

基于多智能体深度强化学习的配电网双时间尺度电压控制策略

赵晶晶, 张超立, 王涵, 盛杰
(上海电力大学电气工程学院, 上海 200090)

摘要: 风电、光伏(photovoltaics, PV)在新型电力系统中的渗透率日益增加, 使得配电网电压波动加剧, 而储能(energy storage, ES)、电动汽车(electric vehicles, EV)对降低配电网电压波动有重要作用。与此同时, 智能电表、智能传感器以及改进的通信网络广泛部署, 可获取的数据量越来越大, 数据驱动技术兴起。提出了一种基于多智能体深度强化学习(multi-agent deep reinforcement learning, MADRL)的配电网双时间尺度有功-无功功率协调的电压控制策略。慢时间尺度下用双深度Q网络算法(double deep Q-network algorithm, DDQN)求解电容器组(capacitor banks, CBs)、有载调压变压器(on-line tap changer, OLTC)与ES有功-无功功率优化问题。快时间尺度下用具有注意力机制的经验增强多智能体柔性参与者-评论家算法(experience augmentation- multi-agent soft actor critic, EA-MASAC)调节PV、风机(wind turbine, WT)、静止无功补偿装置(static var compensator, SVC)的无功功率与EV的有功功率。最后, 在IEEE-33节点系统上验证了所提方法的有效性。

关键词: 数据驱动; 多智能体深度强化学习; 双时间尺度; 电压控制; 功率优化

Distribution Network Dual Time Scale Voltage Control Strategy Based on Multi-Agent Deep Reinforcement Learning

ZHAO Jingjing, ZHANG Chaoli, WANG Han, SHENG Jie

(College of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: The increasing penetration rate of wind power and photovoltaics (PV) in new power systems exacerbates voltage fluctuations in distribution networks, while energy storage (ES) and electric vehicles (EV) play important roles in reducing voltage fluctuations in distribution networks. At the same time, smart meters, smart sensors and improved communication networks are widely deployed, the amount of data available is increasing, and data-driven technology is emerging. This paper proposes a multi-agent deep reinforcement learning (MADRL) based dual time scale active and reactive power coordinated voltage control strategy for distribution networks. Using the double deep Q-network algorithm (DDQN) to solve the optimization problems of capacitor banks (CBs), on line tap transformers (OLTC), and ES active and reactive power at a slow time scale. At a fast time scale, the EA-MASAC algorithm with attention mechanism is used to enhance the reactive power of PV, wind turbines (WT), and static var compensators (SVCs), as well as the active power of EVs. Finally, the effectiveness of the proposed method is verified on an IEEE-33 node system.

Key words: data driven; multi-agent deep reinforcement learning; dual time scales; voltage control; power optimization

0 引言

近年来, 随着大量分布式电源(distributed generation, DG)接入配电网, 电压越限问题频繁出

现, 这时单调控配电网的有功功率资源并不能保证配电网运行的安全性与经济性, 需同时协调配电网有功和无功功率优化^[1]。传统配电网无功功率优化主要是控制有载调压变压器(on-line tap changer, OLTC)、静止无功补偿装置(static var compensator, SVC)、电容器组(capacitor banks, CBs)等无功设备来稳定系统电压。随着光伏(photovoltaics, PV)、风机(wind turbine, WT)大规模并入电网, 电动汽

基金项目: 国家自然科学基金资助项目(52007112)。

Foundation item: Supported by the National Natural Science Foundation of China (52007112).

车兴起,车联网技术(vehicle-to-grid, V2G)愈发成熟, DG与电动汽车也可参与配电网电压调节^[2]。

目前,配电网功率优化方法主要有基于物理模型的优化法和数据驱动法两种。基于物理模型的电压控制方式有3种:集中式、分布式和分散式电压控制^[3]。文献[4-5]分别用模型预测控制、混合整数二阶锥规划等集中式电压控制技术实现了配网有功无功功率优化。关于分布式电压控制方式,文献[6-7]分别用多分块交替方向乘子法、改进交替方向乘子法解决了集中式调度海量通信的问题。分散式电压控制方面,文献[8]采用准实时无功功率优化降低了系统投资成本。但基于物理模型的配网优化方法其计算时间会随着配电网的大小和DG的数量呈指数形式增长^[9]。

为了应对以上挑战,数据驱动技术兴起,其中的深度强化学习(deep reinforcement learning, DRL)算法被广泛用于配电网电压调节^[10]、能源优化管理^[11]、机器人控制^[12]、网络重构^[13]等方面。文献[14]提出一种约束柔性参与者-评论家的非策略安全深度强化学习算法,并将配电网电压控制问题表述为具有离散动作空间的马尔可夫决策过程。但该算法只处理了离散动作控制问题,未考虑连续动作变量。文献[15]考虑离散和连续动作变量,提出一种基于混合随机变量联合分布的随机策略来处理配电网优化运行问题。文献[14-15]所提算法考虑单一智能体来进行配电网有功无功功率优化,但单一智能体算法鲁棒性差,学习效率低。

基于集中训练分散执行(centralized training and decentralized execution, CTDE)框架的MADRL算法可将任务分给多个智能体以实现分布式处理。文献[16]在长、短时间尺度用DQN、AC算法优化离散与连续调节设备。文献[17]用多智能体深度确定策略梯度算法对配电网有功无功功率进行调度。但文献[16-17]所用算法样本采样效率低,训练结果不稳定,这也是目前MADRL算法的通病。

数据增强(data augmentation, DA)技术是在已有数据上生成新数据来人为增加数据,可提高MADRL算法训练效率。DA已广泛用于单一智能体训练中,核心在于产生可靠的样本以扩大原始数据集,但多智能体算法中该技术较少见。文献[18]提出一种新的数据增强技术:经验增强技术(experience augmentation, EA),能提供无偏的生

成数据集,提高样本利用率。

综上所述,本文考虑DG、传统无功调节设备的无功功率调节能力以及ES、EV的有功功率调节能力,从双时间尺度对配电网进行有功无功功率优化。慢时间尺度用DDQN训练离散动作设备,快时间尺度用EA-MASAC训练连续动作设备。最后通过传统的MADRL算法、智能优化算法的算法组合比较,在IEEE 33节点算例上验证了所提算法组合的有效性、优越性。

1 配电网双时间尺度有功-无功功率协调电压控制策略

考虑到不同调节设备参与配电网有功无功功率优化时响应速度不同^[19], OLTC、CBs属于离散调节设备,调节时间在1 h左右; PV、WT、SVC属于连续调节设备,调节时间在1 min左右。图1为基于MADRL的配电网双时间尺度有功-无功功率协调电压控制框架。

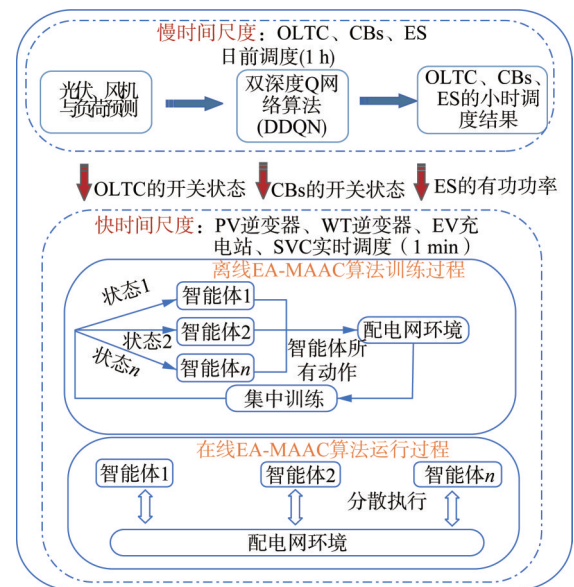


图1 配电网双时间尺度有功-无功功率协调电压控制策略框图

Fig. 1 Diagram of two timescale active and reactive power coordinated voltage control of distribution network

慢时间尺度上, OLTC、CBs和ES根据光伏、风机与负荷的预测值按每小时进行调度,并用DDQN算法进行求解。

快时间尺度上,采用EA-MASAC算法对配电网中的智能体进行离线训练来获得最优电压控制策

略并传给各个智能体,智能体根据本地观测值进行 PV、WT、SVC 的无功功率以及 EV 充电站的有功功率调节。

2 多智能体深度强化学习

深度学习用多层人工神经网络进行建模,侧重于对事物的感知。强化学习由智能体(agent)、环境(environment)、奖励(reward)组成。通过智能体本身的试错机制与周围环境进行交互,从而使得从环境中获得的累计奖励最大化。深度强化学习利用深度学习提取输入数据的抽象特征,并以此为依据进行强化学习,最后求出问题的最优策略^[20]。

