

基于深度强化学习算法的分布式光伏-EV 互补系统智能调度

陈 宁, 李法社, 王 霜, 张慧聪, 唐存靖, 倪梓皓
(昆明理工大学冶金与能源工程学院, 昆明 650031)

摘 要: 针对分布式光伏与电动汽车(electric vehicle, EV)大规模接入电网将对电力系统造成冲击的问题, 通过建立分布式光伏-EV 互补调度模型, 以平抑光伏并网波动、增加 EV 用户经济性为目标, 考虑光伏出力的随机性、负荷功率波动、EV 接入时间及电量随机性、实时电价、电池老化成本等因素, 提出采用梯度随机扰动的改进型近端策略优化算法(gradient random perturbation-proximal policy optimization algorithm, GRP-PPO)进行求解, 通过对模型目标函数的调整, 得到基于不同优化目标的 2 种实时运行策略。通过算例可知, 实时调度策略可有效地平抑并网点功率波动, 调度效果较传统 PPO 算法提高了 3.48%; 策略一以用户的出行需求及平抑并网点功率波动为首要目标, 能够保证用户的 24 h 用车需求, 同时并网点功率稳定率达到 91.84%; 策略二以用户经济效益为首要优化目标, 全天参与调度的 EV 收益可达 82.6 元, 可起到鼓励用户参与调度的目的。

关键词: 分布式光伏; 电动汽车; V2G; 深度强化学习; 实时调度; 近端策略优化

Intelligent Scheduling of Distributed Photovoltaic EV Complementary Systems Based on Deep Reinforcement Learning Algorithm

CHEN Ning, LI Fashe, WANG Shuang, ZHANG Huicong, TANG Cunjin, NI Zihao

(School of Metallurgy and Energy Engineering, Kunming University of Science and Technology, Kunming 650031, China)

Abstract: Aiming at the problem that large-scale access to the grid by distributed photovoltaic and electric vehicles (EVs) will cause an impact on the power system, we develop a distributed PV-EV complementary scheduling model. With the goals of smoothing the fluctuation of PV grid connection and increasing the economy of EV users, the factors such as the stochasticity of PV output, load power fluctuation, stochasticity of EV access time and power, real-time tariffs, battery aging cost and so on are taken into consideration, and an improved proximal policy optimization algorithm with stochastic perturbation of the gradient is proposed to solve the problem. Meanwhile, the model objective function is adjusted to obtain 2 real-time operation strategies based on different optimization objectives. It can be seen from the arithmetic example that the real-time scheduling strategy can effectively smooth outlet power fluctuations, and the scheduling effect is improved by 3.48% compared with the traditional PPO algorithm. Strategy 1 takes the user's travel demand and smoothing the power fluctuation of the parallel network as the primary objectives, which can ensure the user's 24 h demand, and at the same time, the power stabilization rate of the parallel network reaches 91.84%. Strategy 2 takes the user's economic benefit as the primary optimization objective, and the EV benefit of participating in dispatching for the whole day can reach RMB 82.6, which can encourage the users to participate in dispatching.

Key words: distributed photovoltaic; cluster electric vehicle; V2G; deep reinforcement learning; real time scheduling; proximal policy optimization

0 引言

分布式光伏发电以其容量、选址的灵活性, 常常与负荷中心、城市屋顶、闲置空地结合, 实现所

发电力的就近消纳。“双碳”大背景下, 分布式光伏发电发展迅猛, 截至 2022 年底, 国内分布式光伏装机容量已达 157.6 GW, 其新增容量连续两年超过集中式光伏, 发展迅猛^[1]。但是光伏发电量受外界影响大, 供电可靠性低, 稳定性较差, 直接调度难度高, 微网系统靠单一的光伏电源难以保证系统内负荷的平稳运行^[2-3]。

基金资助项目: 国家自然科学基金(52166013)。

Project supported by National Natural Science Foundation of China (52166013).

对光伏设备配套储能装置是一种行之有效的解决办法, 足够的储能装置能够很好地改善光伏发电间歇性、随机性等弊端, 但其存在前期投入成本过大的问题。电动汽车(electric vehicles, EV)作为移动式储能载体, 近年来市场发展迅猛。在政策推动下, 预计到 2030 年, 我国电动汽车保有量将达到 0.95 亿辆^[4], 大规模 EV 随机接入电网同样会给电力系统造成负面影响^[5-6]。随着 EV 数量的增加及 EV 到电网(vehicle to grid, V2G)技术的发展, 分布式光伏和 EV 的互补调度被认为是一种极具潜力的发展模式^[7-8], 城市中分布式光伏与 EV 互补调度不需要额外的设备就能进行工作, 是未来微网系统的重要发展方向。

然而, 在分布式光伏-EV 互补调度模型中, 可参与调度的 EV 数量受接入时间、接入时电量、离开时所需电量等因素制约, 使得可参与调度 EV 数量波动性较大。文献[9]提出了一种计及电动汽车需求响应的风储混合系统日前联合优化调度模型, 但由于 EV 的接入时间、接入时的电量及新能源的出力情况均具有较大的随机性, 因此日前调度模型难以处理实际变化情况。文献[10]提出一种计及电网侧、用户侧需求的集群代理商决策调度策略, 以满足供需两侧的调度意愿, 但该策略未能明确充电电价以及放电激励, 对用户收益没有加以讨论, 难以激发 EV 用户参与调度的积极性。文献[11]提出电动车的实时功率分配算法, 以提升调度精度, 但仅考虑了 EV 的充电属性, 未将 EV 的电源属性考虑在内。由于系统内不确定性因素较多, 且光伏功率波动的时间尺度一般为秒到分钟级^[12], 因此找到一种能够实时处理各方面因素变化并做出合理决断的调度方法至关重要。机器学习在计算智能、感知智能和认知智能方面具有较强的处理能力, 非常适合应用在电力系统和综合能源系统中^[13-14]。

深度强化学习(deep reinforcement learning, DRL)将深度学习的感知能力和强化学习的决策能力相结合, 以深度神经网络逼近动作值函数, 是解决多变量复杂问题的有效途径^[15-16]。文献[17]针对可再生能源出力和负荷的不确定性, 开展了对 EV 充放电控制的 DRL 研究, 但所用的深度 Q 网络(deep Q-Network, DQN)算法无法应用于连续动作的控制。文献[18]采用 DRL 技术研究了混合动力电动客车的能源管理策略, 但其以燃油经济性为主要目标, 对普通 EV 参与电力系统调度的参考价值较小。文

