

# 基于强化学习的电动汽车换电站实时调度策略优化

张文昕<sup>1</sup>, 栗然<sup>1</sup>, 臧向迪<sup>1</sup>, 严敬汝<sup>2</sup>, 祝晋尧<sup>3</sup>

(1. 华北电力大学 电气与电子工程学院, 河北 保定 071003;

2. 国网河北省电力有限公司电力科学研究院, 河北 石家庄 050022; 3. 国网石家庄供电公司, 河北 石家庄 050004)

**摘要:**随着电动汽车的应用推广,换电站的调度优化逐渐成为研究热点。传统的基于换电需求预测值的调度策略在实际应用中面临着难以适应动态干扰因素、预测误差累积等问题。为了解决这些问题,提出了一种基于带基线的蒙特卡罗策略梯度法的换电站实时调度策略,用于优化换电站的充放电策略以及响应电池数量。提出了带基线的蒙特卡罗策略梯度强化学习,并为换电站实时调度问题选取合适的状态空间和动作空间;设计了奖励函数对智能体进行离线训练,从电池状态数据、分时电价和排队电动汽车数量中学习得到最优策略网络;在离线训练好的模型基础上进行实时调度策略测试。基于换电站的服务可用率和经济效益验证了所提调度策略的有效性和经济性,算例结果表明所提策略能对电网负荷起到一定的削峰填谷作用。

**关键词:**电动汽车;换电站;强化学习;策略梯度;分时电价;实时调度

**中图分类号:**U 469.72;TM 734

**文献标志码:**A

**DOI:**10.16081/j.epae.202203003

## 0 引言

电动汽车(EV)凭借其低碳环保、灵活可控的特点被普遍认为是提高汽车产业竞争力、转型低碳经济、保障能源安全的重要途径<sup>[1]</sup>。随着EV规模化应用,大量EV集中在电网负荷高峰时段接入电网充电,将进一步增大负荷峰值,带来系统容量不足、影响电能质量等一系列问题。目前,EV的电能补充模式主要包括整车慢充、整车快充、电池更换3种。其中,电池更换模式下EV与电池的可分离特性能使电池充电脱离EV的停驶时间限制,更有利于集中管理电池充放电以避免大规模EV随机充电对电网运行造成的不利影响<sup>[2]</sup>。将电池视为电网的分布式微储能单元对其充放电进行有序调控,可以实现削峰填谷<sup>[3]</sup>、促进新能源消纳<sup>[4]</sup>、进行电压和频率调节<sup>[5]</sup>等功能,实现EV与电网的双向友好互动<sup>[6]</sup>。同时,相比于电池充电站,在相同的服务容量下电池换电站能为EV运营商提供更多的经济收益<sup>[7]</sup>。故研究在换电站并入电网的情况下站内电池的充放电优化调度已成为智能电网发展的必然趋势。

已有文献从不同的角度对上述问题进行了研究,大多先预测EV在未来时刻的换电需求,然后基于预测结果进行调度决策。文献[8-9]提出了充/换电站的日前调度优化模型,对次日整天的充放电计划进行优化。但日前调度只能对次日计划进行粗略预测,无法考虑EV通勤行为、交通状态、能源消耗带来的不确定性。而且这种提前预测十分依赖于对具体应用场景内换电需求的物理过程建模,无法简单推广至更多的场景,泛化能力较弱。还有一部分研

究将1d划分为多个时段,并在每个时段内进行独立决策,以实现实时调度。文献[10-11]首先进行日前预测,然后在日内每个时段求解换电站的最优决策;文献[12-13]建立了换电站和电网的双层调度模型,先预测每个时段的换电需求,然后对上、下层问题分别进行迭代求解,得到换电站与电网的最优决策。但是求解每个时段的优化问题需要进行大量的迭代计算,算力消耗大且成本高,难以支撑大规模充放电策略的在线计算。上述研究均先进行换电需求预测,并将预测值视为真实值进行调度,这不但需要建立额外的预测模块,而且该预测模块也会带来误差累积,使得算法无法适应真实的换电需求。此外,上述实时调度都是通过分时段独立计算来实现“伪”实时决策,相邻时段的决策之间没有关联,这可能会导致决策振荡问题。

为了解决以上问题,本文提出了基于带基线的蒙特卡罗策略梯度法的换电站实时调度策略优化方法。策略梯度法<sup>[14]</sup>属于无模型(model-free)的强化学习算法,无模型主要体现在:算法不依赖于对环境的准确建模,不需要对换电需求和EV用户行为进行预测,也不需要人为提供启发式的决策规则。因此,决策的制定不依赖于任何对未发生事件的假设。决策器通过与环境的交互来学习调度策略以寻求回报最大化。在进行训练时,采用蒙特卡罗采样对当前状态-动作之后时段的回报均值进行无偏估计,然后使用该估计值指导决策器的更新。相比于基于预测的方法只考虑预测值的情况,策略梯度法实际上考虑了当前状态-动作之后所有可能出现的情形,使得算法具有较强的对换电需求不确定性的适应能力。同时强化学习的数学基础理论是马尔可夫决策过程(MDP),使得相邻时段的决策是相互关联的,这在理

论上能够缓解决策振荡问题。

本文所提换电站实时调度策略包含充放电策略和参与调度电池数量两部分。首先,给出策略梯度强化学习的框架,然后确定换电站实时调度问题的状态空间、动作空间、奖励函数、决策器,构建基于蒙特卡罗策略梯度法的换电站实时调度模型。该模型以各时段换电站的充放电状态、参与调度电池数量为控制变量,综合考虑电池数量、电量等约束条件对实时调度最优策略进行求解。此外,本文还讨论了不同充电功率对调度策略的影响及电池选择方式对换电站服务可用率、经济效益和电网基础负荷的影响。

## 1 问题构建

换电站的结构示意图如图 1 所示,其主要由充放电电机、配电变压器、动力电池组及其存放仓、更换装置等部分构成。

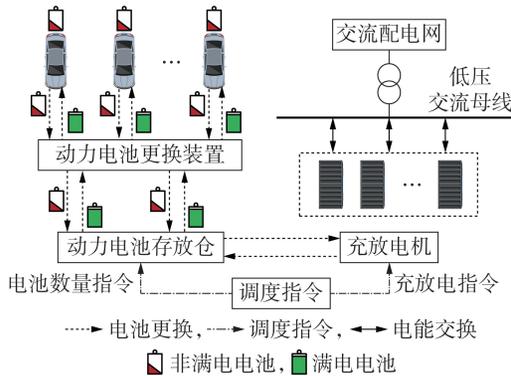


图 1 换电站的结构示意图

Fig.1 Structure diagram of battery swapping station

图 1 描述了换电站的工作模式:到站 EV 在动力电池更换装置上完成电池更换,卸载的电池被存入动力电池存放仓等待调度指令。调度指令分为充放电指令、电池数量指令两部分,其中充放电指令决定充放电电机的充放电状态,电池数量指令决定新放入插槽的电池数量。接收到调度指令后,被选中的电池接入充放电电机。当电池充满电或放电至允许下限或充放电指令改变时,将电池从充放电电机上卸下,并重新存入动力电池存放仓等待下一次调度指令。

