

基于多智能体强化学习的两阶段电压控制策略

张涛¹, 郝正航¹, 徐玉韬², 马启鹏³, 李超¹, 杨玉杰¹

(1. 贵州大学电气工程学院, 贵阳 550025; 2. 贵州电网有限责任公司电力科学研究院, 贵阳 550002; 3. 国网宁波供电公司, 浙江 宁波 315000)

摘要: 随着大量分布式光伏接入配电网, 配电网在应对网络重构和源荷储不确定性等方面面临较大挑战。因此提出一种主动配电网两阶段电压控制策略, 第一阶段对主动配电网联络开关进行集中控制, 以小时为调度周期并以网损最小为目标进行网络重构, 建立混合整数二阶锥规划模型进行求解。第二阶段对光伏和储能系统进行实时电压控制, 将实时电压控制问题转换为马尔科夫博弈过程(Markov game process, MGP)并实行多智能体建模, 采用离线训练-在线运行的方法。相比于传统的两阶段均采用数学规划的方法, 所提控制策略不依赖于精确的配网潮流模型, 对通信要求低、求解速度更快。最后在改进的IEEE 33节点系统算例验证了所提控制策略的有效性。

关键词: 分布式光伏; 分布式储能; 配电网重构; 多智能体强化学习; 电压控制

Two-Stage Voltage Control Strategy Based on Multi-Agent Reinforcement Learning

ZHANG Tao¹, HAO Zhenghang¹, XU Yutao², MA Qipeng³, LI Chao¹, YANG Yujie¹

(1. Electrical Engineering College, Guizhou University, Guiyang 550025, China; 2. Electric Power Research Institute of Guizhou Power Grid Co., Ltd., Guiyang 550002, China; 3. State Grid Ningbo Power Supply Company, Ningbo, Zhejiang 315000, China)

Abstract: With a large number of distributed photovoltaic (PV) access to the distribution network, the distribution network faces greater challenges in dealing with network reconstruction and source-load-storage uncertainty, etc. Therefore, a two-stage voltage control strategy for active distribution networks is proposed. In the first stage, the contact switch of the active distribution network is centrally controlled. The network reconstruction is carried out with the goal of minimizing the network loss in an hourly scheduling period, and a mixed integer second-order cone planning model is established to solve the problem. In the second stage, the real-time voltage control of photovoltaic and energy storage systems is carried out, and the real-time voltage control problem is converted into the Markov game process (MGP). The multi-agent model is implemented and the offline training-online operation method is adopted. Compared with the traditional two-stage mathematical planning approach, the control strategy proposed does not rely on an accurate distribution network power flow model, has low communication requirements and a faster solution speed. Finally, the effectiveness of the proposed control strategy is verified by the improved IEEE 33-node calculation examples.

Key words: distributed photovoltaic; distributed energy storage; distribution network reconstruction; multi-agent reinforcement learning; voltage control

0 引言

党的二十大报告提出, 要积极稳妥推进碳达峰碳中和, 加快建设新能源体系^[1-2]。而分布式光伏

发电具有清洁环保、运维方便、就近供电和经济实用等优点, 在未来的新能源体系中将是必不可少的^[3]。但分布式光伏发电具有随机性和不可调度性, 可能会造成电网电压波动、越限等问题; 随着分布式能源在配网中渗透率的增大, 其电压波动、越限等问题将更为严重^[4-7]。因此, 分布式光伏接入主动配电网导致的电压越限问题亟需解决。

基金项目: 国家自然科学基金资助项目(522670031001610)。

Foundation item: Supported by the National Natural Science Foundation of China (522670031001610).

国内外针对主动配电网电压控制方法已经开展了广泛研究,主要可分为集中式控制、分布式控制和多时间尺度控制^[8-10]。集中式控制通过采集整个配电网电气信息来实现调度,各个终端采集信息传送给中央处理器,中央处理器完成最优潮流计算后发出调度指令给各终端控制器。但配电网优化是非线性优化问题,通常将非凸问题转化为凸优化问题能够提高求解效率^[11-14]。文献[13]提出网络重构与无功电压协调优化的混合整数凸规划模型,将配电网中的非凸模型转化为具有凸特性的混合整数二阶锥规划模型,以此来求得最优解。分布式控制采用配电网局部信息与相邻区域的信息交互实现对配电网电压的控制,主要采用交替方向子乘法^[15-17]。文献[17]构建基于交替方向子乘法的多区域电压分布式协同优化控制框架,基于残差平衡原理和松弛技术提出了一种加速的交替方向子乘法,能够降低网络损耗以及解决配电网电压越限风险。多时间尺度控制方法利用主动配电网中机械式设备的调整时间较长和电力电子式设备的调整时间较短的特性来实现长-短时间尺度优化调度^[18-20]。文献[20]提出基于 Fisher 时段划分的配电网源-网-荷-储协调优化控制策略,通过日前优化、日内滚动优化和区域自治优化实现三层电压控制,能够有效快速地解决电压越限问题。

以上方法均是基于精确的物理模型进行验证,而配电网实际运行的参数难以获取,并且以上方法的计算量庞大,在一定程度上限制了这些方法的应用。近年来,人工智能算法依靠无模型和数据驱动等优势逐渐被应用到主动配电网电压控制中,通过离线训练-在线运行的方式将漫长的计算过程转移到线下^[21-25]。文献[23]提出配电网动态网络重构,使用 Q-learning 和深度确定性策略梯度算法(deep deterministic policy gradient, DDPG)进行智能体训练,结果表明其执行速度更快有效地解决了电压越限问题。但将所有调节设备看作是单智能体时,随着系统中调节设备的增多,会造成动作空间维数的爆炸式增长,因而导致整个模型很难得到最优解。文献[24-25]提出采用多智能体深度确定性策略梯度算法,使各智能体之间相互协作,结果表明相比于将所有可控设备视为一个单智能体其训练效果更好速度更快,但该算法存在 Q 值过估计以及训练不稳定等问题,最终收敛至一个次优解,导致最后的

优化结果较差。

鉴于此,为了充分发挥配网重构以及分布式光伏、储能系统的电压协调控制作用,本文提出了一种基于两阶段多智能体强化学习的配网重构与光储系统无功电压协调优化相统一的电压控制策略。首先,在第一阶段,配电网以网损最小化为目标和以小时为调度周期进行网络重构,第二阶段实行多智能体分布式光储系统无功电压调节。使用双延迟深度确定性策略梯度(twin delayed deep deterministic policy gradient algorithm, TD3)算法对多智能体进行训练,实现多个智能体之间的相互协调优化,最终实现对电压的控制。与基于数学模型的传统数学规划和启发式算法求解相比,本文提出的方法不需要具体的配电网物理模型,精度更高,求解运行速度极快,且扩展性更好。最后在 IEEE 33 节点算例仿真验证了所提出的两阶段多智能体强化学习的配网重构与光储无功电压协调优化相统一的电压控制策略的可行性。

