (16)$$

在实际应用中,SAC 为了解决 Q 值的过高估计问题,会引入两个 Q 网络( $Q_\theta^k$ )和两个目标 Q 网络( $Q_\theta^{k^*}$ )。式(13)和式(14)中的式子可以重写为:

$$J_Q(\theta^k) = \mathbb{E}_{(s_t, a_t) \sim \mathcal{D}} \left[ \frac{1}{2} (Q_\theta^k - \min_k Q_\theta^{k^*})^2 \right], k \in \{1, 2\}$$

$$J_\pi(\phi) = \mathbb{E}_{s_t \sim \mathcal{D}, a_t \sim \pi_\phi} [\alpha \log \pi_\phi - \min_k Q_\theta^k]$$

图 2 给出了 SAC 与 PHNN 结合训练的过程,当电动汽车接入充电桩后,SAC-PHNN 会接管其充放电控制。每一周期内生成相应的充放电功率,并根据式(11)、式(12)转换为 PHNN 的输入,得到预测的下一周期的电池健康度,并与状态和动作一起输入回报函数,得到对应的奖励后存入经验回放缓存。当车辆的算力空闲时,从缓存中取出曾经的状态转移,训练 SAC 算法。

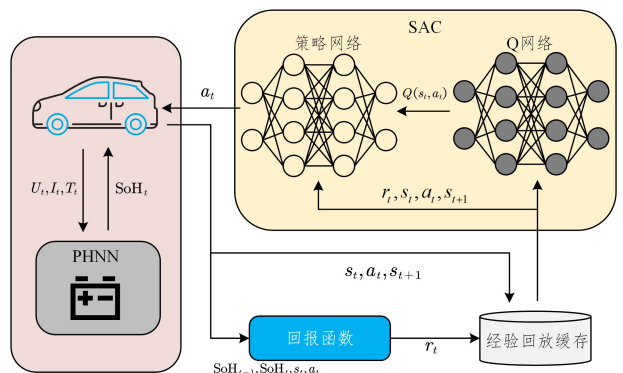


图 2 基于 PHNN 和 SAC 的电动汽车充放电控制系统

Fig. 2 Charge/discharge system of EVs based on PHNN and SAC

## 4 实验及分析

### 4.1 实验参数设置

实验软件环境为 Python3.8.15 和 Pytorch1.7.1,硬件环境为 Intel Xeon Silver 4310 CPU、128GB RAM 和 NVIDIA 4090 GPU。采用新英格兰 ISO 采集的小时电价数据<sup>[91]</sup>来进行仿真测试。用户的出行习惯参考了加利福尼亚的电动汽车

出行记录。本文车辆的状态转移和充放电控制均以一小时为时间隙。车辆初始的  $SoH$  从  $[0, 7, 1]$  中随机采样,初始的  $ISoC$  从  $[0, SoH]$  中随机采样,用户期盼的车辆离开时的  $ISoCd$  从  $[0.5, SoH]$  中随机采样;电池容量  $C$  设置为  $30\text{ kW}\cdot\text{h}$ ;充放电效率  $\eta$  设置为  $0.98$ 。电动汽车的电池通常由多个小电芯串联组成,本文假设电芯型号为 18650,充放电电流  $I_t$  固定为  $2\text{ A}$ ,温度  $T_t$  固定为  $298.15\text{ K}(25^\circ\text{C})$ ,电池由 4300 个 18650 电芯构成。PHNN 的输入为过去 8 个周期的预测  $SoH$ 。奖励函数的权重系数分别设置为 7,25 和 40。