作为电网与 EV 用户的中间环节,换电站的利润来自对 EV 用户换电所得换电收益、对电网馈电所得馈电收益与电网购电成本之间的差额。在保证换电服务可用率的情况下,换电站通过响应分时电价对站内电池进行统一充放电调度,能减少电网购电成本,增大对电网馈电所得收益,从而获得更大的利润空间。因此,换电站的实时调度策略应着眼于安排合理的充放电时间以及接入电网的电池数量。

### 1.1 强化学习概述

强化学习是一类求解序列决策问题的有效算

法,其基本框架见附录 A 图 A1。强化学习框架的主要角色为智能体和环境。时段  $i$  环境所处的状态为  $s_i$ ,智能体在某策略下根据该状态执行动作  $a_i$ ,该动作会影响环境的状态,使得环境在下一个时段转移到状态  $s_{i+1}$ 。同时,智能体会获得环境的反馈,又被称为奖励  $r_{i+1}$ 。该奖励在一定程度上衡量了在状态  $s_i$  下执行动作  $a_i$  的优劣。若智能体的某个策略获得了环境的正向奖励,则之后智能体产生该策略的趋势会加强。基于此,在与环境不断交互的过程中,智能体以“试错”的方式进行学习,根据奖励调整策略使整个过程的累计奖励达到最大。

在强化学习中,环境的数学基础是 MDP,即环境的状态转移只与当前状态有关,而与历史状态无关。MDP 由一个四元组构成,可表示为  $(S, A, P_{ss'}^a, R_s^a)$ 。其中,  $S$  为状态空间集;  $A$  为动作空间集;  $P_{ss'}^a$  为状态转移概率,表示在状态  $s$  下执行动作  $a$  后转移到状态  $s'$  的概率分布;  $R_s^a$  为奖励函数,表示在状态  $s$  下执行动作  $a$  获得的奖励。

### 1.2 带基线的蒙特卡罗策略梯度强化学习

为了解决传统换电站调度策略优化方法对预先给定数据、换电需求预测值依赖性高的问题,本文采用策略梯度强化学习对换电站实时调度策略进行优化。将连续时间以时间间隔  $\Delta t$  进行离散化,智能体在每个时段根据当前状态进行 1 次决策。智能体在环境上运行  $I$  个时段后得到 1 条经验轨迹  $\tau$ ,其具体构成为  $\langle s_0, a_0, r_0, s_1, a_1, r_1, \dots, s_{I-1}, a_{I-1}, r_{I-1}, s_I \rangle$ 。

本文预解决的换电站实时调度问题属于无模型的强化学习,无法得知状态转移概率  $P_{ss'}^a$ 。智能体的动作策略用  $\pi_\theta(a|s)$  描述,表示在状态  $s$  的条件下选择动作  $a$  的概率,其中  $\theta$  为参数。最优动作策略的参数  $\theta$  可根据式 (1) 所示优化问题得到。

$$\max_{\theta} J(\theta) = E_{\tau \sim p_{\theta}(\tau)} [G_0(\tau)] = E_{\tau \sim p_{\theta}(\tau)} \left[ \sum_{i=0}^{I-1} \gamma^i r_i(\tau) \right] \quad (1)$$

式中:  $J(\theta)$  为目标函数,即在动作策略  $\pi_\theta$  下智能体初始状态的回报期望;  $G_0(\tau)$  为初始状态的回报,即轨迹  $\tau$  中所有通过折扣因子衰减的奖励总和;  $\gamma \in [0, 1]$  为折扣因子,当  $\gamma=1$  时,  $G_0(\tau)$  为轨迹  $\tau$  中的奖励总和;  $r_i(\tau)$  为轨迹  $\tau$  中时段  $i$  的奖励;  $E_{\tau \sim p_{\theta}(\tau)}[\cdot]$  为数学期望,  $\tau \sim p_{\theta}(\tau)$  表示轨迹  $\tau$  在动作策略  $\pi_\theta$  下的分布为  $p_{\theta}(\tau)$ 。

采用带基线的蒙特卡罗估计  $J(\theta)$  对  $\theta$  的梯度如式 (2) 所示,具体推导过程见附录 B。

$$\nabla_{\theta} J(\theta) \approx \frac{1}{N} \sum_{j=1}^N \sum_{i=0}^{I-1} \nabla_{\theta} \ln \pi_{\theta}(a_i | s_i) (G_i(\tau_j) - V_{\pi_{\theta}}(s_i)) \quad (2)$$

式中: “ $\approx$ ” 表示等号右侧为左侧的无偏估计;  $N$  为从  $p_{\theta}(\tau)$  中采样的轨迹数量;  $\tau_j$  为采样的第  $j$  条轨迹;  $G_i(\tau_j)$  为轨迹  $\tau_j$  中时段  $i$  的回报;  $V_{\pi_{\theta}}(s_i)$  为动作策略

$\pi_\theta$ 下状态 $s_i$ 的价值。

引入基线虽然不改变策略梯度的期望,但降低了梯度的方差,使算法更加稳定。所以,可采用梯度上升<sup>[15]</sup>的方法优化 $J(\theta)$ 。算法的伪代码见附录C。

### 1.3 状态空间

在强化学习中,状态代表了智能体所能感知到的环境信息。状态通常由1个高维向量表示,其每一个元素表示环境的一个可观测特征,状态向量全体组成状态空间。状态空间的选取标准是尽可能选择与智能体决策相关的最小环境特征子集。基于该标准,智能体在决策时既有充分的依据,又避免了冗余特征带来的过拟合和额外计算开销问题。

本文选择6个典型特征构成1个状态向量 $s_i = [n_i^f, n_i^w, n_i^p, e_i^c, e_i^d, v_i^l]$ ,其元素分别为时段 $i$ 的可用电池数量、不可用电池数量、在插电池数量(在充电与在放电的数量之和)、购电电价、上网电价、排队等待EV数量。

### 1.4 动作空间

本文所提换电站的实时调度策略包含充放电状态和参与调度的电池数量这2个动作指令。充放电状态用一个二进制变量 $\delta$ 表示, $\delta=1$ 时表示充放电机拨入充电档位, $\delta=0$ 时表示充放电机拨入放电档位。由于换电站的电池库存众多,将具体插入的电池数量作为动作之一会使动作空间过大,导致模型难以训练。为了减小动作空间,本文将 $u$ 块电池作为一个动作整体,即在 $\Delta t$ 时间内接入充放电机的电池数量只能为0或 $u$ 。这样,参与调度的电池数量可用一个二进制变量 $\mu$ 表示,当 $\mu=1$ 时表示充放电机新接入 $u$ 块电池,当 $\mu=0$ 时表示充放电机无新接入的电池。因此,一个完整的动作可表示为2维向量 $\mathbf{a}=[\delta, \mu]$ 。