1 两阶段配电网电压控制策略框架

基于传统全局信息的集中式电压控制方法对配电网进行调度,通常需要全部实时的配电网参数信息,因而需要大量的计算以及实时通信,但现阶段配电网通信设施落后,基于传统全局信息的集中式电压控制方法的适用性受限。主动配电网联络开关不能实时动作,但如今的电力电子设备能够实现快速连续调节。鉴于此,本文提出一种基于多智能体强化学习的两阶段主动配电网电压控制策略,第一阶段对配电网进行小时网络重构,第二阶段对光储设备进行实时调度。在对配电网进行网络重构的基础上对光储设备进行实时电压调节,能够进一步缓解电压越限和网络损耗。所提控制策略框架如图 1 所示。

第一阶段针对配电网联络开关以小时为调度周期并以网损最小化为目标进行网络重构。在第二阶段中将光伏逆变器与储能设备作为控制对象进行建模,为降低通信要求和由通信延迟造成的控制性能下降进行多智能体分区协调控制。离线训练时使用历史数据进行多智能体强化学习得到相应的控制策略,在训练完成后将训练好的智能体导入配电网环境中在线运行,最后实现对配电网电压的实时控制。

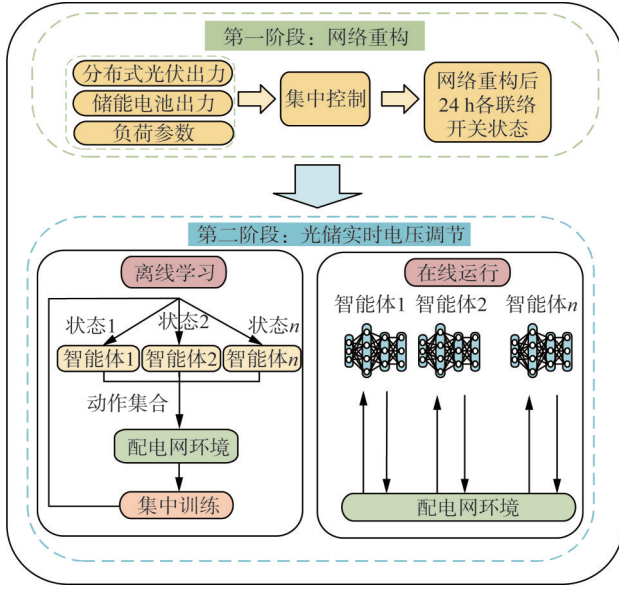


图 1 两阶段电压控制框图

Fig. 1 Block diagram of two-stage voltage control

2 第一阶段配网重构模型

2.1 目标函数

在第一阶段主动配电网重构模型中以网损最小化为目标, 目标函数为:

$$\min \sum_{i=1}^n P_{i2} \quad (1)$$

式中: P_{i2} 为配电网的网络损耗; i 为网络中第 i 个节点编号; n 为总节点数量。

2.2 约束条件

1) 配电网辐射状网络结构约束。

$$\begin{cases} \sum_{l=1}^m \alpha_l = n \\ \beta_{ij} + \beta_{ji} = \alpha_i \\ \sum_{j \in N(i)} \beta_{ij} = 1 \\ \beta_{0j} = 0 \\ \beta_{ji} \in \{0, 1\}, i = 1, 2, \dots, n, j \in N(i) \\ 0 \leq \alpha_l \leq 1, l = 1, 2, \dots, m \end{cases} \quad (2)$$

式中: α_l 为支路状态变量, 支路连通时, $\alpha_l = 1$, 支路断开时, $\alpha_l = 0$; β_{ij} 和 β_{ji} 为二进制变量被用做支路 α_l 的连接状态, 当 $\beta_{ij} = 1$ 时, 节点 i 是节点 j 的父节点, 当 $\beta_{ji} = 1$ 时, 节点 j 是节点 i 的父节点; l 为支路总数量; $N(i)$ 为与节点 i 连接的所有节点的集合; m 为总支路数量。

2) 配电网潮流方程约束。

$$\begin{cases} P_{i1} + P_{i, dg} - P_{i, load} = \\ \sum_{j \in N(i)} \alpha_{ij} [-G_{ij} U_i^2 + U_i U_j (G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij})] \\ Q_{i1} + Q_{i, dg} - Q_{i, load} = \\ \sum_{j \in N(i)} \alpha_{ij} [-B_{ij} U_i^2 + U_i U_j (G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij})] \end{cases} \quad (3)$$

式中: P_{i1} 、 $P_{i, dg}$ 、 $P_{i, load}$ 分别为在节点 i 处根节点输送的有功功率、分布式电源的有功出力 and 有功负荷; Q_{i1} 、 $Q_{i, dg}$ 、 $Q_{i, load}$ 分别为节点 i 处根节点输送的无功功率、分布式电源的无功出力和无功负荷; U_i 为节点 i 的电压; θ_{ij} 为节点 i 与 j 之间的相角差; α_{ij} 为二进制变量, $\alpha_{ij} = 1$ 时节点 i 与 j 联通, $\alpha_{ij} = 0$ 时节点 i 与 j 断开; G_{ij} 和 B_{ij} 分别为线路 ij 的电导和电纳矩阵。

3) 节点电压安全约束。

$$U_{\min} \leq U_i \leq U_{\max} \quad (4)$$

式中 $U_{\min}$ 、 $U_{\max}$ 分别为节点 i 处电压 U_i 的最小值、最大值。

4) 支路电流约束。

$$I_{ij}^2 \leq I_{ij, \max}^2 \quad (5)$$

式中: I_{ij} 为支路 ij 的电流; $I_{ij, \max}$ 为支路 ij 允许流过的最大电流。

5) 分布式光伏运行约束。

$$\begin{cases} 0 \leq P_{i, dg} \leq \bar{P}_{i, dg} \\ Q_{i, t, dg} = P_{i, t, dg} \tan \theta_{i, t, dg} \\ \sqrt{P_{i, t, dg}^2 + Q_{i, t, dg}^2} \leq S_{i, dg} \end{cases} \quad (6)$$

式中: $\bar{P}_{i, dg}$ 为节点 i 处分布式光伏有功出力的最大值; $P_{i, t, dg}$ 和 $Q_{i, t, dg}$ 分别为 t 时刻节点 i 处分布式光伏的有功和无功出力; $\theta_{i, t, dg}$ 为 t 时刻节点 i 处分布式光伏的功率因数角; $S_{i, dg}$ 为节点 i 处分布式光伏容量。在本文中光伏出力的最小值设为 0, 最大值受光照强度等影响, 且均以恒功率因数运行。