献[19]针对单台 EV, 提出了考虑 EV 用户行为随机性的 DRL 实时调度方法, 但所提方法用奖惩函数表征用户需求, 无法保证用电需求得到满足。文献[20]以 DRL 技术研究 EV、换电站电池、电力系统三者之间的调度平衡问题, 实现了大规模 EV 与电力系统之间的平衡调度, 但换电站受到成本高、电池标准难以统一等因素制约, 使得这种模式难以普及。光伏-EV 互补发电系统的调度问题是一个复杂的动态问题, 由于系统内在因素, 例如光伏发电量、负荷需求电量、EV 接入系统时间, 电量、EV 离开系统时间, 及离开时所需的电量等不确定性较高, 随机性较强, 因此传统方法难以准确建模和优化。而深度强化学习能有效处理动态复杂环境, 适应实时变化, 从而可提高系统调度性能, 是实现 EV 参与电力系统调度的有效途径, 但现阶段深度强化学习在 V2G 中的应用仍然较少, 且未能明确用户参与调度的经济性及满足用户出行需求。

因此, 本文以分布式光伏-EV 互补发电系统为研究对象, 以 V2G 技术为落脚点, 充分考虑光伏发电的不确定性及 EV 充放电的随机性, 以就近消纳光伏、减小并网节点功率波动、最大化经济收益、满足 EV 用户电量需求为目标, 构建光伏-EV 互补模型, 并针对传统近端策略优化算法(proximal policy optimization algorithm, PPO)在复杂优化任务上容易提前陷入局部最优解且难以跳出的缺陷, 提出以梯度随机扰动(gradient random perturbation, GRP)的改进型近端策略优化算法训练智能体, 从而优化 EV 的运行策略, 使其能够根据系统内关键信息决定当前时刻 EV 充放电幅度, 进而实现分布式光伏-EV 的智能互补调度。本研究内容的实施将城市 EV 变成移动储能模块, 不仅能够减小新能源及 EV 大规模接入电网对电网稳定性的影响, 也能在充分消纳光伏的基础上为 EV 用户创造一定的经济收益。

1 光伏-电动汽车互补系统描述

在微网系统中, 商业楼宇能够安装光伏面板并提供 EV 充放电设施, 同时商业楼宇的用电峰谷与光伏的发电峰谷存在偏差, 因此本文的主要研究对象为装有光伏面板及配备 EV 充电设施的商业楼宇, 其示意图如图 1 所示。

分布式光伏所发电量优先满足系统内负荷使用, 所发电量盈余时给系统内 EV 充电, 仍有余量则就近并网; 所发电量不足以供给系统内负荷使用

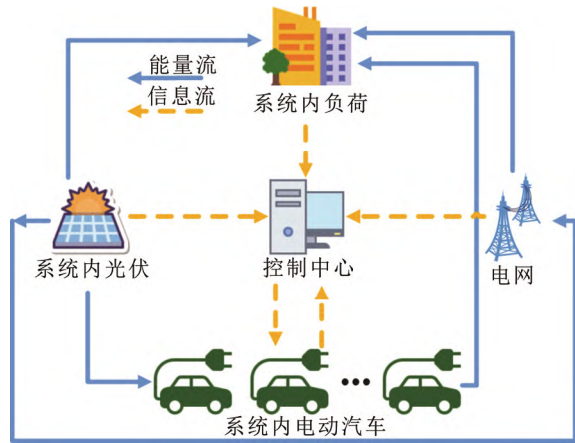


图1 光伏-电动汽车互补系统

Fig.1 PV-EV complementary system

时, EV 通过 V2G 技术对系统进行供能, 若还不够则从电网获取支持, 以此达到分布式光伏所发电能的 100% 消纳, 进而减小电网功率波动, 并为用户创造一定经济收益等目标。

1.1 电动汽车充放电的数学模型

系统内部 EV 在不同时段承担储能、电源 2 种角色。对于 EV 用户来说, 电池的剩余电量在用户使用时长达到需求值是必要条件。

电动车工作在充电模式下的数学模型为:

$$P_t^{\text{charge}} = \frac{1}{\eta_{\text{charge}}} \sum_{i=1}^m \frac{R_i^{\text{initial}} - R_i^{\text{initial}-\Delta t}}{\Delta t} \quad (1)$$

式中: P_t^{charge} 为电动汽车 t 时刻的总充电功率, kW; η_{charge} 为 EV 的平均充电效率; m 为系统内电动车数量; R_i^{initial} 为第 i 辆 EV 当前的剩余电量, kWh; $R_i^{\text{initial}-\Delta t}$ 为第 i 辆 EV Δt 时间前的剩余电量, kWh; $\Delta t = 15 \text{ min}$ 。

电动车工作在供电模式下的数学模型为:

$$P_t^{\text{electric}} = \eta_{\text{electric}} \sum_{i=1}^n \frac{R_i^{\text{initial}} - R_i^{\text{initial}-\Delta t}}{\Delta t} \quad (2)$$

式中: P_t^{charge} 为电动汽车 t 时刻的总供电功率, kW; η_{electric} 为 EV 的平均供电效率; n 为可参与调度的 EV 数量。

系统内 EV 的总电量变化计算式为:

$$R_t^{\text{all}} = R_{t-\Delta t}^{\text{all}} + R_{t-\Delta t}^{\text{EVa}} - R_{t-\Delta t}^{\text{EVb}} + \int_{t-\Delta t}^t (P_t^{\text{charge}} - P_t^{\text{electric}}) dt \quad (3)$$

式中: R_t^{all} 为 t 时刻系统内 EV 的总体剩余电量, kWh; $R_{t-\Delta t}^{\text{all}}$ 为 $t-\Delta t$ 时刻系统内 EV 的总体剩余电量, kWh; $R_{t-\Delta t}^{\text{EVa}}$ 为 Δt 时间内接入系统 EV 的剩余电量, kWh; $R_{t-\Delta t}^{\text{EVb}}$ 为 Δt 时间内离开系统 EV 的剩余电