### 1.5 奖励函数

智能体的目标是最大化累计奖励总和,通过人为设计合适的奖励函数可以引导智能体给出合理的换电站调度策略。

本文考虑换电站基于分时电价的需求响应模式,认为换电站的充放电行为是通过电价引导的。因此设计奖励函数 $r_i$ 包括时段 $i$ 对EV用户换电所得收益 $r_i^1$ 、对电网馈电所得收益 $r_i^2$ 和服务可用率惩罚项 $r_i^3$ ,如式(3)所示。

$$r_i = \beta_1 r_i^1 + \beta_2 r_i^2 + \beta_3 r_i^3 \quad (3)$$

式中: $\beta_1$ 、 $\beta_2$ 、 $\beta_3$ 为权重值,均属于超参数,可通过网格搜索(grid search)确定其具体取值。

### 1.6 决策器

本文以 $\Delta t$ 为时间间隔对连续的时间进行离散化,换电站在每个时段根据当前状态进行1次决策。换电站的决策是随机的,由条件概率 $\pi_\theta(\mathbf{a}|s)$ 描述。

为了学习动作和状态之间的复杂依赖关系,本

文调度策略的决策器 $\pi_\theta(\mathbf{a}|s)$ 采用神经网络的形式,被称为策略网络。本文假设充放电状态指令 $\delta$ 和参与调度的电池数量指令 $\mu$ 在给定状态 $s$ 的情况下条件独立,即满足式(4)。

$$\pi_\theta(\mathbf{a}|s) = \pi_{\theta^\delta}(\delta|s) \pi_{\theta^\mu}(\mu|s) \quad (4)$$

式中: $\pi_{\theta^\delta}(\delta|s)$ 、 $\pi_{\theta^\mu}(\mu|s)$ 分别为状态 $s$ 下选择充放电状态指令 $\delta$ 、参与调度的电池数量为 $\mu$ 的概率,这2个概率均使用带2个隐层的神经网络进行建模,并且在2个隐层进行参数共享。在输出概率时,采用Sigmoid函数对概率值进行归一化处理。决策器的神经网络结构示意图如图2所示。

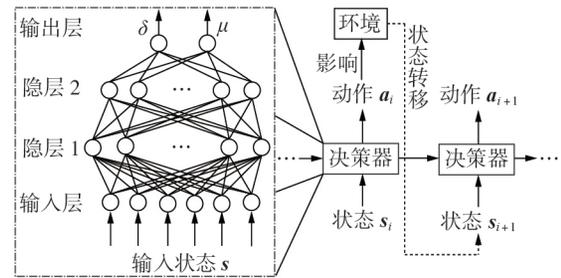


图2 决策器的神经网络结构示意图

Fig.2 Neural network structure diagram of decision maker

策略网络的目标是在保证服务可用率的同时,尽可能最大化对电网馈电所得收益和对EV用户换电所得收益。该优化目标即为策略梯度强化学习的目标,因此策略网络的最优参数可由式(5)所示优化问题给出。

$$\max_{\theta} J(\theta) = E_{\tau \sim p_{\theta}(\tau)} \left[ \sum_{i=0}^{l-1} \beta_1 r_i^1(\tau) + \beta_2 r_i^2(\tau) + \beta_3 r_i^3(\tau) \right] \quad (5)$$

式中: $r_i^1(\tau)$ 、 $r_i^2(\tau)$ 、 $r_i^3(\tau)$ 分别为轨迹 $\tau$ 中时段 $i$ 对EV用户换电所得收益、对电网馈电所得收益、服务可用率惩罚项。

在策略网络离线训练完成后,通常提取贪心策略 $\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmax}}(\mathbf{a}|s)$ 作为最终的确定性策略。而已有研究大多先预测下一时段的状态 $s^p$ ,然后根据该预测状态进行决策,得到 $\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmax}}(\mathbf{a}|s^p)$ 。相较而言,本文所提方法不需要建立额外的预测模块,避免了预测模块带来的误差。

## 2 换电站实时调度策略的数学形式

本节首先描述状态与环境交互之后进行状态转移的数学规则,然后根据换电站的收益明确奖励函数的具体计算过程。

### 2.1 电池的状态转移

在任意时段,换电站内的电池总数量恒为 $M$ 块。

换电站内的电池有以下 4 种状态:①可用电池,即动力电池存放仓中的满电电池,可支持换电服务,数量为  $n_i^f$  块;②不可用电池,即动力电池存放仓中的非满电电池,数量为  $n_i^w$  块;③在充电电池,即接入充放电机且正在充电的电池,数量为  $n_i^c$  块;④在放电电池,即接入充放电机且正在放电的电池,数量为  $n_i^d$  块。同一时段充放电机只能处于充电或放电中的一种状态,因此  $n_i^c$  和  $n_i^d$  中至少有 1 个数值为 0。电池状态转移关系示意图如图 3 所示。

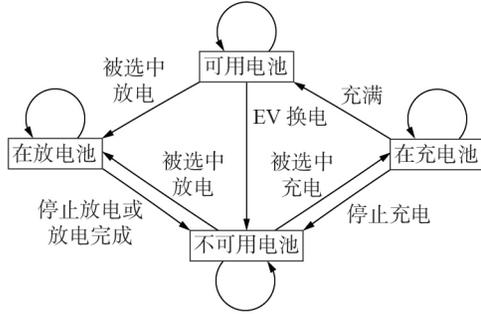


图 3 电池状态转移关系示意图

Fig.3 Schematic diagram of battery state transition

电池状态的转移关系可表示为:

$$\begin{cases} n_{i+1}^f = n_i^f + \Delta n_i^{cf} - \Delta n_i^{fd} - v_i^s \\ n_{i+1}^w = n_i^w + \Delta n_i^{dw} + \Delta n_i^{cw} - \Delta n_i^{wc} - \Delta n_i^{wd} + v_i^s \\ n_{i+1}^c = n_i^c + \Delta n_i^{cw} - \Delta n_i^{wc} - \Delta n_i^{cf} \\ n_{i+1}^d = n_i^d + \Delta n_i^{fd} + \Delta n_i^{wd} - \Delta n_i^{dw} \\ v_{i+1}^l = v_i^l + v_i^c - v_i^s \end{cases} \quad (6)$$

式中:  $\Delta n_i^{kk'}$  为时段  $i$  从  $k$  状态转移到  $k'$  状态的电池数量,  $k, k' \in \{f, w, c, d\}$  且  $k \neq k'$ ;  $v_i^s$  为时段  $i$  接受换电服务的 EV 数量;  $v_i^l, v_i^c$  分别为时段  $i$  排队等待换电、新参与排队的 EV 数量。

## 2.2 电池的荷电状态

电池的荷电状态 SOC(State Of Charge)是指电池在一定的放电倍率下,剩余电量与相同条件下额定容量的比值。本文将电池的恒流-恒压两阶段充放电过程简化为恒功率过程, SOC 的计算公式为:

$$Q_{i+1} = \begin{cases} Q_i + P_c \Delta t \eta_c / C & \delta = 1 \\ Q_i - P_d \Delta t / (\eta_d C) & \delta = 0 \end{cases} \quad (8)$$

式中:  $Q_i$  为时段  $i$  电池的 SOC;  $P_c, P_d$  分别为充、放电功率;  $\eta_c, \eta_d$  分别为充、放电效率;  $C$  为电池额定容量。