6) DESS 运行约束。

$$\begin{cases} 0 \leq P_{i, t}^+ \leq y_i, 0 \leq P_{i, t}^- \leq y_i \\ E_{i, t} = E_{i, t-1} + \eta_i^c P_{i, t}^+ - (1/\eta_i^d P_{i, t}^-) \\ z_i \cdot S_{i, \min}^{\text{soc}} \leq E_{i, t} \leq z_i \cdot S_{i, \max}^{\text{soc}} \\ E_{i, t=T} = z_i \cdot S_{i, t=0}^{\text{soc}} \end{cases} \quad (7)$$

式中: $P_{i, t}^+$ 、 $P_{i, t}^-$ 分别为在 t 时刻节点 i 处 DESS 的充电和放电有功功率; y_i 为节点 i 处 DESS 的功率容量; $E_{i, t}$ 为 t 时刻存储在节点 i 处 DESS 的能量; η_i^c 、 η_i^d 分别为节点 i 处 DESS 的充电、放电效率; z_i 为节

点 i 处 DESS 的能量容量; $S_{i, \min}^{\text{SOC}}$ 、 $S_{i, \max}^{\text{SOC}}$ 分别为节点 i 处 DESS 的最小、最大荷电状态极限。

2.3 模型的凸化处理

一阶段配电网重构的数学模型描述的是一个混合整数非线性问题, 其中潮流约束具有非线性以及非凸性, 不仅求解难度大, 而且很难求得最优解。因此可将潮流方程进行凸化松弛, 得到其二阶锥的形式, 就能将原问题转化为混合整数凸规划问题, 因而能够保证高效地求得全局最优解。首先根据二阶锥松弛定义以下中间变量。

$$V_i = U_i / \sqrt{2} \quad (8)$$

$$\delta_{ij} = U_i U_j \cos(\theta_i - \theta_j) \quad (9)$$

$$\chi_{ij} = U_i U_j \sin(\theta_i - \theta_j) \quad (10)$$

式中: V_i 、 δ_{ij} 、 χ_{ij} 分别为根据配电网特点而定义的中间变量; θ_i 、 θ_j 分别为节点 i 和 j 的功角。

将式(8)–(10)代入原目标函数和潮流约束方程中, 式(3)可变形为:

$$\begin{cases} P_{i1} + P_{i, \text{dg}} - P_{i, \text{load}} = P_{i2} \\ Q_{i1} + Q_{i, \text{dg}} - Q_{i, \text{load}} = Q_{i2} \end{cases} \quad (11)$$

其中,

$$\begin{cases} P_{i2} = \sum_{j=1}^n \alpha_{ij} (G_{ij} \delta_{ij} + B_{ij} \chi_{ij}) \\ Q_{i2} = \sum_{j=1}^n \alpha_{ij} (B_{ij} \chi_{ij} - G_{ij} \delta_{ij}) \end{cases} \quad (12)$$

$$2V_i V_j \geq \delta_{ij}^2 + \chi_{ij}^2 \geq 0, \chi_{ij} \geq 0 \quad (13)$$

式中 P_{i2} 、 Q_{i2} 分别为配电网的有功损耗和无功损耗。

通过上述的等价变形, 式(13)可以写成标准的二阶锥形式。

$$\left\| \begin{bmatrix} V_i + V_j, 2\delta_{ij}, 2\chi_{ij} \end{bmatrix}^T \right\| \leq V_i - V_j \quad (14)$$

经过式(11)–(14)的凸变换, 原可行域已经松弛成凸可行域, 最后使用商业求解器 GUROBI 进行求解。

3 第二阶段基于多智能体强化学习的电压控制策略

3.1 多智能体强化学习

首先, 每个智能体与环境进行交互完成强化学习训练, 智能体通过感知环境的状态再对环境进行选择动作, 对所观测到执行该动作后的结果进行相应调整, 最终通过完整训练的智能体能够对当前环

境做出最优动作策略, 以此来获得最大奖励值。多智能体强化学习的模型可以描述为马尔可夫博弈过程, 可表示为 $\langle n, s, a_1, \dots, a_n, p, \gamma, r_1, \dots, r_n \rangle$, 其中 n 为智能体的数量, $s = s_1 \times \dots \times s_n$ 为状态空间, s_n 为第 n 个智能体的状态, $a = a_1 \times \dots \times a_n$ 为每个智能体的动作空间, a_n 为第 n 个智能体的动作, p 为状态转移函数, $p: s \times a_1 \times \dots \times a_n \times s \rightarrow [0, 1]$; γ 为累计奖励折扣系数, $\gamma \in [0, 1]$; r_n 为第 n 个智能体做完一个互动所得的奖励; $r_i(s_i, a_1, \dots, a_n, s_{i+1})$ 为智能体 i 在状态 s_i 时执行联合动作所得到的奖励 r_i , 而奖励函数的设置通常与环境 and 目标有关。对于每个智能体而言, 它们都有一个共同的目标即最大化奖励回报。用期望可表示为:

$$\max_{\pi_i} E_{a_{i,t} \sim \pi_i, s_{t+1} \sim P(s_{t+1}|s_t, a_t)} \left[\sum_{t=0}^T \gamma r_i, t \right] \quad (15)$$

式中: E 为求取的期望值; π_i 为智能体 i 的策略; $a_{i,t}$ 为智能体 i 在 t 时刻的动作; $P(s_{t+1}|s_t, a_t)$ 为智能体在状态 s_t 时执行动作 a_t 状态转移到 s_{t+1} 的概率; T 为总时间步长。

如今多智能体强化学习算法大多都是基于 Actor-Critic 框架。在该框架下, 为了解决多个智能体交互的问题需要 MADRL 同时学习一个策略和一个值函数。策略网络 Actor 用于指导智能体如何选择动作, 评论网络 Critic 用于对这一动作进行价值评估。其大概流程为智能体观察到当前环境状态 s , 生成一个动作策略 $\pi_i = (a_i, l_{s_i})$, 配电网作为环境在收到智能体给出的动作策略后将再次计算配电网潮流并向每个智能体返还更新状态。

3.2 双延迟深度确定性策略梯度算法

3.2.1 TD3算法的介绍

深度确定性策略梯度算法 (deep deterministic policy gradient, DDPG) 在配电网分布式光储电压协调控制中的表现很好, 但是依然存在一些问题。DDPG 算法是根据 DQN (deep q-learning) 思想进行改进而得出的, 因此其很容易因为 Q 值过估计而造成次优策略的问题。鉴于此, 提出一种双延迟深度确定性策略梯度算法的电压控制策略, 该算法基于 AC 框架, 通过建立两个目标 Critic 网络来产生两个评估值 Q'_1 和 Q'_2 , 并同时学习两个网络, 用最小化的方法选择 Q 值, 从而避免 Q 值被过估。其次是延