单一智能体强化学习任务可用马尔可夫决策(Markov decision process, MDP)进行建模^[21]。MDP 由 (S, A, P, R) 四元素构成,其中 S 为状态空间; A 为动作空间; P 为状态转移概率; R 为奖励空间。

策略 π 表示智能体某一状态下采取确定或随机动作的概率分布。确定与随机策略函数如式(1)——(2)所示。

$$a = \pi(s) \quad (1)$$

$$\pi(S|a) = P(S|a) \quad (2)$$

式中: $a = \pi(s)$ 为智能体在状态 s 下采取策略 π 得到的确定性动作 a ; $\pi(S|a)$ 为智能体在状态 s 下采取随机性策略 π 得到随机性动作 a 的概率。

智能体从当前状态到最终状态获得的奖励之和定义为回报(return) G_t , 引入折扣系数 γ , 折扣回报表述为:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots, \gamma \in [0, 1] \quad (3)$$

式中: G_t 为智能体在第 t 次动作时得到的总回报; R_t 为智能体在第 t 次动作时获得的奖励。

这里引入值函数来判定策略 π 的好坏。状态与动作有相应的值函数, 分别对应式(4)——(5)。

$$v_\pi(s) = E_\pi(G_t|s) \quad (4)$$

$$q_\pi(s, a) = E_\pi(G_t|s, a) \quad (5)$$

式中: $v_\pi(s)$ 、 $q_\pi(s, a)$ 分别为智能体在状态 s 下采取策略 π 得到的状态与动作值函数; $E_\pi(\cdot)$ 为智能体在策略 π 下的数学期望。

强化学习核心任务在于找到 MDP 中的最优策略 π^* , 即:

$$v_{\pi^*}(s) = \max_{\pi} v_\pi(s), s \in S \quad (6)$$

式中: $v_{\pi^*}(s)$ 为智能体在状态 s 下采取最优策略 π^* 得到的最优状态值函数; S 为所有智能体状态的集合。

多智能体强化学习任务则是由马尔可夫博弈 $[M, S, A_1, A_2, \dots, A_M, R_1, R_2, \dots, R_M]$ 描述, 其中 M 为智能体的数量, $[A_1, A_2, \dots, A_M]$ 为所有智能体的动作集合, $[R_1, R_2, \dots, R_M]$ 为所有智能体从初始状态到最终状态的奖励集合。

3 基于 DDQN 算法的配电网慢时间尺度电压控制模型

3.1 数学模型

配电网慢时间尺度电压控制优化目标为配电网一天内网络损耗成本、电压偏差成本、离散设备的动作成本以及 ES 充放电成本 F_{low} 最小, 即:

$$F_{\text{low}} = \min \sum_{t=1}^T [c_{\text{loss}} P_{\text{Loss}, t} + \xi \sum_{i=1}^n \left| \frac{V_{i,t} - V_{i,N}}{V_{\text{max}} - V_{\text{min}}} \right| + C_{\text{OLTC}, t} + C_{\text{CBs}, t} + C_{\text{ES}, t}] \quad (7)$$

式中: T 为 24 h; t 为决策数量 ($t=1, 2, 3, \dots, 24$); c_{loss} 为网损系数, 设为 0.5 元/kWh; $P_{\text{Loss}, t}$ 为第 t 次决策时配电网网损; ξ 为电压偏差成本系数, 设定为 500 元/p. u.; $V_{i,t}$ 、 $V_{i,N}$ 为第 t 次决策时节点 i 实际电压与额定电压; V_{max} 与 V_{min} 分别为配电网安全运行的电压上下限, 设为 1.05 p. u. 与 0.95 p. u.; n 为配电网节点数量; $C_{\text{OLTC}, t}$ 、 $C_{\text{CBs}, t}$ 与 $C_{\text{ES}, t}$ 分别为第 t 次决策时 OLTC、CBs 的动作成本与 ES 充放电成本。 $P_{\text{Loss}, t}$ 、 $C_{\text{OLTC}, t}$ 、 $C_{\text{CBs}, t}$ 、 $C_{\text{ES}, t}$ 如式(8)——(11)所示。

$$P_{\text{Loss}, t} = r_{ij} \cdot (I_{ij,t})^2 \quad (8)$$

$$C_{\text{OLTC}, t} = \lambda_{\text{OLTC}} \sum_{m=1}^{M_{\text{OLTC}}} |T_{m,t} - T_{m,t-1}| \quad (9)$$

$$C_{\text{CBs}, t} = \delta_{\text{CBs}} \sum_{w=1}^{W_{\text{CBs}}} |C_{w,t} - C_{w,t-1}| \quad (10)$$

$$C_{\text{ES}, t} = c_{\text{ES}} \sum_{y=1}^{Y_{\text{ES}}} (P_{y,t}^{\text{disc}} + P_{y,t}^{\text{char}}) \quad (11)$$

式中: r_{ij} 为节点 i 、 j 间电阻; λ_{OLTC} 、 δ_{CBs} 分别为 OLTC、CBs 单位调节成本, 设为 2 元/次、1 元/次; M_{OLTC} 、 W_{CBs} 分别为 OLTC 与 CBs 数量; $T_{m,t}$ 为第 m 台 OLTC 第 t 次决策时的挡位; $C_{w,t}$ 为第 w 台 CB 第 t 次决策时的挡位; c_{ES} 为 ES 充放电成本系数, 设为 0.1 元/kWh; Y_{ES} 为 ES 数量; $P_{y,t}^{\text{char}}$ 、 $P_{y,t}^{\text{disc}}$ 分别为第 y

台ES第 t 次决策期间的充放电功率。

约束条件包括潮流约束、电压约束、OLTC与CBs分接头调约束、ES约束,如式(12)–(18)所示。

$$P_i = V_i \sum_{j=1}^n V_j (G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}) \quad (12)$$

$$Q_i = V_i \sum_{j=1}^n V_j (G_{ij} \sin \delta_{ij} - B_{ij} \cos \delta_{ij}) \quad (13)$$

$$V_{i, \min} < V_{i, \Delta t} < V_{i, \max} \quad (14)$$

$$\sum_{t=1}^T |T_{m, t+1} - T_{m, t}| \leq N_{\text{OLTC}, m, \max} \quad (15)$$

$$\sum_{t=1}^T |C_{w, t+1} - C_{w, t}| \leq N_{\text{CBs}, w, \max} \quad (16)$$

$$E_{i, \Delta t+1} = E_{i, \Delta t} + \eta_{\text{char}} P_{i, \Delta t}^{\text{char}} - \frac{P_{i, \Delta t}^{\text{disc}}}{\eta_{\text{disc}}} \quad (17)$$

$$S_{\min}^{\text{OC}} \leq S_{i, \Delta t}^{\text{OC}} \leq S_{\max}^{\text{OC}} \quad (18)$$

式中: P_i 、 Q_i 分别为注入节点 i 的有功、无功功率; V_i 、 V_j 、 G_{ij} 、 B_{ij} 、 δ_{ij} 分别为线路 ij 首末电压、电导电纳以及首末电压相角差; $V_{i, \max}$ 、 $V_{i, \min}$ 分别为节点 i 的电压上下限; $V_{i, \Delta t}$ 为 Δt 时刻节点 i 电压; $T_{m, t}$ 为第 m 台OLTC第 t 次决策时的档位; $C_{w, t}$ 为第 w 台CB第 t 次决策时的档位; $N_{\text{OLTC}, m, \max}$ 、 $N_{\text{CBs}, w, \max}$ 分别为一天内第 m 个OLTC与第 w 个CB动作次数上限; $E_{i, \Delta t}$ 为节点 i 处ES在 Δt 时刻的电量; η_{char} 、 η_{disc} 分别为ES充放电效率; $P_{i, \Delta t}^{\text{char}}$ 、 $P_{i, \Delta t}^{\text{disc}}$ 、 $S_{i, \Delta t}^{\text{OC}}$ 分别为节点 i 处ES在 Δt 时刻的充放电功率与荷电状态; $S_{\max}^{\text{OC}}$ 、 $S_{\min}^{\text{OC}}$ 分别为ES电量上下限。

3.2 慢时间尺度电压控制马尔可夫博弈过程

慢时间尺度下将OLTC、CBs、ES视为智能体。

1) 定义智能体 m 第 t 次决策时的智能体状态空间为 $s_{m, t}$ 。

$$s_{m, t} = [P_{m, t}, Q_{m, t}, V_{m, t}, T_{m, t}, C_{m, t}, N_{\text{OLTC}, m, t}, N_{\text{CBs}, m, t}, E_{m, t}] \quad (19)$$