在 SAC 框架中,Q 网络以如下配置构建:由 3 层全连接隐藏层构成,每层含有 128 个神经元。在各隐藏层后,采用修正线性单元(ReLU)激活函数以实现输出的约束。Q 网络的输入由状态  $s_t$  以及动作  $a_t$  所构成,最后通过一层线性层,得到最终的输出 Q。策略网络的构建如下:包含 4 层全连接隐藏层,各层均含有 128 个神经元。策略网络的输入为当前状态  $s_t$ ,通过两层平行线性层生成输出,分别对应平均值( $\mu$ )与标准偏差( $\sigma$ )。经验回放缓存( $\mathcal{D}$ )的规模被设定为 10000,批处理大小设置为 512。Q 网络、策略网络和温度系数的学习率分别设置为  $10^{-4}$ ,  $3\times 10^{-4}$  和  $2\times 10^{-4}$ ,奖励的折扣系数  $\gamma$  设置为  $0.98$ 。整个 SAC 的训练共迭代 10000 个回合。

对于 PHNN,本研究中使用的是 NASA 卓越预测中心的数据集<sup>[32]</sup>。在室温  $25^\circ\text{C}$  下,一组锂电池经历了充电、放电和阻抗 3 个独立操作的循环。电池最初以  $1.5\text{ A}$  的恒定电流充电,直到电压达到  $4.2\text{ V}$ ;然后以恒定电压充电,直到充电电流达到  $20\text{ mA}$ ;最后电池以连续  $2\text{ A}$  的速度放电,直到电压达到校准值。该数据集被广泛用于电池容量损失的相关工作中,以验证所提出的模型在锂离子电池中的 SOH 预测或 estimation 性能。

在 PHNN 网络中,采用如下配置:机器学习模块由 3 层全连接隐藏层构成,逐层的神经元数量从 8 递增至 128;随后是 LSTM 神经网络,含有 128 个神经元;最后通过一层线性层,得到最终的输出  $D_t$ 。模型驱动模块  $F_p(U_t, I_t)$  由 5 层全连接层组成,逐层的神经元数量从 2 递增至 128,再递减至 1。模型驱动模块中的剩余部分已在 2.1 节中详细说明。

### 4.2 PHNN 损耗预测性能分析

为评估所提出 PHNN 网络模型在预测锂电池容量损耗方面的性能,本实验选取了编号 B05 循环数据的前 50% 作为训练集,然后在后 50% 循环数据上进行测试。模型网络预测输出的  $SoH$  变化曲线与数据集曲线如图 3 所示,在进行测试的 83 个周期中,预测输出的  $SoH$  曲线与数据集的真实曲线在整体趋势上保持高度一致,其平均绝对误差为  $0.65\%$ 。这充分证明了所提出的 PHNN 在锂电池损耗预测方面的准确性。

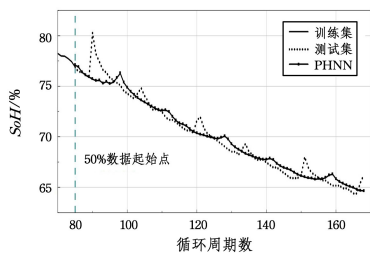


图 3 锂电池损耗预测曲线

Fig. 3 Predicted curves of lithium battery capacity loss

### 4.3 SAC 收敛性分析

SAC 算法的训练过程呈现出如图 4 所示的回报增长

趋势,在经过约 1000 轮的迭代后逐渐趋于收敛。图中的阴影区域展示了训练过程中实际回报变化的曲线,而实线部分则代表经过平滑处理后的回报曲线。尽管由于采样数据的随机性,实际回报曲线呈现出较大的波动,但经过平滑处理后的回报曲线逐渐呈上升趋势,最终达到相对平稳的状态。这表明 SAC 算法能够使智能体逐步学习并形成稳定的充放电控制策略。

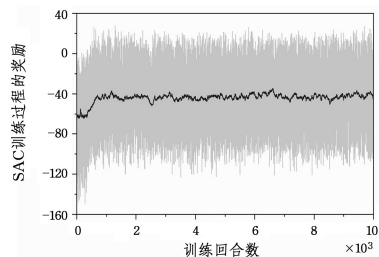


图 4 SAC 训练过程中的奖励曲线

Fig. 4 Training reward of SAC

### 4.4 电动汽车充放电模拟

为了评估所提出算法在电动汽车充放电控制策略生成方面的有效性,本文构建了一个连续 5 个工作日的持续性出行计划。在电动汽车行驶过程中,其每小时电能消耗率为  $5\%$ 。模拟结果如图 5、图 6 所示,结果表明电动汽车倾向于在低电价时进行充电,并在高电价时进行放电。电价通常反映电力网络的负荷状态,电动汽车在充放电的同时参与电力网络调控,从而实现了负荷峰值的削减和负荷谷值的填补,从而降低用户的充电成本。鉴于奖励函数中融入了用户对所需电量的期望,充放电策略调度还确保在用户出行时预留足够的电量,以满足其需求。