迟策略更新, 目标网络和当前网络的不同步更新, 在一定程度上可以减少累计误差, 以此来降低方差。第三是目标策略平滑, 通过引入噪声使得估计值更加平滑。

3.2.2 TD3算法的设计流程

把当前状态 s_t 作为策略网络的输入, 把控制决策视作动作 $a_t = \pi(s_t) + \varepsilon$ 的输出, 其中 $\pi(s_t)$ 为在 t 时刻状态为 s_t 下执行的策略。该模型根据状态 s_t 和动作 a_t 预测在下一个时刻的状态转移 s'_{t+1} 以及在当前环境下获得的奖励 r'_t , 并将状态转移和奖励反馈给智能体。智能体通过与配电网环境的连续交互, 得到预测的动作数据, 并将这些数据存放在经验回放池中, 并从中抽取一定样本数据用于策略和价值优化。策略网络参数根据最大化累计期望回报进行更新, 价值网络则根据最小化估值和目标之间的差值进行更新。

1) 策略网络的更新。TD3算法是基于确定性策略梯度的算法, 因此策略网络为 $\pi(s_t; \theta)$, 优化策略网络的参数 θ 以此达到最大化策略函数 $J_\pi(\pi)$, 其值可表示为:

$$J_\pi(\pi) = E_{s \sim \kappa^\pi} [Q^\pi(s_t, \pi(s_t; \theta))] \quad (16)$$

式中: $J_\pi(\pi)$ 为能够获得的期望回报; κ^π 为当前状态在策略 π 下的分布; $Q^\pi(s_t, \pi(s_t; \theta))$ 为在当前状态下执行策略 π 动作的价值。采用梯度下降的方法来更新策略函数 $J_\pi(\pi)$ 中的相关参数。

$$\nabla_{\theta^\pi} J_\pi(\pi) = E_{s \sim \kappa^\pi} [\nabla_{a_t} Q^\pi(s_t, a_t) | a_t = (a_{t,1}, \dots, a_{t,n}) \cdot \nabla_{\theta^\pi} \pi(s_t | \theta^\pi)] \quad (17)$$

式中: θ^π 为当前主网络参数; θ^π 为目标主网络参数。

2) 价值网络的更新。价值网络是用来对策略网络所做出的动作进行价值评估的, 动作价值函数 $Q(s_t, a_t)$ 用贝尔曼方程可表示为:

$$Q(s_t, a_t) = E[r_t + \gamma Q(s'_{t+1}, a'_{t+1})] \quad (18)$$

最小化贝尔曼残差近似动作函数为:

$$J_Q(\omega) = E[r_t + \gamma Q(s'_{t+1}, a'_{t+1} | \omega^Q) - Q(s_t, a_t | \omega^Q)] \quad (19)$$

式中: $J_Q(\omega)$ 为一个参数化的动作值函数; ω^Q 为当前 Q 网络的网络参数; ω^Q 为目标 Q 网络的网络参数; $Q(s_t, a_t | \omega^Q)$ 为当前动作 a_t 的动作一值函数。

在 $t+1$ 时刻的价值网络输出为:

$$y_t = r_t + \gamma Q(s'_{t+1}, a'_{t+1} | \omega_j^{Q'}) \quad (20)$$

在智能体和环境的交互过程中中心化动作价值函数 Q 的损失函数为:

$$L(\omega) = \frac{1}{N} \sum_{j=1}^N [(y_t - Q(s_t, a_t | \omega^Q))^2] \quad (21)$$

式中: $L(\omega)$ 为损失函数用于衡量模型预测值与实际值之间的差异; N 为训练样本的数量; y_t 为目标 Q 值。由于 Q 值网络在训练中存在过估和 DDPG 策略导致的目标估计值方差偏大的问题, TD3 算法使用的双 Q 网络和目标策略平滑正则化予以解决, 双 Q 网络是指在两套完全相同的网络架构下的 Q 值网络, 选取其中小的值作为目标价值, 以此来抑制价值网络造成的过估计问题, 表示为:

$$y_t = r_t + \gamma \min_{j=1,2} \{Q(s'_{t+1}, a'_{t+1} | \omega_j^{Q'})\} \quad (22)$$

式中: $\omega_j^{Q'}$ 是指两个完全不同目标 Q 网络的网络参数。为防止网络的过拟合, 通过向目标动作中添加正态分布噪声并求平均值来近似动作期望, 本质是一种正则化方法, 采用噪声平滑掉 Q 峰值。因此在下一个状态 s_{t+1} 的动作为:

$$a'_{t+1} = \pi(s'_{t+1}) + \varepsilon \quad (23)$$

$$\varepsilon \sim \text{clip}[N(0, \sigma), -c, c] \quad (24)$$

式中: ε 为随机噪声; $\text{clip}(\cdot)$ 为裁剪函数用于限制 $N(0, \sigma)$ 的值在一定范围内; $N(0, \sigma)$ 为正态分布, σ 为方差; c 为噪声裁剪的限值。

在整个训练过程中 Critic 网络的参数根据训练阶段来更新, 但是目标网络不参与整个训练过程, 而其参数随着估值网络的参数进行迭代更新直到结果收敛。TD3 算法的神经网络结构如图 2 所示。

3.3 计及采样优先级的 TD3 算法

由于样本的 TD 误差越大对方向梯度的计算影响就越大, 因此被采样到的概率就越大, 从而造成经验池中具有较高价值的信息没有被采样到, 导致最后收敛至一个次优解。为了使得算法收敛更快并提高训练效率, 本文采用计及采样优先级的深度确定性策略梯度 (prioritized experience relay MATD3, PER MATD3) 算法。PER MATD3 算法考虑不同样本的目标 Q 网络的 Q 值和当前 Q 网络的 Q 值所得到的时序差分误差 (temporal-difference error, TD 误差) 对神经网络反向传播作用不同, 通

