

DOI: 10.19783/j.cnki.pspc.240117

考虑分区与模仿学习的深度强化学习配电网电压优化策略

李士丹¹, 李航¹, 李国杰¹, 韩蓓¹, 徐晋¹, 李玲², 王宏韬³

(1. 电力传输与功率变换控制教育部重点实验室(上海交通大学), 上海 200240; 2. 上海沛可科技有限公司, 上海 200240; 3. 国网浙江省电力有限公司嘉兴供电公司, 浙江 嘉兴 314000)

摘要: 现有深度强化学习(deep reinforcement learning, DRL)方法在解决配电网电压优化问题时, 存在信用分配难、探索效率低等问题, 在模型训练速度和优化效果等方面表现欠佳。为此, 结合配电网分区降损与模仿学习的思想, 提出一种基于指导信号的多智能体深度确定性策略梯度(guidance signal based multi-agent deep deterministic policy gradient, GS-MADDPG)的电压优化方法。首先, 将电动汽车(electric vehicles, EV)集群、分布式电源(distributed generations, DG)和无功调节装置作为决策智能体, 构建强化学习优化模型。然后, 通过配电网分区, 解耦多智能体的外部奖励, 并结合模仿学习, 利用指导信号引入内部奖励, 帮助智能体快速寻优。最后, 基于改进 IEEE 33 节点系统进行算例测试。结果表明, 所提电压优化策略较传统 DRL 方法具有更高的样本利用率, 实现了更稳定的收敛及更高的模型训练效率, 提升了配电网电压的优化效果。

关键词: 配电网电压优化; 深度强化学习; 分区降损; 模仿学习; 指导信号

Voltage optimization strategy for a distribution network based on deep reinforcement learning considering regionalization and imitation learning

LI Shidan¹, LI Hang¹, LI Guojie¹, HAN Bei¹, XU Jin¹, LI Ling², WANG Hongtao³

(1. Key Laboratory of Control of Power Transmission and Conversion, Ministry of Education (Shanghai Jiao Tong University), Shanghai 200240, China; 2. Shanghai PeiKe Technology Co., Ltd., Shanghai 200240, China; 3. Jiaying Power Supply Company, State Grid Zhejiang Electric Power Co., Ltd., Jiaying 314000, China)

Abstract: The current deep reinforcement learning (DRL) method has some issues with voltage optimization, such as challenging credit allocation and low exploration efficiency. These all lead to poor performance in model training speed and optimization effect. Considering regionalization and imitation learning, a voltage optimization strategy based on the guidance signal-based multi-agent deep deterministic policy gradient (GS-MADDPG) is proposed. First, electric vehicle (EV) clusters, distributed generation (DG) and reactive power regulators are taken as decision agents to build the reinforcement learning optimization model. Secondly, the external reward is decoupled through regionalization of the distribution network, and combined with imitation learning, an internal reward is introduced through the guidance signal to help agents search for optimization quickly. Finally, an example test is conducted on the improved IEEE 33-node distribution network. The results indicate that the proposed voltage optimization strategy has higher sample utilization, more stable convergence, and higher model training efficiency than the traditional DRL method, and improves the voltage optimization effect.

This work is supported by the National Key Research and Development Program of China (No. 2022YFE0105200).

Key words: voltage optimization of distribution network; deep reinforcement learning; zoned loss reduction; imitation learning; guidance signal

0 引言

深化电力体制改革, 构建以新能源为主体的新型电力系统, 是我国推动碳达峰、碳中和目标实施

作出的重大决策部署。以新能源发电为代表的分布式电源(distributed generation, DG)渗透率的提升以及电动汽车(electric vehicles, EV)持有率的增加, 都增大了配电网中的源荷不确定性, 这将导致局部过载、电能质量下降、短时电压降超标等问题, 对配电网容量、自动化水平、可靠性和安全性均提出了更高的要求^[1-4]。因此, 探索 DG 和 EV 在配电网中

基金项目: 国家重点研发计划项目资助(2022YFE0105200); 国网浙江省电力有限公司科技项目资助(5211JX230004)

的电压优化控制策略, 优化配电网中的电压越限和网络损耗, 具有重要的现实意义。

一方面, DG 并网逆变器在运行过程中具有无功补偿的能力, 可作为理想的电压调节设备参与配电网的无功调节^[5-6]。另一方面, 随着车网交互技术(vehicle-to-grid, V2G)的发展, EV 参与配电网电压调节的能力得到了许多国内外学者的关注。现有研究表明, 在住宅区与工业区, 用户 EV 长时间处于停放状态, 具有一定的充放电功率调控空间。经聚合后的 EV 集群可作为一种分布式储能单元参与电网调控, 通过引导其有序充放电, 在负荷高峰时减少充电功率、在低负荷时增加充电功率, 可以实现削峰填谷、促进新能源消纳、支撑配电网电压实时调控的目的^[7-8]。

配电网电压优化问题通常被描述为基于最优潮流的拥有众多目标、变量及约束的混合整数非线性规划(mixed-integer nonlinear programming, MINLP)问题^[9]。目前, 常用的求解方法包括二阶锥规划(second-order cone programming, SOCP)^[10-11]等运筹学优化算法, 粒子群优化^[12-13]等启发式搜索算法, 处理不确定性的模糊优化^[14]、鲁棒优化^[15]等算法。然而, 传统基于模型的电压优化方法, 往往高度依赖于对未来的源荷曲线、EV 充电需求等的预测, 其求解精度与计算速度难以满足当下配电网系统对于实时优化控制的要求。

近年来, 由数据驱动的深度强化学习(deep reinforcement learning, DRL)算法在优化决策领域的优势逐渐显著^[16]。例如, 文献[17]采用深度 Q 网络(deep Q-network, DQN)算法, 制定了微网内 EV 充电桩的实时调度策略, 有效减少了运行费用和电网功率波动。文献[18]基于 DQN 算法, 构建 Π 型阻抗匹配网络的多参数最优求解模型, 减少了匹配运算的时间。文献[19]采用卷积神经网络挖掘电网运行特征, 并利用 DQN 算法构建映射模型, 确定机组调整方案, 从而消除线路过载。文献[20]利用 DQN 算法学习负荷与光伏出力信息, 对配电网进行动态重构与联合优化。文献[21]结合双重深度 Q 学习算法, 以最小化并网微网购电费用为目标, 制定了储能设备的控制策略。文献[22]将柔性动作-评价(soft actor critic, SAC)算法与图神经网络结合, 对综合能源系统进行优化调度, 以最小化运行成本。文献[23]采用深度确定性策略梯度(deep deterministic policy gradient, DDPG)算法实时控制海上风力涡轮机的转子转速, 实现了风机与光伏发电的互补运行。

在主动配电网电压优化控制领域, 文献[24]针对配电网电压管控量测数据不足的情况, 提出了一种基于强化学习的分散式就地电压管控方法, 并采

用多智能体深度确定性策略梯度(multi-agent DDPG, MADDPG)算法进行训练。文献[25]通过同时迭代训练双时间尺度的优化模型, 制定调节设备的出力计划。文献[26]提出了双层多智能体强化学习算法, 分别训练微网中连续和离散动作的智能体组, 并利用迁移学习, 减少了训练的时间。文献[27]在 MADDPG 算法中引入经验增强和注意力机制, 提升了模型的训练效率。然而, 上述研究缺乏对 EV 集群在配电网优化中可调度潜力的研究, 且采用的算法普遍存在信用分配难、探索效率低等问题, 在模型训练速度和优化效果等方面表现欠佳。

