

DOI: 10.19783/j.cnki.pspc.240122

多时间尺度协同的配电网分层深度强化学习电压控制策略

祁向龙, 陈健, 赵浩然, 张文, 张柯雨

(电网智能化调度与控制教育部重点实验室(山东大学), 山东 济南 250061)

摘要: 光伏和电动汽车的大量接入给配电网电压控制带来了严峻挑战。为此, 提出了一种多时间尺度协同的配电网分层深度强化学习电压控制策略。首先, 考虑电动汽车用户需求响应特性, 在日前协同调度分组投切电容器制定分时电价引导用户改变充电行为实现有序充电。其次, 在日内构建具有双层结构的深度强化学习策略, 上层以长时间尺度通过电价调整引导用户响应实时价格激励以调节电动汽车充电功率, 下层以短时间尺度对光伏逆变器和无功补偿装置进行控制, 通过双层策略实现对具有不同时间响应特性资源的实时协同调控以降低系统电压偏差。最后, 通过对改进的 IEEE 33 节点系统进行算例分析验证了所提策略的有效性。

关键词: 电动汽车; 多时间尺度; 电压控制; 多智能体; 深度强化学习

Multi-time scale cooperative voltage control strategy of a distribution network based on hierarchical deep reinforcement learning

QI Xianglong, CHEN Jian, ZHAO Haoran, ZHANG Wen, ZHANG Keyu

(Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan 250061, China)

Abstract: The access of a high proportion of photovoltaic energy and a large number of electric vehicles imposes severe challenges to the voltage control of a distribution network. This paper presents a multi-time scale cooperative voltage control strategy based on hierarchical deep reinforcement learning. First, considering the demand response characteristics of electric vehicle users, the day-ahead time-of-use price is formulated to guide users to change their charging behaviors to achieve orderly charging of coordinated scheduling of switching capacitors. Secondly, a deep reinforcement learning strategy with a two-layer structure is constructed within the day. The upper layer guides users to respond to real-time price incentives through price adjustment to adjust electric vehicle charging load over a long time-scale, and the lower layer controls photovoltaic inverters and reactive power compensation devices on a short time-scale. By the two-layer strategy, the real-time coordinated regulation of resources with different time response characteristics is realized to reduce system voltage deviation. Finally, the effectiveness of the proposed strategy is verified on the improved IEEE 33-bus system.

This work is supported by the National Key R&D Program of China (No. 2018YFA0702200).

Key words: electric vehicle; multi-time scale; voltage control; multi-agent; deep reinforcement learning

0 引言

由于在节能减排方面的优势, 分布式光伏 (photovoltaic, PV) 和电动汽车 (electric vehicle, EV) 的发展对加快电力系统能源转型和实现我国“双碳”目标具有重要意义^[1-2]。随着配电网中 PV 和 EV 数量的不断增加, 系统峰谷差增大、反向潮流导致电压越限等问题频繁出现, 给配电网的运行控制造成

了一系列影响^[3-4]。因此, 研究大规模 PV 和 EV 接入下配电网的电压控制问题并减少其负面影响是亟需解决的关键问题之一。

当前国内外学者多以有载调压变压器、静止无功补偿装置、PV 等作为主要调压设备进行配电网电压控制, 并进行了大量研究^[5-6]。文献[7]充分考虑 PV 逆变器无功功率输出与实时下垂控制的协同, 提出一种分布式混合时间尺度的电压控制方法。文献[8-9]分别考虑了 PV 有功无功协调的集中式和分布式电压控制策略, 能够实现良好的电压控制效果。

基金项目: 国家重点研发计划项目资助 (2018YFA0702200)

文献[10]建立了两级电压控制框架,通过本地控制器进行信息共享实现分布式控制。

但上述研究中大多仅考虑源侧和网侧的调压设备,忽视了负荷侧可调节资源(如EV等)对配电网调压的作用。对于通过EV改善配电网电压水平等方面国内外学者进行了相关研究。文献[11]考虑EV充电站无功补偿能力建立了综合下垂控制和电压支撑控制的电压调节策略,有效改善了电能质量。文献[12]考虑EV充电功率因素可调节特性进行无功补偿,以制定最优电压控制策略。考虑EV有功可调节能力,文献[13]提出一种对充电站有功和逆变器无功协调的电压控制策略。文献[14]建立了EV和分布式电源的分层分区调压框架,满足了配电网电压的安全运行。除上述研究外,有学者考虑通过价格激励的方式对EV进行调节并得到了优化控制策略。文献[15]考虑分时电价影响提出一种EV动态充电调度策略。文献[16]以实时电价为手段对EV的充放电行为进行优化,建立了最小化充电费用的最优充放电策略。文献[17]综合考虑价格型需求响应和激励型需求响应建立了EV充放电调度策略。通过价格信号等手段调节EV,并与其他调压设备协同进行电压调节,能够有效改善配电网电压水平,但以上研究缺乏对于这方面的考虑。

上述文献多采用基于模型的控制方法,虽然目前基于模型的优化方法能够满足电压控制要求,但这些方法严重依赖于精确的系统模型和良好的通信条件,随着系统复杂度增加和设备规模扩大,问题求解和控制难度剧增。因此,受数据驱动等方法的启发,国内外学者尝试采用深度强化学习(deep reinforcement learning, DRL)方法,并通过模拟与系统仿真环境进行交互的过程,得到控制器的电压控制策略^[18-20]。文献[20-21]考虑有载调压变压器、分组电容器、PV等调压设备,提出了一种两阶段电压控制策略,通过基于物理模型和数据驱动的方法进行问题求解。文献[22]计及PV调节潜力进行分区,并通过多智能体DRL(multi-agent DRL, MADRL)算法实现就地电压控制,缓解了配电网电压波动。文献[23]考虑不同电压调节设备特性,分别从不同时间尺度通过MADRL进行协同控制,以缓解电压越限问题。由于对EV进行调节的时间尺度较长,与PV等装置协同进行电压控制可认为是一种具有不同时间响应特性的复合空间的序贯决策问题,通过分层强化学习方法能够实现对上述高维复合动作空间的感知、探索和学习,得到实时协同决策的最优策略。

为此,本文考虑PV、无功补偿装置与EV的时间响应特性差异,基于分层深度强化学习提出了一

种多时间尺度协同的配电网电压控制策略。首先,考虑EV用户的不同出行特征,协同调度分组投切电容器(switching capacitor, SC)并制定分时电价以引导用户改变充电行为,实现日前EV的有序充电,降低系统网损。其次,为减小系统运行电压偏差,将日内电压调节过程分为长时间尺度控制和短时间尺度控制,并构建具有双层结构的深度强化学习策略。上层以长时间尺度(小时级)通过实时电价引导用户响应实时价格激励以调节EV充电功率,下层以短时间尺度(分钟级)控制PV逆变器和静止无功补偿装置(static var compensator, SVC)的无功功率,从而实现对不同时间响应特性资源的协同调控。最后,通过在改进的IEEE 33节点系统上进行仿真,验证了所提策略的有效性。

1 EV响应价格信号概率模型

大量电动汽车接入配电网后进行无序充电易形成新的负荷高峰,加剧系统峰谷差,导致电压偏差增大,通过价格信号对EV进行引导能有效调动用户响应的积极性,优化系统运行状态,并为电压控制提供可调节能力。

1.1 EV充电模型

EV充电时长主要受EV用户的出行需求、出行习惯等因素影响,因为EV处于停泊状态时才进行充电行为,所以EV的充电时长由到达时刻、离开时刻、到达时电池的荷电状态(state of charge, SOC)和离开时期望的SOC共同决定^[24]。本文通过正态分布函数对EV到达时刻和离开时刻的概率分布进行表示。

$$f(t) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) & 0 < t < \mu + 12 \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-24-\mu)^2}{2\sigma^2}\right) & \mu + 12 < t < 24 \end{cases} \quad (1)$$

$$f(t) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t+24-\mu)^2}{2\sigma^2}\right) & 0 < t < \mu - 12 \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) & \mu - 12 < t < 24 \end{cases} \quad (2)$$

式中: μ 为EV抵达/离开充电站时刻的期望值; σ 为对应的标准差。当 $0 \leq \mu < 12$ 时,通过式(1)表示,当 $12 \leq \mu < 24$ 时,通过式(2)表示。EV抵达充电站时刻的SOC也可通过上述公式进行表示。