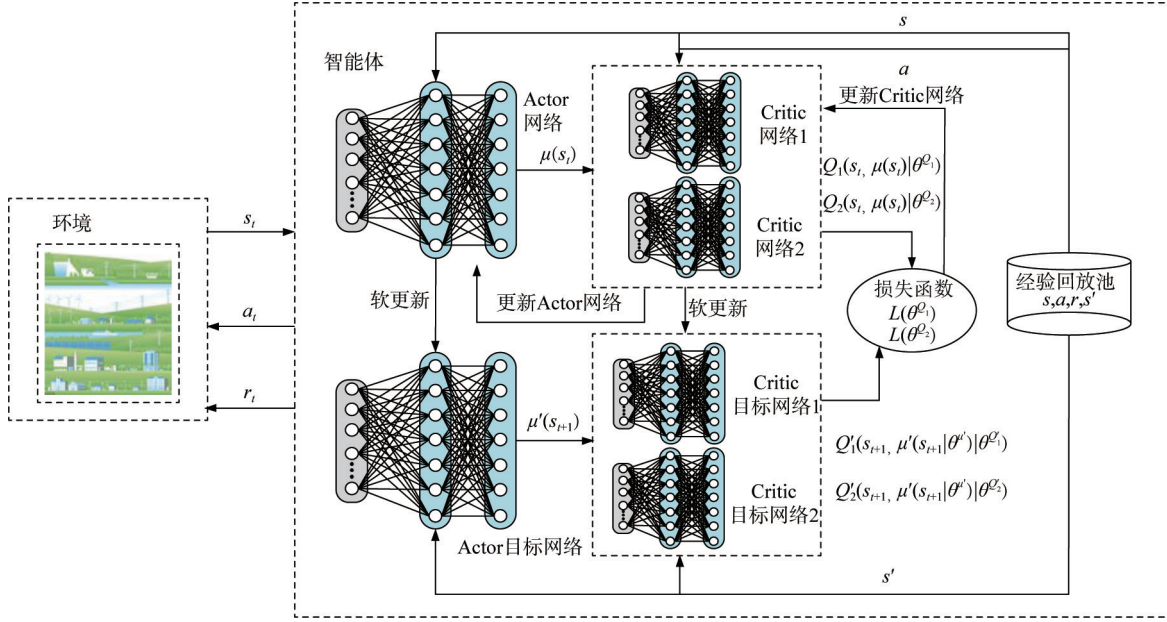


图2 TD3算法神经网络架构

Fig. 2 TD3 algorithm neural network architecture

过给每个样本设置一个正比于TD误差绝对值的优先级参数 δ ，将其存放在经验池当中，使得具有较高价值的信息被优先采样。

3.4 实时电压控制及马尔可夫博弈设计

在进行光储电压协调控制时将配电网看作是环境，将光储设备视作能够与环境进行实时交互的智能体。状态空间 s 为检测到的配网电压信息，动作空间 a 为光储协调控制装置的调度命令集合，因而可将原来的光储协调控制问题转化为实时的马尔可夫博弈过程。

定义第 i 个智能体在第 t 时刻的状态观测值为 $s_{i,t}$ ， U_t 为测量到节点的电压向量。 n 个智能体在第 t 时刻的状态集合为 $s_t = \{s_{1,t}, s_{2,t}, \dots, s_{n,t}\}$ 。光储设备具备实时连续调节能力，能够满足目前配电网实时电压调节要求。本文定义光伏和储能的逆变器为一个智能体，若配电网后期增加分布式能源可以将其视为新的智能体并进行训练即可，故而其扩展性好。设智能体 i 在 t 时刻的动作为 $a_{i,t} \in a$ ，则智能体的无功/有功出力为 $a_i = \{a_{i,pv}, a_{i,de}\}$ 。第 n 个智能体在 t 时刻的动作集合 $a_t = \{a_{1,t}, a_{2,t}, \dots, a_{i,t}, a_{n,t}\}$ 。在整个学习训练过程当中，多智能体根据当前的状态 s 和动作 a 的共享观察结果，来选择最优动作，在整个交互过程中配电网的运行约束将自然被满足。基于光储设备的无功/有功协调的电压调节，能够在—

阶段重构的基础上进一步降低网损、改善电压分布，因此可将目标函数定义为：

$$\min F = \sum_t (\lambda_1 P_{\text{loss},t} + \lambda_2 (V_{i,t} + \eta)) \quad (25)$$

$$V_{i,t} = \sum_{j \in L} [\text{ReLU}^2(V_{\min} - V_{j,t}) + \text{ReLU}^2(V_{j,t} - V_{\max})] \quad (26)$$

式中： F 为系统网损和各节点电压越限惩罚值之和； $\text{ReLU}(\cdot)$ 为线性整流函数； λ_1 为网损系数； $P_{\text{loss},t}$ 为在 t 时刻配电网线路的网损； λ_2 为节点电压越限惩罚系数； η 为电压越限附加惩罚值； L 为除根节点外的节点总数； $V_{i,t}$ 为 t 时刻节点 i 的电压越限量。定义第 i 个智能体在 t 时刻观测到的奖励为 $r_{i,t}$ ， n 个智能体在 t 时刻的奖励集合为 $r_t = \{r_{1,t}, r_{2,t}, \dots, r_{i,t}, \dots, r_{n,t}\}$ ，整个系统中的智能体通过相互合作的方式进行无功/有功电压优化。强化学习的目标就是累积奖励最大化，因此，本文将目标函数转换为强化学习中的马尔可夫决策过程中的单步奖励时，将其设计为相反数，如式(27)所示。

$$r_{i,t} = -\lambda_1 P_{\text{loss},t} - \lambda_2 (V_{i,t} + \eta) \quad (27)$$

4 算例分析

4.1 算例系统介绍

本文仿真测试的硬件平台配置为：NVIDIA RTX 3050T Intel(R) Core(TM) i5-12500H CPU，所

有仿真均是基于该硬件平台实现。采用如图 3 所示的 IEEE 33 节点系统作为仿真算例。设定基准电压为 12.66 kV, 根节点电压标幺值为 1.00 p. u., 系统安全运行的电压标幺值范围在 0.95 p. u. 到 1.05 p. u. 之间。分布式光伏和储能的安装点如图 3 所示。安装容量 PV1 和 PV4 均为 1.25 MW, PV2 和 PV5 均为 1.2 MW, PV3 和 PV6 均为 1 MW, 储能 1—3 容量分别为 0.3、0.4、0.3 MW。选取贵州省某地的光伏历史数据, 负荷数据来自原始的 IEEE 33 节点算例的空间负荷分布。训练集和测试集分别为 10 d 和 5 d 光伏历史数据, 测试集中的光伏日前预测曲线和负荷日前预测曲线如图 4 所示。本文将一天划分为 24 个时段, 每个时段为 1 h, 并以小时为单位进行两阶段电压优化。