式中: $P_{m, t}$ 、 $Q_{m, t}$ 、 $V_{m, t}$ 分别为智能体 m 第 t 次决策时各节点有功、无功功率矩阵以及电压矩阵; $T_{m, t}$ 、 $C_{m, t}$ 分别为智能体 m 第 t 次决策时OLTC与CBs所处档位; $N_{\text{OLTC}, m, t}$ 、 $N_{\text{CBs}, m, t}$ 分别为OLTC与CBs累计动作数量; $E_{m, t}$ 为智能体 m 第 t 次决策时ES电量。

2) 定义智能体 m 第 t 次决策时的智能体动作空间为 $a_{m, t}$ 。

$$a_{m, t} = [T_{m, t}, C_{m, t}, P_{m, t}^{\text{ES}}] \quad (20)$$

式中 $P_{m, t}^{\text{ES}}$ 为智能体 m 第 t 次决策时的ES充放电状态, ES充电时, $P_{m, t}^{\text{ES}} = P_{m, t}^{\text{char}}$, 放电时, $P_{m, t}^{\text{ES}} = -P_{m, t}^{\text{disc}}$ 。

3) 定义智能体 m 第 t 次决策时的智能体奖励函数为 r_t 。

$$r_t = -(c_{\text{loss}} P_{\text{Loss}, t} + \xi \sum_{i=1}^n \left| \frac{V_{i, t} - V_{i, N}}{V_{\max} - V_{\min}} \right| + C_{\text{OLTC}, t} + C_{\text{CBs}, t} + C_{\text{ES}, t}) - \phi [\sigma(V_{i, \min} < V_{i, t} < V_{i, \max}) + \sigma(N_{\text{OLTC}, m, t} \leq N_{\text{OLTC}, m, \max}) + \sigma(N_{\text{CBs}, w, t} \leq N_{\text{CBs}, w, \max})] \quad (21)$$

式中: ϕ 为惩罚系数, 设为1 000; $\sigma(\cdot)$ 为判断函数, 若电压及离散设备动作次数没越限, 判断函数的值为0, 反之为1。

4 基于EA-MASAC算法的配电网快时间尺度电压控制模型

4.1 数学模型

配电网快时间尺度电压控制优化目标为配电网一个周期内的网损成本、电压偏差成本、PV与WT有功削减成本、EV充电站运行成本最小, 总成本 F_{fast} 表述为:

$$F_{\text{fast}} = \min_{\Delta t=0}^T [c_{\text{loss}} P_{\text{Loss}, \Delta t} + \xi \sum_{i=1}^n \left| \frac{V_{i, \Delta t} - V_{i, N}}{V_{\max} - V_{\min}} \right| + c_{\text{PV}} \sum_{i=1}^n P_{i, \Delta t}^{\text{PV}, \text{dec}} + c_{\text{WT}} \sum_{i=1}^n P_{i, \Delta t}^{\text{WT}, \text{dec}} + c_{\text{TOU}} \sum_{x=1}^M P_{\text{char}, x, \Delta t}^{\text{EV}}] \quad (22)$$

式中: n 为配电网节点数量; $V_{i, \Delta t}$ 为节点 i 在 Δt 时刻实际电压; c_{PV} 、 c_{WT} 分别为PV、WT有功功率削减成本系数, 设为700元/MWh与800元/MWh; c_{TOU} 为分时电价; $P_{i, \Delta t}^{\text{PV}, \text{dec}}$ 、 $P_{i, \Delta t}^{\text{WT}, \text{dec}}$ 分别为节点 i 在 Δt 时刻PV与WT的有功功率削减量; $P_{\text{char}, x, \Delta t}^{\text{EV}}$ 为第 x 个EV充电站在 Δt 时刻的充电功率; M 为EV充电站数量。

约束条件包括PV、WT有功-无功功率约束, SVC无功功率约束以及EV充电站有功功率约束, 如式(23)–(30)所示。

$$0 \leq P_{i, \Delta t}^{\text{PV}} \leq P_{i, \Delta t}^{\text{PV}, \text{MPPT}} \quad (23)$$

$$P_{i, \Delta t}^{\text{PV}} = P_{i, \Delta t}^{\text{PV}, \text{MPPT}} - P_{i, \Delta t}^{\text{PV}, \text{dec}} \quad (24)$$

$$0 \leq P_{i, \Delta t}^{\text{PV}, \text{dec}} \leq P_{i, \Delta t}^{\text{PV}, \text{MPPT}} \quad (25)$$

$$|Q_{i, \Delta t}^{\text{PV}}| \leq \sqrt{(S_{i, \Delta t}^{\text{PV}})^2 - (P_{i, \Delta t}^{\text{PV}})^2} \quad (26)$$

$$0 \leq P_{i, \Delta t}^{\text{WT}} \leq P_{i, \Delta t, \max}^{\text{WT}} \quad (27)$$

$$(P_{i, \Delta t}^{\text{WT}})^2 + (Q_{i, \Delta t}^{\text{WT}})^2 \leq (S_{i, \Delta t}^{\text{WT}})^2 \quad (28)$$

$$Q_{i, \Delta t, \min}^{\text{SVC}} \leq Q_{i, \Delta t}^{\text{SVC}} \leq Q_{i, \Delta t, \max}^{\text{SVC}} \quad (29)$$

$$0 \leq P_{\text{char}}^{\text{EV}} \leq P_{\text{char}, \max}^{\text{EV}} \quad (30)$$

式中: $P_{i, \Delta t}^{\text{PV}}$ 、 $P_{i, \Delta t}^{\text{WT}}$ 分别为节点 i 在 Δt 时刻 PV 与 WT 的有功功率; $Q_{i, \Delta t}^{\text{PV}}$ 、 $Q_{i, \Delta t}^{\text{WT}}$ 分别为节点 i 在 Δt 时刻 PV 与 WT 的无功功率; $P_{i, \Delta t}^{\text{PV, MPPT}}$ 为节点 i 在 Δt 时刻 PV 处于 MPPT 模式下的有功功率; S^{PV} 为 PV 额定容量, 一般为装机容量的 1.0~1.1 倍^[22]; $S_{i, \Delta t}^{\text{WT}}$ 为节点 i 在 Δt 时刻 WT 的视在功率; $Q_{i, \Delta t, \max}^{\text{SVC}}$ 、 $Q_{i, \Delta t, \min}^{\text{SVC}}$ 分别为节点 i 在 Δt 时刻 SVC 无功功率的上下限; $P_{\text{char}}^{\text{EV}}$ 、 $P_{\text{char}, \max}^{\text{EV}}$ 分别为 EV 充电站的充电功率及其上限。

4.2 快时间尺度电压控制马尔可夫博弈过程

快时间尺度下将 PV、WT、EV 充电站、SVC 作为智能体。

1) 定义智能体 m 在 Δt 时刻的智能体状态空间为 $s_{m, \Delta t}$

$$s_{m, \Delta t} = [P_{m, \Delta t}, Q_{m, \Delta t}, V_{m, \Delta t}] \quad (31)$$

式中 $P_{m, \Delta t}$ 、 $Q_{m, \Delta t}$ 、 $V_{m, \Delta t}$ 分别为智能体 m 在 Δt 时刻各节点有功、无功功率矩阵以及电压矩阵。

2) 定义智能体 m 在 Δt 时刻的智能体动作空间为 $a_{m, \Delta t}$

$$a_{m, \Delta t} = [Q_{m, \Delta t}^{\text{PV}}, Q_{m, \Delta t}^{\text{WT}}, Q_{m, \Delta t}^{\text{SVC}}, P_{m, \Delta t}^{\text{PV, dec}}, P_{m, \Delta t}^{\text{WT, dec}}, P_{m, \Delta t}^{\text{EV, char}}] \quad (32)$$

式中: $Q_{m, \Delta t}^{\text{PV}}$ 、 $Q_{m, \Delta t}^{\text{WT}}$ 、 $Q_{m, \Delta t}^{\text{SVC}}$ 分别为智能体 m 在 Δt 时刻 PV、WT 以及 SVC 的无功功率; $P_{m, \Delta t}^{\text{PV, dec}}$ 、 $P_{m, \Delta t}^{\text{WT, dec}}$ 、 $P_{m, \Delta t}^{\text{EV, char}}$ 分别为智能体 m 在 Δt 时刻 PV、WT 的有功功率削减量以及 EV 充电站的充电功率。