因此,根据上述出行特征能够得到EV充电时长为

$$T_{ch}^{ev} = \begin{cases} \frac{E_{max}^{ev}(S_{oc\ le} - S_{oc\ ar})}{P_{max}^{ev}}, \frac{E_{max}^{ev}(E_{max}^{ev}(S_{oc\ le} - S_{oc\ ar}))}{P_{max}^{ev}} \leq T_{st}^{ev} \\ T_{st}^{ev}, \frac{E_{max}^{ev}(E_{max}^{ev}(S_{oc\ le} - S_{oc\ ar}))}{P_{max}^{ev}} > T_{st}^{ev} \end{cases} \quad (3)$$

式中: T_{ch}^{ev} 为 EV 在停泊时的充电时间; E_{max}^{ev} 为 EV 的电池容量; P_{max}^{ev} 为 EV 的额定充电功率; $S_{oc\ ar}$ 为抵达充电站时 EV 的 SOC; $S_{oc\ le}$ 为离开充电站时 EV 期望电池的 SOC; T_{st}^{ev} 为 EV 的停泊时长, 若停泊时长小于所需充电时长, 则该 EV 不会改变充电行为。

1.2 EV 响应日前分时电价概率模型

分时电价是基于价格的需求侧响应的一种, 考虑用电成本, 用户的用电需求与该时段的电价呈负相关, 电价越高则该时段用电量越小, 通过峰谷价差能够调整用户改变充电行为的概率, 实现对充电负荷的控制。

EV 用户能够通过智能控制终端对充电行为进行自主规划, 在不响应分时电价的情况下, 用户到达充电站后即开始充电, 当电量达到期望值时停止^[25], 若用户响应分时电价, 则最优充电时段根据以下原则确定: (1) 保证 EV 在停车状态下能够充电至期望电量, 改变充电时段后总充电费用最小; (2) 考虑用车需求, 尽早结束充电行为。

所以, 响应分时电价前后某 EV 进行充电的时段变化如图 1 所示。

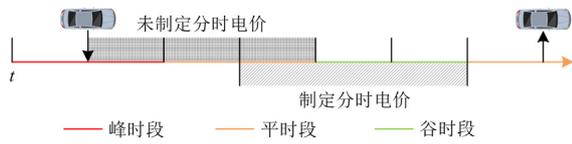


图 1 某 EV 充电时段变化

Fig. 1 Changes in charging periods for an EV

因为 EV 用户响应分时电价前后的充电费用不同, 所以两者的差值会影响用户改变充电行为的意愿, 差值越大激励水平越高, 用户的响应程度越高, 通过引入价格弹性系数来表征充电费用差值与用户响应价格信号概率间的关系。

$$p_{rob} = \varepsilon \left(\frac{c_{ch}^{ev} - c_{ch}^{ev'}}{T_{ch}^{ev}} \right) \quad (4)$$

$$\begin{cases} c_{ch}^{ev} = \omega_{peak} h_{peak}^{ev} + \omega_{flat} h_{flat}^{ev} + \omega_{valley} h_{valley}^{ev} \\ c_{ch}^{ev'} = \omega_{peak} h_{peak}^{ev'} + \omega_{flat} h_{flat}^{ev'} + \omega_{valley} h_{valley}^{ev'} \\ h_{peak}^{ev} + h_{flat}^{ev} + h_{valley}^{ev} = h_{peak}^{ev'} + h_{flat}^{ev'} + h_{valley}^{ev'} = T_{ch}^{ev} \end{cases} \quad (5)$$

式中: p_{rob} 表示 EV 响应分时电价的概率; c_{ch}^{ev} 和 $c_{ch}^{ev'}$

分别为 EV 不响应和响应分时电价的充电费用; ω_{peak} 、 ω_{flat} 和 ω_{valley} 分别为峰、平、谷时段的电价; h_{peak}^{ev} 、 h_{flat}^{ev} 和 h_{valley}^{ev} 分别为不响应分时电价时属于峰、平、谷时段的小时数; $h_{peak}^{ev'}$ 、 $h_{flat}^{ev'}$ 和 $h_{valley}^{ev'}$ 分别为响应分时电价时属于峰、平、谷时段的小时数; ε 为价格弹性系数, $\varepsilon > 0$ 。通过 EV 响应概率经过蒙特卡洛模拟法能够得到各 EV 充电时段, 对各 EV 的充电功率进行累加, 即可得到各时刻各充电站充电负荷曲线。

1.3 EV 响应日内实时电价概率模型

日前分时充电电价制定后, EV 用户从自身充电需求出发能够自主响应分时电价改变充电行为, 但容易导致用户在谷电价时段进行集中充电造成配电网出现电压问题。从日内角度考虑, 通过调整实时电价能够引导 EV 快速响应为配电网电压控制提供功率支撑。

该时刻更新后的实时电价只影响该时刻接入充电站的 EV 用户, 不影响已经接入充电站开始充电和还未接入充电站的用户。根据式(4), 可以得到 EV 响应分时电价的预期平均价格激励 λ_{ev} 为

$$\lambda_{ev} = \frac{c_{ch}^{ev} - c_{ch}^{ev'}}{T_{ch}^{ev}} \quad (6)$$

当 EV 达到充电站时, 该时刻分时电价 ω_t^{lou} 与实时电价 ω_t 的差值作为实时价格激励 λ_t , 其会影响用户改变充电行为的概率, 引入实时价格激励和预期平均价格激励的比值以及 Sigmoid 函数来更新当前 EV 改变充电行为的概率。

$$\kappa_t = \frac{\lambda_t}{\lambda_{ev}} = \frac{\omega_t^{lou} - \omega_t}{\lambda_{ev}} \quad (7)$$

$$p'_{robt} = \frac{\varepsilon}{1 + \beta e^{\alpha(\kappa_t + 1)}} \left(\frac{c_{ch}^{ev} - c_{ch}^{ev'}}{T_{ch}^{ev}} \right) \quad (8)$$

式中: κ_t 为 t 时刻对 EV 的激励系数, 当实时价格激励为正时, 系数为正, EV 改变充电行为的概率降低, 当实时电价激励为负时, 系数为负, EV 改变充电行为概率增大; α 和 β 为 Sigmoid 函数的参数, 决定曲线陡度和曲线在纵轴的位置; p'_{robt} 为 t 时刻更新实时价格激励后 EV 改变充电行为的概率。同 1.2 节, 通过更新后的概率进行蒙特卡洛模拟, 即可得到该时刻各充电站充电负荷功率。

2 多时间尺度协同的电压控制策略

多时间尺度协同的电压控制策略如图 2 所示, 包括日前和日内两个优化阶段。

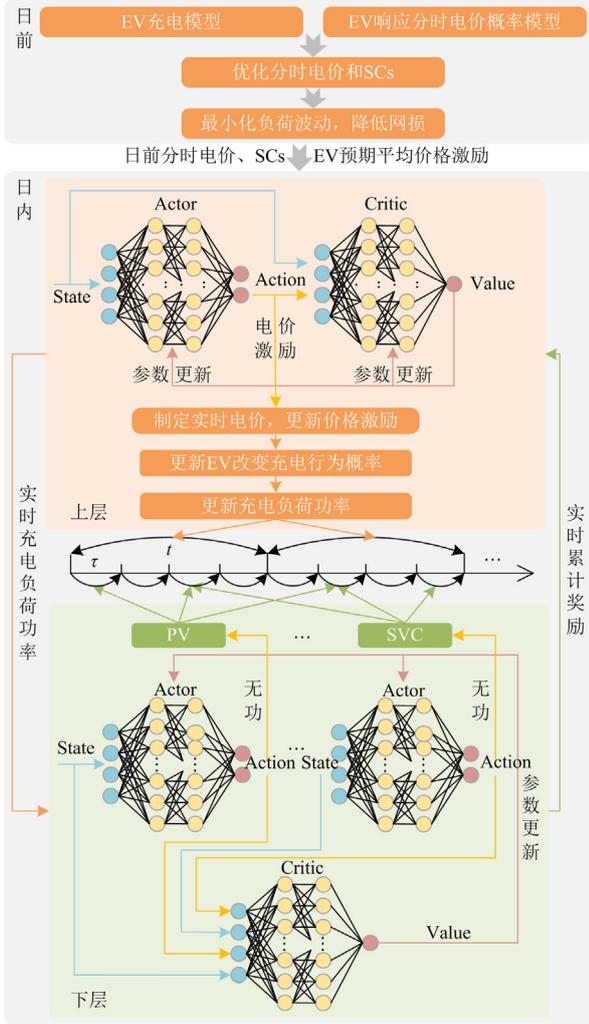


图2 多时间尺度协同的电压控制策略

Fig. 2 Multi-time scale coordinated voltage control strategy

在日前优化阶段,基于1.1节和1.2节优化分时电价引导EV有序充电,能够有效降低配电网负荷波动。由于负荷曲线方差与系统网损成正比,所以日前以最小化网损为优化目标,并通过协同调度SC能够有效减少有功损耗^[26]。日前定价的意义在于在日前实现EV的有序充电,并为日内制定电价激励提供预期平均价格激励基础。