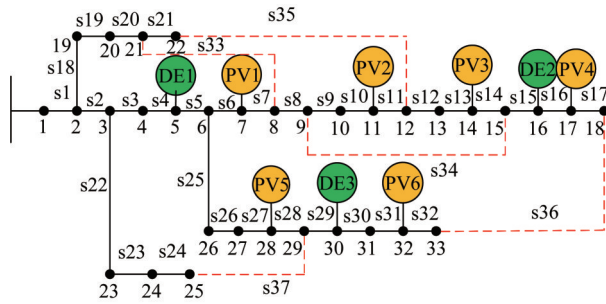


图 3 IEEE 33 节点

Fig. 3 IEEE 33 nodes

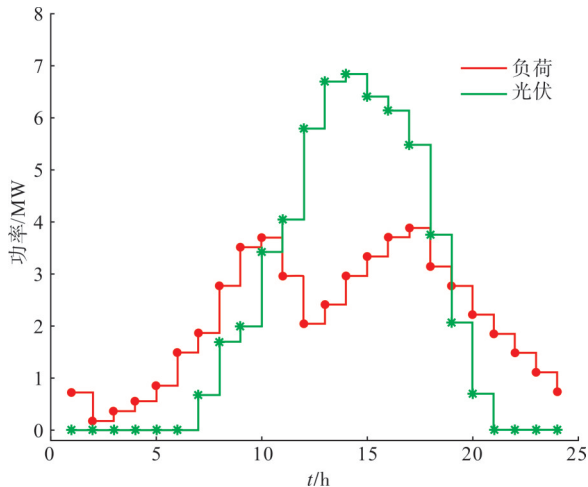


图 4 负荷和光伏预测出力

Fig. 4 Load and PV forecast outputs

本文中强化学习的参数设置如下。每个智能体网络结构均是相同的, Actor 和 Critic 网络结构各包含 1 个输入层、3 个隐藏层和 1 个输出层。Actor 网络和 Critic 网络的学习率均为 0.05, 奖励折扣因子

γ 为 0.99, 软更新速度因子为 0.000 1。目标 Q 值估计因子 μ 被设置为 0.95。经验回放池容量为 10 000, 训练最大回合数为 24, 训练总步长为 20 000。Actor 网络和 Critic 网络的输入层和隐藏层的激活函数均为整流线性单元函数, 输出层的激活函数均为双曲正切函数。

4.2 仿真方案和结果分析

为验证所提算法的有效性, 将 PER MATD3 算法与以下几种方法进行对比, 且以下方案均在同一配电网环境中运行。

方案 1: 原始方法, 不采取任何调压措施。

方案 2: 考虑两阶段电压优化, 使用商业求解器 GUROBI 求解。

方案 3: 不考虑第一阶段网络重构, 第二阶段采用 PER MATD3 算法对光储进行协调电压优化。

方案 4: 考虑第一阶段网络重构, 第二阶段采用 MADDPG 算法对光储进行协调电压优化。

方案 5: 考虑第一阶段网络重构的作用, 第二阶段采用 PER MATD3 算法对光储进行协调电压优化。

在同一环境中对智能体进行训练, 方案 3、4 和 5 的训练累计奖励如图 5 所示。在训练刚开始时由于智能体不成熟导致电压越限以及网损偏大使得奖励值较低, 在训练 4 000 轮后, PER MATD3 算法收敛, 其收敛后的平均奖励值为 -62.96, 说明智能体通过训练学得了电压控制策略, 电压不再越限。而用 MADDPG 算法训练智能体时训练 13 000 轮后才渐渐收敛, 且收敛精度较差, 其收敛后的平均累计奖励值为 -65.48, 比方案 5 的平均累计奖励值小 -2.52。从图中还可以看出方案 3 的累计平均奖励值为 -91.51, 相比方案 5 有一阶段网络重构, 其奖励值小了 -25.55, 说明网络重构能够进一步降低电压越限的风险和网损。

图 6 为在没有采取任何调压措施时各节点各时段的电压分布图。在 14 h 光照强度最大时有多个节点电压越限, 最大值达到 1.077 p. u., 电压严重越限, 且电压分布极不均匀。