量, kWh。

1.2 优化模型

系统的优化目标是在就近消纳分布式光伏发电的前提下, 通过 EV 的充放电动作减小并网点功率波动, 并在实时电价模式下, 考虑电池循环充放电成本, 最大化 EV 用户的经济收益, 同时将功率波动转化为经济惩罚, 得出如下优化目标:

$$E_t = (\partial_t + \rho) |P_t^{\text{electric}}| \Delta t - \partial_t P_t^{\text{charge}} \Delta t - P_t \Delta t B^{\text{ageing}} - \lambda_t^{\text{wave}} \quad (4)$$

$$E = \sum_{t=1}^T E_t \quad (5)$$

式中: E_t 为 t 时刻的收益, 元; T 为运行周期; 优化目标是最大化运行周期 T 内的收益 E ; ∂_t 为 t 时刻的实时电价, 元/kWh; ρ 为放电补贴, 元/kWh; P_t 在充电模式下为 P_t^{charge} , 在供电模式下为 P_t^{electric} ; B^{ageing} 为电池老化损失, 元/kWh; $(\partial_t + \rho) |P_t^{\text{electric}}| \Delta t$ 为车辆在供电模式下收益; $\partial_t P_t^{\text{charge}} \Delta t$ 为车辆在充电模式下的支出; $P_t \Delta t B^{\text{ageing}}$ 为循环充放次数增加造成的电池老化损失; λ_t^{wave} 为功率波动转化的经济惩罚量, 其定义为:

$$\lambda_t^{\text{wave}} = \begin{cases} \varphi_1 \Delta P_t^3, & 20 < \Delta P_t \\ \varphi_2 \Delta P_t^2, & 5 < \Delta P_t \leq 20 \\ 0, & 0 < \Delta P_t \leq 5 \end{cases} \quad (6)$$

式中: φ_1 、 φ_2 为功率波动惩罚系数, 元/kW; ΔP_t 表示 Δt 时间内的功率波动量, 定义为:

$$\Delta P_t = |P_t^{\text{grid}} - P_{t-\Delta t}^{\text{grid}}| \quad (7)$$

式中: P_t^{grid} 为 t 时刻并网点交互的功率, 其表达式见式(8)。

$$P_t^{\text{grid}} = P_t^{\text{pv}} - P_t^{\text{load}} - P_t \quad (8)$$

式中: P_t^{pv} 为 t 时刻的分布式光伏发电功率值; P_t^{load} 为 t 时刻系统内负荷功率值; P_t 为 t 时刻 EV 的发出或吸收功率, kW, 定义为:

$$P_t = \begin{cases} P_t^{\text{charge}}, & P_t^{\text{pv}} \geq P_t^{\text{load}} \\ P_t^{\text{electric}}, & P_t^{\text{pv}} < P_t^{\text{load}} \end{cases} \quad (9)$$

1.3 约束条件

分布式光伏-电动汽车互补系统内部应满足以下几个约束条件。

1) 可参与调度 EV 充放电功率约束

$$P_{\text{max}}^{\text{electric}} \leq P_t \leq P_{\text{max}}^{\text{charge}} \quad (10)$$

2) 系统内 EV 的总电量约束

$$0.7mR^{\text{full}} \leq R_t^{\text{all}} \quad (11)$$

式中: mR^{full} 为系统内 m 辆 EV 满电时的总电量, 总电量约束保证系统内单张 EV 电量满足需求。

2 深度强化学习建模

深度强化学习完美地结合了深度学习的感知能力与强化学习的决策能力, 可以直接根据输入的信息进行控制, 是一种更接近人类思维方式的人工智能方法^[21-22]。

2.1 系统调度问题转化为强化学习任务

深度强化学习最初被应用于游戏领域并取得极大的成功。但是光伏-EV 互补模型不同于游戏模型, 其将互补模型中的调度问题转化为强化学习任务, 是深度强化学习应用于互补模型的关键。对于一个深度强化学习任务来说, 最重要的 2 个条件是智能体(agent)以及其交互的环境。

本文所研究的分布式光伏-EV 互补系统是深度强化学习的环境, EV 的充放电动作作为智能体的动作 a_t , 其定义为:

$$a_t = p_t \quad (12)$$

在分布式光伏-EV 互补系统环境中, 在任意时刻 t , 分布式光伏的发电量 P_t^{pv} 、负荷需求 P_t^{load} 、实时电价 ∂_t 、可参与调度电动车数量 n_t 、系统内 EV 的总电量 R_t^{all} 作为状态。其状态 s_t 定义为:

$$s_t = \{ P_t^{\text{pv}}, P_t^{\text{load}}, \partial_t, n_t, R_t^{\text{all}} \} \quad (13)$$

在动作 a_t 的作用下, 环境中的状态产生变化, 并反馈一定的奖励值, 由优化目标体现为:

$$g(a_t | s_t) = E_t \quad (14)$$

电动车的充放电动作目标是使累积回报值最大化, 但是其动作策略必须在环境的约束条件之内, 超出约束条件的动作应该给予一定惩罚, 故必须把约束条件合理的转化为奖励函数的一部分。EV 充放电功率约束体现在智能体对动作幅值的限制上, 系统内 EV 的总电量约束以惩罚函数的形式表示为:

$$\lambda_t^R = \begin{cases} \eta(0.7mR^{\text{full}} - R_t^{\text{all}})^2, & 0.7mR^{\text{full}} > R_t \\ 0, & 0.7mR^{\text{full}} \leq R_t \end{cases} \quad (15)$$

式中: η 为电量惩罚系数, 其取值与模型相关。

智能体与环境交互过程中累积回报为:

$$G = g(a_t | s_t) + \mu g(a_{t+1} | s_{t+1}) + \mu^2 g(a_{t+2} | s_{t+2}) + \dots + \mu^{T-t} g(a_T | s_T) \quad (16)$$

式中: μ 为奖励折扣系数, 其取值范围为 $[0, 1]$ 。深度强化学习中, 为了加快模型的收敛速度, 当前的

奖励值往往比未来的奖励值权重更大, 故对未来奖励值做折扣。

对于策略, 从初始状态 s_t 执行动作 a_t 的探索过程所获得的回报可由动作价值函数 $Q_\omega(s, a)$ 表示为:

$$Q_\omega(s, a) = E^\omega(G_t | s_t = s, a_t = a) \quad (17)$$

故分布式光伏-电动汽车互补模型的智能调度问题转化为:

$$g(a_t | s_t) = E_t - \lambda_t^R \quad (18)$$

2.2 深度强化学习算法

PPO 算法采用近端策略优化的方法, 可以限制策略的变化范围, 避免过大的策略更新, 从而避免算法陷入不稳定状态, 进而增强鲁棒性^[23]。

本文在环境的训练过程中, 采用经验回放机制, 在线处理得到转移样本 $e_t = \{s_t, a_t, g_t, s_{t+1}\}$, 在每个时间步 t , 将智能体与环境交互得到的转移样本存放到记忆单元 $D = (e_1, e_2, \dots, e_t)$ 之中, 在训练时从记忆单元随机抽取小批量的转移样本。随机采样的方式降低了样本之间的关联性, 提升了算法的稳定性。