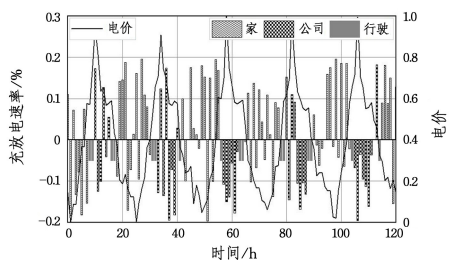


图 5 电动汽车充放电模拟

Fig. 5 Charging/discharging simulation of electric vehicle

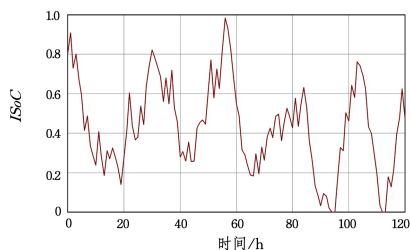


图 6 电动汽车充放电模拟中的 ISoC 变化

Fig. 6 ISoC changes in charging/discharging simulation of electric vehicle

### 4.5 电动汽车电池健康度分析

电动汽车电池的寿命常以循环次数来衡量,其充放电过程中持续的电能交换将加速电池的老化。此外,电池的使用寿命可能与充放电过程中的电压和电流参数相互关联。过多的充放电循环将导致电池寿命的缩短,从而促使用户提前进

行电池更换,进而增加整体的使用成本负担。为了评估引入 PHNN 模型后,充放电策略对电池寿命的影响,本研究在前文所述的实验设置基础上,模拟了一个连续进行了 180 个工作日的充放电过程,并利用 PHNN 模型绘制了电池  $SoH$  衰减曲线。研究对比了基于电池损耗模型的 SAC 算法与忽略电池损耗的 SAC 算法对电池健康状况的影响,具体结果如图 7 所示。基于电池损耗模型的 SAC 算法在充放电过程中能够实现更小幅度的  $SoH$  衰减,这表明本研究提出的算法在优化用户电池成本方面表现出更为卓越的性能。

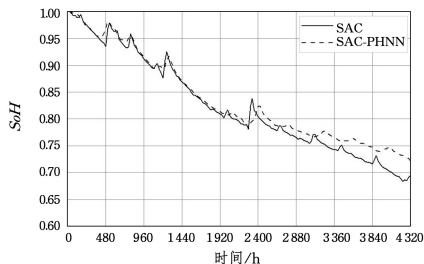


图 7 SAC 和 SAC-PHNN 控制下电池的  $SoH$  下降曲线  
Fig. 7  $SoH$  decline curves of battery controlled by SAC and SAC-PHNN

#### 4.6 算法对比

本小节选取了几种常见的强化学习算法与 PHNN 结合来做电动汽车充放电控制,并对比它们的效果。我们选取了被广泛使用的 DDPG 和 PPO(Proximal Policy Optimization) 算法<sup>[33]</sup>,这两种算法在类似的时序控制问题中皆有良好的效果。

所有的算法都基于 4.3 小节设计的 180 天的连续充放电控制进行仿真控制。我们对比两个指标,分别为累积奖励和电池健康度。累积奖励为车辆充放电控制中奖励函数的累加值,电池健康度即为电池最终的  $SoH$ ,两个指标越大说明算法的控制效果越好。结果如表 1 所列,结果 SAC-PHNN 算法在累积奖励和电池健康度两个维度皆优于其余两种算法。这表明本文提出的算法在电动汽车充放电控制方面具有优异的效果。