针对上述问题, 本文结合配电网分区降损与模仿学习的思想, 提出了一种基于指导信号的 MADDPG(guidance signals based MADDPG, GS-MADDPG)算法, 以制定配电网电压优化控制策略。首先, 考虑 EV 的实时可调度潜力, 基于闵可夫斯基求和, 将 EV 集群聚合为广义储能设备; 其次, 将 EV 集群、DG 并网逆变器以及无功调节装置作为决策智能体, 构建部分可观的马尔可夫决策过程(partially observable Markov decision process, POMDP); 然后, 基于 MADDPG 算法, 通过配电网分区, 解耦多智能体的外部奖励, 解决信用分配难问题, 并结合模仿学习, 利用指导信号引入内部奖励, 解决探索效率低问题; 最后, 在改进的 IEEE 33 节点系统进行了算例仿真和分析。结果表明, 本文提出的 GS-MADDPG 算法可以提高 DRL 算法在配电网优化问题中的样本利用效率, 实现了更高的模型训练效率及收敛稳定性, 提升了配电网电压优化效果。

1 基于闵可夫斯基求和的 EV 集群建模

1.1 单辆 EV 的充放电边界模型

EV 的接入、离开时刻, 接入时荷电状态(state of charge, SOC)以及用户需求 SOC 由用户向充电站申报, 考虑 EV 在每个决策时段内的充放电功率恒定, 可建立 EV 充放电边界模型, 如式(1)和式(2)所示。

$$\begin{cases} 0 \leq P_{ch,n}^{EV}(t) \leq P_{ch,n,max}^{EV}(t)x_{ch,n}^{EV}(t) \\ 0 \leq P_{dis,n}^{EV}(t) \leq P_{dis,n,max}^{EV}(t)x_{dis,n}^{EV}(t) \\ 0 \leq x_{ch,n}^{EV}(t) + x_{dis,n}^{EV}(t) \leq x_n^{EV}(t) \end{cases} \quad (1)$$

$$\begin{cases} s_n^{EV}(t) = s_n^{EV}(t-1) + (\eta P_{ch,n}^{EV}(t) - \frac{1}{\eta} P_{dis,n}^{EV}(t))\Delta t \\ s_{n,min}^{EV}(t) \leq s_n^{EV}(t) \leq s_{n,max}^{EV}(t), t \in [t_n^{EV,arrive}, t_n^{EV,leave}] \\ s_n^{EV}(t_n^{EV,arrive}) = s_{n,ini}^{EV}, s_n^{EV}(t_n^{EV,leave}) = s_{n,need}^{EV} \end{cases} \quad (2)$$

式中: $P_{ch,n}^{EV}(t)$ 和 $P_{dis,n}^{EV}(t)$ 分别表示 t 时刻第 n 辆 EV 的充、放电功率; $P_{ch,n,max}^{EV}(t)$ 和 $P_{dis,n,max}^{EV}(t)$ 分别为 t 时

刻 EV 的最大充、放电功率; $x_{ch,n}^{EV}(t)$ 和 $x_{dis,n}^{EV}(t)$ 分别表示 t 时刻 EV 的充、放电状态, 处于充/放电状态时有 $x_{ch,n}^{EV}(t)=1, x_{dis,n}^{EV}(t)=0 / x_{ch,n}^{EV}(t)=0, x_{dis,n}^{EV}(t)=1$;

$x_n^{EV}(t)$ 为表征 t 时刻 EV 接入状态的布尔变量, 取 1 时有 EV 处于接入状态, 取 0 时则有 EV 处于非接入状态; $s_n^{EV}(t)$ 表示 t 时刻 EV 的 SOC; $s_{n,min}^{EV}(t)$ 和 $s_{n,max}^{EV}(t)$ 分别为 EV 所允许的最小、最大电量; $t_n^{EV,arrive}$ 和 $t_n^{EV,leave}$ 分别为 EV 的接入、离开时刻; $s_{n,ini}^{EV}$ 和 $s_{n,need}^{EV}$ 分别为 EV 用户初始 SOC 和需求 SOC; η 为充放电效率; Δt 为单位时间间隔。

本文采用文献[28]提出的边界模型作为 EV 双向充放电的约束条件, 构成 SOC 的能量上界和能量下界。

考虑 EV 参与配电网优化控制通常需要由聚合商进行统一管理^[29], 因此将充电站(charging station, CS)视为一个调度单元, 对站内所有停放的 EV 进行聚合, 构成一个具有较大容量和充放电功率的广义储能设备(generalized energy storage, GES), 统一参与电力系统调度。

1.2 EV 集群建模

闵可夫斯基求和是一种适用于欧几里得空间的加法, 表达式如式(3)所示, 其实质是多个定义域相同的变量空间的膨胀集^[30]。

$$A \oplus B = \{a+b | a \in A, b \in B\} \quad (3)$$

式中: A 和 B 为两个变量空间; a 和 b 分别为 A 和 B 中的元素; $A \oplus B$ 表示 A 和 B 的闵可夫斯基求和。

为使单辆 EV 模型具有闵可夫斯基可加性, 需将式(1)中布尔变量 $x_n^{EV}(t)$ 的定义域延拓至相同的调度时段可行域下。由此, 可以得到如式(4)所示的 EV 集群模型。

$$\left\{ \begin{array}{l} P_{v,ch}^{CS}(t) = \sum_{n \in N_v^{EV}} P_{ch,n}^{EV}(t), P_{v,dis}^{CS}(t) = \sum_{n \in N_v^{EV}} P_{dis,n}^{EV}(t) \\ P_{v,ch,max}^{CS}(t) = \sum_{n \in N_v^{EV}} P_{ch,n,max}^{EV}(t) x_{ch,n}^{EV}(t) \\ P_{v,dis,max}^{CS}(t) = \sum_{n \in N_v^{EV}} P_{dis,n,max}^{EV}(t) x_{dis,n}^{EV}(t) \\ S_v^{CS}(t) = \sum_{n \in N_v^{EV}} s_n^{EV}(t) \\ S_{v,max}^{CS}(t) = \sum_{n \in N_v^{EV}} s_{n,max}^{EV}(t) x_n^{EV}(t) \\ S_{v,min}^{CS}(t) = \sum_{n \in N_v^{EV}} s_{n,min}^{EV}(t) x_n^{EV}(t) \\ \Delta S_v^{CS}(t) = \sum_{n \in N_v^{EV}} s_{n,ini}^{EV} x_n^{EV}(t) (x_n^{EV}(t) - x_n^{EV}(t-1)) - \\ \sum_{n \in N_v^{EV}} s_{n,need}^{EV} x_n^{EV}(t-1) (x_n^{EV}(t-1) - x_n^{EV}(t)) \end{array} \right. \quad (4)$$

式中: N_v^{EV} 表示站内所有 EV 的集合; $P_{v,ch}^{CS}(t)$ 和 $P_{v,dis}^{CS}(t)$ 分别表示 t 时刻第 v 个 CS 的充、放电功率; $P_{v,ch,max}^{CS}(t)$ 和 $P_{v,dis,max}^{CS}(t)$ 分别为 t 时刻第 v 个 CS 的最大充、放电功率; $S_v^{CS}(t)$ 表示 t 时刻第 v 个 CS 的电量; $S_{v,max}^{CS}(t)$ 和 $S_{v,min}^{CS}(t)$ 分别为第 v 个 CS 所允许的最大、最小电量; $\Delta S_v^{CS}(t)$ 为 t 时刻因站内 EV 接入状态变化引起的 CS 电量变化量。EV 集群模型的参数集合表示为 $\{S_v^{CS}(t), P_{v,ch,max}^{CS}(t), P_{v,dis,max}^{CS}(t), \Delta S_v^{CS}(t)\}$, 该集合决定了 EV 集群的可调度潜力。