为提升算法性能, 避免算法陷入局部最优解, 在智能体动作 a_t 上附加动作噪声 γ 得到随机动作 $\hat{a}_t$ (式(19)), 从而有助于探索困难问题的最优解。

$$\hat{a}_t = a_t + \nu\gamma \quad (19)$$

式中: ν 为动作噪声衰减因子, 其取值范围为 $[0, 1]$, 训练刚开始进行时, 其取值较大, 起到全面探索环境的作用, 随着训练的进行, 其值逐渐衰减, 减少动作 $\hat{a}_t$ 因为噪声带来的误差。

PPO 算法通常使用 Adam 优化器^[23], 它是一种自适应学习率的梯度下降算法。Adam 优化器会根据梯度的方向和大小自适应地调整学习率。Adam 的更新计算式可以表示为:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla J(\theta_t) \quad (20)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla J(\theta_t))^2 \quad (21)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (22)$$

$$\beta_2 \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (23)$$

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \quad (24)$$

式中: m_t 是梯度 1 阶矩的指数加权移动平均, 用于估计梯度的均值; v_t 是梯度 2 阶矩的指数加权移动平均, 用于估计梯度的方差; $\nabla J(\theta_t)$ 是当前步骤的梯度(损失函数关于权值的导数); β_1 是 Adam 的超

参数, 分别控制 1 阶矩和 2 阶矩的衰减程度; α 是学习率, 控制了权值更新的步长; ε 是一个小常数, 用于数值稳定性, 防止除零错误。

虽然 Adam 优化器在多数情况下表现出色, 但在复杂的优化任务上容易提前陷入局部最优解且难以跳出。

为提升 Adam 优化器的性能, 本文提出采用梯度随机扰动的改进型近端策略优化算法 (GRP-PPO), 即在模型的前向传播和反向传播时计算梯度后加入一个随机的噪声向量, 如式(25)所示。

$$\nabla \hat{J}(\theta_t) = \nabla J(\theta_t) + n_t \quad (25)$$

式中: $\nabla \hat{J}(\theta_t)$ 是增加噪声后的当前步骤梯度; n_t 是随机噪声向量, 满足均值为 0, 方差为 0.05 的高斯分布, 维度与当前步骤梯度相同。

GRP-PPO 算法的算法流程如图 2 所示。

2.3 DRL 过程

基于 DRL 方法的分布式光伏-电动汽车互补调度模型调度流程图如图 3 所示, 具体实施步骤如下。

1) 首先对参数进行初始化。

2) 随机读入训练集中某一天的分布式光伏发电量 P_t^{PV} 、负荷需求电量 P_t^{load} 、实时电价 ∂_t 、可参与调度电动车数量 n_t 及总电量 R_t^{all} 。

3) 与环境交互获得状态 s_t , 由策略网络得到动作 a_t , 在动作上叠加噪声如式(19)所示, 得到随机动作 $\hat{a}_t$, 以动作 $\hat{a}_t$ 与环境交互得到奖励 g_t 与下一时刻状态 s_{t+1} 。

4) 利用经验回放机制将样本 $e_t = (s_t, \hat{a}_t, r_t, s_{t+1})$ 存储到经验记忆库中。

5) 判断记忆库是否存满, 若存满, 则从中抽取批量数据更新网络参数, 再更新状态和累积奖励。

6) 更新时间, 判断当前训练回合是否结束, 若结束则输出当前回合的累积奖励, 再重复步骤 2) 一步骤 6), 否则重复步骤 3) 一步骤 6)。

3 算例分析

本文利用 Python 3.9 和 Pytorch 架构实现改进型 GRP-PPO 算法对分布式光伏-EV 互补调度问题的求解。为验证 DRL 在互补调度模型优化问题上的效果, 本文选取一个装机容量为 2 MW 的分布式光伏发电系统及 EV 接入的商业楼宇作为算例进行仿真计算, 取优化时间步长为 15 min^[24]。

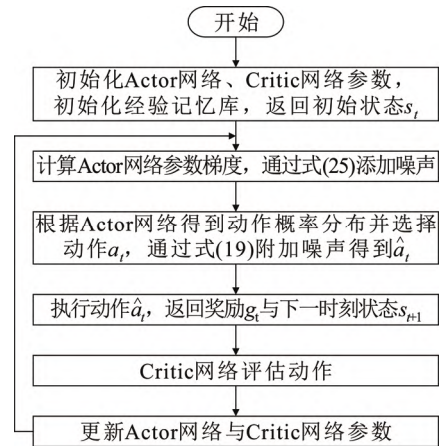


图 2 GRP-PPO 算法流程

Fig.2 GRP-PPO algorithm process

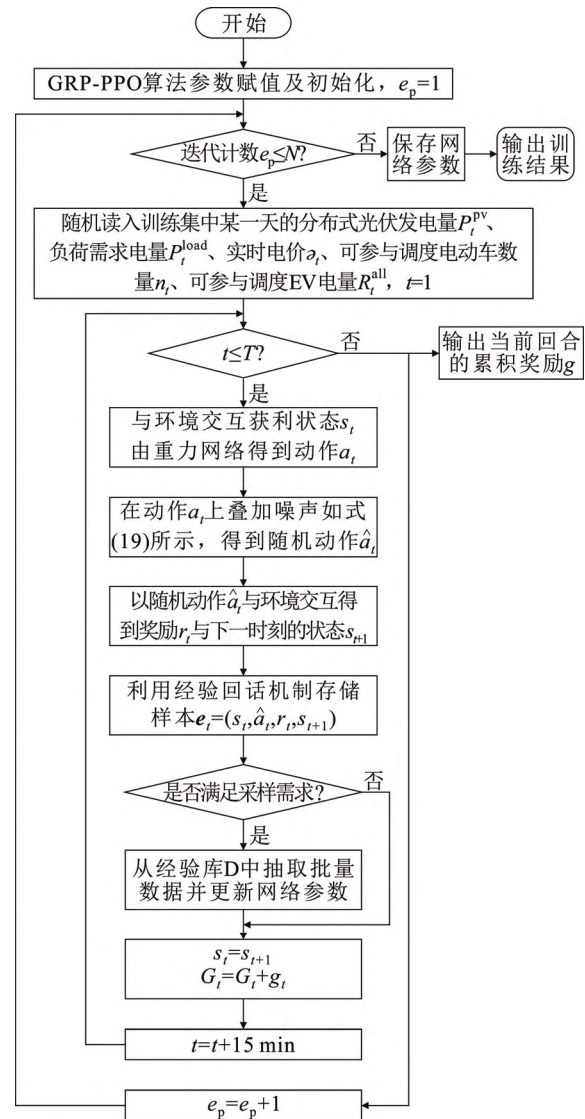


图 3 分布式光伏-电动汽车智能调度流程图

Fig.3 Distributed photovoltaic electric vehicle intelligent dispatching flow chart