在日前优化的基础上,所提出的日内优化根据EV、PV和SVC的响应时间尺度分为上、下两层,以系统电压偏差最小为目标构建具有双层结构的深度强化学习策略。其中,基于日前分时电价得到的EV预期平均价格激励,上层以长时间尺度通过对电价进行实时调整更新实时价格激励,改变用户心理预期,通过影响用户改变充电行为的概率调节具有粗粒度时间的EV充电功率;下层则以短时间尺度控制具有细粒度时间的PV和SVC。

2.1 日前优化策略

根据1.2节所述,改变分时电价能够改变EV响应后的充电负荷功率 $P_{i,t}^{\text{ev, ch}}$,通过优化分时电价能够减小系统净负荷波动、降低网损,所以日前优化模型的目标函数和相关约束为

$$f = \min \sum_{i=1}^{T_d} \sum_{l=1}^L I_{l,t}^2 r_l \quad (9)$$

式中: T_d 为日内小时数; L 为配电网的线路数量; $I_{l,t}$ 为 t 时刻线路 l 的电流; r_l 为线路 l 的电阻。

$$\begin{cases} \sum_{l(i,:)} P_{l,t} + P_{i,t}^{\text{pv}} + P_{i,t}^{\text{sub}} = \sum_{l(i,:)} (P_{l,t} + I_{l,t}^2 r_l) + P_{i,t}^{\text{load}} + P_{i,t}^{\text{ev, ch}} \\ \sum_{l(i,:)} Q_{l,t} + Q_{i,t}^{\text{pv}} + Q_{i,t}^{\text{svc}} + Q_{i,t}^{\text{sub}} + Q_{i,t}^{\text{sc}} = \\ \sum_{l(i,:)} (Q_{l,t} + I_{l,t}^2 x_l) + Q_{i,t}^{\text{load}} \end{cases} \quad (10)$$

$$U_{i,t}^2 = U_{j,t}^2 + 2(P_{l,t} r_l + Q_{l,t} x_l) + I_{l,t}^2 (r_l^2 + x_l^2) \quad (11)$$

$$\begin{cases} 2P_{l,t} \\ 2Q_{l,t} \\ \|I_{l,t}^2 - U_{i,t}^2\| \end{cases} \leq I_{l,t}^2 + U_{i,t}^2 \quad (12)$$

$$\begin{cases} U_{\min}^2 \leq U_{i,t}^2 \leq U_{\max}^2 \\ I_{\min}^2 \leq I_{l,t}^2 \leq I_{\max}^2 \\ P_{\min}^{\text{sub}} \leq P_{i,t}^{\text{sub}} \leq P_{\max}^{\text{sub}} \\ Q_{\min}^{\text{sub}} \leq Q_{i,t}^{\text{sub}} \leq Q_{\max}^{\text{sub}} \\ 0 \leq P_{i,t}^{\text{pv}} \leq P_{\max}^{\text{pv}} \\ (Q_{i,t}^{\text{pv}})^2 + (P_{i,t}^{\text{pv}})^2 \leq (S^{\text{pv}})^2 \\ -Q_{\max}^{\text{svc}} \leq Q_{i,t}^{\text{svc}} \leq Q_{\max}^{\text{svc}} \end{cases} \quad (13)$$

$$Q_{i,t}^{\text{sc}} = \sum_v n_{i,v,t} q_{i,t}^{\text{sc}} \quad (14)$$

$$\sum_{t=1}^{T_d-1} |n_{i,v,t+1} - n_{i,v,t}| \leq n_{i,v,\max} \quad (15)$$

式中: $l(i,:)$ 和 $l(:,i)$ 分别为以节点 i 为首端和末端的线路集合; $P_{l,t}$ 和 $Q_{l,t}$ 分别为 t 时刻线路 l 的有功功率和无功功率; $P_{i,t}^{\text{pv}}$ 和 $Q_{i,t}^{\text{pv}}$ 分别为 t 时刻接入节点 i 的PV有功功率和无功功率; $Q_{i,t}^{\text{sc}}$ 为 t 时刻接入节点 i 的SC无功功率; $P_{i,t}^{\text{sub}}$ 和 $Q_{i,t}^{\text{sub}}$ 分别为 t 时刻接入节点 i 的变电站有功功率和无功功率; $P_{i,t}^{\text{load}}$ 和 $Q_{i,t}^{\text{load}}$ 分别为 t 时刻接入节点 i 的负荷有功功率和无功功率; $Q_{i,t}^{\text{svc}}$ 为 t 时刻接入节点 i 的SVC无功功率; $P_{i,t}^{\text{ev, ch}}$ 为 t 时刻节点 i 处充电站的充电功率; x_l 为线路 l 的电抗; $U_{i,t}$ 为 t 时刻节点 i 的电压; U_{\min} 和 U_{\max} 为节点电压

限值; I_{\min} 和 I_{\max} 为线路能够通过电流的限值; P_{\min}^{sub} 、 P_{\max}^{sub} 、 Q_{\min}^{sub} 和 Q_{\max}^{sub} 分别为变电站的有功功率和无功功率限值; P_{\max}^{PV} 为 PV 能够发出的最大有功功率; S^{PV} 为光伏逆变器的额定容量; Q_{\max}^{SVC} 为 SVC 的无功功率限值。式(14)和式(15)为 SC 无功输出和运行约束; $n_{i,v,t}$ 为 t 时刻第 v 个 SC 的投切动作; $q_{i,t}^{\text{sc}}$ 为 SC 投切一次的额定无功功率; $n_{i,v,\max}$ 为第 v 个 SC 最大动作次数。

由于关于 SC 的约束中存在非线性项, 所以将式(14)、式(15)通过线性化方法转化为式(16)。

$$\begin{cases} Q_{i,t}^{\text{sc}} = \sum_v m_{i,v,t} q_{i,t}^{\text{sc}} \\ y_{i,v,t} = m_{i,v,t} + m_{i,v,t+1} - 2z_{i,v,t} \\ z_{i,v,t} \leq m_{i,v,t} \\ z_{i,v,t} \leq m_{i,v,t+1} \\ z_{i,v,t} \geq m_{i,v,t} + m_{i,v,t+1} - 1 \\ \sum_{t=1}^{T_i} y_{i,v,t} \leq n_{i,v,\max} \end{cases} \quad (16)$$

式中: $y_{i,v,t}$ 为二进制变量, 表示 t 时刻第 v 个 SC 相邻两时刻是否进行投切操作; $m_{i,v,t}$ 为二进制变量, 表示 t 时刻第 v 个 SC 档位的状态; $z_{i,v,t}$ 为二进制变量, $z_{i,v,t} = m_{i,v,t} \cdot m_{i,v,t+1}$ 。

2.2 日内基于分层 DRL 的电压控制策略

2.2.1 单/多智能体深度强化学习

深度学习 (deep learning, DL) 和强化学习 (reinforcement learning, RL) 是机器学习 (machine learning, ML) 中两个重要的组成领域。DL 基于人工神经网络实现, 具有非常强的数据特征表达能力; 而 RL 能够设置智能体和环境, 通过智能体与环境进行交互得到真实反馈实现学习进化, 而 DRL 充分结合了两者的优势, 在强化学习过程中使用神经网络实现端到端的优化学习。

DRL 中智能体对序列决策过程进行不断重复探索以学习到最优策略, 通常单智能体强化学习过程可以构造为马尔科夫策略过程 (Markov decision process, MDP), 而多智能体强化学习需要通过马尔科夫博弈过程 (Markov game process, MGP) 进行描述。无论是 MDP 还是 MGP, 智能体的元素均包括 $\langle \mathcal{S}, \zeta, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, 其中: \mathcal{S} 为智能体能够感知的当前状态; ζ 为智能体的策略; \mathcal{A} 为智能体在当前状态下根据策略做出的动作; \mathcal{P} 为状态转移函数, 即在当前状态和动作下转移到下一状态的概率; \mathcal{R} 为智能体在环境中执行动作后能够获得的奖励; γ 表示奖励折扣因子, 表示未来的奖励对价值函数的

影响。智能体与环境进行交互获得轨迹 τ , 核心目的是最大化期望回报 $J(\zeta)$ 。

$$\max J(\zeta) = \max_{\tau} \int_{\tau} \mathcal{P}(\tau | \zeta) \mathcal{R}(\tau) = \max_{\tau \sim \zeta} \mathbb{E}_{\tau \sim \zeta} [\mathcal{R}(\tau)] \quad (17)$$