3) 定义智能体 m 在 Δt 时刻的智能体奖励函数为 $r_{\Delta t}$

$$r_{\Delta t} = -(c_{\text{loss}} P_{\text{Loss}, \Delta t} + \xi \sum_{i=1}^n \left| \frac{V_{i, \Delta t} - V_{i, N}}{V_{\max} - V_{\min}} \right| + c_{\text{PV}} \sum_{i=1}^n P_{i, \Delta t}^{\text{PV, dec}} + c_{\text{WT}} \sum_{i=1}^n P_{i, \Delta t}^{\text{WT, dec}} + c_{\text{TOU}} \sum_{x=1}^M P_{\text{char}, x, \Delta t}^{\text{EV}}) - \phi [\sigma (V_{i, \min} < V_{i, t} < V_{i, \max})] \quad (33)$$

5 DDQN 算法与 EA-MASAC 算法基本原理及训练过程

5.1 DDQN 算法

双深度 Q 网络算法 (DDQN) 由 H. Van Hasselt 等人在双 Q-learning 算法的基础上提出^[23]。DDQN 算法在寻找最优策略时采用经验回放机制将智能体 t 时刻的状态 s_t 、动作 a_t 、即时奖励 r_t 以及 $t+1$ 时刻

的状态 s_{t+1} 存储到经验回放池中。回放池储存满后, DDQN 每次训练智能体都会从中随机抽取小批次数据, 并用梯度下降法对参数 θ_e 进行更新, 且每隔一段时间将参数 θ_e 拷贝给目标网络, 最后计算两个网络输出值的均方差函数的梯度是否达到最小, 达到最小值则输出最优动作。

DDQN 算法原理框图如图 2 所示。

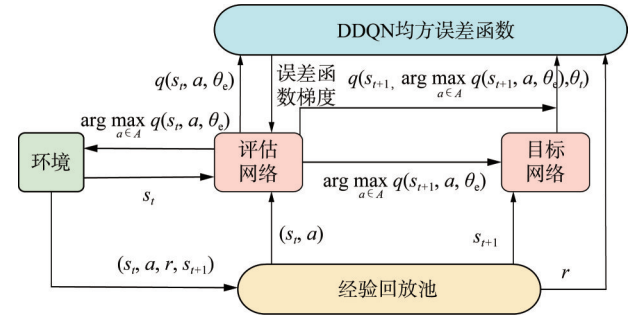


图 2 DDQN 原理框图流程

Fig. 2 Principle frame diagram of DDQN

5.2 EA-MASAC 算法

5.2.1 柔性参与者-评论家 (SAC) 算法

SAC 算法由 Haarnoja^[24] 等人提出, 是一种引入了最大熵鼓励机制的离线策略优化算法。利用最大熵机制, 智能体可以探索的范围更广, 可以选择的动作更多。SAC 算法有 1 个 actor (参与者) 网络与 4 个 critic (评论家) 网络。

SAC 算法目标可表述为:

$$\pi_*(s|a_t) = \arg \max_{\pi} \sum_{t=1}^T E_{(s_t, a_t) \sim p_{\pi}} [r(s_t, a_t) + \varpi H(\pi(s_t|\cdot))] \quad (34)$$

式中: $\arg \max_{\pi}$ 为智能体采取策略 π 时价值函数的最大值; $(s_t, a_t) \sim p_{\pi}$ 为策略 π 的状态动作轨迹; $E_{(s_t, a_t) \sim p_{\pi}}$ 为智能体在策略 π 下的状态动作轨迹所对应的数学期望; $r(s_t, a_t)$ 为智能体在状态 s_t 下采取动作 a_t 得到的奖励; ϖ 为熵的温度系数, 其值越大, 智能体探索能力越强; $H(\pi(s_t|\cdot))$ 为策略 π 在状态 s_t 下的熵。

SAC 算法通过柔性策略迭代方式搭配贝尔曼 backup 算子 T^{π} 进行策略更新。

$$T^{\pi} q(s_t, a_t) \rightarrow r(s_t, a_t) + \gamma E_{s_{t+1} \sim p_{\pi}} (v(s_{t+1})) \quad (35)$$

$$v(s_t) = E_{a_t \sim \pi} [q(s_t, a_t) - \log \pi(s_t|a_t)] \quad (36)$$

式中: $v(s_t)$ 、 $q(s_t, a_t)$ 分别为智能体在状态 s_t 下采取策略 π 得到的状态与动作值函数; $E_{s_{t+1} \sim p_{\pi}}$ 为智能

体在策略 π 下的状态轨迹所对应的数学期望; $E_{a_i \sim \pi}[\cdot]$ 为智能体在策略 π 下采取动作 a_i 得到的数学期望。

5.2.2 注意力机制与经验增强技术

注意力机制是查询向量(query)到匹配向量(key)-目标向量(value)的映射^[25], 可让一个智能体将其他智能体的信息融入到自己值函数的估计中, 以此判断其他智能体对自己的重要性。

经验回放机制在 DRL 算法中非常重要, 回放池需要收集的训练样本数量的期望为:

$$E(N_{\text{sample}}) = \frac{n}{T} \cdot \frac{b}{v} \quad (37)$$

式中: N_{sample} 为训练所需的样本数量; n/T 为训练网络时的更新频率, 即每 T 次迭代采样 n 组数据更新

训练模型; v 为读取回放池样本的速度; b 为批数据的大小, 即每次训练从回放池采样的样本数量。

为了防止算法过拟合, 在样本数量期望不变的前提下更新频率与批数据大小要尽可能大, 这时增加读取速度有利于经验回放。文献[18]将这种技术称为经验增强(EA)技术, 该技术依据智能体之间收集样本的对称性, 使得智能体的状态和动作空间探索更加快速、全面。

EA 技术可描述为: 找到一个置换矩阵集 $\mathbf{Z}$, 其中的每个矩阵按一定的顺序对回放池中的原始经验 (s_i, a, r, s_{i+1}) 进行转换。