3.1 参数设置

光伏-电动汽车互补模型如图 1 所示, 其中 EV 的仿真参数参考文献[25]设置, 如表 1 所示, 其余主要参数如表 2 所示, GRP-PPO 算法的超参数如表 3 所示。

实时电价 ∂_t 采用云南省一般工商业及其他用电电价的分时标准, 其峰谷分时电价时段划分为: 峰时段为 09:00—12:00、18:00—23:00; 平时段为 07:00—09:00、12:00—18:00; 谷时段为 00:00—07:00、23:00—24:00。其中, 峰时段、平时段、谷时段电价分别为 0.574、0.404、0.234 元/kWh。

本算例以某实际光伏电站的出力数据及某一实际负荷做算例分析, 系统内部实时可参与调度 EV 数量及电量参照美国 La Canada 国家研究实验室公开数据集设定^[28], 采取一年中前 300 d 的数据作为训练集, 后 65 d 的数据作为测试集。3 月 1 日至 3 月 7 日光伏发电量、负荷需求电量的变化趋势如图 4 所示。

分别采用 PPO 算法和 GRP-PPO 算法以图 3 所示流程进行训练, 模型训练过程如图 5 所示。结果表明: ①总训练步数为 96 000 步, 改进前后的算法总体趋势接近, 其中: 1—23 000 步为探索阶段, 智能体在这个阶段以随机动作对环境进行探索; 23 001—51 000 步为学习阶段, 智能体在这个阶段逐步做出正确的动作, 通过对神经网络中各个神经元权重的迭代更新, 使得累计奖励越来越大; 51 001—96 000 步模型已经收敛且较为平稳, 说明模型已经可以针对变化的环境做出正确动作使累积奖励最大。②改进前后算法训练结果的差异主要表现在探索阶段和收敛阶段。在探索阶段, 由于 GRP-PPO 策略网络的梯度加入了一个随机噪声, 导致其寻找策略的速度慢于 PPO 算法且初期波动较大; 但是当模型处于收敛阶段时, 由于 PPO 算法的策略梯度缺乏探索多样性, 因此容易陷入局部最优且难以跳出。GRP-PPO 采用的随机噪声梯度使得模型在收敛阶段时仍能够进行小范围的探索, 从而易于跳局部最优解, 其调度效果较传统 PPO 算法提高了 3.48%。

3.2 模型效果测试

为了验证模型的调度性能及在随机变化环境中持续调度的稳定性, 以下采取连续 2 d 的数据对模型进行测试。其调度前后的并网节点功率变化情况如图 6 所示。由图 6 可知, 由于分布式光伏的发电峰谷与负荷用电需求的峰谷不一致, 导致系统并

表 1 EV 充放电仿真参数设置

Table 1 EV charging and discharging simulation parameter setting

参数	数值
电池容量/kWh	80
最大充电功率/kW	24
最大放电功率/kW	24

表 2 模型的主要参数

Table 2 Main parameters of the model

参数	数值
$\eta_{\text{charge}}/\%$	92
$\eta_{\text{electric}}/\%$	92
$\beta_{\text{ageing}}/(\text{元}\cdot\text{kWh}^{-1})$	0.12 ^[26]
$\rho/(\text{元}\cdot\text{kWh}^{-1})$	2 ^[27]
T/h	24
$\Delta t/\text{min}$	15

表 3 GRP-PPO 算法的超参数

Table 3 Hyper parameters of PPO

参数	数值
初始学习率	0.01
经验记忆库容量	20 000
总步数	96 000

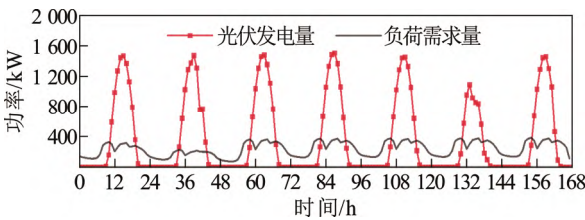


图 4 一周光伏及负荷变化趋势

Fig.4 Trend of photovoltaic and load changes in one week

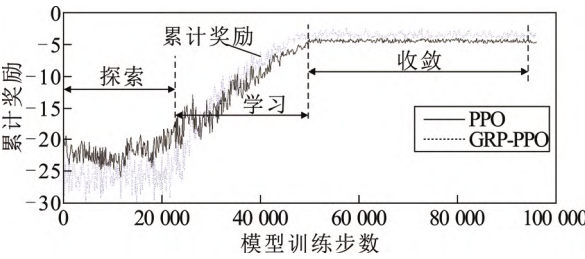


图 5 模型训练过程

Fig.5 Model training process

网点功率波动较大; 在大多数情况下, 分布式光伏所发电能大于负荷需求, 因此会出现较大的功率波动。在电动车的充放电动作下, 并网点的功率波动

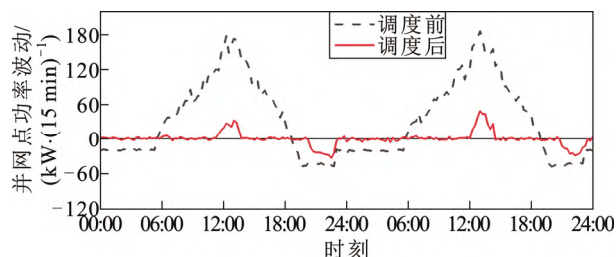


图6 调度前后并网点功率波动情况对比(策略一)

Fig.6 Comparison of power fluctuation of parallel network before and after dispatching (strategy 1)

情况得到明显的改善,在绝大部分时段并网点功率波动控制在 $5(\text{kW}/(15\text{ min}))$ 以内,晚上 21:00—23:00 时段调度效果不明显,主要原因是系统内可参与调度的 EV 数量减少了。

图7为系统内不同数量的V2G充电桩对并网点功率波动情况的影响趋势图。图7表明,系统内存在双向V2G充电桩时,并网点功率波动得到缓解,且随着可参与调度EV数量的提升,调度的效果越好。当系统内存在120个双向V2G充电桩时,并网点功率波动降低至 $\pm 5(\text{kW}/(15\text{ min}))$ 范围内,达到了减缓功率波动的目标。