式中: $J(\zeta)$ 为智能体在策略 ζ 下获得的期望回报; $\mathcal{P}(\tau | \zeta)$ 为智能体在策略 ζ 下以轨迹 τ 进行交互的状态转移函数; $\mathcal{R}(\tau)$ 为智能体以轨迹 τ 进行交互得到的奖励函数; $\mathbb{E}_{\tau \sim \zeta}$ 为在策略 ζ 下以轨迹 τ 进行交互得到奖励的期望。

2.2.2 近端策略优化算法

近端策略优化 (proximal policy optimization, PPO) 算法是强化学习中的基于策略梯度算法的一种, 通过智能体根据当前策略与环境交互生成数据并进行重要性采样, 然后利用随机梯度上升优化目标函数改进策略参数。

PPO 算法基于行动者-评价者 (Actor-Critic) 网络框架, 维护一个策略网络和一个值函数网络。智能体根据 Actor 网络选择动作输出, 而 Critic 网络用于指导 Actor 网络选择能够获得最大奖励的动作。与确定性策略梯度算法的 Actor 网络不同, PPO 算法采用随机分布的策略函数, 通过 Actor 网络输出的分布上进行采样得到动作值, 所以 PPO 算法具有天然探索环境的优势, 不需要对动作增加其他扰动。

PPO 算法中为了最大化 $J(\zeta)$, 需要计算策略梯度 ∇ 并用于更新参数 θ 。

$$\begin{aligned} \nabla_{\theta} J(\zeta) &= \int_{\tau} \nabla_{\theta} \mathcal{P}_{\theta}(\tau | \zeta) \mathcal{R}(\tau) = \\ & \int_{\tau} \mathcal{P}_{\theta}(\tau | \zeta) \nabla_{\theta} \log(\mathcal{P}_{\theta}(\tau | \zeta)) \mathcal{R}(\tau) = \\ & \mathbb{E}_{\tau \sim \zeta} [\nabla_{\theta} \log(\mathcal{P}_{\theta}(\tau | \zeta)) \mathcal{R}(\tau)] = \\ & \mathbb{E}_{\tau \sim \zeta} [\mathcal{R}(\tau) \sum \nabla_{\theta} \log(\mathcal{P}_{\theta}(\tau | \zeta))] \end{aligned} \quad (18)$$

式中: ∇_{θ} 为计算参数为 θ 时的策略梯度; \mathcal{P}_{θ} 为在参数 θ 下的状态转移函数。

在实际计算中, 为提高更新效率和样本使用率, PPO 算法能够通过多次小批量采样进行参数更新, 并通过未来折扣回报 $Q_{\zeta}(s, a)$ 替换奖励 $\mathcal{R}(\tau)$, 则梯度更新为

$$\nabla_{\theta} J(\zeta) = \sum \mathcal{R}(\tau) \nabla \log(\zeta_{\theta}(a | s)) = \sum Q_{\zeta}(s, a) \nabla \log(\zeta_{\theta}(a | s)) \quad (19)$$

式中: a 为智能体的动作; s 为智能体的状态; ζ_{θ} 表示参数为 θ 时智能体的策略; $\zeta_{\theta}(a | s)$ 为智能体在状态 s 时输出动作 a 的策略 ζ_{θ} ; $Q_{\zeta}(s, a)$ 为智能体在状态 s 时以策略 ζ 为指导输出动作 a 的未来折扣奖励。

进一步根据一个 batch 中第 u 步的经验进行梯度计算, 得

$$\nabla_{\theta} J_u(\zeta) = \sum_{\mathbb{E}_{a_u \sim \zeta}} Q_{\zeta}(s, a) \nabla \log(\zeta_{\theta}(a_u | s_u)) = \mathbb{E}_{a_u \sim \zeta} [Q_{\zeta}(s, a) \nabla \log(\zeta_{\theta}(a_u | s_u))] \quad (20)$$

则更新后的参数 θ' 为

$$\theta' = \theta + lr \cdot \nabla_{\theta} J_u(\zeta) \quad (21)$$

式中: J_u 为第 u 步的期望回报; a_u 为智能体在第 u 步的动作; s_u 为智能体在第 u 步的状态; lr 为网络参数更新的学习率。

在训练过程中, 若更新前后两个策略参数相差过大, 则训练过程不稳定, 容易导致训练失败, 为减小新旧两个策略间差异较大带来的影响, PPO 算法中衍生出 PPO-惩罚和 PPO-截断两种形式, 本文采用基于 PPO-截断的算法, 对应的优化目标函数为

$$J_u(\theta) = \mathbb{E}_{a_u \sim \zeta'(a_u | s_u)} [\min Q_{\zeta'}(s, a) \frac{\zeta(a_u | s_u)}{\zeta'(a_u | s_u)}, \text{clip}\left(\frac{\zeta(a_u | s_u)}{\zeta'(a_u | s_u)}, 1 + \chi, 1 - \chi\right) Q_{\zeta'}(s, a)] \quad (22)$$

式中: ζ' 为智能体更新后的策略; χ 为截断超参数, 一般设置为 0.2。

2.2.3 上层电压控制策略

因为确定了实时电价后该时刻 EV 充电功率随即确定, 所以上层智能体只需决策每一长时间间隔的实时电价, 上层强化学习中智能体的元素如下所述。

1) 状态空间 \mathcal{S}_n : 上层智能体在间隔 t_1 能够观测到的状态 ($s_n \in \mathcal{S}_n$), 包括当前时间间隔的光伏有功功率 $P_{i,n}^{\text{pv}}$ 、负荷有功功率 $P_{i,n}^{\text{load}}$ 、负荷无功功率 $Q_{i,n}^{\text{load}}$ 、充电站 EV 到达数量和当前时间间隔 t_1 。

2) 动作空间 \mathcal{A}_n : 智能体在间隔 t_1 输出动作 $a_n \in \mathcal{A}_n$, 并经过线性映射后得到实时电价。

$$\omega_n = a_n \Delta \lambda + \omega_t^{\text{lou}} \quad (23)$$

式中: ω_n 为上层间隔 t_1 的实时电价; $\Delta \lambda$ 为实时电价激励的可调节范围。

3) 奖励空间 \mathcal{R}_n : 上层智能体输出时间间隔 t_1 的电价激励后, 下层智能体在该时段 t_1 中以更短的时间步长进行连续控制, 并将该间隔内奖励之和的负值作为上层智能体获得的奖励。

$$\mathcal{R}_n = - \sum_{t_s=1}^{T_n} \sum_{i=1}^{N_b} \left| \frac{U_{i,t_s} - U_0}{U_0} \right| - \zeta (|U_{i,t_s} - U_{\max}|^+ + |U_{i,t_s} - U_{\min}|^-) \quad (24)$$

式中: T_n 为上层智能体一个决策间隔中包含的下层智能体的决策步长总数; N_b 为配电网节点数; U_{i,t_s} 为节点 i 在时刻 t_s 的电压; U_0 为系统额定电压; ζ 为电压越限惩罚系数; “ $| \cdot |^+$ ”、“ $| \cdot |^-$ ”分别为计算最

大值和最小值的绝对值函数。

2.2.4 下层电压控制策略

下层智能体控制 PV 和 SVC 的无功功率, 并采用基于中心化训练去中心化执行 (centralized training with decentralized execution, CTDE) 框架的多智能体 PPO (multi-agent PPO, MAPPO) 算法得到控制策略。在 CTDE 框架中, 每个智能体的 Actor 网络仅能观察到自身状态而非全局状态, 而 Critic 网络能观测到每个智能体的状态。所以在训练过程中, 所有智能体的 Critic 网络共享一套参数, 通过中心化的 Critic 网络对每个智能体 Actor 网络进行指导; 而在执行过程中, 每个智能体仅通过自身的观测信息和 Actor 网络完全独立做出动作。在框架下训练好的 PV 和 SVC 控制策略能够保证设备间的协作, 将策略部署至本地控制器后能够使设备不需要进行信息共享, 仅根据就地信息进行实时控制。

下层强化学习中各智能体的元素如下所述。

1) 状态空间 \mathcal{S}_{i,t_s} : 对于第 i 个智能体, 它在间隔 t_1 中的 t_s 时刻的状态 ($s_{i,t_s} \in \mathcal{S}_{i,t_s}$), 包括当前时刻的光伏有功功率 P_{i,t_s}^{pv} 、负荷有功功率 P_{i,t_s}^{load} 、负荷无功功率 Q_{i,t_s}^{load} 和当前时刻, 由于上层以较长时间间隔进行决策, 所以下层当前时刻为上层当前间隔 t_1 中的时刻 t_s 。除上述状态外, 下层的各个智能体应能够观测到上层智能体在时间间隔 t_1 做出的决策 a_n 。