2 配电网电压优化控制问题建模

2.1 马尔可夫决策过程设计

EV 集群和 DG 的接入增加了配电网的不确定性, 使得传统数学优化方法难以满足当前配电网优化控制对精确性和实时性的需求。而 DRL 算法能够有效解决上述不足, 实现数据驱动和实时优化决策。使用 DRL 算法通常需要将问题建模为马尔可夫决策过程, 又由于受到通信时延、隐私保护的限制, 以及分散就地电压管控的需求, 各智能体仅能观测到接入节点的本地信息, 因此本文将配电网电压优化问题构建为部分可观的马尔可夫决策过程。

本文考虑的调节设备包括静止无功发生器(static var generator, SVG)、DG 逆变器以及 EV 集群充电站, 其中, 假定 DG 逆变器仅利用其无功容量参与调压控制, 不考虑削减有功出力。在一天 24 h 的优化周期内, 对 SVG、DG 逆变器无功输出以及 CS 充放电功率进行实时优化, 优化时间尺度为 15 min。定义 SVG、光伏逆变器和 EV 集群充电站 CS 为强化学习框架中的智能体, 每台设备(充电站)均可视为 1 个单独的智能体, 对应的观测状态、动作集合和奖励函数定义如下。

1) 状态空间

各智能体的观测状态空间 $s_{k,t}^{node}$ 均包含对应接入节点的信息, 有 $s_{k,t}^{node} = (t, P_{k,t}, Q_{k,t}, V_{k,t-1}), k \in N_{IR}$, 即第 k 个智能体接入节点在 t 时刻的负荷有功 $P_{k,t}$ 、负荷无功 $Q_{k,t}$, 以及 $t-1$ 时刻的电压幅值 $V_{k,t-1}$, N_{IR} 为智能体个数, 它是 SVG 智能体个数 N_{SVG} 、PV 智能体个数 N_{PV} 、充电站智能体个数 N_{CS} 的总和, 有 $N_{IR} = N_{SVG} + N_{PV} + N_{CS}$ 。

另外, 为确保智能体动作的合理选取, CS 和 PV 智能体分别需要添加观测指标 $s_{v,t}^{CS}$ 、 $s_{\phi,t}^{PV}$, 如式(5)所示。

$$\begin{cases} S_{v,t}^{\text{CS}} = (S_v^{\text{CS}}(t), P_{v,\text{ch,max}}^{\text{CS}}(t), P_{v,\text{dis,max}}^{\text{CS}}(t), \Delta S_v^{\text{CS}}(t)) \\ S_{\varphi,t}^{\text{PV}} = (P_{\varphi}^{\text{PV}}(t)), \varphi \in N_{\text{PV}}, v \in N_{\text{CS}} \end{cases} \quad (5)$$

式中: $S_v^{\text{CS}}(t)$ 、 $P_{v,\text{ch,max}}^{\text{CS}}(t)$ 、 $P_{v,\text{dis,max}}^{\text{CS}}(t)$ 、 $\Delta S_v^{\text{CS}}(t)$ 均为第 v 个充电站智能体的 EV 集群模型参数; $P_{\varphi}^{\text{PV}}(t)$ 为第 φ 个 PV 智能体在 t 时刻的有功出力。

系统状态空间集合 S_t 为各智能体观测空间的集合, 如式(6)所示。

$$\begin{cases} S_t = \{s_t^w, s_t^{\varphi}, s_t^v | w \in N_{\text{SVG}}, \varphi \in N_{\text{PV}}, v \in N_{\text{CS}}\} \\ s_t^w = (s_{k,t}^{\text{node}}, s_t^{\varphi} = (s_{k,t}^{\text{node}}, s_{\varphi,t}^{\text{PV}}), s_t^v = (s_{k,t}^{\text{node}}, s_{v,t}^{\text{CS}}) \end{cases} \quad (6)$$

2) 动作空间

本文中, 系统动作空间集合 A_t 表示 t 时刻所有智能体联合动作, 它由各智能体的个体输出功率组成, 并分别受到对应时刻输出功率的上下限约束: SVG 的可调无功功率 $Q_{w,t}$ 受到最大可调容量的限制; PV 的可调无功功率 $Q_{\varphi,t}$ 受装机容量 $S_{\varphi,t}$ 和输出有功功率 $P_{\varphi,t}$ 的影响, 有 $|Q_{\varphi,t}| \leq \sqrt{S_{\varphi,t}^2 - P_{\varphi,t}^2}$; 充电站的有功功率 $P_{v,t} = P_{v,\text{ch}}^{\text{CS}}(t) - P_{v,\text{dis}}^{\text{CS}}(t)$, 它受到式(4)中最大充、放电功率的约束。综合得到动作空间集合 $A_t = \{Q_{w,t}, Q_{\varphi,t}, P_{v,t} | w \in N_{\text{SVG}}, \varphi \in N_{\text{PV}}, v \in N_{\text{CS}}\}$, 其中, 为避免梯度消失和梯度爆炸问题, 需将各智能体的动作经 Min-Max 归一化至 $[-1,1]$ 区间, 再输入给神经网络。其中, 归一化过程确保了各智能体的输出功率满足上下限约束。特别地, 为了避免陷入局部最优解, 在评论家网络输出 $\mu(s_{k,t} | \theta^{\mu})$ 的基础上, 通过式(7)添加随机噪声, 构成实际的输出动作 $a_{k,t}$, 以探索潜在的最优策略。

$$a_{k,t} = \mu(s_{k,t} | \theta^{\mu}) + \mathcal{N}(0, \sigma) \quad (7)$$

式中: σ 表示动作噪声系数, 它会随着训练的进行等比例减小; $\mu(s_{k,t} | \theta^{\mu})$ 为评论家网络根据输入状态 $s_{k,t}$ 的输出动作, θ^{μ} 为评论家网络的网络参数; $\mathcal{N}(0, \sigma)$ 为正态分布噪声。

3) 奖惩函数

一般地, 在配电网优化问题中, 各可调度单元具有相同的目标, 即避免配电网各节点电压越限的同时, 降低全网的网络损耗。因此, 在强化学习框架下, 各智能体共享同一个全局奖励 r_t , 表示为

$$r_t = -C_{\text{ls}} P_{\text{ls}}(t) - C_{\beta} R_V(t) \quad (8)$$

$$R_V(t) = \sum_{i \in N_i} [\max(0, \underline{V} - V_i(t)) + \max(0, V_i(t) - \bar{V})] \quad (9)$$

式中: C_{ls} 和 C_{β} 分别为边际网损系数和电压越限附加成本系数; $P_{\text{ls}}(t)$ 和 $R_V(t)$ 分别表示配电网总线损

和节点电压越限值; N_i 为配电网节点集合; $V_i(t)$ 为第 i 个节点的电压; \bar{V} 和 \underline{V} 分别为节点电压上下限。

然而, 多智能体共享全局奖励的设置, 无法量化每个智能体对全局奖励的贡献程度, 容易导致“惰性智能体”的出现, 使得部分智能体动作探索不充分, 最终的决策策略容易陷入局部最优解。针对这一问题, 本文基于配电网分区降损的思想, 将全局奖励按配电网分区解耦, 各智能体的奖励 $r_{t,k}^o$ 仅反映所处配电网分区的网络损耗和电压越限情况。

$$r_{t,k}^o = -C_{\text{ls}} P_{\text{ls}}(t, g(k)) - C_{\beta} R_V(t, g(k)) \quad (10)$$

式中, $g(k)$ 为由智能体编号到智能体所在分区的映射。

2.2 结合模仿学习的奖惩函数设计

传统 DRL 算法在训练初期的随机探索阶段采样效率较低, 导致模型训练时间普遍较长; 另外, 考虑配电网运行的实际场景, 完全随机的样本探索可能会造成大面积的节点电压越限, 并不契合实际的运行要求。因此, 如何用更少的样本、更快地获得行之有效的优化控制策略, 是近年来的研究热点。