为了避免过充过放对电池造成的损害,需要对电池的充放电限度进行约束。限制接入充放电机电池的 SOC 上、下限分别为 95%、20%, 即:若电池的 SOC 达到 95%, 则视为满电电池;若电池的 SOC 低于 20%, 则不再参与放电。满电电池和不再参与放电的电池均从充放电机上拔出并存入动力电池存放仓。

## 2.3 换电站收益

换电站收益包括对 EV 用户换电所得收益和对电网馈电所得收益。

1) 对 EV 用户换电所得收益。

时段  $i$  换电站对 EV 用户换电所得收益  $r_i^1$  为向 EV 用户提供电能的收益与换电服务费用之和, 可表示为:

$$r_i^1 = \sum_{l=1}^{n_i^f} (Q_l^c - Q_l^o) C e + v_i^s e^s \quad (9)$$

式中:  $Q_l^c$  为第  $l$  辆 EV 更换新电池的 SOC;  $Q_l^o$  为第  $l$  辆 EV 卸载旧电池的 SOC;  $e$  为换电站对 EV 用户提供电能的单价;  $e^s$  为单次换电服务费用。

2) 对电网馈电所得收益。

时段  $i$  换电站对电网馈电所得收益  $r_i^2$  可表示为:

$$r_i^2 = (n_i^d e_i^d P_d / \eta_d - n_i^c e_i^c P_c \eta_c) \Delta t \quad (10)$$

## 2.4 排队意愿和服务可用率

对 EV 用户的排队意愿进行如下假设:当动力电池存放仓内有可用电池但动力电池更换装置无空闲时, EV 用户愿意排队等待;当动力电池存放仓内无可用电池时, EV 用户拒绝排队等待电池充满电即直接离开, 被视为未成功换电<sup>[16]</sup>。在这种排队意愿下, 排队等待换电的 EV 数量满足式(7)。定义日服务可用率  $\xi$  为 1 d 内成功换电的 EV 数量  $v^s$  与到站 EV 总数量  $v^c$  之比, 如式(11)所示。

$$\xi = \frac{v^s}{v^c} = \frac{\sum_{i=1}^I v_i^s}{\sum_{i=1}^I v_i^c} \quad (11)$$

则奖励函数中时段  $i$  的服务可用率惩罚项  $r_i^3$  可定义为:

$$r_i^3 = \varepsilon_1 v_i^s - \varepsilon_2 (v_i^l + v_i^c - v_i^s) \quad (12)$$

式中:  $\varepsilon_1, \varepsilon_2$  分别为换电成功奖励系数、换电失败惩罚系数, 均为正数。设置该惩罚项的目的在于:惩罚换电站一直向电网馈电, 鼓励其尽可能多地满足 EV 用户的换电需求, 以提高换电站的日服务可用率。  $\varepsilon_1$  和  $\varepsilon_2$  均为超参数, 可通过网格搜索确定其具体取值。

## 3 算例分析

为了验证本文所提调度策略的有效性, 设计了以下 2 组实验进行仿真分析:①可视化动作策略, 以验证基于带基线的蒙特卡罗策略梯度法的实时调度策略能否通过合理安排充放电时间和接入电池数量来增加获利空间, 进一步讨论不同充电功率对调度策略的影响;②给出 2 种电池选择方式, 讨论不同的电池选择方式对换电站经济收益、服务可用率、电网基础负荷的影响。

### 3.1 实验设置

#### 3.1.1 算例参数设置

本文选取小型换电站作为算例, 并进行了如下假设:①EV 使用同一型号的电池, 且电池的充放电功率恒定;②EV 换电过程所需的时长服从正态分布

$N(10, 2)$ ; ③到站换电的EV电池的剩余SOC服从(20%, 30%)范围内的均匀分布; ④每个时段内到达换电站的EV数量是一个服从泊松分布 $P(\lambda)$ 的随机变量。算例从泊松分布中进行采样是对真实EV到达换电站的场景进行仿真, 可使用真实数据对本文所提方法进行训练。

换电站的相关参数以及到站EV数量泊松分布的 $\lambda$ 取值分别见附录D表D1和表D2。购电电价参考国网北京市电力公司的峰谷分时电价<sup>[17]</sup>, 假设峰时段和平时段的上网电价为购电电价的80%, 由于不鼓励在谷时段向电网馈电, 谷时段的上网电价设为0。分时电价数据见附录D表D3。

### 3.1.2 电池的选择方式

当换电站接收到动作指令 $\mu$ 选择 $u$ 块电池接入充放电机时, 面临具体电池的选择问题。由2.2节可知, 在同一充放电功率下, 电池的SOC不同, 其充电耗时和放电空间也不同。选择不同的电池响应动作指令 $\mu$ 可能会对后续的调度指令、换电站的经济效益产生影响, 故设定以下2种电池选择方式进行探讨。

#### 1) “高先充低先放”方式。

“高先充低先放”方式是换电站的常用电池选择方式, 即为了尽可能在相同时间内获得更多的满电可用电池, 优先选择电量高的电池充电; 同时, 为了尽量维持高电量电池的数量, 优先选择电量低的电池放电。直观来看, 这种选择方式能够最大限度地满足EV用户的换电需求<sup>[18]</sup>。具体而言, 在充电模式下, 即当 $\delta=1$ 时, 选择不可用电池中SOC最大的 $u$ 块电池进行充电; 在放电模式下, 即当 $\delta=0$ 时, 选择不可用电池中SOC最小的 $u$ 块电池进行放电。

#### 2) 随机选择方式。

随机选择方式是指换电站接收到动作指令 $\mu$ 后在动力电池存放仓内随机选择电池以响应指令。

### 3.1.3 训练过程

本文设置决策时间间隔 $\Delta t=2\text{ min}$ , 即将1 d分为720个时段。对于策略网络 $\pi_{\theta}(\mathbf{a}|s)$ 而言, 其输入维度为状态空间的维度6, 输出维度为动作空间的维度2。采用带2个隐层的全连接神经网络对策略网络进行建模, 2个隐层的维度分别为12、10, 激活函数为ReLU函数。对策略网络进行更新时采用随机梯度下降SGD(Stochastic Gradient Descent)算法, 学习率设为 $10^{-3}$ 。设置折扣因子 $\gamma=0.6$ , 采样轨迹长度为1000。服务可用率惩罚项中换电成功奖励系数 $\varepsilon_1=30$ , 换电失败惩罚系数 $\varepsilon_2=50$ 。

选取最大迭代轮数为1000进行训练, 轨迹的平均奖励变化曲线如图4所示。为了更好地呈现平均奖励的变化趋势, 利用基于移动窗口的加权平均算法对平均奖励进行平滑除噪, 在长度为60的滑动窗

口内对3阶多项式进行最小二乘拟合得到滑动平均奖励。由图可以看出: 平均奖励存在一定的局部振荡, 这是因为采用蒙特卡罗采样估计策略网络的梯度带来了误差; 平均奖励随着迭代轮数的增大呈现较稳定的上升趋势, 并逐渐收敛稳定。