2) 动作空间 \mathcal{A}_{i,t_s} : 第 i 个智能体在间隔 t_1 中的 t_s 时刻的动作为 $a_{i,t_s} \in \mathcal{A}_{i,t_s}$, 因为智能体控制 PV 或 SVC, 所以需要将智能体的动作输出线性映射为设备真实无功功率输出。

$$Q_{i,t_s}^{\text{pv}} = a_{i,t_s} \sqrt{(S^{\text{pv}})^2 - (P_{i,t_s}^{\text{pv}})^2} \quad (25)$$

$$Q_{i,t_s}^{\text{svc}} = a_{i,t_s} Q_{\max}^{\text{svc}} \quad (26)$$

3) 奖励空间 \mathcal{R}_{i,t_s} : 各智能体目标函数相同, 配电网各节点电压偏差的负值为

$$\mathcal{R}_{i,t_s} = - \sum_i^{N_b} \left| \frac{U_{i,t_s} - U_0}{U_0} \right| - \zeta (|U_{i,t_s} - U_{\max}|^+ + |U_{i,t_s} - U_{\min}|^-) \quad (27)$$

式中, U_{i,t_s} 为节点 i 在间隔 t_1 中的 t_s 时刻的电压。

3 算例仿真

3.1 仿真设置

以改进的 IEEE 33 节点配电网为例, 进行算例分析, 如图 3 所示。系统中 PV 安装在节点 13、19 和 27 上, 每个节点安装容量均为 1.5 MW, SVC 安

装在节点 11、23 和 30 上, 每个节点 SVC 容量均为 0.4 Mvar, SC 安装在节点 14 和 31 上, 共有两个档位, 每档为 0.1 Mvar, 每日动作次数上限为 5, EV 充电站设置在节点 15、21 和 29 上。PV 和负荷曲线如附录 A 图 A1 和图 A2 所示, 取自比利时某地实际运行数据^[27], 其中选取光伏 7 天监测数据用于验证本文所提策略的有效性, 随机挑选 5 天数据为训练集, 其余 2 天数据为测试集^[28]。具有不同出行目的的 EV 相关特征如表 1 所示, 设置各充电站 EV 数量均为 700 辆, 电池额定容量为 36 kWh, 额定充电功率为 3 kW, 其中 EV1、EV2 和 EV3 中回家、工作和娱乐的比例分别为 4:2:1、2:4:1 和 1:1:5, EV 所有参数通过蒙特卡洛模拟得到, 电动汽车离开时期望的 SOC 为 90%。

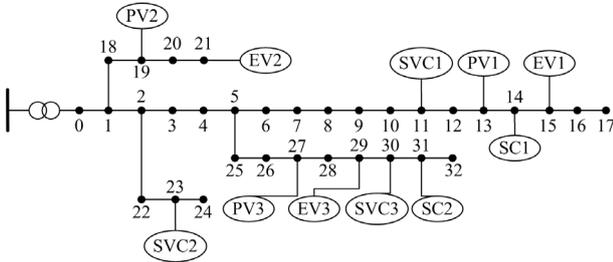


图 3 改进的 IEEE 33 节点配电网

Fig. 3 Improved IEEE 33-bus distribution network

表 1 EV 出行特征

Table 1 EV travel characteristics

出行目的	达到时间分布		离开时刻分布		到达时 SOC	
	μ	σ	μ	σ	μ	σ
回家	19:00	1.5	07:30	1.0	0.4	0.1
工作	08:30	1.0	17:00	1.0	0.5	0.1
娱乐	12:00	1.0	20:00	1.5	0.6	0.1

本文中采用的强化学习算法均基于 Python3.9.13 的 Pytorch1.12.1 神经网络框架进行实现, 上层 PPO 算法和下层 MAPPO 算法中的网络结构和超参数设置如表 2 所示。

3.2 日前优化结果

划分峰时段(06:00—10:00, 18:00—22:00)、平时段(22:00—次日 02:00, 10:00—12:00, 16:00—18:00)和谷时段(02:00—06:00, 12:00—16:00), 设置峰、平、谷时段分时电价可调节范围分别为 1.2~1.5、0.9~1.2、0.6~0.9 元/kWh。通过对 2.1 节中的问题进行求解, 可以得到峰、平、谷时段的电价优化结果分别为 1.26、1.14 和 0.74 元/kWh, EV 响应分时电价后充电站 EV1、EV2 和 EV3 的充电功率如图 4 所示, SCs 的调度结果如附录 A 图

A3 所示, 满足日最大调节次数。

表 2 DRL 网络结构和超参数设置

Table 2 Network structure and hyperparameters of DRL

参数	数值	
	PPO	MAPPO
Actor 网络		
网络层数	3	3
隐藏层神经元个数	128	64
输出层激活函数	Tanh 函数	Tanh 函数
学习率	0.001	0.001
Critic 网络		
网络层数	4	4
隐藏层神经元个数	128	64
输出层激活函数	Tanh 函数	Tanh 函数
学习率	0.001	0.001
折扣因子	0.99	0.99
训练回合数	24	96

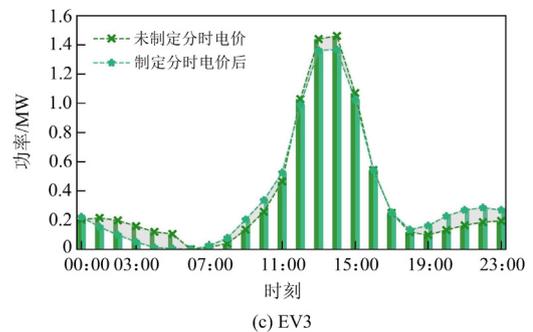
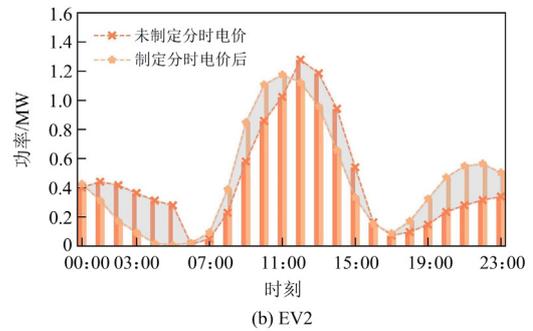
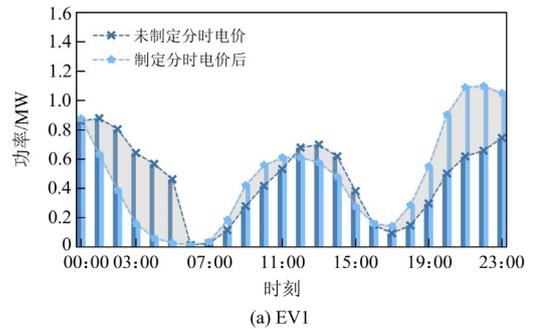


图 4 各充电站充电负荷曲线

Fig. 4 Charging load curve of each charging station

设置分时电价后，由于 06:00—11:00 时段的电价较高，而 12:00—15:00 时段的电价为谷时段电价，所以在 06:00—11:00 时段抵达充电站的 EV 用户改变自身充电行为的意愿更高，而在 12:00—15:00 时段抵达充电站的用户更愿意选择即到即充以降低充电成本。同样地，由于 16:00—次日 02:00 时段的分时电价较高，而 02:00—05:00 时段的电价为谷时段电价，EV 用户在保证充电要求的基础上选择电价更低的时段进行充电。如图 4(a)、图 4(b)所示，充电站 EV1 和 EV2 中以回家和工作为目的的用户数量较多，所以在用户响应价格信号后充电负荷转移，新的充电负荷高峰出现在 11:00—14:00 和 01:00—04:00 时段。对比图 4(b)，由于 EV3 中大部分用户在谷时段抵达充电站，且谷时段电价能够吸引其他时段充电负荷转移到该时段，所以该时段负荷功率增大。

未进行日前优化时，系统有功功率损耗为 2865.08 kWh，而优化后，系统有功功率损耗为 2343.71 kWh，降低了 18.2%，由以上结果可知，日前通过协调调度 SCs 和优化分时电价能够有效减小负荷波动和减少有功损耗，实现了可再生能源的充分利用。

3.3 基于分层 DRL 的实时电压控制结果

设置上层强化学习中智能体个数为 1 对电价进行实时调节，实时电价可调节范围为 0.2 元/kWh，调节时间尺度为 1 h；下层强化学习中智能体个数为 6，分别控制 3 个 PV 和 3 个 SVC 的无功功率，控制时间尺度为 15 min。通过附录 A 图 A4 对智能体进行训练，两个智能体得到的奖励值相同且随着训练不断进行逐渐收敛，智能体的策略趋于稳定，如图 5 所示。