$$(s_i, a, r, s_{i+1}) \rightarrow \mathbf{Z}(s_i, a, r, s_{i+1}) \quad (38)$$

5.3 MADRL 算法训练过程

MADRL 算法训练流程如图 3 所示。

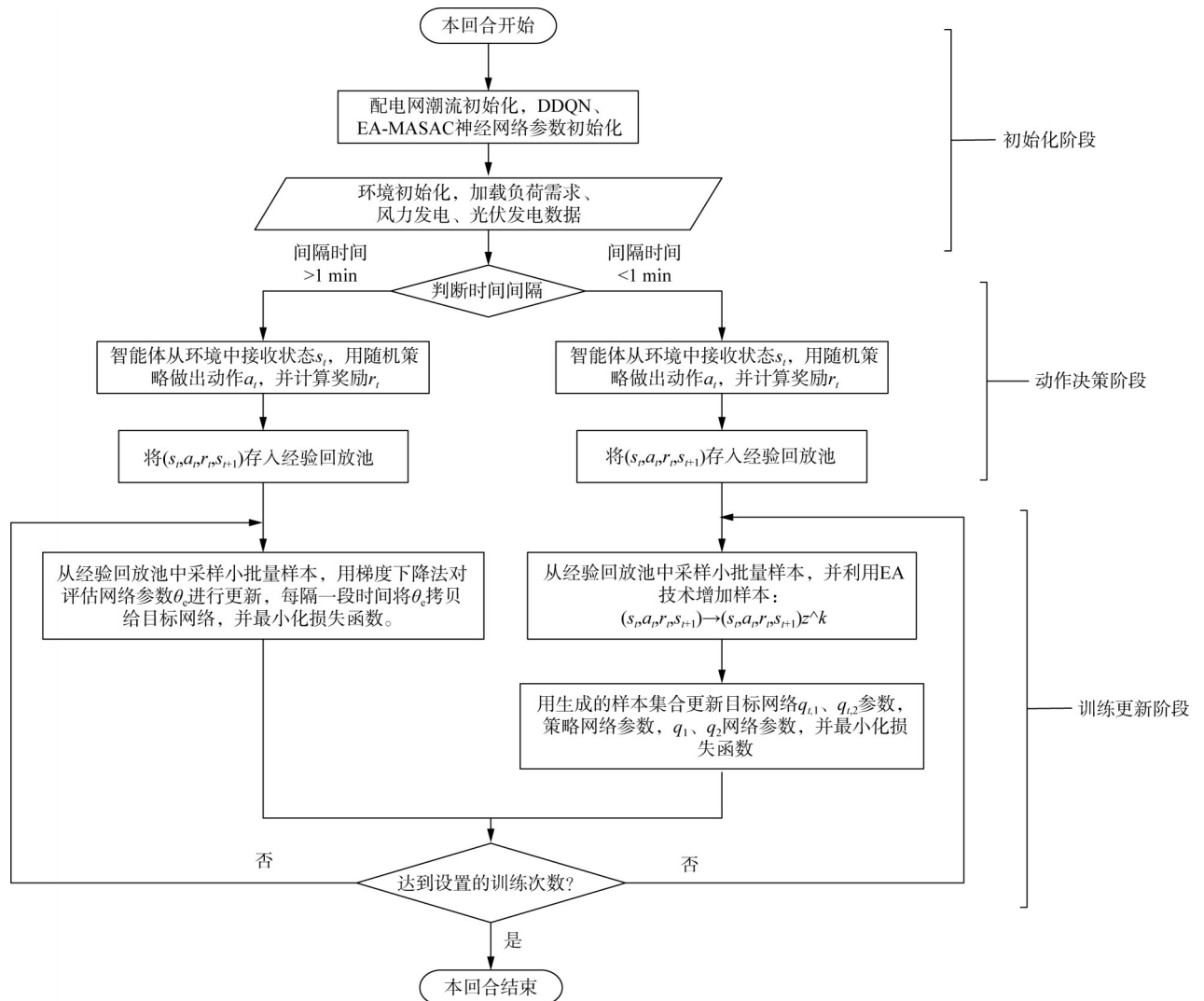


图3 配电网双时间尺度 MADRL 算法训练流程

Fig. 3 Training process of dual time scale MADRL algorithm in distribution network

6 算例分析

6.1 算例系统介绍及参数设置

本文提出的配电网双时间尺度有功-无功功率协调电压控制策略在改进后的 IEEE-33 节点的配电系统上进行测试，如图 4 所示。

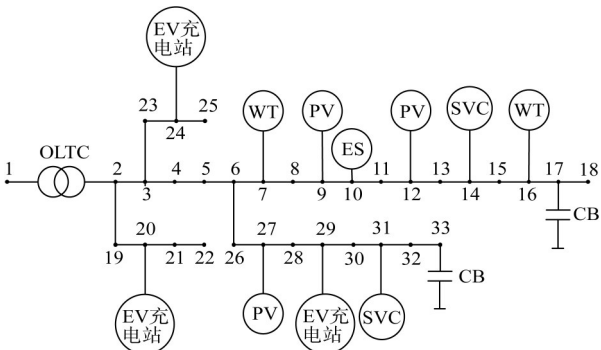


图 4 改进的 IEEE-33 节点配电网系统
Fig. 4 Improved IEEE-33 node distribution network system

系统的基准电压为 12.66 kV，EV 充电站包括家庭区、工作区与公共区充电站。各 EV 电池充电效率为 0.9，最大充电功率为 7 kW。本文对 1 000 辆 EV 进行蒙特卡洛抽样来产生 3 种 EV 充电站的充电功率曲线，如图 5 所示。

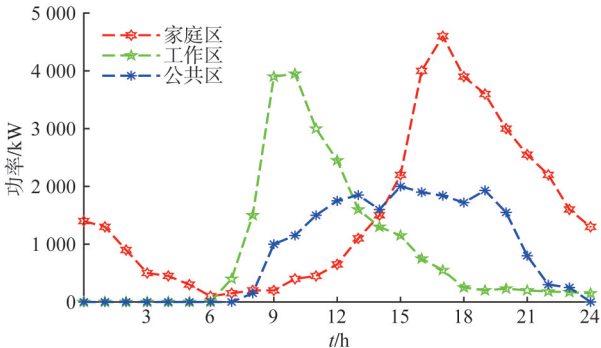


图 5 3 种电动汽车充电站充电功率曲线
Fig. 5 Three types of charging power curves for electric vehicle charging stations

系统中设备具体参数与分时电价^[26]如表 1—2 所示。

本文选择上海某地区某月 PV、WT 的实测数据，选取 20 d 作为强化学习训练集，其余 10 d 作为强化学习测试集。训练集被用在中心式学习阶段训练多智能体算法，当训练完成后，利用测试集数据在分布式执行阶段评估所提算法控制性能。测试集

表 1 IEEE-33 节点系统各设备参数

Tab. 1 Parameters of IEEE-33 node system equipments				
设备	参数	数值	运行限制	安装节点
OLTC	电压分接头档位	$\pm 5 \times 1\%$	5	1、2 之间
CB	每个档位可调节的无功功率/Mvar	3×0.5	5	17、33
ES	额定容量/kWh	500	SOC 值: 0.1~0.8	10
PV	最大有功功率/MW	0.6		9、12、27
WT	最大有功功率/MW	0.6		7、16
SVC	无功功率调节范围/Mvar	-0.5~0.5		14、31
EV 充电站	额定容量/kWh	50		20、24、29

表 2 分时电价

Tab. 2 Time-of-use tariffs	
时段	电价/(元·kWh ⁻¹)
(11:00—13:00, 16:00—17:00)	1.529 5
(10:00—11:00, 13:00—15:00, 18:00—21:00)	1.400 2
(7:00—10:00, 15:00—16:00, 17:00—18:00, 21:00—23:00)	0.874 5
(23:00—0:00, 0:00—7:00)	0.374 8

的 PV、WT 预测曲线如图 6—7 所示。

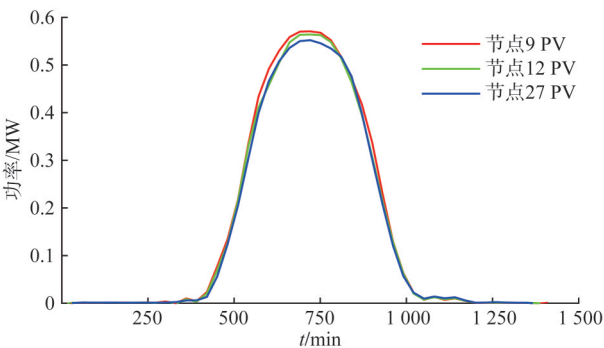


图 6 PV 预测曲线
Fig. 6 Photovoltaic forecast curves

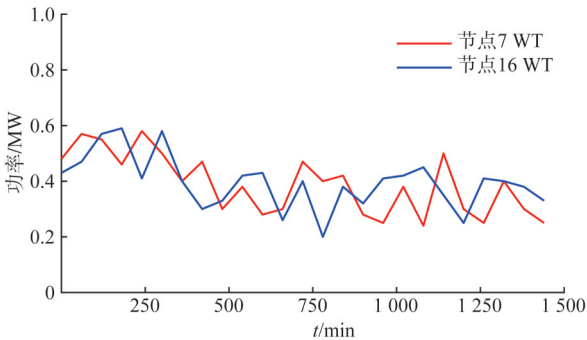


图 7 WT 预测曲线
Fig. 7 Wind turbine forecast curves

各个节点的负荷根据原始 IEEE33 节点系统中的空间负荷分布, 加入标准差 5 % 的随机高斯噪声, 用来模拟负荷需求的不确定性。负荷变化率曲线如图 8 所示。