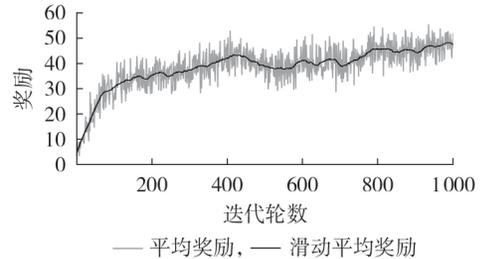


图4 平均奖励的变化曲线

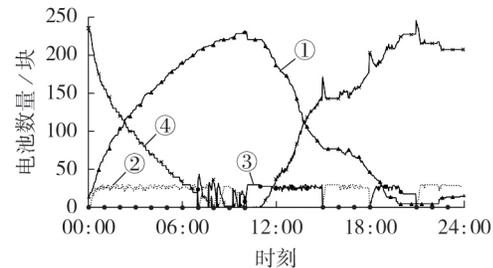
Fig.4 Change curves of average reward

## 3.2 不同充电功率下调度策略对比分析

在3.1节实验设置的基础上改变充电功率, 对比分析充电功率为6.6 kW和15 kW时的调度结果。对不同充电功率的策略网络进行训练和测试的过程中, 电池均采用“高先充低先放”选择方式。

### 3.2.1 充电功率为6.6 kW时的调度结果可视化

当充电功率为6.6 kW时, 测试日4种状态的电池数量如图5所示。



①可用电池, ②在充电, ③在放电, ④不可用电池

图5 充电功率为6.6 kW时4种状态的电池数量

Fig.5 Number of batteries in four states when charging power is 6.6 kW

结合图5、附录D表D2和表D3, 可得到如下结论。

1) 谷时段(00:00—07:00、23:00—24:00)的在充电电池数量基本保持在25~30块, 这是因为谷时段的购电电价低且上网电价也较低, 换电站通过学习得到“错峰充电”策略以降低充电费用, 同时也为当天的换电需求增加可用电池储备, 避免在未来时段由于没有足够的可用电池导致有换电需求的EV离开而带来的惩罚费用。在该策略下, 00:00—07:00时段内可用电池数量持续快速上升, 从15块增加到195块; 23:00—24:00时段内可用电池数量从8块增加到15块。

2) 峰时段(10:00—15:00、18:00—21:00)的在放电电池数量基本保持在25~30块, 这是因为峰时段的

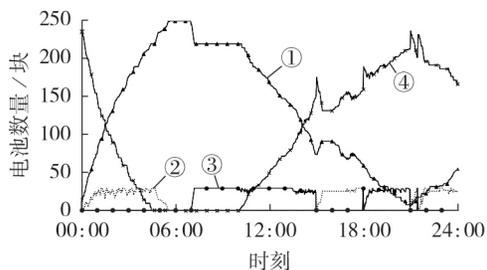
上网电价高,且换电站内有足够的可用电池储备支撑换电需求,换电站为了扩大利润空间,在峰时段选择集中放电以增加对电网的馈电收益。对比表 D2 可见,12:00—14:00、18:00—19:00 存在 2 个换电需求高峰,因此 10:00—15:00 时段内可用电池数量由 225 块迅速减少到 90 块,18:00—21:00 时段内可用电池数量由 43 块减少到 5 块。

3) 平时段为 07:00—10:00、15:00—18:00 以及 21:00—23:00。07:00—10:00 时段内可用电池数量增幅减缓,这是因为平时段的购电电价较谷时段更高,且该时段的可用电池储备较充足,换电站虽仍选择充电但充电力度减弱。07:00—10:00 时段内可用电池数量由 195 块增加到 225 块,占站内电池总数量的 90%。为了满足 15:00—18:00、21:00—23:00 时段内的换电需求,站内的充放电机接近满载。对比表 D2 可知,15:00—18:00 时段的换电需求较大,可用电池数量由 90 块减少到 43 块;21:00—23:00 时段内换电需求相对较少,持续为电池充电导致可用电池数量少量增加。

综上可知,在不同的时段采用本文所提方法可为换电站做出合理的决策。由图 5 可见:充电站内全天均有可用电池库存,表明本文所提调度策略能够满足换电需求;该测试日结束时所剩可用电池数量与 00:00 时刻的初始可用电池数量基本持平,可以认为学习所得策略网络具有可持续运用的能力。

### 3.2.2 充电功率为 15 kW 时的调度结果可视化

当充电功率为 15 kW 时,测试日 4 种状态的电池数量如图 6 所示。



① 可用电池, ② 在充电电池, ③ 在放电电池, ④ 不可用电池

图 6 充电功率为 15 kW 时 4 种状态的电池数量  
Fig.6 Number of batteries in four states when charging power is 15 kW

分析图 6 可得如下结论。

1) 在谷时段 00:00—07:00 内,可用电池数量由 10 块增加到 250 块,即换电站能在该时段内完成站内所有库存电池的充电过程。

2) 在平时段 07:00—10:00 内,换电站选择满载放电,这是因为平时段的上网电价与谷时段的购电电价之间存在电价差,此时换电站能通过放电获得利润;而图 5 中由于电池只能慢速充电,07:00 时刻只有 195 块可用电池,因此换电站仍选择充电来增

加可用电池库存。

3) 在时段 10:00—24:00 内,图 6 中的调度策略与图 5 总体一致,均能保证站内全天都有可用电池库存,保障能够满足换电需求。

对比图 5 和图 6 可知,2 种不同充电功率下的调度策略整体类似,但较大的充电功率可以缩短电池的充电时间,能在同样满足换电需求的基础上,有更大的“低充高放”利润空间。可见,充电功率的大小会间接影响换电站的调度策略,而本文所提换电站实时调度策略具有一定的通用性。

### 3.3 电池选择方式对比分析

在充电功率为 6.6 kW 的策略网络下,分析电池的“高先充低先放”和随机选择方式对日服务可用率、换电站收益以及电网基础负荷的影响。

#### 1) 对日服务可用率的影响。

在训练好的模型参数下分别使用 2 种电池选择方式对策略网络重复测试 1 000 次,得到的日服务可用率分布直方图见附录 D 图 D1。由图可见:在“高先充低先放”选择方式下,该策略网络的日服务可用率大多能达到 90% 以上,其中一半以上的测试能完全满足换电需求,可认为在该策略网络下电池采取“高先充低先放”选择方式能保障换电站的正常运行;而在随机选择方式下,该策略网络的日服务可用率主要分布在(60%, 90%)范围内,为了使电池随机选择方式也能满足换电需求,需增加换电站的电池储备数量和充放电机数量,这样就增加了硬件成本。

#### 2) 对换电站收益的影响。

在同一策略网络下,采用 2 种电池选择方式分别测试 1 000 次后,可得换电站日收益结果见附录 D 表 D4,日收益分布直方图见附录 D 图 D2。由表 D4 可见,“高先充低先放”选择方式下换电站的平均日收益更高。由图 D2 可见,“高先充低先放”选择方式下的换电站日收益总体比随机选择方式下的日收益更高。这是因为“高先充低先放”选择方式能满足更多的换电需求,提高了从 EV 用户处获得的收益。