系统内20辆EV的电量曲线如图8所示。可以看出,EV在光伏发电大于负荷需求电量的时段进行充电动作,在光伏发电小于负荷需求电量的时段进行放电动作,EV参与调度的全流程绝大部分时间段电量保持在约束值 $70\%R^{\text{full}}$ 以上,以保证EV用户全时段用车的需求。系统内不同时刻有不同电量的EV接入或离开,EV接入系统后迅速开始充电达到 $80\%R^{\text{full}}$,并减小充电功率充至 $95\%R^{\text{full}}$ 左右。

在图8中可以看到,在电价位于谷时段时,EV仍在进行放电动作,考虑为模型对功率波动处罚较大的缘故,减小模型对功率波动的惩罚并加大放电动作的收益,再次训练模型得到策略二。在策略二的调度下系统内20辆EV的电量曲线如图9所示,调度前后的并网点功率变化情况如图10所示。在策略二的调度下,系统内的EV在牺牲了一定功率稳定性的情况下,学会了在电价位于谷时段时进行充电动作,在电价位于峰时段且负荷缺口时进行更大幅度的放电动作,因此经济性得到提高。

参与调度的EV用户在系统内充电费用按 $0.2\text{元}/\text{kWh}$ 收取。表4为在策略一调度下系统内5辆EV的接入、离开时间和电量,以及其参与调度的收益,全天在系统内参与调度的EV当天收益为47.4元。

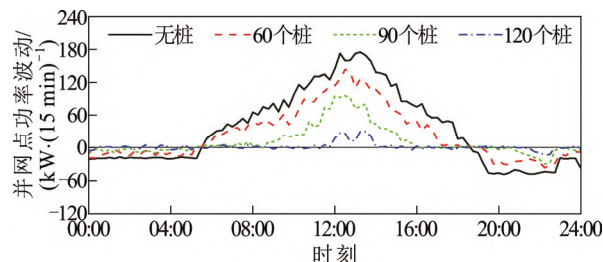


图7 系统内充电桩数量对并网点功率影响

Fig.7 Influence of the number of charging piles in the system on the power of the grid connection point

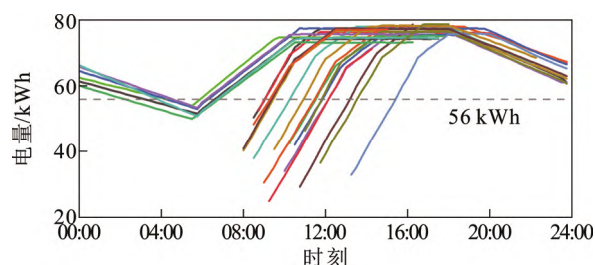


图8 系统内20辆EV的电量曲线(策略一)

Fig.8 Power curves for 20 EVs in the system (strategy 1)

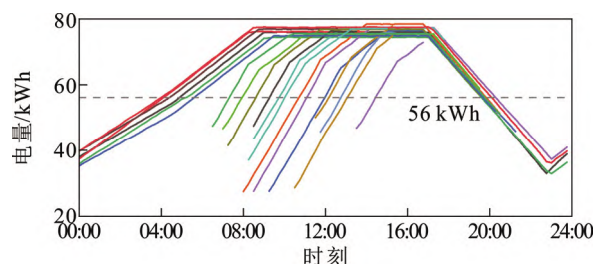


图9 系统内20辆EV的电量曲线(策略二)

Fig.9 Power curves for 20 EVs in the system (Strategy 2)

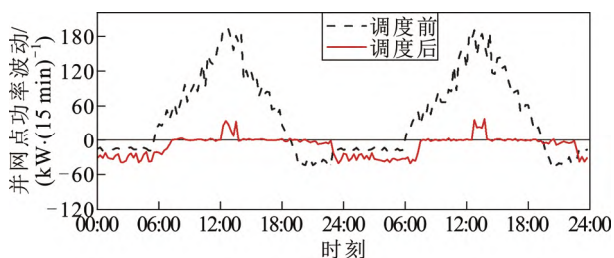


图10 调度前后并网点功率波动情况对比(策略二)

Fig.10 Comparison of power fluctuation of parallel network before and after dispatching (strategy 2)

表5为在策略二调度下系统内5辆EV的接入、离开时间和电量,以及其参与调度的收益,全天在系统内参与调度的EV当天收益为82.6元,相比策略一经济性提升明显,接入系统参与调度的EV均获得一定的经济收益,在放电时间段参与调度的EV

表 4 各辆 EV 收益(策略一)

Table 4 Revenue per EV (strategy 1)

EV	时间, 电量/kWh		原充电 费/元	系统内充 电费/元	放电补 贴/元	收益/ 元
	接入	离开				
EV8	全天				47.4	47.4
EV22	08:00, 40.8	23:45, 62.9	12.7	4.4	24.4	32.7
EV38	09:00, 30.5	22:15, 67.4	21.2	7.4	21.6	35.4
EV62	10:45, 29.2	16:15, 78.7	28.4	9.8		18.6
EV75	11:45, 36.5	18:00, 78.8	24.3	8.5		15.8

经济性获得显著提高, 有助于提升用户参与调度的积极性。

3.3 对比分析

为验证本文提出的基于 GRP-PPO 算法的智能调度方法的有效性, 将基于 GRP-PPO 的调度方法、序列二次规划(sequential quadratic programming, SQP)调度方法和优势动作评论(advantage actor critic, A2C)调度方法进行对比。以连续 2 d 的数据为例对 3 种方法的调度结果进行比较。由于 SQP 法解决调度问题时需要对其未来的源/荷数据进行预测, 因此将训练集中数据作为其准确预测值, 并通过改变数据精度的方法改变其预测精度以测试其对调度性能的影响。

表 6 给出了 3 种方法调度情况下并网节点的功率波动情况, 并网节点功率稳定率定义为运行周期内并网节点减小的功率波动值比原功率波动值。基于 GRP-PPO(策略一)的调度方法稳定率较基于 A2C 的调度方法高 14.49%, 调度效果优异; 即使在主要满足经济性的情况下, 策略二功率稳定性也有一定的提升。采用 SQP 进行调度, 预测精度为 100%时, 并网节点的功率稳定性最高, 且随着预测精度的下降, 其调度结果下降明显, 由于预测技术的限制, 其预测精度不能到达 100%; 当预测精度为 90%时, 其并网节点功率稳定率低于基于 GRP-PPO(策略一)的调度方法, 且在决策时间方面, 基于 GRP-PPO 算法的智能调度方法在线决策时间仅为 0.178 s 左右, 相较于 SQP 其在线决策效率大幅提升。