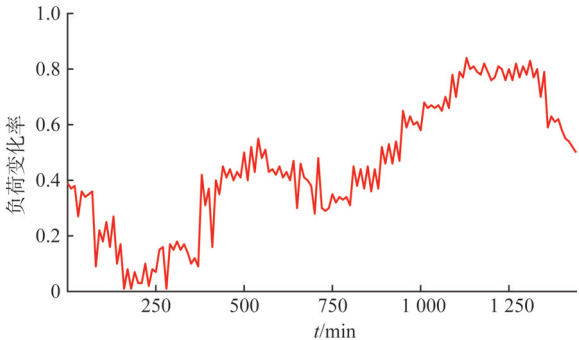


图 8 负荷变化率曲线

Fig. 8 Curve of load variation rates

6.2 MADRL 训练过程

本文潮流计算部分在 MATLAB 上完成, MADRL 算法在 Python 的 Pytorch 神经网络框架上完成。具体参数如表 3 所示。

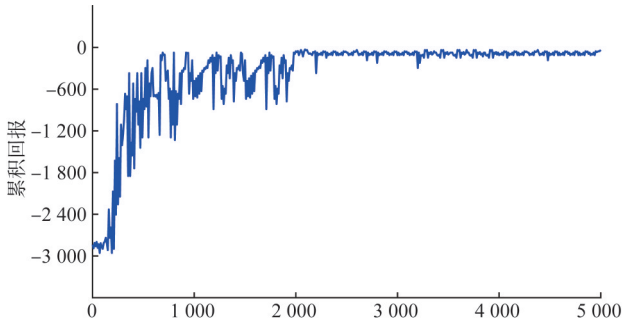
表 3 MADRL 参数设置

Tab. 3 Parameter settings of MADRL

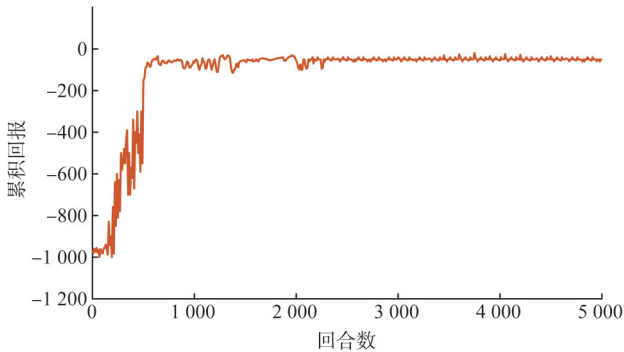
算法	参数	数值
DDQN	折扣系数	0.95
	学习率	0.001
	批数据大小	64
	经验回放池规模	10 000
	探索率初始值	0.1
	总探索步数/步	120 000
	每回合最大步数/步	24
	最大回合数量/回合	5 000
	隐藏层激活函数	ReLU
	隐藏层神经元数量/个	256
EA-MASAC	折扣系数	0.99
	actor, critic 网络学习率	0.001
	批数据大小	256
	经验回放池规模	106
	总探索步数/步	7 200 000
	每回合最大步数/步	1 440
	最大回合数量/回合	5 000
	隐藏层激活函数	ReLU
	隐藏层神经元数量/个	256
	软更新系数	0.001
	熵的温度系数	0.02

其中 DDQN 算法总探索步数取 120 000, 且每隔 24 步对目标网络参数进行更新, 一天为 1 个回合, 1 个回合包含 24 步数对应 24 h, 最大回合数取 5 000。EA-MASAC 算法总探索步数取 7 200 000, 1 个回合包含 1 440 步数对应每分钟, 最大回合数取 5 000。

图 9 为双时间尺度 MADRL 算法训练结果。



(a) DDQN 算法训练结果



(b) EA-MASAC 算法训练结果

图 9 算法训练结果

Fig. 9 Training results of Algorithm

从图 9 可见, 两算法智能体的累积回报在训练回合达到 200 前没有明显上升, 这时的智能体处于探索阶段, 深度神经网络的参数没有更新。智能体的动作、状态都存入到经验池中。200 回合后, 累积回报大幅度上升, 智能体开始训练, 深度神经网络的参数开始更新以适应环境变化。这时智能体积累的经验已经存满了经验池, 智能体能做出越来越优秀的决策, 使得网损与各类成本越来越小。DDQN 算法与 EA-MASAC 算法的累积回报分别在 2 000、1 500 回合以后逐渐收敛, 此时训练完成。

6.3 双时间尺度优化结果分析

如图 10 所示, OLTC 一天内的动作次数为 4, 两 CB 一天内的动作次数分别为 5、3, 均未超过最大调节次数。

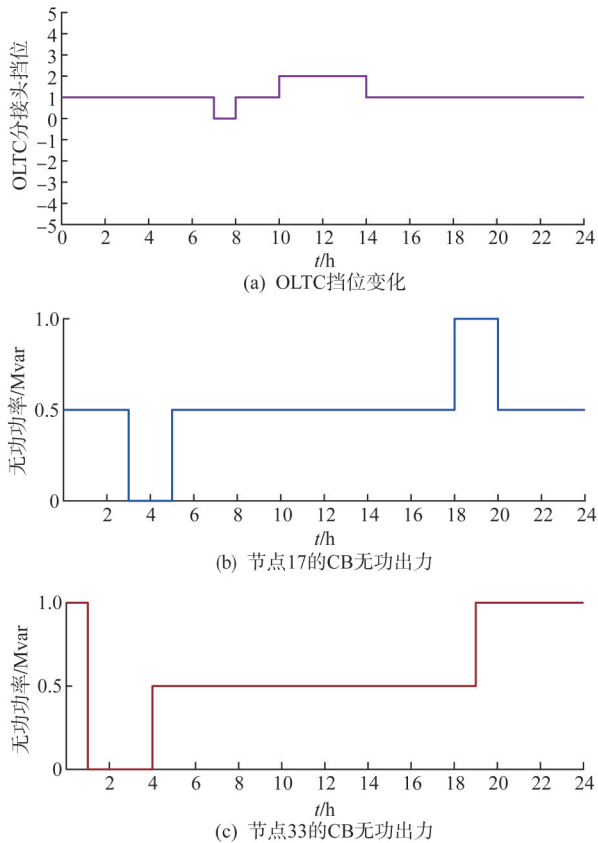


图 10 各个离散设备日前调度结果

Fig. 10 Daily scheduling results of each discrete equipment

如图 11(a)所示, 10:00—19:00ES 持续充电, 其余时段处于放电。因为 10:00—19:00, PV、WT 发电量大, 多余电量被 ES 储存以减少功率倒送, 其余时段 PV、WT 发电量低且负荷需求大。ES 的 SOC 变化曲线如图 11(b)所示, 在 0.1~0.8 之间波动, 满足约束。

连续动作设备日内调度结果如图 12 所示。

如图 12(a)所示, 在 720 决策步长左右, 两 SVC 都吸收无功功率, 平抑因 PV、WT 发电引起的电压波动。如图 12(b)所示, 家庭区、工作区、公共区充电站充电功率分别集中在晚高峰、早高峰、中下午时段。如图 12(c)—(d)所示, PV 与 WT 在各自出力较大时削减更多功率来降低电压偏移。

6.4 算法有效性分析

为了验证快时间尺度下 EA-MASAC 算法的有效性, 保持慢时间尺度 DDQN 算法不变, 将 EA-MASAC 算法与粒子群算法 (particle swarm optimization, PSO) 以及多智能体深度确定性策略梯度 (multi-agent deep deterministic policy gradient,

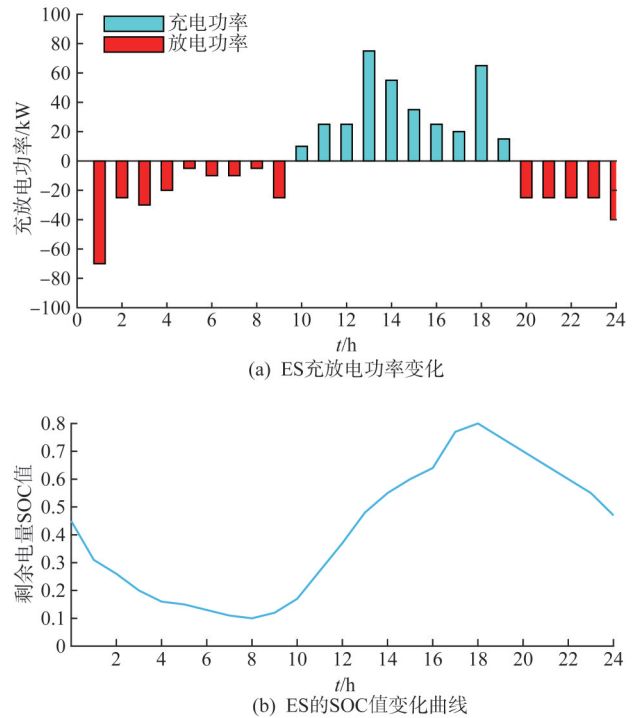


图 11 ES 日前调度结果

Fig. 11 Daily scheduling results of ES

MADDPG)算法进行对比。MADDPG 算法中 critic 网络学习率取 0.002, 其余参数与 EA-MASAC 算法保持一致。

如图 13 所示, 因为 EA-MASAC 算法使智能体在训练前样本数量较小时可获得更多的经验来实现更高的采样效率和更快的收敛速度。所以 EA-MASAC 算法相比较 PSO、MADDPG 算法累积回报更高, 且比 MADDPG 算法收敛更快, 训练效果更好。

此外, 为了说明所提出的双时间尺度算法组合在电压优化方面的有效性, 将以下 3 种方案与本文方法对比。

方案一: 采用双时间尺度 PSO 算法^[27]。

方案二: 采用双时间尺度遗传 (genetic algorithm, GA) 算法^[28-29]。

方案三: 慢时间尺度保持 DDQN 不变, 快时间尺度用 MADDPG。

1) 网络损耗分析

不同方案下配电网一天内各时刻网损变化如图 14 所示。方案一、方案二、方案三以及本文方法的日平均网损为 50.13 kW、47.17 kW、46.04 kW、44.38 kW, 可见本文方法在降低网损方面的优

