

面向主动配电网的安全多智能体深度强化学习电压优化控制

梅铭洋¹, 寇鹏^{1,2,3}, 张智豪¹, 梁得亮^{1,2,3}

(1. 西安交通大学电气工程学院, 710049, 西安; 2. 西安交通大学陕西省智能电网重点实验室, 710049, 西安;
3. 西安交通大学电力设备电气绝缘国家重点实验室, 710049, 西安)

摘要: 针对主动配电网电压优化控制中模型不确定性和通信代价大的问题,提出了一种基于灵敏度矩阵安全的多智能体深度强化学习(SMS-MADRL)算法。该算法利用安全深度强化学习,应对主动配电网的固有不确定性,并采用多智能体结构实现通信代价较小的分布式控制。首先,将电压优化控制问题描述为受约束的马尔可夫博弈(CMG);然后,对无功功率进行适当修改,通过分析节点电压的变化得到灵敏度矩阵,进而与主动配电网环境进行交互,训练出若干可以独立给出最优无功功率指令的智能体。与现有多智能体深度强化学习算法相比,该算法的优点在于给智能体的动作网络增添了基于灵敏度矩阵的安全层,在智能体的训练和执行阶段保证了主动配电网的电压安全性。在IEEE 33节点系统上的仿真结果表明:所提出的算法不仅能够满足电压约束,而且相较于多智能体深度确定性策略梯度(MADDPG)算法,网络损耗减少了4.18%,控制代价减少了70.5%。该研究可为主动配电网的电压优化控制提供理论基础。

关键词: 主动配电网;电压优化控制;多智能体深度强化学习;安全深度强化学习

中图分类号: TM714.3 **文献标志码:** A

DOI: 10.7652/xjtuxb202312016 **文章编号:** 0253-987X(2023)12-0157-11

Safe Multi-Agent Deep Reinforcement Learning for Optimal Voltage Control in Active Distribution Networks

MEI Mingyang¹, KOU Peng^{1,2,3}, ZHANG Zhihao¹, LIANG Deliang^{1,2,3}

(1. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. Shaanxi Key Laboratory of Smart Grid, Xi'an Jiaotong University, Xi'an 710049, China;

3. State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In order to tackle the challenges arising from model uncertainty and high communication costs of optimal voltage control in active distribution networks, a sensitivity matrix based safe multi-agent deep reinforcement learning (SMS-MADRL) algorithm was proposed. The algorithm effectively addressed the inherent uncertainties associated with the active distribution networks by safe deep reinforcement learning and enabled the distributed control at lower communication costs through the multi-agent structure. In this paper, the optimal voltage control problem was formulated as a constrained Markov game (CMG). Then, the sensitivity matrix is derived by making appropriate adjustments to the reactive power and analyzing the voltage variations. Sub-

sequently, a set of agents that can independently give optimal reactive power commands were developed and trained via the interaction with the active distribution network. Notably, compared with the existing multi-agent deep reinforcement learning algorithms, the approach proposed herein adds a sensitivity matrix based safety layer to the action network of each agent and ensures the voltage safety of the active distribution network during the training and execution of the agents. The results of the simulation on the IEEE 33-bus system show that the proposed SMS-MADRL algorithm not only satisfies the voltage constraints, but also reduces the network power loss by 4.18% and the control costs by 70.5%, compared with the MADDPG algorithm. This study lays a theoretical foundation for the optimal voltage control in active distribution networks.

Keywords: active distribution network; optimal voltage control; multi-agent deep reinforcement learning; safe deep reinforcement learning

近年来,越来越多的新能源发电机组以分布式电源(distributed generator, DG)形式接入配电网,形成具有潮流双向流动特点的主动配电网^[1]。潮流双向流动给配电网运行的安全性和经济性带来了挑战,其中最为突出的就是电压波动和网络损耗问题^[2]。主动配电网内的各类型无功补偿设备为解决这些问题提供了硬件资源,但充分利用这些硬件资源的前提,是建立合理的电压优化控制方法,从而对各个无功补偿设备进行高效的协调控制^[3]。

按照被控对象的建模思路进行划分,现有主动配电网的电压优化控制策略可分为基于机理模型控制策略和无模型控制策略两大类。其中,基于机理模型的控制策略首先从物理背景出发,建立主动配电网的电压和损耗机理模型,然后利用模型预测控制^[4](model predictive control, MPC)、随机优化^[5]、启发式优化^[6]、二阶锥规划^[7]等方法,依托机理模型求解出各无功补偿设备的最优控制指令。不难发现,此类策略的效能严重依赖于配电网模型的准确性。然而,实际配电网的物理参数,尤其是线路参数,一般难以准确获得,相应的模型准确度也难以保证^[8],这给基于机理模型控制策略的应用带来了障碍。