表 1 几种算法的累积奖励和电池健康度

Table 1 Accumulated rewards and battery health of several algorithms

	累积奖励	电池健康度
DDPG-PHNN	-9012	0.71
PPO-PHNN	-7963	0.73
SAC-PHNN	-7559	0.73

**结束语** 本文针对电动汽车的充放电控制问题,提出了一种基于 PHNN 和 SAC 的电动汽车充放电控制算法。该算法设计了一个新的奖励函数,通过调整充放电的功率,使得车辆能够在用电低谷时储存电力,并在用电高峰时释放电力以缓解电网压力。同将电动汽车充放电过程中的电池损耗也纳入考虑范畴,保证用户在此期间的经济收益。对比实验表明,本文提出的基于 PHNN 的 SAC 在控制电动汽车充放电时对电池造成的损耗更小,具有更强的实际意义。

#### 参考文献

[1] ZHANG X,GAO F,GONG X,et al. Comparison of climate change impact between power system of electric vehicles and in-

ternal combustion engine vehicles[C]// Advances in Energy and Environmental Materials:Proceedings of Chinese Materials Conference 2017 18th. 2018:739-747.

- [2] LOPES J A P,SOARES F J,ALMEIDA P M R. Integration of electric vehicles in the electric power system[C]// Proceedings of the IEEE. 2010:168-183.
- [3] KEMPTON W,TOMIC J. Vehicle-to-grid power implementation:From stabilizing the grid to supporting large-scale renewable energy[J]. Journal of Power Sources, 2005, 144 (1): 280-294.
- [4] SOVACOL B K,KESTER J,NOEL L,et al. Actors, business models, and innovation activity systems for vehicle-to-grid (V2G) technology:A comprehensive review[J]. Renewable and Sustainable Energy Reviews, 2020, 131:109963.
- [5] BALTHAZAR P,HARRI P,PRATER A,et al. Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics[J]. Journal of the American College of Radiology, 2018, 15(3): 580-586.
- [6] CAO Y,WANG H,LI D,et al. Smart online charging algorithm for electric vehicles via customized actor-critic learning[J]. IEEE Internet of Things Journal, 2021, 9(1): 684-694.
- [7] WU D,ZENG H,LU C,et al. Two-stage energy management for office buildings with workplace EV charging and renewable energy[J]. IEEE Transactions on Transportation Electrification, 2017, 3(1): 225-237.
- [8] SORTOMME E,EL-SHARKAWI M A. Optimal combined bidding of vehicle-to-grid ancillary services[J]. IEEE Transactions on Smart Grid, 2011, 3(1): 70-79.
- [9] HAN S,HAN S,SEZAKI K. Development of an optimal vehicle-to-grid aggregator for frequency regulation[J]. IEEE Transactions on Smart Grid, 2010, 1(1): 65-72.
- [10] HUANG Q,JIA Q S, QIU Z, et al. Matching EV charging load with uncertain wind: A simulation-based policy improvement approach[J]. IEEE Transactions on Smart Grid, 2015, 6(3): 1425-1433.
- [11] CHUNG H M,MAHARJAN S,ZHANG Y,et al. Intelligent charging management of electric vehicles considering dynamic user behavior and renewable energy:A stochastic game approach [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(12): 7760-7771.
- [12] ALSABBAGH A,WU B,MA C. Distributed electric vehicles charging management considering time anxiety and customer behaviors[J/OL]. IEEE Transactions on Industrial Informatics, 2020, 17 (4): 2422-2431. [https://www.eia.gov/conference/2008/conf\\_pdfs/Tuesday/Hogan.pdf](https://www.eia.gov/conference/2008/conf_pdfs/Tuesday/Hogan.pdf).
- [13] WAN Z,LI H,HE H,et al. Model-free real-time EV charging scheduling based on deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2018, 10(5): 5246-5257.
- [14] ZHANG F, YANG Q, AN D. CDDPG: A deep-reinforcement-learning-based approach for electric vehicle charging control[J]. IEEE Internet of Things Journal, 2020, 8(5): 3075-3087.
- [15] LI H,WAN Z,HE H. Constrained EV charging scheduling based on safe deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 11(3): 2427-2439.
- [16] YAN L,CHEN X,ZHOU J,et al. Deep reinforcement learning for continuous electric vehicles charging control with dynamic