在采用本文方案 5 的方法后, 各节点各时段的电压分布如图 7 所示, 从图 7 可以看出各时段的节点电压均未越限。

14 h 光照强度最大时采用各个方案优化后的电压如图 8 所示, 从图中 8 得出, 在不采用任何调压

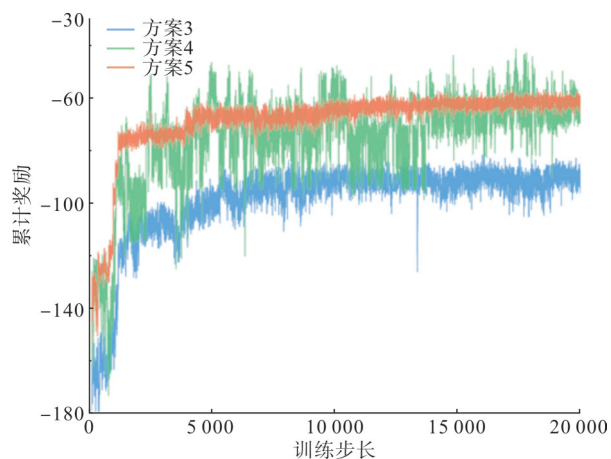


图5 累计奖励

Fig. 5 Cumulative rewards

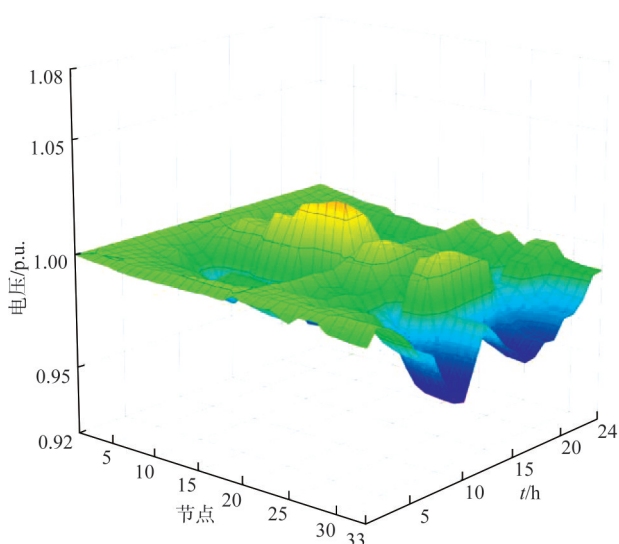


图7 方案5各个时段电压

Fig. 7 Voltage at each time in scheme 5

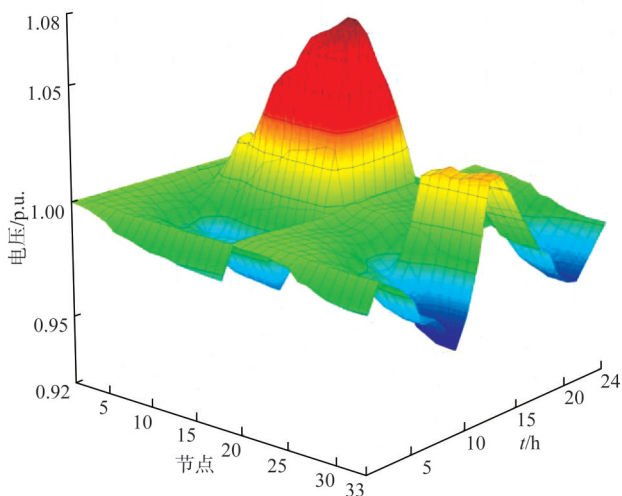


图6 方案1各个时段电压

Fig. 6 Voltages at each time in scheme 1

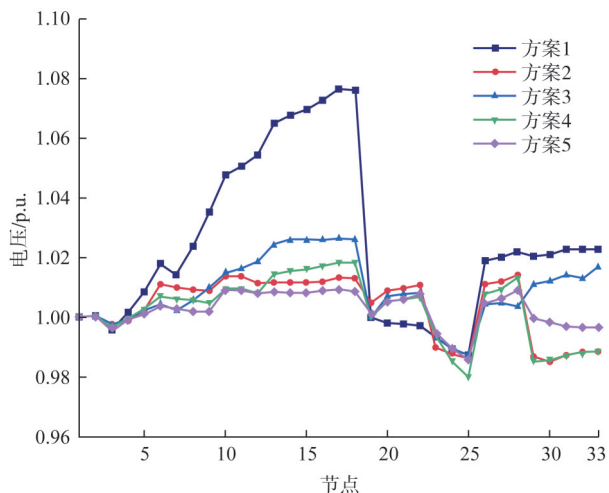


图8 14时各节点电压

Fig. 8 Each node voltages at 14h

措施时节点 11—18 的电压越限, 节点 17 的电压达到了 1.077 p. u., 电压严重越限, 在采用调压措施后各节点电压均在安全范围内, 且本文所提方法要比其他方法优化后的电压更加接近 1.0 p. u.。

在采用各方案优化后日均网损如图 9 所示。从图 9 可以看出在进行方案 2 的两阶段电压控制后, 相比没有进行任何优化的方案 1 网损下降了 70.3%, 可见在进行两阶段电压控制优化后能够大大降低网损。采用方案 4 优化后的网损相比方案 1 的网损下降了 71.9%; 采用方案 5 优化后的网损相比方案 1 的网损降低了 72.7%; 在不进行一阶段网络重构时方案 3 的网络相比方案 1 降低了 59.2%; 方案 5 在进行两阶段电压控制时的网损进一步降低

13.5%, 因此在一阶段采用网络重构能够进一步降低网损。

由于方案 1 没有采取任何电压控制策略, 因此其运行时间不做表述。从表 1 中可知, 将方案 2 的求解误差设为 10%, 需要 285.01 s 求解完成; 方案 3、4 和 5 的运行时间均小于 0.4 s, 可见基于人工智能算法的两阶段电压控制策略的在线运行速度远快于基于数学规划的两阶段电压控制策略的在线运行速度。

5 结论

针对分布式光伏接入主动配电网导致的电压越

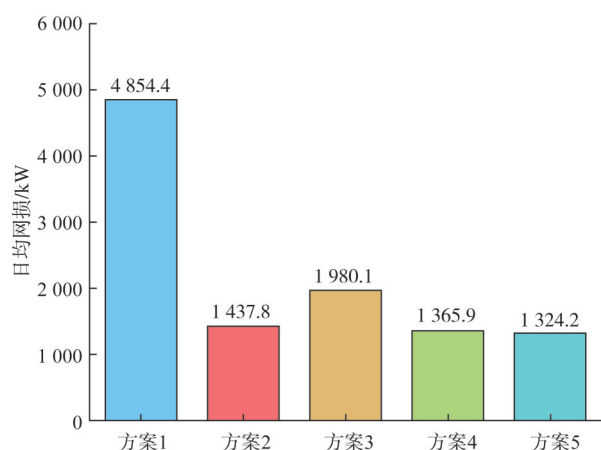


图9 日均网损

Fig. 9 Average daily line losses

表1 各方案运行时间

Tab. 1 Running time for each scenario

方案	平均运行时间/s
2	285.01
3	0.380 4
4	0.381 2
5	0.380 4

限问题,本文提出了基于两阶段网络重构和光储系统电压协调控制策略,能够在一定程度上解决电压越限问题,但是基于传统数学规划求解的两阶段电压协调控制策略运行求解时间慢、精度较低,因此本文提出了基于多智能体强化学习的两阶段电压控制策略,通过离线训练-在线运行的方法将漫长的求解时间转移到线下,用训练好的智能体对环境进行在线控制运行,相比于传统控制策略降低了运行时间并进一步降低了网络损耗。

本文所提出的基于PER MATD3算法训练的智能体有效解决了MADDPG算法训练时 Q 值过估和累计误差较大而造成优化结果较差等问题,能够进一步降低网损以及提高训练效率和稳定性。

参考文献

- [1] 王勋.“双碳”目标下新型电力系统发展路径[J].电力与能源, 2023,44(1):7-9.
WANG Xun. Development path of new power system based under the “Dual Carbon” goal[J]. Power & Energy, 2023, 44(1):7-9.
- [2] 张珍珍,吕清泉,张健美.“双碳”目标下分布式光伏发电技术的研究进展及展望[J].太阳能,2023,345(1):17-21.
ZHANG Zhenzhen, LÜ Qingquan, Zhang Jianmei, et al. Research progress and prospect of distributed PV power generation technology under the goal of emission peak and carbon neutrality[J]. Solar Energy, 2023, 345(1):17-21.
- [3] WECKX S, DRIESEN J. Optimal local reactive power control by PV inverters[J]. IEEE Transactions on Sustainable Energy, 2016,7(4):1624-1633.
- [4] 王颖,文福拴,赵波,等.高密度分布式光伏接入下电压越限问题的分析与对策[J].中国电机工程学报,2016,36(5):1200-1206.
WANG Ying, WEN Fujian, ZHAO Bo, et al. Analysis and countermeasures of voltage violation problems caused by high-density distributed photovoltaics[J]. Proceedings of the CSEE, 2016,36(5):1200-1206.
- [5] 王斌,张裕,罗晨.高比例光伏接入的配电网分布式运行优化技术[J].南方电网技术,2024,18(6):121-130.
WANG Bin, ZHANG Yu, LUO Chen. Distributed operation optimization technology for distribution network with high penetration of photovoltaics [J]. Southern Power System Technology, 2024, 18(6): 121-130.
- [6] 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 grid technology, 2018, 42(1):220-229.
- [7] 颜湘武,王晨光,贾焦心,等.基于小交流信号的低压配电网中分布式光伏辅助调压控制策略[J].中国电力,2022,55(5):134-142.
YAN Xiangwu, WANG Chenguang, JIA Jiaoxin, et al. Distributed PV auxiliary voltage control strategy in low voltage distribution network based on small AC signals [J]. Electric Power, 2022,55(5):134-142.
- [8] 李征,王文宾,韩胜峰,等.考虑无功支撑的分布式光伏接入配电网电压适应性研究[J].电力系统保护与控制,2022,50(11):32-41.
LI Zheng, WANG Wenbin, HAN Shengfeng, et al. Voltage adaptability of distributed photovoltaic access to a distribution network considering reactive power support [J]. Power System Protection and Control, 2022,50(11):32-41.
- [9] 刘洪波,高旭升,刘庸.基于改进人工鱼群算法的主动配电网日前两阶段优化调度[J].广东电力,2023,36(11):122-129.
LIU Hongbo, GAO Xusheng, LIU Yong. Two-stage day-ahead optimal dispatching of active distribution network based on improved artificial fish swarm algorithm [J]. Guangdong Electric Power, 2023, 36(11): 122-129.
- [10] 黄立滨,蔡海青,顾浩瀚,等.计及分布式光伏和储能主动支撑的配电网日前日内协调优化运行策略[J].南方电网技术,2024,18(8):51-60.
HUANG Libin, CAI Haiqing, GU Haohan, et al. Coordinated optimal strategy of day-ahead and intra-day operation of distribution network considering the active support of distributed photovoltaic and energy storage system [J]. Southern Power System Technology, 2024, 18(8):51-60.
- [11] LI Peng, JI Haoran, WANG Chengshan, et al. Coordinated control method of voltage and reactive power for active distribution networks based on soft open point. [J] IEEE Transaction Sustan Energy, 2017,8(4): 1430-1442.
- [12] LIN Baiquan, JIANG Tao, LI Fangxing, et al. Distributed energy storage planning in soft open point based active distribution networks incorporating network reconfiguration and DG