实际上, 配电网现有的控制方法, 如基于预测的二阶锥规划等, 虽然精确性较低, 但相较于 DRL 算法在决策初期完全随机的动作探索, 仍具有一定的参考价值。因此, 本文结合模仿学习的思想, 将上述基于预测信息的二阶锥规划结果作为指导信号, 加速 DRL 算法的训练过程。

模仿学习是一种监督学习方法, 其能够从人类专家或者已经训练好的智能体演示中学习, 学习效率高, 收敛速度快, 但其收敛效果受限于专家样本, 缺乏探索能力。本文将模仿学习和强化学习相结合, 从模仿学习中学习初步策略, 通过缩小搜索空间来提高学习效率; 随后, 应用强化学习继续探索, 以适应高度不确定性的应用需求。

具体地, 利用指导信号引入内部奖励 $r_{t,k}^i$, 如式(11)所示。

$$r_{t,k}^i = -C_{\gamma,k} \max((a_{t,k} - \pi_{t,k})^2 - A_{\sigma,k}^2, 0) \quad (11)$$

式中: $a_{t,k}$ 和 $\pi_{t,k}$ 分别为 t 时刻第 k 个智能体的实际输出功率和指导信号给出的输出结果; $A_{\sigma,k}$ 为第 k 个智能体允许的最大动作探索值; $C_{\gamma,k}$ 为第 k 个智能体的内部奖励系数, $C_{\gamma,k}$ 越大, 则智能体更倾向于模仿指导信号的行为, 反之则更鼓励智能体探索环境, 寻找更好的策略。

外部奖励式(10)和内部奖励式(11)共同构成了各智能体的奖励函数 $R_{t,k}$, 即式(12), 实现了智能体的快速寻优。

$$R_{t,k} = r_{t,k}^o + r_{t,k}^i \quad (12)$$

3 基于 GS-MADDPG 的决策算法

3.1 MADDPG 算法概述

MADDPG 算法是一种被广泛应用的多智能体强化学习算法, 其包含演员网络和评论家网络。为了保证学习过程的稳定性, 演员网络和评论家网络中又同时包含两个网络结构相同、但参数不同的神经网络, 分别称为估值网络和目标网络。估值网络参数随训练过程不断更新, 而目标网络参数则每隔一段时间利用估值网络的参数进行软更新(soft update)。

演员网络的作用是在某一状态下决定选取何种动作, 其参数更新遵循式(13)。

$$\nabla_{\theta} J(\theta^{\mu_k}) = E_{s_t, a_t \sim D} [\nabla_{\theta^{\mu_k}} \mu_k(s_{k,t} | \theta^{\mu_k}) \cdot \nabla_{a_{k,t}} Q_k^{\mu}(s_t, a_t | \theta^{\mu_k})] \quad (13)$$

式中: D 为经验回放缓冲区, 存放所有智能体的历史经验; $Q_k^{\mu}(s_t, a_t)$ 为智能体 k 的动作价值函数; θ^{μ_k} 和 θ^{Q_k} 分别为演员网络和评论家网络中的估值网络参数; $J(\theta^{\mu_k})$ 为智能体 k 的状态价值函数。值得注意的是, $Q_k^{\mu}(s_t, a_t)$ 在计算时不仅考虑了智能体 k 的动作, 也将其他智能体的动作加以考虑, 因此可以在任意智能体策略发生改变时, 保证整体环境的稳定。

评论家网络的作用是评估策略好坏, 其参数更新目标为最小化损失函数, 损失函数表示为

$$L(\theta^{Q_k}) = E_{s_t, a_t, r_t, s_{t+1}} [(Q_k^{\mu}(s_t, a_t | \theta^{Q_k}) - y_k)^2] \quad (14)$$

其中

$$y_k = R_{t,k} + \gamma Q_k^{\mu}(s_{t+1}, a'_{t+1} | \theta^{Q_k}) \quad (15)$$

式中: y_k 为时序差分目标; γ 为奖励折扣因子; θ^{Q_k} 为评论家网络中的目标网络参数; $Q_k^{\mu}(s_{t+1}, a'_{t+1} | \theta^{Q_k})$ 为第 k 个目标评论家网络对应的动作价值函数。

演员网络和评论家网络中目标网络参数 θ^{μ_k} 、 θ^{Q_k} 采取的软更新策略可表示为

$$\theta^{\mu_k} \leftarrow (1 - \tau)\theta^{\mu_k} + \tau\theta^{\mu_k} \quad (16)$$

$$\theta^{Q_k} \leftarrow (1 - \tau)\theta^{Q_k} + \tau\theta^{Q_k} \quad (17)$$

式中, τ 为软更新系数, 且 $\tau \ll 1$ 。

3.2 GS-MADDPG 算法流程

在 MADDPG 算法的基础上, 本文引入配电网分区降损和指导信号的思想, 提出 GS-MADDPG 算法。GS-MADDPG 算法采用中心化训练、去中心化执行(centralized training with decentralized execution,

CTDE)的架构, 即在配电网中央控制器进行集中训练, 随后将训练好的演员网络参数下放至各调节设备的边缘计算装置中, 用于分布式执行。表 1 给出了 GS-MADDPG 算法在训练阶段的整体求解流程。在实际配电网中, 可以利用传感器采集配电网历史运行数据, 构成各调节设备和配电网环境交互的样本集, 也即经验回放缓冲区 D , 从而进行离线训练。

表 1 GS-MADDPG 算法求解流程

Table 1 Solution process of GS-MADDPG algorithm

GS-MADDPG 算法	
1:	参数赋值及初始化, 生成指导信号, 开始集中式训练
2:	for 回合=1: 总回合数
3:	重置环境, 蒙特卡洛抽样生成相关日数据
4:	for 时间步=1: 总周期数
5:	各智能体观测当前状态, 根据式(7)选取动作
6:	环境根据智能体选取的动作, 生成外部、内在奖励和下一时间步的状态
7:	存入当前样本 (s_t, a_t, r_t, s_{t+1}) , 更新经验回放缓冲区
8:	end for
8:	if 回合数大于采样回合数 and 回合数达到更新周期
9:	从经验回放缓冲区中随机抽取批量样本进行训练
10:	更新各演员网络、评论家网络的参数
11:	end if
12:	end for
13:	训练结束, 保存并下发各智能体的演员网络参数, 用于分布式执行

4 算例分析

4.1 算例设置

为了验证本文提出的优化策略的有效性, 本文在修改后的 IEEE 33 节点算例上进行了测试。该算例系统含有 2 个 EV 充电站、3 个分布式 PV 和 2 个 SVG, 如图 1 所示, 各调节设备的安装节点与容量参数如表 2 所示, 其中, CS 对应的参数表示充电站内所有充电桩以额定功率充电时的最大输出功率。参考文献[31]将配电网划分为 4 个分区, 采用蒙特卡洛模拟法构造系统运行数据, 基准电压标幺值为 1.00 p.u., 节点电压安全范围为 [0.95, 1.05] p.u.。分布式光伏出力预测如附录 A 图 A1 所示, 设置不确定波动区间为预测值的 80%~100%; 各节点的负荷根据原始 IEEE 33 节点系统中的空间负荷分布, 加入 $\pm 10\%$ 的噪声, 模拟负荷需求的不确定性, 负荷变化率见附录 A 图 A2。

假定每个 EV 充电站日均参与调度的 EV 数量遵循 $U(75,100)$ 均匀分布, 考虑 EV 用户到达与离开时间遵循正态分布, 均值为 09:00 和 20:00。EV

蓄电池容量为 24 kWh, 充电效率为 0.95, 最大充放电功率为 6 kW。调度区间设定为 00: 00—24: 00 共 24 h, 调度周期间隔为 15 min, 在每个调度时刻, 利用 pypower 工具包计算配电网潮流, 得到实时的线路网络损耗。POMDP 参数设定如表 3 所示。