- user behaviors[J]. IEEE Transactions on Smart Grid, 2021, 12(6):5124-5134.
- [17] ARORA P, WHITE R E, DOYLE M. Capacity fade mechanisms and side reactions in lithium-ion batteries[J]. Journal of the Electrochemical Society, 1998, 145(10):3647.
- [18] SONG Y, PENG Y, LIU D. Model-based health diagnosis for lithium-ion battery pack in space applications[J]. IEEE Transactions on Industrial Electronics, 2020, 68(12):12375-12384.
- [19] HU X, YUAN H, ZOU C, et al. Co-estimation of state of charge and state of health for lithium-ion batteries based on fractional-order calculus[J]. IEEE Transactions on Vehicular Technology, 2018, 67(11):10319-10329.
- [20] LI Y, WEI Z, XIONG B, et al. Adaptive ensemble-based electrochemical-thermal degradation state estimation of lithium-ion batteries[J]. IEEE Transactions on Industrial Electronics, 2021, 69(7):6984-6996.
- [21] FU Y, XU J, SHI M, et al. A fast impedance calculation-based battery state-of-health estimation method[J]. IEEE Transactions on Industrial Electronics, 2021, 69(7):7019-7028.
- [22] ZHANG Y, LIU Y, WANG J, et al. State-of-health estimation for lithium-ion batteries by combining model-based incremental capacity analysis with support vector regression[J]. Energy, 2022, 239:121986.
- [23] REN L, DONG J, WANG X, et al. A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life[J]. IEEE Transactions on Industrial Informatics, 2020, 17(5):3478-3487.
- [24] QIN Y, YUEN C, YIN X, et al. A Transferable Multistage Model With Cycling Discrepancy Learning for Lithium-Ion Battery State of Health Estimation[J]. IEEE Transactions on Industrial Informatics, 2022, 19(2):1933-1946.
- [25] XU P, WANG C, YE J, et al. State-of-Charge Estimation and Health Prognosis for Lithium-Ion Batteries Based on Temperature-Compensated Bi-LSTM Network and Integrated Attention Mechanism[J]. IEEE Transactions on Industrial Electronics, 2023, 66(10):8773-8783.
- [26] FAN Y, XIAO F, LI C, et al. A novel deep learning framework for state of health estimation of lithium-ion battery[J]. Journal of Energy Storage, 2020, 32:101741.
- [27] DUBARRY M, QIN N, BROOKER P. Calendar aging of commercial Li-ion cells of different chemistries-A review[J]. Current Opinion in Electrochemistry, 2018, 9:106-113.
- [28] LI K, ZHOU P, LU Y, et al. Battery life estimation based on cloud data for electric vehicles[J]. Journal of Power Sources, 2020, 468:228192.
- [29] WANG J, LIU P, HICKS-GARNER J, et al. Cycle-life model for graphite-LiFePO4 cells[J]. Journal of Power Sources, 2011, 196(8):3942-3948.
- [30] NING G, HARAN B, POPOV B N. Capacity fade study of lithium-ion batteries cycled at high discharge rates[J]. Journal of Power Sources, 2003, 117(1/2):160-169.
- [31] ISO New England. Real-time maps and charts [EB/OL]. <https://www.iso-ne.com/isoexpress/>.
- [32] SAHA, B. AND GOEBEL, K. Moffett Field, CA, USA; NASA Ames Prognostics Data Repository [EB/OL]. <http://ti.arc.nasa.gov/project/prognostic-data-repository>.
- [33] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[J]. arXiv:1707.06347, 2017.



**LU Yue**, born in 1989, postgraduate, engineer. His main research interests include power supply marketing and comprehensive energy.



**WANG Qiong**, born in 1983, doctoral candidate, engineer. Her main research interests include the development of software and hardware for charging stations and power-side infrastructure, alongside the integration of advanced security initiatives in vehicle networking platforms.