与基于机理模型的控制策略不同,无模型控制策略不依赖于主动配电网的机理模型,而是利用数据交互构建电压控制策略。近年来广受关注的深度强化学习(deep reinforcement learning, DRL)算法便是无模型控制策略的突出代表,其通过智能体与环境的交互过程构建控制策略,可自适应被控对象参数的不确定性^[9]。现有 DRL 算法的研究中,已有将其应用于电力系统电压调节的报道。文献[10]建立了一种集中式深度强化学习算法,并通过控制并联电容器的投切实现了电压控制。文献[11]提出了一种多时间尺度配电网在线无功优化算法,针

对具有不同响应时间的多种无功补偿设备,设计双时间尺度分别对有载调压变压器和静止无功发生器进行优化控制。从控制系统架构角度看,以上算法皆属于单智能体架构,其集中式控制逻辑与配电网中无功补偿设备的分布式属性存在一定矛盾。对此,文献[12]提出了一种基于多智能体深度确定性策略梯度(multi-agent deep deterministic policy gradient, MADDPG)^[13]的电压控制算法,可实现各无功补偿设备的分散式控制,在提升电压控制容错的同时也降低了通信代价。

纵览以上 DRL 电压控制策略,尽管多智能体架构相较于单智能体能更好地适应主动配电网的分布式特性,但二者均存在明显的不足之处,即无法有效处理电压控制问题中的约束条件^[14-19]。而主动配电网电压控制的一个重要目标,就是将各节点电压维持在可行范围之内,属于典型的约束最优控制问题。对此,现有 DRL 策略大多采用奖励整定法^[20-23],将电压偏移作为惩罚项纳入智能体的奖励函数,从而间接地将电压约束条件融入控制策略设计。此类算法的局限性在于,电压偏移惩罚项所对应惩罚系数的整定极其依赖于经验,若惩罚系数过小,则难以保证满足电压约束,若惩罚系数过大,则会增大控制代价并恶化配电网损耗优化效果。

针对以上问题,本文提出了一种基于灵敏度矩阵的安全多智能体深度强化学习算法(sensitivity matrix based safe multi-agent deep reinforcement learning, SMS-MADRL),可实现主动配电网中各无功补偿设备的分布式协同优化控制。与现有算法相比,该算法的优点在于将安全层概念引入各智能体设计,并在设计中融入主动配电网的电压灵敏度矩阵,显式地处理各节点的电压约束,从机理上保证了主动配电网的电压安全性。

1 主动配电网电压优化控制问题

1.1 主动配电网结构

采用辐射状拓扑的修改后 IEEE 33 节点系统作为主动配电网研究案例^[24],搭建主动配电网环境如图 1 所示。系统中接入了分布式电源和无功补偿装置,分布式电源全部为永磁同步发电机组,无功补偿装置为混合式配电变压器(hybrid distribution transformer, HDT)^[25]。相较于传统变压器,HDT 变压器可以为主动配电网的中压侧提供电压调节能力^[4]。针对 HDT 变压器的分散式安装特点,根据安装位置可将主动配电网分为 3 个子网络。

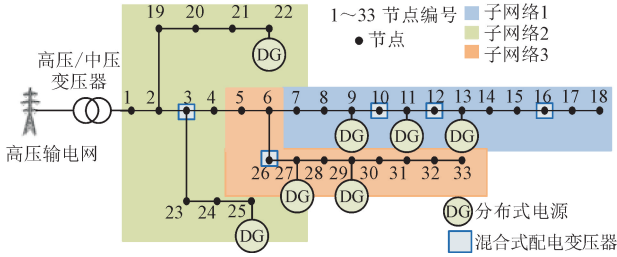


图 1 主动配电网环境

Fig. 1 The active distribution network environment

1.2 混合式配电变压器

所采用的配电网电压调节设备为混合式配电变压器 HDT,定义中压侧有功功率为 P_H ,无功功率为 Q_H ,并满足下列关系

$$P_H + jQ_H = -\dot{V}_M^2 Y_A^* + \dot{V}_M \dot{I}_C^* Y_B^* \quad (1)$$

$$P_H + jQ_H = -\dot{V}_M^2 (Y_C^* + Y_D^*) + \dot{V}_M \dot{V}_C Y_D^* \quad (2)$$