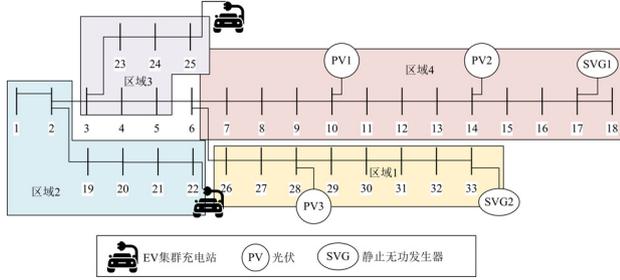


图 1 IEEE 33 配电网算例系统

Fig. 1 IEEE 33 distribution network system

表 2 调节设备参数

Table 2 Parameters of regulation devices

调节设备	容量参数	安装节点
PV	0.55 MW	10,14,28
CS	0.6 MW	22,25
SVG	-0.6~0.6 Mvar	17,33

表 3 POMDP 参数设定

Table 3 Parameters of POMDP algorithm

基本参数	取值
边际网损系数 C_{ls}	48 元/MWh
电压越限附加成本系数 C_{β}	800 元/p.u.
CS 内部奖励系数 $C_{\gamma,k}$	6×10^{-10} 元/W ²
PV、SVG 内部奖励系数 $C_{\gamma,k}$	6×10^{-11} 元/W ²
CS 最大探索边界 $A_{\sigma,k}$	40 kW
PV、SVG 最大探索边界 $A_{\sigma,k}$	0.2 MW

表 4 为 GS-MADDPG 算法的超参数设定。其中, 充电站智能体的动作涉及到配电网电压优化和 EV 用户充电两个层面, 而充电行为具备时间效益, 故设其折扣因子为 0.99; 而由于 PV 和 SVG 智能体只服务于配电网电压优化, 其动作不具备连续意义上的长期效益, 故设其折扣因子为 0。

4.2 训练结果

为了验证所提方法的有效性, 分别采用传统 MADDPG 算法、基于分区降损的 MADDPG 算法 (regionalization based MADDPG, RG-MADDPG) 以及所提 GS-MADDPG 算法的控制方法训练各智能体, 得到相应的控制策略, 其中, 前 400 个回合为随机采样, 各智能体动作完全随机, 不进行动作策略学习。每隔 10 个回合进行 1 次评估, 计算平均奖励值, 并选取 5 个随机种子进行重复训练, 观察训

表 4 GS-MADDPG 算法超参数设定

Table 4 Hyperparameters of GS-MADDPG algorithm

基本参数	取值
总训练回合次数	1200
采样回合数	400
演员网络学习率	1×10^{-4}
评论家网络学习率	1×10^{-3}
充电站智能体折扣因子	0.99
其他智能体折扣因子	0
软更新因子	1×10^{-2}
经验回放缓冲区大小	1×10^6
批处理大小	128
动作噪声系数	1
隐藏层结构	3
隐藏层神经元个数	(256,64,64)

练过程的稳定性。此外, 本文选取“无控制方法”作为参考, 即在同一典型日场景下, 配电网中充电站采用无序充电方式输出功率, PV 和 SVG 设备输出的无功功率均为零。基于上述 4 种方法的模型训练奖励曲线见图 2, 图中阴影部分表示平均奖励的标准偏差。

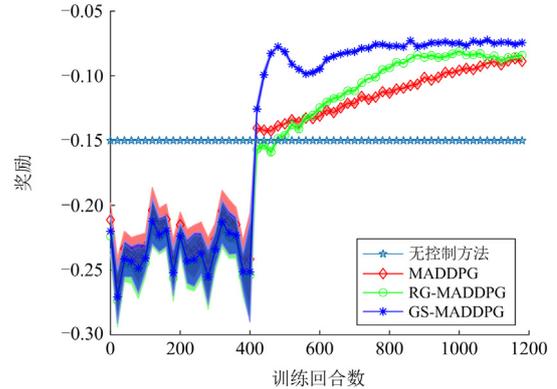


图 2 训练奖励曲线

Fig. 2 Training curves of reward

由图 2 可知, 本文所提 GS-MADDPG 算法在收敛稳定性、训练速度方面均具有最佳表现。前 400 个回合作为初始样本库, 各智能体动作选取随机, 且不具有协同性, 控制效果较差, 奖励值低于无控制时的表现。400 个回合后, 各智能体开始从经验回放缓冲区中学习历史经验, 动作策略开始趋优, 奖励曲线迅速上升, 随后收敛至稳定。其中, 无控制方法的奖励为 -0.1518, 基于传统 MADDPG 算法、基于 RG-MADDPG 算法以及基于 GS-MADDPG 算法的控制方法分别第 1160 个回合、第 990 个回合和第 820 个回合左右收敛, 最终收敛奖励值分别为 -0.086、-0.082、-0.074, 相较于无控制方法, 强化

学习算法的收敛奖励值分别提升了 43.3%、46.0%、51.3%。而相较于传统 MADDPG 算法, 分区降损思想的引入, 由于细分了各智能体对全局奖励的贡献程度, 一定程度上加快了训练的速度, 但最终收敛奖励值提升不大, 仍容易陷入局部最优解; 而指导信号的引入, 则进一步提升了动作探索的效率。

4.3 测试集验证结果

保存并提取上述训练好的策略模型参数, 并在测试集上进行有效性验证。增加基于典型日的日前集中式优化方法, 测试指标包括电压越限率、日内网络损耗两项, 其中, 电压越限率是指节点电压超过安全范围的时间占比。表 5 展示了基于 5 种方法的测试指标的平均值和标准差。

表 5 测试集结果

Table 5 Results of test dataset

控制策略	电压越限率/%		网络损耗/MW	
	平均值	标准差	平均值	标准差
无控制	16.94	3.252	0.078 89	0.0097
MADDPG	0.6281	0.6796	0.050 88	0.0056
RG-MADDPG	0.1326	0.2699	0.048 54	0.0055
GS-MADDPG	0	0	0.034 24	0.0055
日前集中式优化	0.9091	1.3245	0.039 83	0.0058

由表 5 可知, 相较于传统 MADDPG 算法, 本文所提 GS-MADDPG 算法具有更好的稳压降损效果。当调节设备不参与优化控制时, 配电网平均有 16.94% 的电压越限率, 且网络损耗较大。当采用基于 MADDPG 算法、基于 RG-MADDPG 算法、基于 GS-MADDPG 算法的优化控制方法后, 节点电压越限和网络损耗均得到了明显改善, 且基于 GS-MADDPG 算法得到的优化效果最佳, 电压越限率为 0%, 平均网络损耗为 0.034 24 MW。而基于典型日的日前集中式优化方法, 由于日内源、荷数据与日前预测具有偏差, 整体效果不如本文所提算法。上述结果验证了本文所提算法的优越性, 通过配电网分区降损和指导信号的引入, 提升了算法的性能表现。

4.4 典型日验证结果

选取空间负荷为 100% 的典型日运行场景, 对上述 5 种方法的优化效果进行分析。图 3 展示了基于所提 GS-MADDPG 方法进行优化控制的电压分布情况, 该场景下各可调单元的有功无功出力曲线见附录 A 图 A3—图 A5。

由图 3 可知, 基于 GS-MADDPG 方法进行优化控制, 配电网各节点在全部时段内的电压值均处于 $[0.95, 1.05]$ p.u. 的安全范围内。其中, 18 号节点位于线路末端, 受到的支撑较弱, 发生电压越限的可能

性较大。因此, 将 18 号节点作为代表性节点, 对比 5 种优化方法的电压幅值(标么值)曲线, 并观察配电网系统损耗情况, 分别见图 4 和图 5。其中, 在典型日场景下, 基于典型日的日前集中式优化即基于全局信息的集中式优化具有理论最优的优化结果。