#### 3) 对电网基础负荷的影响。

为了验证本文所提调度策略对电网负荷削峰填谷的作用,基于文献[19]中给出的电网基础负荷,得到 2 种电池选择方式下计及换电站负荷前、后的电网总负荷,如图 7 所示。由图可知,在 2 种电池选择方式下,电网基础负荷叠加换电站负荷后,01:00 时刻的负荷谷值有明显的增大,12:00 时刻的负荷峰值有明显的减小,21:00 时刻的第 2 个负荷峰值有少许减小并稍微后移。总体而言,2 种电池选择方式在换电站基于价格的需求响应模式下均能减小全天的负荷峰谷差,其中“高先充低先放”选择方式的削峰填谷效果略优于随机选择方式,但总体差别并不明显。可见,本文所提考虑换电站需求响应模式的实时调度策略利用站内电池的储能特性能够对电网负

荷起到一定的削峰填谷作用。

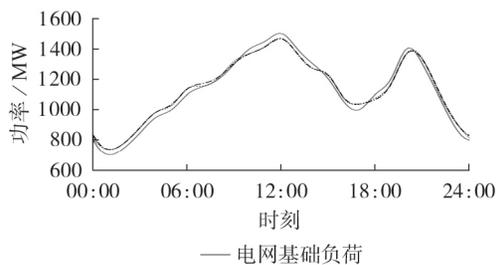


图7 计及换电站负荷前、后的电网总负荷曲线  
Fig.7 Total power grid load curves with and without battery swapping station load

图7 计及换电站负荷前、后的电网总负荷曲线

Fig.7 Total power grid load curves with and without battery swapping station load

上述结果表明,本文所提换电站实时调度策略在“高先充低先放”电池选择方式下能够满足换电需求,获得较大的经济收益,且具有削峰填谷的作用。

### 3.4 训练与决策时间

本文算例均在 Intel (R) Core (TM) i5-9300H CPU@2.40 GHz 的计算机上基于 Python 3.8 和 PyTorch 1.5.1 编程实现。在此基础上,1次训练过程耗时约为 26 min,策略网络的 1次前向传播仅需 5 ms 左右。本文所提带基线的蒙特卡罗策略梯度法的计算复杂度与状态空间、动作空间、神经网络大小有关。但相较于计算复杂度随 EV 数量指数增长的传统方法而言,本文方法在求解换电站实时调度问题上仍有明显的优势,降低了对算力的要求,在实际部署时也可以减少硬件设备成本。

## 4 结论

本文提出了基于强化学习的换电站实时调度策略优化方法,所得主要结论如下。

1)提出了带基线的蒙特卡罗策略梯度法进行 EV 换电站实时调度,测试结果表明所提策略可在保证换电站服务可用率的情况下获得更多的放电收益,也可在长时间尺度下持续应用。同时,训练得到的换电站调度策略能在不同时段做出合理的决策。

2)探讨了 6.6 kW 和 15 kW 这 2 种充电功率下训练得到的换电站调度策略,发现这 2 种充电功率下的调度策略相似,但较大的充电功率能够增大换电站“低充高放”的利润空间,可认为本文所提方法具有一定的通用性。

3)探究了“高先充低先放”和随机选择 2 种电池选择方式对换电站服务可用率、经济效益、电网基础负荷的影响,发现“高先充低先放”选择方式能够在满足更大服务可用率的同时,获得更大的经济效益,同时对电网负荷进行削峰填谷的效果更好。但关于电池的最优选择方式还有待进一步研究。

4)本文所提模型在换电站实时调度问题上具有

通用性,可通过简单修改进行问题迁移。例如:通过在奖励函数中增加可再生能源消纳率的奖惩项,使其向换电站与可再生能源协同消纳问题迁移;将动作空间中的二进制充放电状态变量修改为多值功率档位变量,使其向多档功率充放电问题迁移等。

附录见本刊网络版(<http://www.epae.cn>)。

### 参考文献:

- [1] 于东民,杨超,蒋林洳,等. 电动汽车充电安全防护研究综述[J]. 中国电机工程学报,2022,42(6):2145-2164.  
YU Dongmin, YANG Chao, JIANG Linru, et al. Review on safety protection of electric vehicle charging[J]. Proceedings of the CSEE, 2022, 42(6): 2145-2164.
- [2] 崔岩,胡泽春,段小宇. 考虑充电需求空间灵活性的电动汽车运行优化研究综述[J]. 电网技术,2022,46(3):981-994.  
CUI Yan, HU Zechun, DUAN Xiaoyu. Review on the electric vehicles operation optimization considering the spatial flexibility of electric vehicles charging demands[J]. Power System Technology, 2022, 46(3): 981-994.
- [3] 王敏,吕林,向月. 计及 V2G 价格激励的电动汽车削峰协同调度策略[J]. 电力自动化设备,2022,42(4):27-33,85.  
WANG Min, LÜ Lin, XIANG Yue. Coordinated scheduling strategy of electric vehicles for peak shaving considering V2G price incentive[J]. Electric Power Automation Equipment, 2022, 42(4): 27-33, 85.
- [4] 田园园,廖清芬,徐雨田,等. 基于有序充电策略的换电站及分布式电源多场景协调规划方法[J]. 电力自动化设备,2017,37(9):62-69.  
TIAN Yuanyuan, LIAO Qingfen, XU Yutian, et al. Multi-scenario coordinated planning method of EV battery-swapping station and distributed generation based on coordinated charging strategy[J]. Electric Power Automation Equipment, 2017, 37(9): 62-69.
- [5] 苏粟,李家浩,李泽宁,等. 考虑用户需求的电动汽车虚拟同步机辅助调频控制策略[J]. 电力自动化设备,2021,41(11):40-47.  
SU Su, LI Jiahao, LI Zening, et al. Auxiliary frequency regulation control strategy based on virtual synchronous machine for electric vehicles considering user demand[J]. Electric Power Automation Equipment, 2021, 41(11): 40-47.
- [6] 全慧,李相俊,张杨,等. 快充电站多类型应用方式的并网影响及控制技术综述[J]. 中国电力,2021,54(1):89-95,103.  
QUAN Hui, LI Xiangjun, ZHANG Yang, et al. A review of grid-connection impact and control technology of FCS multi-type application modes[J]. Electric Power, 2021, 54(1): 89-95, 103.
- [7] ZHANG T Y, CHEN X, YU Z, et al. A Monte Carlo simulation approach to evaluate service capacities of EV charging and battery swapping stations[J]. IEEE Transactions on Industrial Informatics, 2018, 14(9): 3914-3923.
- [8] 程杉,魏绍彬,赵子凯,等. 考虑电动汽车时空接入随机性的储能电站有序充放电分散式优化[J]. 电力自动化设备,2021,41(6):28-35.  
CHENG Shan, WEI Zhaobin, ZHAO Zikai, et al. Decentralized optimization of ordered charging and discharging for charging-storage station considering spatial-temporal access randomness of electric vehicles[J]. Electric Power Automation Equipment, 2021, 41(6): 28-35.
- [9] 陈立兴,黄学良. 高速公路充电站电动汽车有序充电策略[J].