式中:带顶标的变量 $\dot{x}$ 表示 x 的正弦交流流量; $\dot{V}_M$ 为中压侧电压; $\dot{V}_C$ 为控制侧电压; $\dot{I}_C$ 为控制侧电流; $*$ 表示复共轭; Y_A^* 、 Y_B^* 、 Y_C^* 、 Y_D^* 分别为与阻抗相关的系数,其推导过程参见文献[4, 25]。

如图 2 所示,式(1)可以视为复平面内圆心为 $-\dot{V}_M^2 Y_A^*$,半径为 $\dot{V}_M \dot{I}_C |Y_B|$ 的圆。同样地,式(2)可

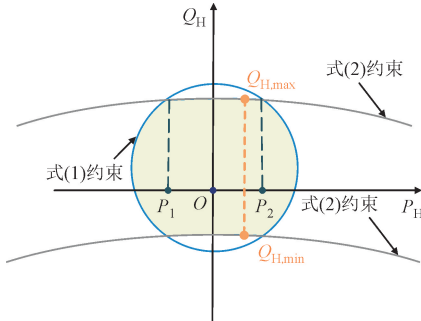


图 2 HDT 变压器的有功-无功功率范围

Fig. 2 Feasible range of the active power and reactive power of the HDT

以视为复平面内以 $\dot{V}_M^2 (Y_C^* + Y_D^*)$ 为圆心,以 $\dot{V}_M \dot{V}_C |Y_D^*|$ 为半径的圆。HDT 变压器的可控无功功率范围如图 2 中的浅绿色部分所示, P_1 、 P_2 是式(1)约束和式(2)约束的交点对应的有功功率。当 P_H 满足 $P_1 \leq P_H \leq P_2$ 时,无功功率的上限 $Q_{H,max}$ 和下限 $Q_{H,min}$ 可视为固定值,则 Q_H 的取值范围为

$$Q_{H,min} \leq Q_H \leq Q_{H,max} \quad (3)$$

1.3 电压优化控制问题表述

主动配电网电压优化控制的目标可归纳为:将各节点电压的波动范围限制在允许范围内,同时尽量减少网络损耗和控制代价。据此,评价指标 r_t 和优化目标 J_r 可表示为

$$r_t = - \left[\beta \frac{\Phi_{loss,t}}{\Phi_{loss}^*} + (1 - \beta) \frac{\Phi_{H,t}}{\Phi_H^*} \right] \frac{\Delta t}{T} \quad (4)$$

$$\max J_r = \sum_t \gamma^{t-1} r_t \quad (5)$$

式中: $\Phi_{loss,t}$ 为 t 时刻的网络损耗; $\Phi_{H,t}$ 为 t 时刻的控制代价; Φ_{loss}^* 和 Φ_H^* 分别为平均网络损耗和平均控制代价归一化因子; β 为加权系数, β 越大,则优化问题更倾向于最小化网络损耗,反之,则更倾向于最小化控制代价; Δt 为控制周期; T 为仿真总时长; γ 为折现系数,一般取接近 1 的值,体现了控制策略对更近时刻的重视程度。

由于主动配电网一般具有辐射状拓扑,故可用一个无向图 $\Pi(\mathcal{N}, \mathcal{E})$ 对其进行描述。其中, $\mathcal{N}$ 表示主动配电网各节点的集合, $\mathcal{N}_H \subset \mathcal{N}$ 表示安装有 HDT 变压器节点的集合, $(i, j) \in \mathcal{E}$ 表示节点 i 与节点 j 通过输电线相连。由此, t 时刻网络损耗 $\Phi_{loss,t}$ 和控制代价 $\Phi_{H,t}$ 的计算式可写为

$$\Phi_{loss,t} = \text{Re} \left(\sum_{(i,j) \in \mathcal{E}} (\dot{V}_{i,t} - \dot{V}_{j,t})^2 Y_{ij} \right) \quad (6)$$

$$\Phi_{H,t} = \sum_{i \in \mathcal{N}_H} |Q_{H,i,t}| \quad (7)$$

式中: $Q_{H,i,t}$ 为 t 时刻安装在节点 i 的 HDT 变压器的无功功率; Y_{ij} 为节点 i 与节点 j 间输电线的导纳; $\dot{V}_{i,t}$ 和 $\dot{V}_{j,t}$ 分别为 t 时刻节点 i 和 j 的电压相量,具体表达式可写为

$$\dot{V}_{i,t} = V_{i,t} \angle \theta_{i,t} \quad (8)$$

$$\dot{V}_{j,t} = V_{j,t} \angle \theta_{j,t} \quad (9)$$

式中: $V_{i,t}$ 和 $V_{j,t}$ 分别为 t 时刻节点 i 和节点 j 的电压幅值; $\angle \theta_{i,t}$ 和 $\angle \theta_{j,t}$ 分别为 t 时刻节点 i 和节点 j 的电压相位。