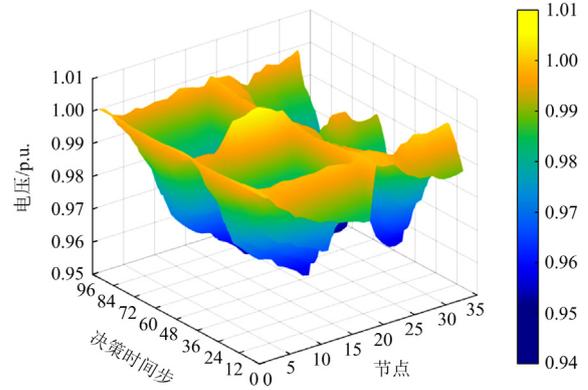


图 3 算例优化后典型日电压分布情况

Fig. 3 Voltage profile after optimization on the typical day

由图 4 可知, 相较于传统 MADDPG 算法, 本文所提 GS-MADDPG 算法实现了更好的稳压效果。当不采取任何调控手段时, 节点电压在 06:00—10:00 时段、16:00—22:00 时段低于安全下限, 且电压波动剧烈。当采用 3 种基于 MADDPG 的方法进行调控时, 节点电压整体抬升, 规避了节点电压越限的风险。

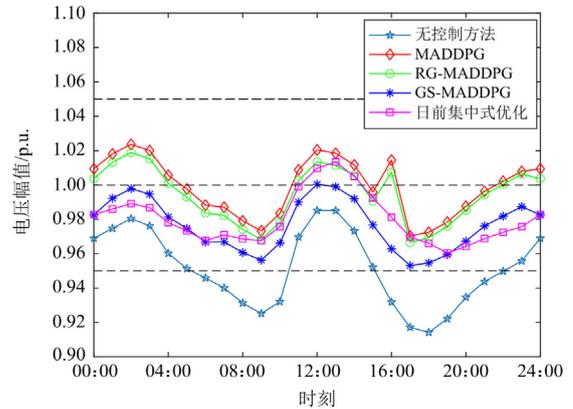


图 4 典型日 18 号节点电压曲线

Fig. 4 Voltage profile of bus 18 on the typical day

图 5 展示了 5 种优化方法对系统网损调控的结果。在无控制方法、MADDPG、RG-MADDPG、GS-MADDPG 和日前集中式优化方法控制下, 典型日平均网络损耗分别为 0.078 22、0.050 26、0.047 95、0.037 85、0.033 65 MW。本文所提 GS-MADDPG 方法的功率损耗调控效果仅次于基于准确预测信息

的日前集中式优化结果。

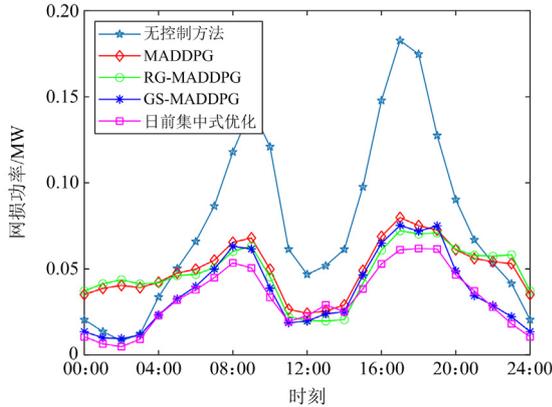


图 5 典型日功率损耗曲线
Fig. 5 Power loss profile on the typical day

4.5 极端场景验证结果

为了验证所提 GS-MADDPG 方法的泛用性和鲁棒性, 选取典型日 110%的空间负荷作为极端场景进行测试, 以模拟日内源、荷数据与日前预测具有偏差的情况。分别对比了 GS-MADDPG 算法、基于典型日的日前集中式优化以及基于极端场景全局信息的集中式优化得到的 18 号节点电压曲线与日功率损耗曲线, 分别见图 6 和图 7。

由于日内实际源、荷数据与日前预测具有偏差, 基于典型日的日前集中式优化结果并不能很好地适用于极端运行场景, 仍会有节点电压越限的情况发生; 而所提 GS-MADDPG 算法仍能起到较好的稳压效果。另外, 在 GS-MADDPG、日前集中式优化和集中式优化方法控制下, 极端场景平均网络损耗分别为 0.048 54、0.049 57、0.042 05 MW, GS-MADDPG 算法的网络损耗相比日前集中式优化, 进一步降低了 2.1%。

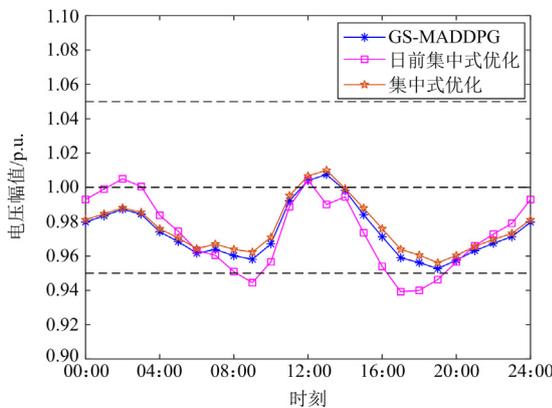


图 6 极端场景 18 号节点电压曲线
Fig. 6 Voltage profile of bus 18 on the extreme scene

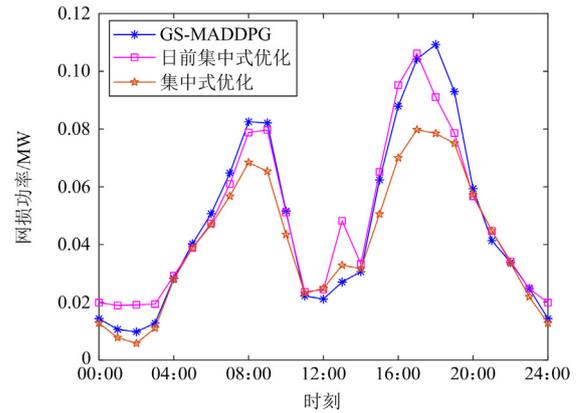


图 7 极端场景功率损耗曲线
Fig. 7 Power loss profile on the extreme scene

通过对比电压曲线和功率损耗曲线, 本文所提算法在典型日场景和极端场景下均具有近似于集中式最优控制的效果。然而, 集中式控制虽然能给出理论最优解, 但严重依赖于完备的配电网全局信息, 且求解速度较慢, 难以进行在线部署。而 GS-MADDPG 算法本质上是一种就地控制策略, 在缺乏量测数据和可靠通信条件的区域配电网中同样适用, 具有更高的应用潜力。

综上, 通过典型日和极端场景下的控制效果对比, 进一步验证了本文所提 GS-MADDPG 算法在精确性和实时性上的优越性。

5 结论

为解决传统 DRL 算法在配电网电压优化问题中存在的信用分配难、探索效率低等问题, 本文基于配电网分区降损和模仿学习的思想, 提出了 GS-MADDPG 算法。通过算例仿真分析, 验证了所提算法能够应对 EV 充电负荷与新能源出力的不确定性, 相比传统方法具有更好的控制效果与更高的应用潜力。

1) 基于闵可夫斯基求和, 将充电站视为一个调度单元, 对站内停放的 EV 进行聚合, 在满足站内所有 EV 充电需求的前提下, 避免了求解维度灾, 实现了统一调度。

2) 基于 MADDPG 算法, 通过配电网分区, 解耦多智能体的外部奖励, 解决信用分配难问题, 并结合模仿学习, 利用指导信号引入内部奖励, 解决探索效率低问题。在改进 IEEE 33 节点系统进行的算例训练结果表明, 本文所提算法具有样本利用效率高、收敛稳定性高的优势。