- 电力自动化设备,2019,39(1):112-117,126.  
CHEN Lixing, HUANG Xueliang. Ordered charging strategy of electric vehicles at charging station on highway [J]. Electric Power Automation Equipment, 2019, 39(1):112-117, 126.
- [10] 陈凯炎, 牛玉刚. 基于V2G技术的电动汽车实时调度策略[J]. 电力系统保护与控制, 2019, 47(14):1-9.  
CHEN Kaiyan, NIU Yugang. Real-time scheduling strategy of electric vehicle based on vehicle-to-grid application[J]. Power System Protection and Control, 2019, 47(14):1-9.
- [11] 王彬, 郭文鑫, 李世明, 等. 基于短期预测信息和长期值函数近似的大规模电动汽车实时随机优化调度算法[J]. 电力系统保护与控制, 2019, 47(24):47-56.  
WANG Bin, GUO Wenxin, LI Shiming, et al. Real-time charging optimization for large-scale electric vehicles based on short term forecast information and long term value function approximation[J]. Power System Protection and Control, 2019, 47(24):47-56.
- [12] 曹一家, 刘易珠, 阙凌燕, 等. 换电站与电网协调的多目标双层实时充放电调度方法[J]. 电力自动化设备, 2015, 35(4):1-7.  
CAO Yijia, LIU Yizhu, QUE Lingyan, et al. Multi-objective bi-level real-time charging/discharging dispatch with coordination of BSS and grid[J]. Electric Power Automation Equipment, 2015, 35(4):1-7.
- [13] 潘振宁, 张孝顺, 余涛, 等. 大规模电动汽车集群分层实时优化调度[J]. 电力系统自动化, 2017, 41(16):96-104.  
PAN Zhenning, ZHANG Xiaoshun, YU Tao, et al. Hierarchical real-time optimized dispatching for large-scale clusters of electric vehicles[J]. Automation of Electric Power Systems, 2017, 41(16):96-104.
- [14] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine Learning, 1992, 8(3/4):229-256.
- [15] SUTTON R, BARTO A. Reinforcement learning: an introduction [M]. Cambridge, Massachusetts, USA: MIT Press, 1998:113-138.
- [16] WIDRICK R S, NURRE S G, ROBBINS M J. Optimal policies for the management of an electric vehicle battery swap station[J]. Transportation Science, 2018, 52(1):59-79.
- [17] 国网北京市电力公司. 国网北京市电力公司所属电动汽车公共充电设施执行峰谷分时电价[EB/OL]. [2021-05-15]. [http://www.bj.sgcc.com.cn/html/main/col1/column\\_1\\_1.html](http://www.bj.sgcc.com.cn/html/main/col1/column_1_1.html).
- [18] 刘灵恺, 雷霞, 李竹, 等. 电动汽车换电站可用电池组数动态调度策略[J]. 电工技术学报, 2017, 32(22):242-250.  
LIU Lingkai, LEI Xia, LI Zhu, et al. Dynamic scheduling strategy for available battery number of electric vehicle in battery-swap station[J]. Transactions of China Electrotechnical Society, 2017, 32(22):242-250.
- [19] 侯建朝, 胡群丰, 谭忠富. 计及需求响应的风电-电动汽车协同调度多目标优化模型[J]. 电力自动化设备, 2016, 36(7):22-27.  
HOU Jianchao, HU Qunfeng, TAN Zhongfu. Multi-objective optimization model of collaborative WP-EV dispatch considering demand response[J]. Electric Power Automation Equipment, 2016, 36(7):22-27.

#### 作者简介:



张文昕

张文昕(1997—),女,硕士研究生,主要研究方向为电动汽车并网优化调度(**E-mail**: zhangwenxin@ncepu.edu.cn);

栗然(1965—),女,教授,博士,研究方向为新能源并网技术、电力系统分析及运行与控制(**E-mail**: liranlele@163.com)。

(编辑 陆丹)

## Real-time scheduling strategy optimization for electric vehicle battery swapping station based on reinforcement learning

ZHANG Wenxin<sup>1</sup>, LI Ran<sup>1</sup>, ZANG Xiangdi<sup>1</sup>, YAN Jingru<sup>2</sup>, ZHU Jinyao<sup>3</sup>

(1. College of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China;

2. Electric Power Research Institute of State Grid Hebei Electric Power Company, Shijiazhuang 050022, China;

3. State Grid Shijiazhuang Electric Power Company, Shijiazhuang 050004, China)

**Abstract:** With the application and promotion of electric vehicles, the scheduling optimization of battery swapping stations has gradually become a research focus. The traditional scheduling strategies based on the predicted values of swapping demand are faced with some problems in practical application, such as being difficult to adapt to dynamic interference factors and accumulation of prediction errors. In order to solve these problems, a real-time scheduling strategy of battery swapping station based on Monte Carlo policy gradient method with baseline is proposed to optimize the charging and discharging strategy and the number of response batteries of battery swapping station. Monte Carlo policy gradient reinforcement learning with baseline is proposed, and the appropriate state space and action space are selected for real-time scheduling of battery swapping station. The reward function is designed to train the agent off-line, and the optimal strategy network is learned from the battery state data, the time-of-use electricity price and the number of queuing electric vehicles. The real-time scheduling strategy is tested on the basis of off-line trained model. The effectiveness and economy of the proposed scheduling strategy are verified based on battery swapping station's service availability and economic benefit. The results of an example show that the proposed strategy can play a certain role in peak load shifting of power grid.

**Key words:** electric vehicles; battery swapping station; reinforcement learning; policy gradient; time-of-use electricity price; real-time scheduling

## 附录 A

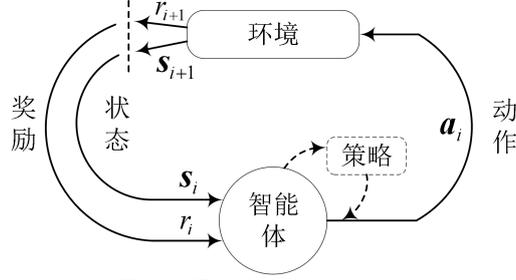


图 A1 强化学习的基本框架  
Fig.A1 Basic structure of reinforcement learning

## 附录 B

轨迹  $\tau$  在策略  $\pi_\theta$  下的分布为  $p_\theta(\tau)$ ， $p_\theta(\tau)$  的具体表达式为：

$$p_\theta(\tau) = p(s_0) \prod_{i=0}^{l-1} \pi_\theta(a_i | s_i) P(s_{i+1} | s_i, a_i) \quad (\text{B1})$$

式中： $p(s_0)$  为初始状态的分布，通常为均匀分布。

由 1.2 节中的式(1)可推导  $J(\theta)$  对  $\theta$  的梯度为：

$$\begin{aligned} \nabla_\theta J(\theta) &= \nabla_\theta E_{\tau \sim p_\theta(\tau)} [G_0(\tau)] = \sum_\tau \nabla_\theta p_\theta(\tau) G_0(\tau) = \\ & \sum_\tau p_\theta(\tau) \nabla_\theta \ln p_\theta(\tau) G_0(\tau) = \\ & E_{\tau \sim p_\theta(\tau)} \left[ \sum_{i=0}^{l-1} (\nabla_\theta \ln \pi_\theta(a_i | s_i) \sum_{i'=0}^{l-1} \gamma^{i'} r(\tau)_{i'+1}) \right] \end{aligned} \quad (\text{B2})$$