主动配电网运行过程中,需满足各节点功率限制、各节点电压限制、各 HDT 变压器功率限制等多项约束条件。主动配电网各节点的功率约束可写为

$$P_{i,t} - jQ_{i,t} = \dot{V}_{i,t}^* \sum_j \dot{V}_{j,t} Y_{ij} \quad \forall (i,j) \in \mathcal{E}, \forall t \quad (10)$$

式中: $P_{i,t}$ 和 $Q_{i,t}$ 分别为 t 时刻节点 i 注入的有功、无功功率; $\dot{V}_{i,t}^*$ 为 $\dot{V}_{i,t}$ 的共轭复数。

主动配电网各节点的电压限制和各 HDT 变压器的无功功率限制可分别表示为

$$V_{\min} \leq V_{i,t} \leq V_{\max}, \quad \forall i \in \mathcal{N}, \forall t \quad (11)$$

$$Q_{H,\min} \leq Q_{H,i,t} \leq Q_{H,\max}, \quad \forall i \in \mathcal{N}_H, \forall t \quad (12)$$

式中: $V_{\max}$ 和 $V_{\min}$ 分别为节点电压幅值的上限和下限; $Q_{H,\max}$ 和 $Q_{H,\min}$ 分别为 HDT 无功功率的上限和下限。

综合式(4)~(10), 主动配电网电压优化控制问题可表述为

$$\begin{cases} \max J_r = \sum_t \gamma^{t-1} r_t \\ \text{s. t. } P_{i,t} - jQ_{i,t} = \dot{V}_{i,t}^* \sum_j \dot{V}_{j,t} Y_{ij}, \quad \forall (i,j) \in \mathcal{E}, \forall t \\ V_{\min} \leq V_{i,t} \leq V_{\max}, \quad \forall i \in \mathcal{N}, \forall t \\ Q_{H,\min} \leq Q_{H,i,t} \leq Q_{H,\max}, \quad \forall i \in \mathcal{N}_H, \forall t \end{cases} \quad (13)$$

1.4 采用受约束的马尔可夫博弈表述的主动配电网电压优化控制问题

如 1.1 小节所述, 在多智能体深度强化学习的框架下, 主动配电网可根据 HDT 变压器的安装位置分为若干子网络, 每个子网络配置一个智能体来控制 HDT 变压器的无功功率^[26]。相应地, 由式(13)表述的主动配电网集中式优化控制问题, 也可被分解为若干并行的子问题, 由所对应子网络中的智能体单独求解, 各智能体通过通讯协调无功功率, 以保证满足约束条件。

基于以上多智能体框架, 主动配电网的电压优化控制问题可以被描述为包含 N 个智能体的受约束的马尔可夫博弈 (constrained Markov game, CMG)^[27]。CMG 将马尔可夫博弈问题延伸为带约束的形式。采用下标 n 表示属于第 n 个子网络的物理量, $i \in \mathcal{N}_n$ 表示属于第 n 个子网络的节点, 则 CMG 可以描述为如下的 7 元组

$$(\mathbf{S}, \{\mathbf{O}_n\}, \{\mathbf{A}_n\}, T, \{R_n\}, C, \gamma), n = 1, 2, \dots, N \quad (14)$$

7 元组各集合或函数的定义描述如下。

$\mathbf{S}$ 为主动配电网的状态集合。状态 $s \in \mathbf{S}$, 且 $s = [\mathbf{P}_L^T, \mathbf{Q}_L^T, \mathbf{P}_G^T, \mathbf{Q}_G^T, \mathbf{V}^T, \boldsymbol{\theta}^T]^T$, 其中向量 $\mathbf{P}_L, \mathbf{Q}_L$ 分别为各节点的有功、无功负荷, 向量 $\mathbf{P}_G, \mathbf{Q}_G$ 分别为各节点分布式电源的有功、无功功率, 向量 $\mathbf{V}, \boldsymbol{\theta}$ 分别为各节点电压的幅值、相位。