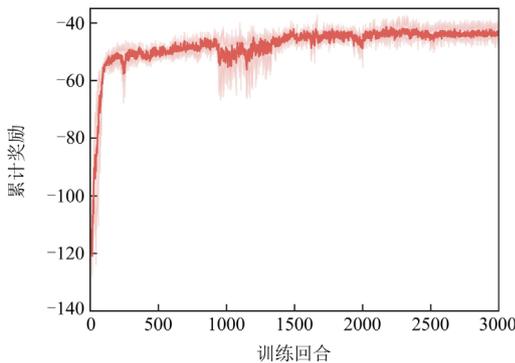
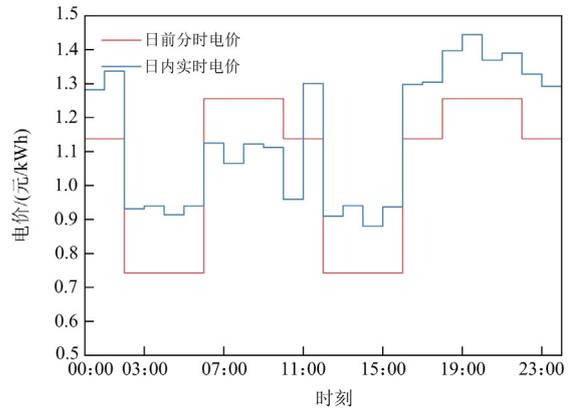


图 5 训练过程中的累计奖励

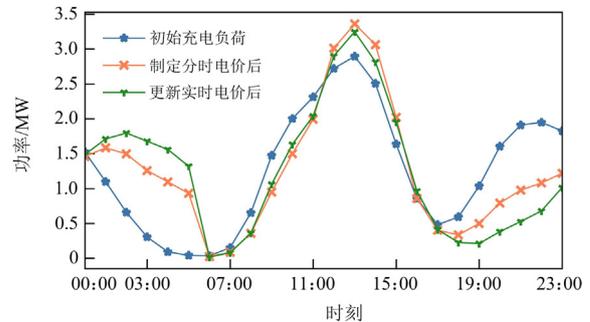
Fig. 5 Cumulative reward during training

通过训练好的策略选取测试集中一天数据进行实时电压控制，得到调整后各时刻的实时电价与充

电功率，如图 6(a)和图 6(b)所示。由附录 A 图 A1 可知，由于夜间 PV 无出力且负荷低谷位于 02:00—06:00 时段，为减小电压偏差，提高 00:00—02:00 时段的电价，实时价格激励为负，引导用户在负荷低谷时多充电，而提高 02:00—06:00 时段的电价是为了引导用户在 07:00—11:00 时段 PV 有功功率输出时进行充电行为。同理，白天 PV 出力较高，所以降低实时电价吸引 EV 充电，实现可再生能源的就地消纳。在 19:00 之后，PV 无出力且此时用电需求较大，出现负荷高峰，所以提高该时段的实时电价，EV 改变充电行为的概率进一步增加，充电功率降低。



(a) 实时电价



(b) 充电功率

图 6 电价对比和充电功率对比

Fig. 6 Electricity price and charging power comparison

经过电压控制策略优化后的配电网各节点电压分布如图 7 所示，在此策略下，系统日内运行电压均在安全运行范围(0.93~1.07)内，能够实现较好的电压控制。将测试集中另一天数据代入计算，得到电压运行结果，如附录 A 图 A5 所示，节点电压也均在安全运行范围内。

为验证实时电压控制策略的有效性，选取以下控制策略进行对比分析。

- 1) 不考虑日前和日内价格信号对 EV 的引导作

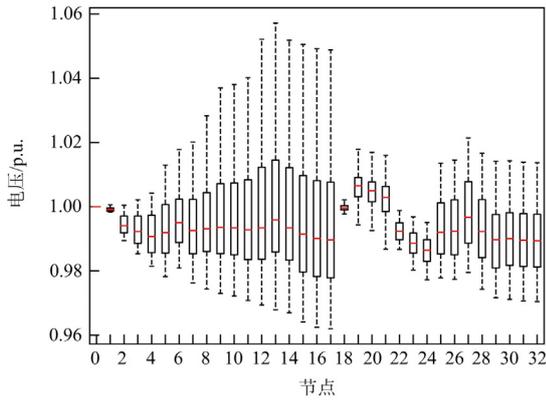


图 7 电压控制策略优化后的节点电压分布

Fig. 7 Node voltage distribution of the optimized voltage control strategy

用, 仅通过 MAPPO 算法对 PV 和 SVC 进行控制;

2) 考虑日前制定分时电价, 不考虑日内实时调整电价对充电功率进行调节, 通过 MAPPO 算法对 PV 和 SVC 进行控制;

3) 考虑日前制定分时电价, 不考虑 PV、SVC 和 EV 间不同时间响应特性, 日内进行实时调节的时间尺度均为 1 h, 通过 MAPPO 算法进行控制;

4) 考虑日前制定分时电价, 日内采用本文所提分层框架下的电压控制策略;

5) 为验证本文所提方法比传统基于模型的方法更具优越性, 选取粒子群(particle swarm optimization, PSO)算法对日内多时间尺度优化问题进行求解。

图 8 为 09:00 时不同控制策略下的系统各节点电压分布。因为策略 1 不制定分时电价, 所有 EV 用户均在抵达充电站后进行充电, 且 09:00 时光伏出力较低, 所以导致电压运行水平较低。与策略 1 相比, 策略 2 在日前制定分时电价后能够降低该时刻充电功率, 但由于日前优化目标为降低系统负荷峰谷差, 所以导致该时段制定的分时电价较高, EV 用户响应程度较高, 仍存在较大电压偏差。基于日前优化结果, 策略 3、4 和 5 在日内对 EV 充电功率进行实时调节, 并控制 PV 和 SVC 无功功率, 与策略 1 和 2 相比能够有效降低电压偏差, 但由于策略 3 中不考虑多时间尺度相互影响, 策略 5 因为 PSO 算法自身寻优机理和变量数量较大导致电压控制效果均不如本文所提策略。

如表 3 和附录 A 图 A6 所示, 对比策略 1, 本文所提策略将电压上限由 1.077 p.u.降低至 1.0589 p.u., 将电压下限由 0.8878 p.u.提高至 0.9649 p.u., 电压运行质量大幅提高, 与策略 1、2、3 和 5 相比, 本文所提策略平均电压偏差分别减小 38.16%、16.07%、

18.97%和 2.08%, 有效改善了系统电压运行水平。

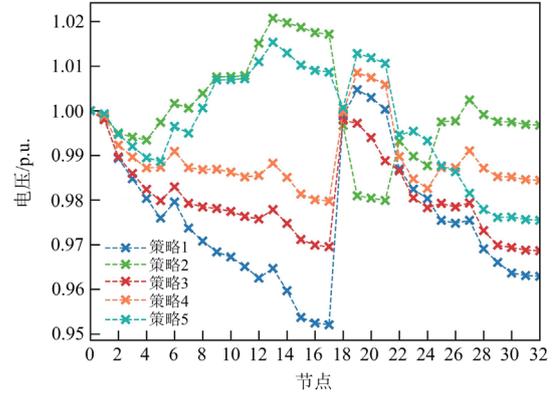


图 8 09:00 时不同策略下的系统各节点电压分布

Fig. 8 Node voltage distribution of different strategies at 9:00

表 3 不同策略下的结果对比

Table 3 Results comparison of different strategies

策略	最大电压/p.u.	最小电压/p.u.	平均电压偏差/p.u.
1	1.0770	0.8878	0.0228
2	1.0895	0.9706	0.0168
3	1.0731	0.9315	0.0174
4	1.0589	0.9649	0.0141
5	1.0796	0.9096	0.0144

4 结论

本文面向高渗透率 PV 和大量 EV 接入的配电网, 充分利用源网荷可调节资源, 提出了配电网分层深度强化学习的电压控制策略, 分别进行了日前分时电价优化和基于分层 DRL 的实时电压控制, 并通过 PPO 算法和 MAPPO 算法得到了电压控制策略。本文所提策略优势如下:

1) 建立了 EV 对价格信号的响应概率模型和基于分时电价的优化模型, 通过日前制定分时电价引导 EV 用户改变充电行为, 实现充电负荷转移, 并通过协同优化 SC 有效降低系统负荷峰谷差, 减少了配电网有功功率损耗;

2) 本文从不同时间尺度进行协同优化控制, 上层从长时间尺度调整实时电价, 更新实时价格激励实现对充电功率的调节, 下层从短时间尺度对 PV 和 SVC 的无功输出进行控制, 分别通过 PPO 和 MAPPO 算法进行决策, 能够将电压控制在安全运行范围内, 有效减小了节点电压偏差;