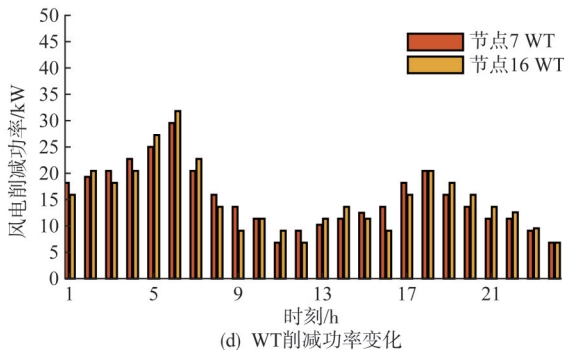
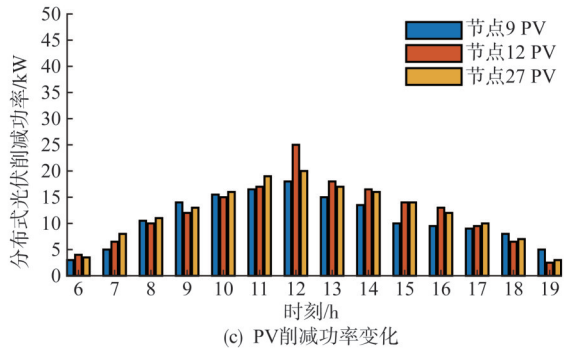
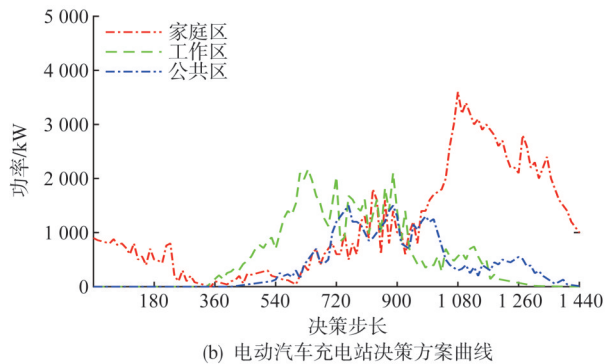
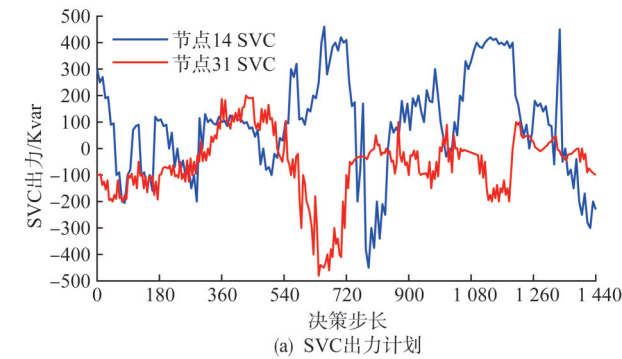


图 12 连续动作设备日内调度结果

Fig. 12 Daily scheduling results of continuous action equipment

2) 电压偏差分析

本文选取 PV、WT 出力较大的 12:00 以及出力较小的 06:00、18:00 这 3 个时刻来对不同方案下 33

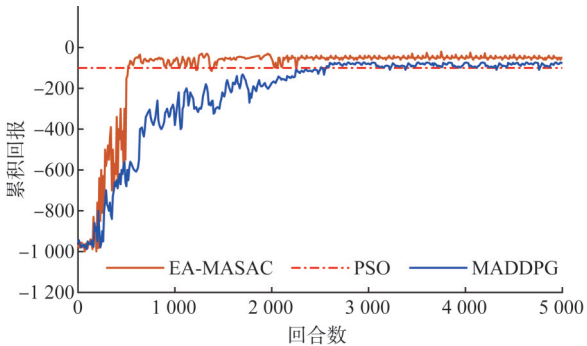


图 13 EA-MASAC 算法与传统优化算法训练结果
Fig. 13 Training results of EA-MASAC and traditional optimization algorithms

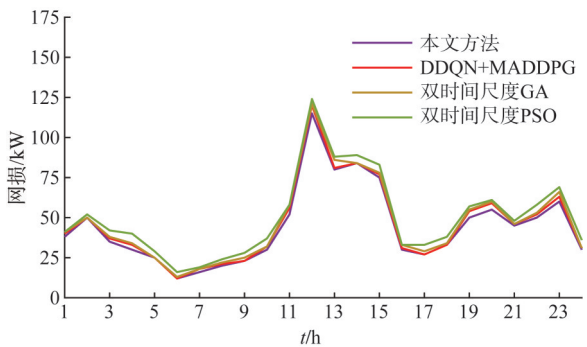


图 14 不同方案下网损变化曲线
Fig. 14 Network loss variation curves under different schemes

节点的电压分布进行对比分析, 分别如图 15(a)—(c) 所示。

12:00、06:00、18:00 时刻不同方案下 33 节点平均电压偏差如表 4 所示。

表 4 12:00、06:00、18:00 时刻各方案各节点平均电压偏差
Tab. 4 Average voltage deviation of each node in each scheme at 12:00, 06:00, 18:00 p.u.

方法	12:00	06:00	18:00
双时间尺度 PSO	0.026 3	0.021 3	0.022 2
双时间尺度 GA	0.025 1	0.020 2	0.021 4
慢时间尺度 DDQN, 快时间尺度 MADDPG	0.024 2	0.018 6	0.020 4
本文方法	0.023 6	0.017 1	0.019 3

可见用本文算法进行电压控制, 各节点平均电压偏差最小, 给配电网安全稳定运行提供了保障。

7 结语

本文基于 MADRL 设计了一个配电网双时间尺度有功-无功功率协调电压控制策略。慢时间尺度

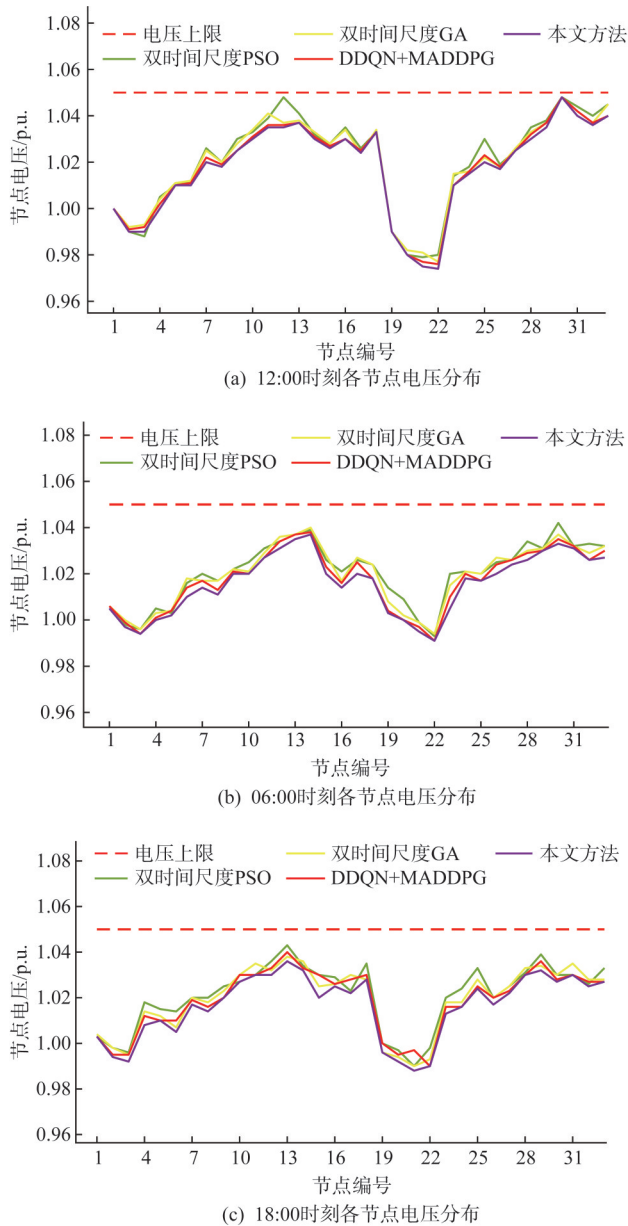


图 15 12:00、06:00、18:00 不同方案下各个节点电压分布
Fig. 15 Voltage distribution of each node under different schemes at 12:00、06:00、18:00

用 DDQN 算法优化离散动作设备，快时间尺度用 EA-MASAC 算法优化连续动作设备。最后在改进的 IEEE-33 节点系统中将本文算法与传统优化算法对比，本文算法配电网网损与节点电压偏差更小，更能有效平抑电压波动。