根据 MDP 的独立性假设，时段  $i$  之前的奖励与时段  $i$  的策略无关，因此式 (B2) 可改写为：

$$\nabla_\theta J(\theta) = E_{\tau \sim p_\theta(\tau)} \left[ \sum_{i=0}^{l-1} \nabla_\theta \ln \pi_\theta(a_i | s_i) G_i(\tau) \right] \quad (\text{B3})$$

式中： $G_i(\tau)$  为轨迹  $\tau$  中时段  $i$  的回报。式(B3)中的期望值难以获取，通常使用蒙特卡洛方法估计后得到式(B4)。

$$\nabla_\theta J(\theta) \approx \frac{1}{N} \sum_{j=1}^N \sum_{i=0}^{l-1} \nabla_\theta \ln \pi_\theta(a_i | s_i) G_i(\tau_j) \quad (\text{B4})$$

式中： $\tau_j$  为从  $p_\theta(\tau)$  中采样的第  $j$  条轨迹；“ $\approx$ ”表示等号右侧为左侧的无偏估计。等式右侧表示采用蒙特卡洛方法用有限数量的采样轨迹对式 (4) 等号右侧的期望的估计值。在得到梯度的估计值后，即可用梯度上升法优化目标函数。然而式(B4)在使用时刻的累计回报奖励作为当前策略下动作价值的无偏估计时，有可能带来较大的噪声和方差，引入基线能够有效降低方差。

## 附录 C

带基线的蒙特卡洛策略梯度算法

- 1: 定义策略逼近器  $\pi_\theta$ ，初始化参数  $\theta$
- 2: 定义价值函数近似器  $v_\omega$ ，初始化参数  $\omega$
- 3: 初始化状态分布  $d$
- 4: **Repeat**
- 5: 根据  $d$  和  $\pi_\theta$  采样  $N$  条轨迹  $\tau_j = \{s_0^{(j)}, a_0^{(j)}, r_0^{(j)}, s_1^{(j)}, \dots\}$
- 6: **for all**  $j = 1, \dots, N$  **do**
- 7: **for all**  $i = 0, 1, \dots$  **do**
- 8: 计算回报  $G_i(\tau_j) = r_{i+1}^{(j)} + \gamma r_{i+2}^{(j)} + \gamma^2 r_{i+3}^{(j)} + \dots$
- 9: 计算优势函数  $A_i^{(j)} = G_i(\tau_j) - v_\omega(s_i^{(j)})$
- 10: 计算更新量  $\Delta\omega = \alpha_\omega A_i^{(j)} \nabla v(s_i^{(j)}, \omega)$
- 11: 计算更新量  $\Delta\theta = \Delta\theta + \alpha_\theta \nabla_\theta \ln \pi_\theta(s_i^{(j)}, a_i^{(j)}) A_i^{(j)}$
- 12: **end for**
- 13: **end for**
- 14: 更新价值参数  $\omega \leftarrow \omega + \Delta\omega$
- 15: 更新策略参数  $\theta \leftarrow \theta + \Delta\theta$
- 16: **Until** 达到一定迭代次数或策略无明显提升

## 附录 D

表 D1 换电站的相关参数

Table D1 Related parameters of battery swapping station

参数	取值
额定充/放电功率 $P_c / P_d$	6.6 kW/6 kW
充/放电效率 $\eta_c / \eta_d$	0.95/0.9
电池容量 $C$	30 kW·h
站内电池总数 $M$	250 块
站内电池更换装置数量	10 台
站内充放电电机数量	30 台
换电零售价 $e$	1.5 元/(kW·h)
换电服务费 $e^s$	10 元/次

表 D2 到站 EV 数量的概率分布参数

Table D2 Probability distribution parameter for number of EVs arriving at station

时段	$\lambda$	时段	$\lambda$	时段	$\lambda$	时段	$\lambda$
00:00—01:00	0	06:00—07:00	6	12:00—13:00	52	18:00—19:00	43
01:00—02:00	0	07:00—08:00	15	13:00—14:00	45	19:00—20:00	28
02:00—03:00	0	08:00—09:00	6	14:00—15:00	40	20:00—21:00	20
03:00—04:00	0	09:00—10:00	11	15:00—16:00	24	21:00—22:00	15
04:00—05:00	0	10:00—11:00	16	16:00—17:00	36	22:00—23:00	7
05:00—06:00	0	11:00—12:00	40	17:00—18:00	41	23:00—24:00	0

表 D3 分时电价

Table D3 Time-of-use electricity price

	时段	购电电价/ [元·(kW·h) <sup>-1</sup> ]	上网电价/ [元·(kW·h) <sup>-1</sup> ]
峰	10:00—15:00, 18:00—21:00	1.0044	0.80
平	07:00—10:00, 15:00—18:00, 21:00—23:00	0.6950	0.56
谷	00:00—07:00, 23:00—24:00	0.3946	0

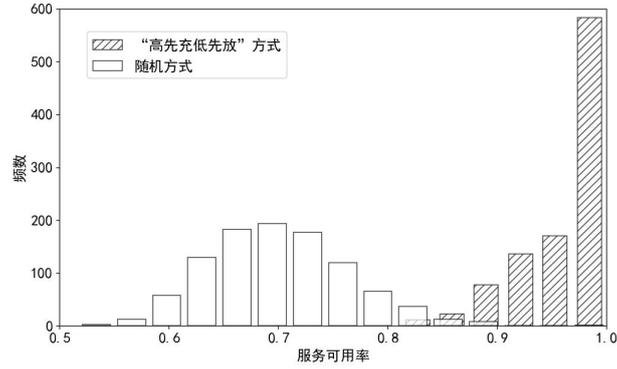


图 D1 服务可用率直方图

Fig.D1 Histogram of service reliability rate

表 D4 不同电池选择方式下的日收益

Table D4 Daily incomes under different battery sorting methods

电池选择方式	平均日收益/元	最大日收益/元	最小日收益/元
“高先充低先放”方式	8413.66	9347.75	7063.33
随机方式	5920.76	7492.88	4321.45

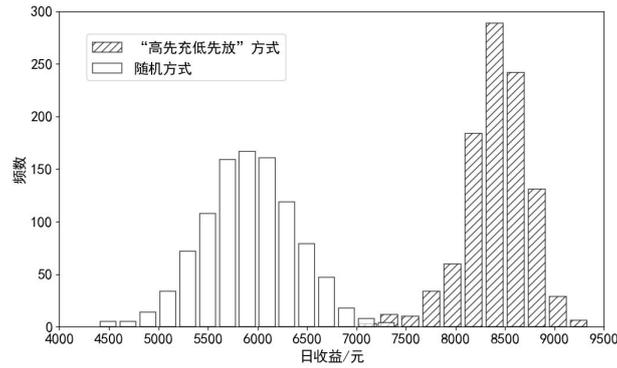


图 D2 日收益分布直方图

Fig.D2 Histogram of daily incomes