3) 在实时电压控制策略中下层 MAPPO 算法采用 CTDE 框架, 离线集中训练过程考虑设备间信息共享, 而在线实时控制中设备仅需就地信息和自身控制策略能够实现独立决策完成全局电压控制, 设备调节速度快且不需要完善的通信设备。

附录 A

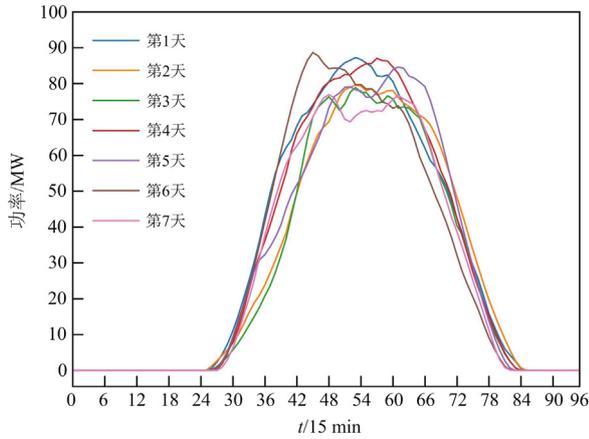


图 A1 光伏数据
Fig. A1 Data of PV

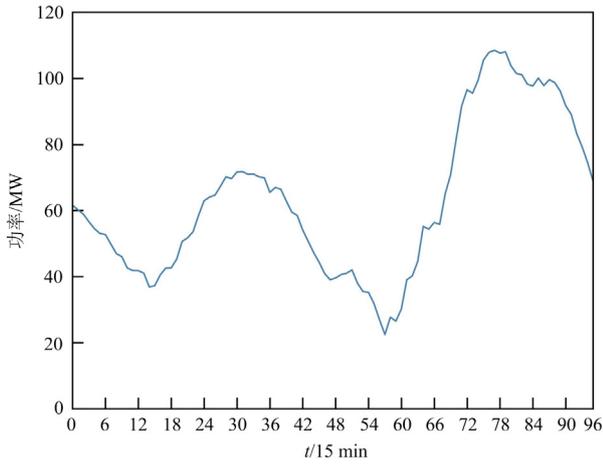


图 A2 负荷数据
Fig. A2 Data of load

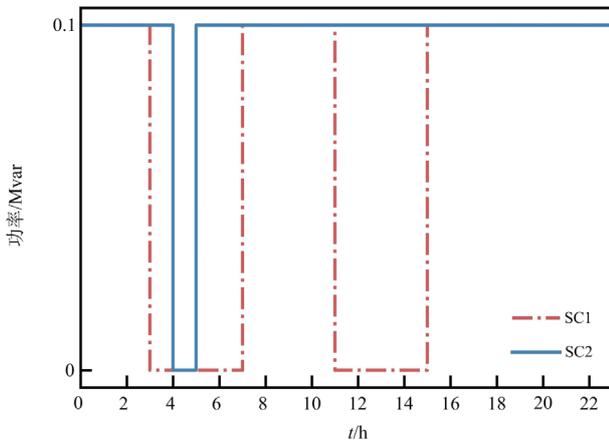


图 A3 SCs 日前调度优化结果
Fig. A3 Optimization results of SCs in day-ahead dispatch

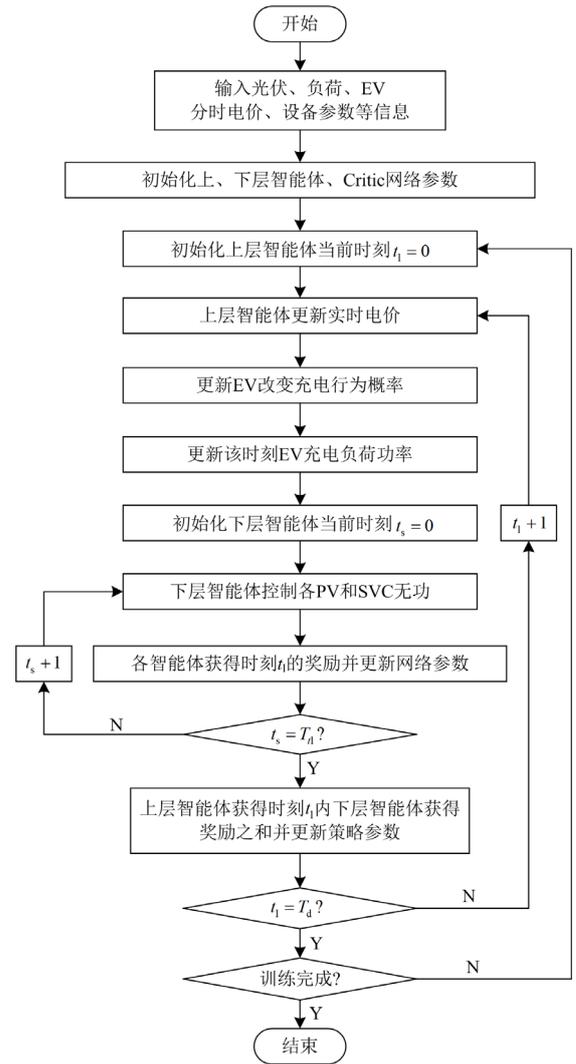


图 A4 分层 DRL 电压控制策略的训练过程
Fig. A4 Training process of hierarchical DRL voltage control strategy