参考文献

- [1] 巨云涛, 陈希, 李嘉伟, 等. 基于分布式深度强化学习的微网群有功无功协调优化调度[J]. 电力系统自动化, 2023, 47(1): 115 – 125.
- [2] 周椿奇, 向月, 童话, 等. 轨迹数据驱动的电动汽车充电需求及 V2G 可调控容量估计[J]. 电力系统自动化, 2022, 46(12): 46 – 55.
- [3] ZHOU Chunqi, XIANG Yue, TONG Hua, et al. Trajectory-data-driven estimation of electric vehicle charging demand and vehicle-to-grid regulable capacity [J]. Automation of Electric Power Systems, 2022, 46(12): 46 – 55.
- [4] SUN X Z, QIU Q. Two-stage volt/var control in active distribution networks with multi-agent deep reinforcement learning method [J]. IEEE Transactions on Smart Grid, 2021, 12(4): 2903 – 2912.
- [5] 任佳依, 顾伟, 王勇, 等. 基于模型预测控制的主动配电网多时间尺度有功无功协调调度[J]. 中国电机工程学报, 2018, 38(5): 1397 – 1407.
- [6] REN Jiayi, GU Wei, WANG Yong, et al. Multi-time scale active and reactive power coordinated optimal dispatch in active distribution network based on model predictive control [J]. Proceedings of the CSEE, 2018, 38(5): 1397 – 1407.
- [7] 郑能, 丁晓群, 郑程拓, 等. 含高比例光伏的配电网有功—无功功率多目标协调优化[J]. 电力系统自动化, 2018, 42(6): 33 – 39, 91.
- [8] ZHENG Neng, DING Xiaoqun, ZHENG Chengtuo, et al. Multi-objective coordinated optimization of active and reactive power for distribution network integrated with high proportion of photovoltaic generation [J]. Automation of Electric Power Systems, 2018, 42(6): 33 – 39, 91.
- [9] 朱鹏程, 刘墨煜, 孙可, 等. 基于多分块交替方向乘子法的蜂巢状配电网分布式优化调度[J]. 中国电力, 2023, 56(6): 90 – 100.
- [10] ZHU Pengcheng, LIU Zhaoyu, SUN Ke, et al. Optimal scheduling of honeycomb distribution network based on BADMM [J]. Electric Power, 2023, 56(6): 90 – 100.
- [11] 赵冬梅, 王浩翔, 陶然. 基于改进交替方向乘子法的输配电网分散协调鲁棒优化调度模型[J]. 电网技术, 2023, 47(3): 1138 – 1151.
- [12] ZHAO Dongmei, WANG Haoxiang, TAO Ran. Decentralized coordination robust optimal scheduling model for transmission and distribution networks based on improved alternating direction method of multipliers [J]. Power System Technology, 2023, 47(3): 1138 – 1151.
- [13] MANBACHI M, SADU A, FARHANGI H, et al. Real-Time co-simulation platform for smart grid volt-var optimization using IEC 61850 [J]. IEEE Transactions on Industrial Informatics, 2016, 12(4): 1392 – 1402.
- [14] GAO Y Q, YU N P. Deep reinforcement learning in power distribution systems: overview, challenges, and opportunities [C]// 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), February 16 – 18, 2021, Washington, DC, USA. New York: IEEE, 2021: 1 – 5.
- [15] 曹迪. 主动配电网电压控制策略研究[D]. 成都: 电子科技大学, 2021.
- [16] LI S C, HU W Y, CAO D, et al. A multi-agent deep reinforcement learning-based approach for the optimization of transformer life using coordinated electric vehicles [J]. IEEE Transactions on Industrial Informatics, 2022, 18(11): 7639 –

- 7652.
- [12] 翁士博. 深度强化学习SAC算法在运动控制中的应用研究[D]. 西安: 西安理工大学, 2020.
- [13] GAO Y Q, SHI J, WANG W, et al. Dynamic distribution network reconfiguration using reinforcement learning [C]// 2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, October 21 – 23, 2019, Beijing, China. New York: IEEE, 1 – 7.
- [14] WANG W, YU N P, GAO Y Q, et al. Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems [J]. IEEE Transactions on Smart Grid, 2020, 11(4): 3008 – 3018.
- [15] LI H P, HE H B. Learning to operate distribution networks with safe deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2022, 13(3): 1860 – 1872.
- [16] 倪爽, 崔承刚, 杨宁, 等. 基于深度强化学习的配电网多时间尺度在线无功优化[J]. 电力系统自动化, 2021, 45(10): 77 – 85.
- NI Shuang, CUI Chenggang, YANG Ning, et al. Multi-timescale online optimization for reactive power of distribution network based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(10): 77 – 85.
- [17] 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.
- [18] HU D E, YE Z H, GAO Y Q, et al. Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization [J]. IEEE Transactions on Smart Grid, 2022, 13(6): 4873 – 4886.
- [19] 周任军, 段献忠, 周晖. 计及调控成本和次数的配电网无功优化策略[J]. 中国电机工程学报, 2005, 25(9): 23 – 28, 157.
- ZHOU Renjun, DUAN Xianzhong, ZHOU Hui. A strategy of reactive power optimization for distribution system considering control action cost and times [J]. Proceedings of the CSEE, 2005, 25(9): 23 – 28, 157.
- [20] 陈卫东, 吴宁, 黄彦璐, 等. 基于深度学习的微电网优化调度辅助决策方法[J]. 南方电网技术, 2022, 16(1): 117 – 126.
- CHEN Weidong, WU Ning, HUANG Yanlu, et al. Auxiliary decision-making method of optimal dispatching for microgrid based on deep learning [J]. Southern Power System Technology, 2022, 16(1): 117 – 126.
- [21] 谈竹奎, 刘斌, 王扬, 等. 基于多智能体协调的柔性温控负荷分布式分层控制策略[J]. 南方电网技术, 2023, 17(7): 45 – 54.
- TAN Zhukui, LIU Bin, WANG Yang, et al. Distributed hierarchical control strategy of flexible thermostatically controlled loads based on multi-agent coordination [J]. Southern Power System Technology, 2023, 17(7): 45 – 54.
- [22] 蔡永翔, 唐巍, 徐鸥洋, 等. 含高比例户用光伏的低压配电网电压控制研究综述[J]. 电网技术, 2018, 42(1): 220 – 229.
- CAI Yongxiang, TANG Wei, XU Ouyang, et al. Review of voltage control research in LV distribution network with high proportion of residential PVs [J]. Power System Technology, 2018, 42(1): 220 – 229.
- [23] HASSELT H V, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning [J]. Computer Science, 2015, 8(4): 2094 – 2100.
- [24] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor [J]. Statistics, 2018, 2(1): 1861 – 1870.
- [25] IQBAL S, SHA F. Actor-attention-critic for multi-agent reinforcement learning [J]. Statistics, 2018, 12(1): 1467 – 5413.
- [26] 赵星宇, 胡俊杰. 集群电动汽车充电行为的深度强化学习优化方法[J]. 电网技术, 2021, 45(6): 2319 – 2327.
- ZHAO Xingyu, HU Junjie. Deep reinforcement learning based optimization for charging of aggregated electric vehicles [J]. Power System Technology, 2021, 45(6): 2319 – 2327.
- [27] 安然. 基于随机矩阵理论的配电网无功优化方法研究[D]. 北京: 北京交通大学, 2020.
- [28] 曹尚. 用于主动配电网协调优化控制的深度强化学习方法及其应用研究[D]. 武汉: 华中科技大学, 2021.
- [29] 谢虎, 杨占杰, 张伟, 等. 基于改进遗传算法的主动配电网优化定价策略研究[J]. 电测与仪表, 2024, 61(7): 116 – 122.
- XIE Hu, YANG Zhanjie, ZHANG Wei, et al. Research on optimal pricing strategy of active distribution network based on improved genetic algorithm [J]. Electrical Measurement & Instrumentation, 2024, 61(7): 116 – 122.

收稿日期: 2024-02-27; 网络首发日期: 2024-05-20

作者简介:

赵晶晶(1980), 女, 副教授, 硕士生导师, 博士, 研究方向为电力系统弹性、新能源发电并网控制技术, jjzhao_sh@163.com;
张超立(1999), 男, 通信作者, 硕士研究生, 研究方向为电力系统运行与控制, zcl1614082329@163.com;
王涵(2001), 男, 硕士研究生, 研究方向为电力系统频率与电压控制, wanghan_suep@163.com。