4 结论

本文针对分布式光伏发电的不确定性与 EV 大规模充电造成电网功率波动大的问题, 以消纳分布式光伏发电量、减缓并网功率波动、为用户创造收益为目标, 提出了分布式光伏-EV 的互补调度

表 5 各辆 EV 收益(策略二)

Table 5 Revenue per EV (strategy 2)

EV	时刻, 电量/kWh		原充电 费/元	系统内充 电费/元	放电补 贴/元	收益/ 元
	接入	离开				
EV15	全天				82.6	82.6
EV21	07:15, 41.5	19:30, 56.8	8.8	3.1	39.6	45.3
EV38	08:00, 27.4	20:30, 51.0	13.5	4.7	54.8	63.6
EV46	13:30, 46.7	16:45, 72.7	14.9	5.2		9.7
EV63	08:15, 37.0	15:00, 76.9	22.9	8.0		14.9

表 6 不同方法调度结果

Table 6 Scheduling results of different methods

调度方法	预测精度/%	功率稳定率	决策时间/s
GRP-PPO(策略一)		91.84	0.178
GRP-PPO(策略二)		79.74	0.174
A2C		77.35	0.187
SQP	100	93.41	
	90	85.42	10.532
	80	76.73	

模型, 利用改进的 PPO(GRP-PPO)深度强化学习算法研究了就近消纳光伏前提下, 分布式光伏-EV 互补系统的实时调度问题, 对比分析了 GRP-PPO 算法的有效性和先进性, 得出以下结论:

1) 对比 PPO 算法, 改进后的 GRP-PPO 算法能够减少模型陷入局部最优且无法跳出的情况。调度效果较传统 PPO 算法提高了 3.48%。GRP-PPO 算法的调度效果优于 A2C 算法和 SQP 算法, 其在线决策时间远小于 SQP 算法, 降低了计算成本。

2) GRP-PPO 算法能够根据系统状态实时处理调度问题, 有效地平抑了并网节点功率波动, 运行时间内并网节点的功率稳定率最高可达到 91.84%, 并为用户创造经济收益, 任意电量的 EV 在任意时刻接入均可得到相应的调度策略。

3) GRP-PPO 算法可以通过对模型目标函数的调整得到基于不同优化目标的优化策略, 从而满足不同需求的优化目标。策略一的调度下参与调度 EV 电量绝大部分时间保持在 70%以上, 能够满足用户的 24 h 用车需求; 策略二调度下参与调度用户的经济效益显著提升, 全天参与调度的 EV 收入达 82.6 元。EV 用户可根据自己的需求选择相应的调度策略, 策略适应性较强, 可起到鼓励用户参与调度的目的。