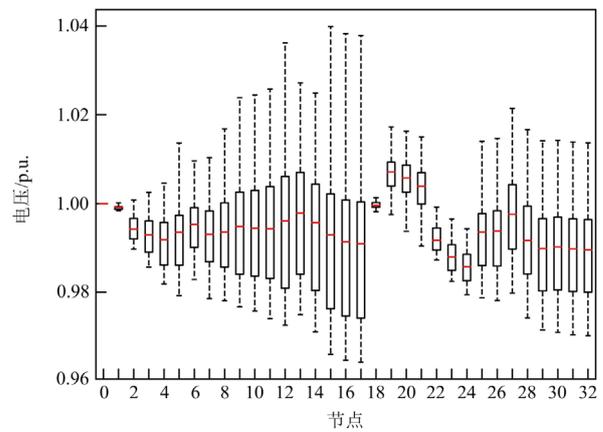


图 A5 节点电压分布
Fig. A5 Node voltage distribution

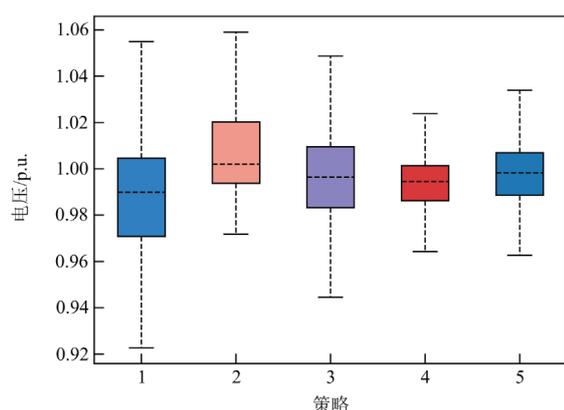


图 A6 不同策略的节点电压

Fig. A6 Node voltage of different strategies

参考文献

- [1] 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.
- [2] 胡志勇, 郭雪丽, 王爽, 等. 考虑响应意愿的电动汽车群-空调集群需求响应策略研究[J]. 电力系统保护与控制, 2023, 51(15): 109-119.
HU Zhiyong, GUO Xueli, WANG Shuang, et al. Demand response strategy for electric vehicles and air conditioners considering response willingness[J]. Power System Protection and Control, 2023, 51(15): 109-119.
- [3] 倪萌, 王蓓蓓, 朱红, 等. 能源互联背景下面向高弹性的多元融合配电网双层分布式优化调度方法研究[J]. 电工技术学报, 2022, 37(1): 208-219.
NI Meng, WANG Beibei, ZHU Hong, et al. Study of two-layer distributed optimal scheduling strategy for highly elastic multi-resource fusion distribution network in energy interconnection environment[J]. Transactions of China Electrotechnical Society, 2022, 37(1): 208-219.
- [4] 张延宇, 饶新朋, 周书奎, 等. 基于深度强化学习的电动汽车充电调度算法研究进展[J]. 电力系统保护与控制, 2022, 50(16): 179-187.
ZHANG Yanyu, RAO Xinpeng, ZHOU Shukui, et al. Research progress of electric vehicle charging scheduling algorithms based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(16): 179-187.
- [5] 曹敬, 金玉龙, 郑涛, 等. 计及分布式电源集群不确定性的配电网分散鲁棒电压控制[J]. 电力系统保护与控制, 2023, 51(22): 155-166.
CAO Jing, JIN Yulong, ZHENG Tao, et al. A decentralized robust voltage control method for distribution networks considering the uncertainty of distributed generation clusters[J]. Power System Protection and Control, 2023, 51(22): 155-166.
- [6] 李可雨, 王峰, 贾红云, 等. 基于光伏逆变器的快速功率控制系统研究及应用[J]. 电力工程技术, 2023, 42(4): 241-247.
LI Keyu, WANG Feng, JIA Hongyun, et al. Research and application of rapid power control system based on photovoltaic inverter[J]. Electric Power Engineering Technology, 2023, 42(4): 241-247.
- [7] 李培师, 吴在军, 张错, 等. 主动配电网分布式混合时间尺度无功/电压控制[J]. 电力系统自动化, 2021, 45(16): 160-168.
LI Peishuai, WU Zaijun, ZHANG Cuo, et al. Distributed hybrid-timescale voltage/var control in active distribution networks[J]. Automation of Electric Power Systems, 2021, 45(16): 160-168.
- [8] XU Xiaoyuan, LI Yunhong, YAN Zheng, et al. Hierarchical central-local inverter-based voltage control in distribution networks considering stochastic PV power admissible range[J]. IEEE Transactions on Smart Grid, 2023, 14(3): 1868-1879.
- [9] 乐健, 王曹, 李星锐, 等. 中压配电网多目标分布式优化控制策略[J]. 电工技术学报, 2019, 34(23): 4972-4981.
LE Jian, WANG Cao, LI Xingrui, et al. The multi-object distributed optimization control strategy of medium voltage distribution networks[J]. Transactions of China Electrotechnical Society, 2019, 34(23): 4972-4981.
- [10] ROBBINS B A, HADJICOSTIS C N, DOMINGUEZ-GARCIA A D. A two-stage distributed architecture for voltage control in power distribution systems[J]. IEEE Transactions on Power Systems, 2013, 28(2): 1470-1482.
- [11] 田增垚, 刘少午, 张强, 等. 电动汽车充电站参与电网调压的建模与控制方法[J]. 南方电网技术, 2022, 16(10): 113-119.
TIAN Zengyao, LIU Shaowu, ZHANG Qiang, et al. Modeling and control methods of charging stations of electric vehicles for voltage regulations[J]. Southern Power System Technology, 2022, 16(10): 113-119.
- [12] 苏粟, 胡勇, 王玮, 等. 基于电动汽车无功补偿的配电网电压调控策略[J]. 电力系统自动化, 2017, 41(10): 72-81.
SU Su, HU Yong, WANG Wei, et al. Voltage regulation strategy for distribution network based on reactive power compensation of electric vehicles[J]. Automation of Electric Power Systems, 2017, 41(10): 72-81.
- [13] 胡丹尔, 李子晨, 彭勇刚, 等. 含电动汽车充电桩的配电网深度强化学习有功-无功协调电压控制策略[J]. 电网技术, 2023, 47(12): 4985-4996.
HU Daner, LI Zichen, PENG Yonggang, et al. Deep reinforcement learning active-reactive coordinated voltage control strategy for distribution network with electric vehicle

- charging[J]. Power System Technology, 2023, 47(12): 4985-4996.
- [14] 业睿. 面向配网电压控制的电动汽车智能体充电调度方法研究[D]. 南京: 东南大学, 2022.
YE Rui. Research on charging scheduling approach for voltage control in distribution networks by electric vehicle agents[D]. Nanjing: Southeast University, 2020.
- [15] WU Weitiao, LIN Yue, LIU Ronghui, et al. Online EV charge scheduling based on time-of-use pricing and peak load minimization: properties and efficient algorithms[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(1): 572-586.
- [16] 麻秀范, 王超, 洪潇, 等. 基于实时电价的电动汽车充放电优化策略和经济调度模型[J]. 电工技术学报, 2016, 31(增刊 1): 190-202.
MA Xiufan, WANG Chao, HONG Xiao, et al. Optimal scheduling of charging and discharging of electric vehicle based on real time price and economic dispatch model[J]. Transactions of China Electrotechnical Society, 2016, 31(S1): 190-202.
- [17] 侯慧, 王逸凡, 赵波, 等. 价格与激励需求响应下电动汽车负荷聚集商调度策略[J]. 电网技术, 2022, 46(4): 1259-1269.
HOU Hui, WANG Yifan, ZHAO Bo, et al. Electric vehicle aggregator dispatching strategy under price and incentive demand response[J]. Power System Technology, 2022, 46(4): 1259-1269.
- [18] 徐晓春, 李佑伟, 戴欣, 等. 基于深度强化学习的馈线-台区两阶段电压优化[J]. 电网与清洁能源, 2023, 39(3): 63-73.
XU Xiaochun, LI Youwei, DAI Xin, et al. Two stage voltage optimization of feeder-station area based on deep reinforcement learning[J]. Power System and Clean Energy, 2023, 39(3): 63-73.
- [19] 潘俊迪, 陈中, 倪纯奕, 等. 基于无模型自适应控制的配电网电压控制方案[J]. 电力工程技术, 2023, 42(5): 128-137.
PAN Jundi, CHEN Zhong, NI Chunyi, et al. Distribution network voltage control scheme based on model-free adaptive control[J]. Electric Power Engineering Technology, 2023, 42(5): 128-137.
- [20] 齐韵英, 许潇, 殷科, 等. 基于深度强化学习的含储能有源配电网电压联合调控技术[J]. 电力建设, 2023, 44(11): 64-74.
QI Yunying, XU Xiao, YIN Ke, et al. Voltage co-regulation technology of active distribution network with energy storage based on deep reinforcement learning[J]. Electric Power Construction, 2023, 44(11): 64-74.
- [21] YANG Qiuling, WANG Gang, SADEGHI A. Two-timescale voltage control in distribution grids using deep reinforcement learning[J]. IEEE Transactions on Smart Grid, 2020, 11(3): 2313-2323.
- [22] 刁伟, 李鹏, 李鹏, 等. 基于深度强化学习的分布式电源就地自适应电压控制方法[J]. 电力系统自动化, 2022, 46(22): 25-31.
XI Wei, LI Peng, LI Peng, et al. Adaptive local voltage control method for distributed generator based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2022, 46(22): 25-31.
- [23] CAO Di, ZHAO Junbo, HU Weihao, et al. Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems[J]. IEEE Transactions on Smart Grid, 13(1): 149-165.
- [24] 程杉, 赵子凯, 陈诺, 等. 计及耦合因素的电动汽车充电负荷时空分布预测[J]. 电力工程技术, 2022, 41(3): 194-201, 208.
CHENG Shan, ZHAO Zikai, CHEN Nuo, et al. Prediction of temporal and spatial distribution of electric vehicle charging load considering coupling factors[J]. Electric Power Engineering Technology, 2022, 41(3): 194-201, 208.
- [25] 史一炜, 冯冬涵, ZHOU E, 等. 基于主从博弈的充电服务商充电引导方法及其定价策略[J]. 电工技术学报, 2019, 34(增刊 2): 742-751.
SHI Yiwei, FENG Donghan, ZHOU E, et al. Stackelberg game based on supervised charging method and pricing strategy of charging service providers[J]. Transactions of China Electrotechnical Society, 2019, 34(S2): 742-751.
- [26] 曾家威. 含电动汽车与光伏的配电网演化动态博弈调度策略研究[D]. 大庆: 东北石油大学, 2023.
ZENG Jiawei. Study on evolutionary dynamic game scheduling strategy for distribution grid with electric vehicles and photovoltaics[D]. Daqing: Northeast Petroleum University, 2018.
- [27] ZHOU E. Solar-PV generation data & load data[EB/OL]. <https://www.elia.be/en/grid-data/power-generation>.
- [28] 胡丹尔, 彭勇刚, 韦巍, 等. 多时间尺度的配电网深度强化学习无功优化策略[J]. 中国电机工程学报, 2022, 42(14): 5034-5045.
HU Daner, PENG Yonggang, WEI Wei, et al. Multi-timescale deep reinforcement learning for reactive power optimization of distribution network[J]. Proceedings of the CSEE, 2022, 42(14): 5034-5045.

收稿日期: 2024-01-26; 修回日期: 2024-05-21

作者简介:

祁向龙(1999—), 男, 硕士研究生, 研究方向为配电网电压控制等; E-mail: qixianglongqx1@163.com

陈健(1986—), 男, 博士, 副教授, 研究方向为配电网与储能的优化规划与运行技术等。E-mail: ejchen@sdu.edu.cn

(编辑 姜新丽)