3) 算例结果表明, 本文所提算法可以在不依赖任何预测信息和通信条件的情况下, 快速给出 DG

逆变器、SVG 和 EV 集群充电站的实时功率决策方案, 有效抑制了节点电压的跌落, 降低了配电网的网络损耗。同时, 本文所提算法在面对复杂场景时也具备更强的鲁棒性。

此外, 本文所提算法能够帮助模型快速收敛至稳定, 在实际工程应用时能够适用于更大规模的配电网优化问题。

附录 A

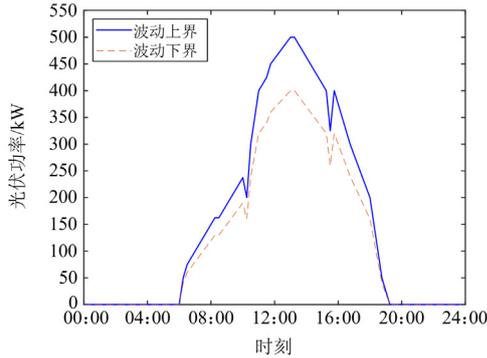


图 A1 光伏出力曲线

Fig. A1 PV output curve

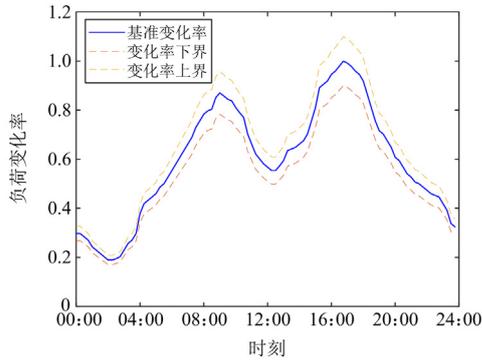


图 A2 负荷变化率曲线

Fig. A2 Load variation rate curve

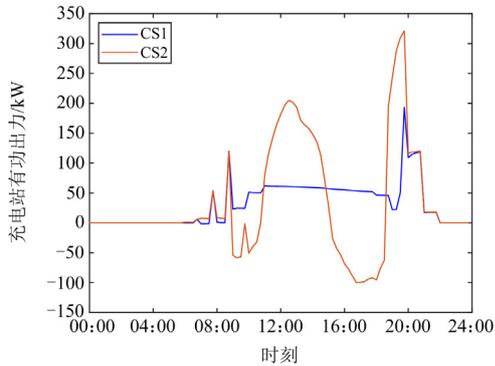


图 A3 充电站有功出力曲线

Fig. A3 Active power of charging stations

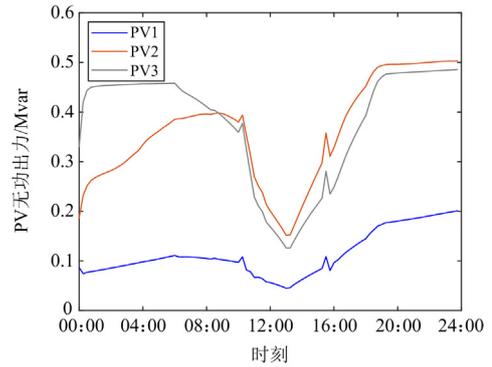


图 A4 PV 无功出力曲线

Fig. A4 Reactive power of PVs

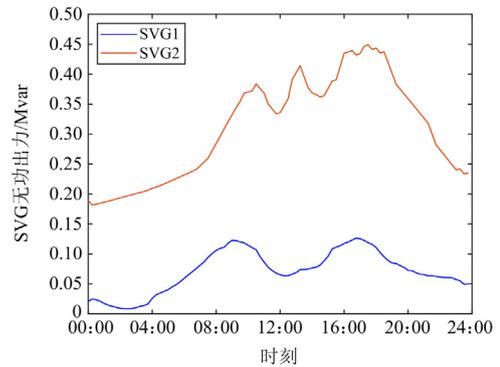


图 A5 SVG 无功出力曲线

Fig. A5 Reactive power of SVGs

参考文献

- [1] 陈国平, 董昱, 梁志峰. 能源转型中的中国特色新能源高质量发展分析与思考[J]. 中国电机工程学报, 2020, 40(17): 5493-5506.
CHEN Guoping, DONG Yu, LIANG Zhifeng. Analysis and reflection on high-quality development of new energy with Chinese characteristics in energy transition[J]. Proceedings of the CSEE, 2020, 40(17): 5493-5506.
- [2] 王智伟, 张振宇, 李征, 等. 提升高渗透率新能源电网承载能力的 DSSC 优化配置[J]. 电力系统保护与控制, 2023, 51(6): 179-187.
WANG Zhiwei, ZHANG Zhenyu, LI Zheng, et al. Optimized DSSC configuration to enhance load-carrying capacity of the new energy grid with high permeability[J]. Power System Protection and Control, 2023, 51(6): 179-187.
- [3] 张旭, 李阳, 刘晓, 等. 含高渗透率新能源的新型交直流储能系统的配电网规划[J]. 南方电网技术, 2022, 16(4): 60-67.
ZHANG Xu, LI Yang, LIU Xiao, et al. Distribution network planning of new AC/DC energy storage system with high penetration new energy[J]. Southern Power System Technology, 2022, 16(4): 60-67.