本文所研究的分布式光伏-EV 互补系统智能调

度策略为电力系统的经济性调度提供了一种新思路,有助于多能耦合情况下智能调度的分析与设计。

参考文献 References

- [1] 国家能源局. 2022 年光伏发电建设运行情况[EB/OL].[2023-02-17]. http://www.nea.gov.cn/2023-02/17/c_1310698128.htm. National Energy Administration. Construction and operation of photovoltaic power generation in 2022[EB/OL]. Retrieved February 17, 2023, from http://www.nea.gov.cn/2023-02/17/c_1310698128.htm.
- [2] DURRANI S P, BALLUFF S, WURZER L, et al. Photovoltaic yield prediction using an irradiance forecast model based on multiple neural networks[J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(2): 255-267.
- [3] 陶 琼, 王德顺, 叶季蕾, 等. 考虑储能配置模式的多数据源融合分布式光伏发电并网接纳分析方法[J]. 高电压技术, 2018, 44(4): 1093-1098.
TAO Qiong, WANG Deshun, YE Jilei, et al. Capacity analysis of distributed photovoltaic generation integrated into power grid considering energy storage configuration mode based on fusion of multiple data sources[J]. High Voltage Engineering, 2018, 44(4): 1093-1098.
- [4] 梁振锋, 祁芙蓉, 王国阳, 等. 电动汽车有序充电优化策略综述[J]. 新型电力系统, 2023, 1(3): 256-271.
LIANG Zhenfeng, QI Furong, WANG Guoyang, et al. Review of optimization strategies for orderly charging of electric vehicles[J]. New Type Power Systems, 2023, 1(3): 256-271.
- [5] 张家美, 孙 凯, 李洪涛, 等. 考虑光-车-氢接入的新型城市配电网规划研究综述与展望[J]. 高电压技术, 2024, 50(3): 1067-1079.
ZHANG Jiamei, SUN Kai, LI Hongtao, et al. Review and prospect of future city distribution networks considering the integration of PV-EV-hydrogen[J]. High Voltage Engineering, 2024, 50(3): 1067-1079.
- [6] 宋晓通, 吕倩楠, 孙 艺, 等. 基于电价引导的电动汽车与综合能源系统交互策略[J]. 高电压技术, 2021, 47(10): 3744-3754.
SONG Xiaotong, LÜ Qiannan, SUN Yi, et al. Interactive strategy of electric vehicles and integrated energy system based on electricity price guidance[J]. High Voltage Engineering, 2021, 47(10): 3744-3754.
- [7] 侯 慧, 王逸凡, 吴细秀, 等. 长时间尺度下计及里程焦虑心理效应的电动汽车充放电调度策略[J]. 高电压技术, 2023, 49(1): 85-93.
HOU Hui, WANG Yifan, WU Xixiu, et al. Electric vehicle charging and discharging scheduling strategy considering psychological effect of mileage anxiety on a long-time scale[J]. High Voltage Engineering, 2023, 49(1): 85-93.
- [8] 肖 丽, 谢尧平, 胡华锋, 等. 基于 V2G 的电动汽车充放电双层优化调度策略[J]. 高压电器, 2022, 58(5): 164-171.
XIAO Li, XIE Yaoping, HU Huafeng, et al. Two-level optimization scheduling strategy for EV's charging and discharging based on V2G[J]. High Voltage Apparatus, 2022, 58(5): 164-171.
- [9] 孙 波, 董 浩, 王 璐, 等. 计及电动汽车需求响应的风储混合系统日前联合优化调度模型[J]. 太阳能学报, 2021, 42(4): 107-114.
SUN Bo, DONG Hao, WANG Lu, et al. Joint optimal scheduling model for wind-storage hybrid system considering demand response of electric vehicle[J]. Acta Energetica Solaris Sinica, 2021, 42(4): 107-114.
- [10] 李怡然, 张 姝, 肖先勇, 等. V2G 模式下计及供需两侧需求的电动汽车充放电调度策略[J]. 电力自动化设备, 2021, 41(3): 129-135, 143.
LI Yiran, ZHANG Shu, XIAO Xianyong, et al. Charging and discharging scheduling strategy of EVs considering demands of supply side and demand side under V2G mode[J]. Electric Power Automation Equipment, 2021, 41(3): 129-135, 143.
- [11] LIU N, CHEN Q F, LIU J, et al. A heuristic operation strategy for commercial building microgrids containing EVs and PV system[J]. IEEE Transactions on Industrial Electronics, 2015, 62(4): 2560-2570.
- [12] OMRAN W A, KAZERANI M, SALAMA M M A. Investigation of methods for reduction of power fluctuations generated from large grid-connected photovoltaic systems[J]. IEEE Transactions on Energy Conversion, 2011, 26(1): 318-327.
- [13] 庞传军, 王 珅, 余建明. 可解释机器学习在电网调控领域中的应用[J]. 电力信息与通信技术, 2024, 22(5): 1-9.
PANG Chuanjun, WANG Shen, YU Jianming. Application of interpretable machine learning in power grid dispatching and control[J]. Electric Power Information and Communication Technology, 2024, 22(5): 1-9.
- [14] 李 涛, 胡维昊, 李 坚, 等. 基于深度强化学习算法的光伏-抽蓄互补系统智能调度[J]. 电工技术学报, 2020, 35(13): 2757-2768.
LI Tao, HU Weihao, LI Jian, et al. Intelligent economic dispatch for PV-PHS integrated system: a deep reinforcement learning-based approach[J]. Transactions of China Electrotechnical Society, 2020, 35(13): 2757-2768.
- [15] QIAO J F, WANG G M, LI W J, et al. An adaptive deep Q-learning strategy for handwritten digit recognition[J]. Neural Networks, 2018, 107: 61-71.
- [16] 江昌旭, 刘晨曦, 林 铮, 等. 基于深度强化学习的电力系统暂态稳定控制策略研究综述[J]. 高电压技术, 2023, 49(12): 5171-5186.
JIANG Changxu, LIU Chenxi, LIN Zheng, et al. Review of power system transient stability control strategies based on deep reinforcement learning[J]. High Voltage Engineering, 2023, 49(12): 5171-5186.
- [17] 李 航, 李国杰, 汪可友. 基于深度强化学习的电动汽车实时调度策略[J]. 电力系统自动化, 2020, 44(22): 161-167.
LI Hang, LI Guojie, WANG Keyou. Real-time dispatch strategy for electric vehicles based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2020, 44(22): 161-167.
- [18] 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.
- [19] WAN Z Q, LI H P, HE H B, et al. Model-free real-time EV charging scheduling based on deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2019, 10(5): 5246-5257.
- [20] 刘敦楠, 王玲湘, 汪伟业, 等. 基于深度强化学习的大规模电动汽车充换电负荷优化调度[J]. 电力系统自动化, 2022, 46(4): 36-46.
LIU Dunnan, WANG Lingxiang, WANG Weiye, et al. Optimal scheduling of electric vehicle load for large-scale battery charging and swapping based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2022, 46(4): 36-46.
- [21] 刘 全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2018, 41(1): 1-27.
LIU Quan, ZHAI Jianwei, ZHANG Zongzhang, et al. A survey on deep reinforcement learning[J]. Chinese Journal of Computers, 2018, 41(1): 1-27.
- [22] 马骋乾, 谢 伟, 孙伟杰. 强化学习研究综述[J]. 指挥控制与仿真, 2018, 40(6): 68-72.
MA Chengqian, XIE Wei, SUN Weijie. Research on reinforcement learning technology: a review[J]. Command Control & Simulation, 2018, 40(6): 68-72.

- [23] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[EB/OL]. arXivorg e-Print archive. [2023-08-28]. <https://arxiv.org/abs/1707.06347>.
- [24] 徐智威, 胡泽春, 宋永华, 等. 充电站内电动汽车有序充电策略[J]. 电力系统自动化, 2012, 36(11): 38-43.
XU Zhiwei, HU Zechun, SONG Yonghua, et al. Coordinated charging of plug-in electric vehicles in charging stations[J]. Automation of Electric Power Systems, 2012, 36(11): 38-43.
- [25] 范培潇, 杨 军, 柯 松, 等. 基于改进 EDRL 的含 V2G 孤岛微电网频率综合控制策略[J]. 电力系统自动化, 2023, 47(20): 23-32.
FAN Peixiao, YANG Jun, KE Song, et al. Integrated frequency control strategy for islanded microgrid with vehicle-to-grid based on improved evolutionary deep reinforcement learning[J]. Automation of Electric Power Systems, 2023, 47(20): 23-32.
- [26] 杨 捷, 郭 凡, 曹子健. 电动汽车储能 V2G 模式的成本与收益分析[J]. 储能科学与技术, 2020, 9(增刊 1): 45-51.
YANG Jie, GUO Fan, CAO Zijian. Cost and benefit analysis of EV energy storage through V2G[J]. Energy Storage Science and Technology, 2020, 9(Supplement1): 45-51.
- [27] 梁振锋, 祁芙蓉, 王德意, 等. 考虑应急电源功能的光储充电站配置方法研究[J]. 电网技术, 2023, 47(8): 3376-3384.
LIANG Zhenfeng, QI Furong, WANG Deyi, et al. Configuration of optical storage & charging and discharging power station considering emergency power function[J]. Power System Technology, 2023, 47(8): 3376-3384.
- [28] LEE Z J, LI T X, LOW S H. ACN-Data: analysis and applications of an open EV charging dataset[C]//Proceedings of the Tenth ACM International Conference on Future Energy Systems. Phoenix, USA: ACM, 2019: 139-149.



CHENG Ning

陈 宁

1999—, 男, 硕士生

主要从事微电网优化调度方面的研究工作

E-mail: 2685910987@qq.com



LI Fashe

李法社(通信作者)

1978—, 男, 博士, 教授, 博导

主要从事生物质能源转化方面的研究工作

E-mail: asan97@qq.com

Ph.D., Professor
Corresponding author

收稿日期 2023-10-12 修回日期 2024-09-20 编辑 何秋萍