- reactive power capability [J]. Applied Energy, 2018 (210): 1082 – 1091.
- [13] 陈永进. 基于混合整数凸规划的有源配电网重构与无功电压协调优化[J]. 电力电容器与无功补偿, 2020, 41(6): 21 – 29.
CHEN Yongjin. A coordinated optimization for active distribution network reconfiguration and volt/var optimization based on mixed integer convex programming [J]. Power Capacitor & Reactive Power Compensation, 2020, 41(6): 21 – 29.
- [14] 刘洪, 徐正阳, 葛少云, 等. 考虑储能调节的主动配电网有功—无功协调运行与电压控制[J]. 电力系统自动化, 2019, 43(11): 51 – 58.
LIU Hong, XU Zhengyang, GE Shaoyun, et al. Coordinated operation of active-reactive power and voltage control for active distribution network considering regulation of energy storage [J]. Automation of Electric Power Systems, 2019, 43(11): 51 – 62.
- [15] CHIANG H D, WANG T, SHENG H. A novel fast and flexible holomorphic embedding power flow method [J]. IEEE Transactions on Power Systems, 2018, 33(3): 255 – 2562
- [16] WANG Yu, ZHAO Tianyang, JU Chengquan, et al. Two-level distributed voltage / var control using aggregated PV inverters in distribution networks [J]. IEEE Transactions on Power Delivery, 2020, 35(4): 1844 – 1855.
- [17] 姜涛, 张东辉, 李雪, 等. 含分布式光伏的主动配电网电压分布式优化控制[J]. 电力自动化设备, 2021, 41(9): 102 – 109, 125.
JIANG Tao, ZHANG Donghui, LI Xue, et al. Distributed optimal control of voltage in active distribution network with distributed photovoltaic [J]. Electric Power Automation Equipment, 2021, 41(9): 102 – 109, 125.
- [18] ZHANG Cuo, XU Yan. Hierarchically-coordinated voltage/var control of distribution networks using PV inverters [J]. IEEE Transactions on Smart Grid, 2020, 11(4): 2942 – 2953.
- [19] 庄慧敏, 张江林, 刘兴茂. 基于多时间尺度的主动配电网双层电压优化协调控制策略[J]. 南方电网技术, 2018, 12(6): 52 – 59.
ZHUANG Huimin, ZHANG Jianglin, LIU Xingmao. Two-level voltage optimal coordination control strategy of active distribution network based on multi timescale [J]. Southern Power System Technology, 2018, 12(6): 52 – 59.
- [20] 柴园园, 赵晓波, 吕超贤, 等. 基于 Fisher 时段划分的配电网源-网-荷-储多时间尺度协调优化调控策略[J]. 电网技术, 2024, 48(4): 1593 – 1606.
CHAI Yuanyuan, ZHAO Xiaobo, LÜ Chaoxian, et al. Coordinated multi-time scale optimal regulation strategy for source-grid-load-storage of distribution network based on fisher period division [J]. Power System echnology, 2024, 48(4): 1593 – 1606.
- [21] 张梓泉, 崔明建, 张程彬, 等. 基于反事实多智能体强化学习和有功无功协同控制的配电网电压优化[J]. 电力系统保护与控制, 2024, 52(18): 76 – 86.
ZHANG Zixiao, CUI Mingjian, ZHANG Chengbin, et al. Active and reactive power coordinated optimal voltage control of a distribution network based on counterfactual multi-agent reinforcement learning [J]. Power System Protection and Control, 2024, 52(18): 76 – 86.
- [22] 沈珺, 柳伟, 李虎成, 等. 基于强化学习的多微电网分布式二次优化控制[J]. 电力系统自动化, 2020, 44(5): 198 – 206.
SHEN Jun, LIU Wei, LI Hucheng, et al. Reinforcement Learning Based Distributed Secondary Optimal Control for Multiple Microgrids [J]. Automation of Electric Power Systems, 2020, 44(5): 198 – 206.
- [23] YANG Qiuling, WANG Gang, SUN Jian, et al. Two-timescale voltage control in distribution grids using deep reinforcement learning [J]. IEEE Transactions on Smart Grid 2020; 11(3): 2313 – 2323
- [24] 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.
- [25] 邓清唐, 胡丹尔, 蔡田田, 等. 基于多智能体深度强化学习的配电网无功优化策略[J]. 电工电能新技术, 2022, 41(2): 10 – 20.
DENG Qingtang, HU Daner, CAI Tiantian, et al. Reactive power optimization strategy of distribution network based on multi-agent deep reinforcement learning [J]. Advanced Technology of Electrical Engineering and Energy, 2022, 41(2): 10 – 20.

收稿日期: 2023-05-30; 网络首发日期: 2024-03-22

作者简介:

张涛(1998), 男, 硕士研究生, 研究方向为新能源并网, 1105563088@qq.com;

郝正航(1972), 男, 通信作者, 教授, 博士, 研究方向为电力系统稳定分析与控制, haozhenghang@163.com;

徐玉韬(1982), 男, 高级工程师, 硕士, 研究方向为交直流互联配电网及微电网, 95616048@qq.com。