- [4] 杨金海, 武家辉, 王海云, 等. 不同渗透率下多种新能源电力系统动态安全域分析[J]. 电力建设, 2022, 43(4): 58-68.
YANG Jinhai, WU Jiahui, WANG Haiyun, et al. Dynamic security region analysis of power system under different penetration rate of new energy[J]. Electric Power Construction, 2022, 43(4): 58-68.
- [5] 熊壮壮. 两阶段鲁棒优化的主动配电网动态无功优化[D]. 郑州: 郑州大学, 2019.
XIONG Zhuangzhuang. Two-stage robust optimization of dynamic reactive power optimization in active distribution network[D]. Zhengzhou: Zhengzhou University, 2019.
- [6] 王永杰, 吴文传, 张伯明, 等. 有功无功协调的主动配电网鲁棒电压控制[J]. 电力系统自动化, 2016, 40(9): 29-34.
WANG Yongjie, WU Wenchuan, ZHANG Boming, et al. Active and reactive power coordinated robust optimization for active distribution networks[J]. Automation of Electric Power Systems, 2016, 40(9): 29-34.
- [7] LE FLOCH C, BELLETTI F, MOURA S. Optimal charging of electric vehicles for load shaping: a dual-splitting framework with explicit convergence bounds[J]. IEEE Transactions on Transportation Electrification, 2016, 2(2): 190-199.
- [8] SCHULLER A, FLATH C M, GOTTWALT S. Quantifying load flexibility of electric vehicles for renewable energy integration[J]. Applied Energy, 2015, 151: 335-344.
- [9] 郑能, 丁晓群, 管志成, 等. 基于场景法的配电网有功-无功协调优化[J]. 电网技术, 2019, 43(5): 1640-1651.
ZHENG Neng, DING Xiaoqun, GUAN Zhicheng, et al. Coordinated optimization of active power and reactive power in distribution network based on scenario method[J]. Power System Technology, 2019, 43(5): 1640-1651.
- [10] 郭清元, 吴杰康, 莫超, 等. 基于混合整数二阶锥规划的新能源配电网电压无功协同优化模型[J]. 中国电机工程学报, 2018, 38(5): 1385-1396.
GUO Qingyuan, WU Jiekang, MO Chao, et al. A model for multi-objective coordination optimization of voltage and reactive power in distribution networks based on mixed integer second-order cone programming[J]. Proceedings of the CSEE, 2018, 38(5): 1385-1396.
- [11] 岳园园, 王主丁, 王辉, 等. 考虑设备调节成本的主动配电网日前优化调度[J]. 中国电力, 2023, 56(8): 136-142.
YUE Yuanyuan, WANG Zhuding, WANG Hui, et al. Day-ahead optimal dispatch for active distribution network considering action cost of devices[J]. Electric Power, 2023, 56(8): 136-142.
- [12] 韦香香. 基于电动汽车与光伏优化配置的配网调度策略研究[D]. 北京: 北京交通大学, 2017.
WEI Xiangxiang. Research on distribution network scheduling strategy based on optimal allocation of electric vehicles and photovoltaic system[D]. Beijing: Beijing Jiaotong University, 2017.
- [13] 佟曦, 侯朗博, 孙昊, 等. 高比例光伏和电动汽车接入配电网的无功优化[J]. 智慧电力, 2023, 51(10): 31-37, 102.
TONG Xi, HOU Langbo, SUN Hao, et al. Reactive power optimization of distribution networks with high proportion of PV and EVs[J]. Smart Power, 2023, 51(10): 31-37, 102.
- [14] LI Yuanzheng, WU Qinghua, JIANG Lin, et al. Optimal power system dispatch with wind power integrated using nonlinear interval optimization and evidential reasoning approach[J]. IEEE Transactions on Power Systems, 2016, 31(3): 2246-2254.
- [15] 崔文倩, 魏军强, 赵云灏, 等. 双碳目标下含重力储能配电网多目标运行优化[J]. 电力建设, 2023, 44(4): 45-53.
CUI Wenqian, WEI Junqiang, ZHAO Yunhao, et al. Multi-objective operation optimization of distribution network with gravity energy storage under double carbon target[J]. Electric Power Construction, 2023, 44(4): 45-53.
- [16] CAO Di, HU Weihao, ZHAO Junbo, et al. Reinforcement learning and its applications in modern power and energy systems: a review[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(6): 1029-1042.
- [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] 胡正伟, 夏思懿, 王文彬, 等. 基于深度强化学习的II型阻抗匹配网络多参数最优求解方法[J]. 电力系统保护与控制, 2024, 52(6): 152-163.
HU Zhengwei, XIA Siyi, WANG Wenbin, et al. Multi-parameter optimal solution method for II-type impedance matching networks based on deep reinforcement learning[J]. Power System Protection and Control, 2024, 52(6): 152-163.
- [19] 孙立钧, 顾雪平, 刘彤, 等. 一种基于深度强化学习算法的电网有功安全校正方法[J]. 电力系统保护与控制, 2022, 50(10): 114-122.
SUN Lijun, GU Xueping, LIU Tong, et al. A deep reinforcement learning algorithm-based active safety correction method for power grids[J]. Power System Protection and Control, 2022, 50(10): 114-122.

- [20] 王子晗, 高红均, 高艺文, 等. 基于深度强化学习的城市配电网多级动态重构优化运行方法[J]. 电力系统保护与控制, 2022, 50(24): 60-70.
WANG Zihan, GAO Hongjun, GAO Yiwen, et al. Multi-level dynamic reconfiguration and operation optimization method for an urban distribution network based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(24): 60-70.
- [21] 梁宏, 李鸿鑫, 张华赢, 等. 基于深度强化学习的微网储能系统控制策略研究[J]. 电网技术, 2021, 45(10): 3869-3877.
LIANG Hong, LI Hongxin, ZHANG Huaying, et al. Control strategy of microgrid energy storage system based on deep reinforcement learning[J]. Power System Technology, 2021, 45(10): 3869-3877.
- [22] 王新, 张良, 任晓龙, 等. 融合图神经网络模型与强化学习的综合能源系统优化调度[J]. 电力系统保护与控制, 2023, 51(24): 102-110.
WANG Xin, ZHANG Liang, REN Xiaolong, et al. Optimal scheduling of integrated energy systems by fusing a graph neural network model and reinforcement learning[J]. Power System Protection and Control, 2023, 51(24): 102-110.
- [23] YIN Xiuxing, LEI Meizhen. Jointly improving energy efficiency and smoothing power oscillations of integrated offshore wind and photovoltaic power: a deep reinforcement learning approach[J]. Protection and Control of Modern Power Systems, 2023, 8(2): 1-11.
- [24] 徐博涵, 向月, 潘力, 等. 基于深度强化学习的含高比例可再生能源配电网就地分散式电压管控方法[J]. 电力系统保护与控制, 2022, 50(22): 100-109.
XU Bohan, XIANG Yue, PAN Li, et al. Local decentralized voltage management of a distribution network with a high proportion of renewable energy based on deep reinforcement learning[J]. Power System Protection and Control, 2022, 50(22): 100-109.
- [25] 李鹏, 姜磊, 王加浩, 等. 基于深度强化学习的新能源配电网双时间尺度无功电压优化[J]. 中国电机工程学报, 2023, 43(16): 6255-6266.
LI Peng, JIANG Lei, WANG Jiahao, et al. Optimization of dual-time scale reactive voltage for distribution network with renewable energy based on deep reinforcement learning[J]. Proceedings of the CSEE, 2023, 43(16): 6255-6266.
- [26] 巨云涛, 陈希. 基于双层多智能体强化学习的微网群分布式有功无功协调优化调度[J]. 中国电机工程学报, 2022, 42(23): 8534-8548.
JU Yuntao, CHEN Xi. Distributed active and reactive power coordinated optimal scheduling of networked microgrids based on two-layer multi-agent reinforcement learning[J]. Proceedings of the CSEE, 2022, 42(23): 8534-8548.
- [27] 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.
- [28] XU Zhiwei, SU Wencong, HU Zechun, et al. A hierarchical framework for coordinated charging of plug-in electric vehicles in China[J]. IEEE Transactions on Smart Grid, 2017, 7(1): 428-438.
- [29] 朱旭, 孙元章, 杨博闻, 等. 考虑不确定性与非完全理性用能行为的电动汽车集群可调度潜力计算方法[J]. 电力自动化设备, 2022, 42(10): 245-254.
ZHU Xu, SUN Yuanzhang, YANG Bowen, et al. Calculation method of EV cluster's schedulable potential capacity considering uncertainties and bounded rational energy consumption behaviors[J]. Electric Power Automation Equipment, 2022, 42(10): 245-254.
- [30] 边晓燕, 孙明琦, 董璐, 等. 计及灵活性聚合功率的源-荷分布式协调调度[J]. 电力系统自动化, 2021, 45(17): 89-98.
BIAN Xiaoyan, SUN Mingqi, DONG Lu, et al. Distributed source-load coordinated dispatching considering flexible aggregated power[J]. Automation of Electric Power Systems, 2021, 45(17): 89-98.
- [31] 刘硕, 郭创新, 冯斌, 等. 基于价值分解深度强化学习的分布式光伏主动电压控制方法[J]. 电力自动化设备, 2023, 43(10): 152-159.
LIU Shuo, GUO Chuangxin, FENG Bin, et al. Active voltage control method of distributed photovoltaic based on value decomposition deep reinforcement learning[J]. Electric Power Automation Equipment, 2023, 43(10): 152-159.

收稿日期: 2024-01-25; 修回日期: 2024-07-10

作者简介:

李士丹(2000—), 男, 硕士研究生, 研究方向为人工智能在电力系统中的应用; E-mail: qingyu143@sjtu.edu.cn

李航(1995—), 男, 通信作者, 博士, 研究方向为人工智能在电力系统中的应用; E-mail: h-li@mail.tsinghua.edu.cn

李国杰(1965—), 男, 博士, 教授, 博士生导师, 研究方向为新能源的控制与接入, 微电网的分析与控制。E-mail: liguojie@sjtu.edu.cn

(编辑 周金梅)