$\{\mathbf{O}_n\} = \{\mathbf{O}_1, \dots, \mathbf{O}_N\}$ 为所有智能体的观察集合, $\mathbf{O}_n$ 为第 n 个智能体所有可能获得的观察集合, 观察 $\mathbf{o}_n \in \mathbf{O}_n$, 且 $\mathbf{o}_n = [\mathbf{P}_{L,n}^T, \mathbf{Q}_{L,n}^T, \mathbf{P}_{G,n}^T, \mathbf{Q}_{G,n}^T]^T$, 即每个智能体的观察为其所在子网络有功、无功功率组成的向量。

$\{\mathbf{A}_n\} = \{\mathbf{A}_1, \dots, \mathbf{A}_N\}$ 为所有智能体的动作集合, $\mathbf{A}_n$ 为第 n 个智能体所有可能动作的集合, 动作 $\mathbf{a}_n \in \mathbf{A}_n$ 且 $\mathbf{a}_n = \mathbf{Q}_{H,n}$, 其中向量 $\mathbf{Q}_{H,n}$ 为第 n 个子网络内所有 HDT 变压器的无功功率输出。

T 为状态转移函数, 表示主动配电网中各智能体在状态 s 下执行动作 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ 后下一时刻状态的概率分布。

$\{R_n\} = \{R_1, \dots, R_N\}$ 为所有智能体的奖励集合, R_n 为第 n 个智能体所有可能获得的奖励集合, $r_n \in R_n$ 为第 n 个智能体的奖励。值得注意的是, 式(4)中传统优化问题的评价指标 r_t , 在 DRL 算法中被称为奖励 r_n ; 同样地, 式(5)中的目标函数 J_r 在 DRL 算法中被称为回报 $J_{r,n}$ 。

$C = \{c_j(s, \mathbf{a}), j = 1, 2, \dots, K\}$ 为约束函数的集合, 其中 K 为约束函数的数量, 每一个约束函数都存在上界值 $c_{j,\max}$, 满足

$$c_j(s, \mathbf{a}) \leq c_{j,\max}, \quad j = 1, 2, \dots, K \quad (15)$$

在主动配电网电压优化控制的背景下, 式(10)表示的节点功率平衡约束包含在主动配电网环境中。因此, CMG 问题只需处理式(11)中的电压约束和式(12)中的无功功率约束, 约束可改写为如下的向量形式

$$\begin{cases} \mathbf{V} + \mathbf{S}_m \Delta \mathbf{Q}_H \leq (\mathbf{V}_{\max} - \boldsymbol{\xi}_1) \\ \mathbf{V} + \mathbf{S}_m \Delta \mathbf{Q}_H \geq (\mathbf{V}_{\min} + \boldsymbol{\xi}_2) \\ Q_{H,\min} \leq Q_H + \Delta Q_H \leq Q_{H,\max} \end{cases} \quad (16)$$

式中: $\mathbf{S}_m$ 为灵敏度矩阵; $\boldsymbol{\xi}_1$ 和 $\boldsymbol{\xi}_2$ 为紧缩项; $\Delta \mathbf{Q}_H$ 为相邻时刻 HDT 变压器无功输出的变化量。

根据上述定义, 式(13)中的主动配电网集中式优化控制问题可以重新表述为式(17)中的 CMG 问题, 即每个智能体的目标为: 在满足主动配电网约束的情况下, 将自身的回报最大化。

$$\begin{cases} \max J_{r,n} = \sum_t \gamma^{t-1} r_t, \quad n = 1, 2, \dots, N \\ \text{s. t. } c_j(s, \mathbf{a}) \leq c_{j,\max}, \quad j = 1, 2, \dots, K \end{cases} \quad (17)$$

式(15)所示的 CMG 问题可通过下一节中介绍的 SMS-MADRL 算法高效求解。

2 SMS-MADRL 算法

2.1 SMS-MADRL 算法原理

为高效求解式(15)所示的 CMG 问题, 本小节提出了 SMS-MADRL 算法。SMS-MADRL 算法是

MADDPG 算法的延伸,是一种采用集中式训练、分布式执行范式的深度强化学习算法。通过该算法,可以训练出若干独立智能体,能够基于本地观测信息独立地给出最优控制动作。与 MADDPG 算法相比,SMS-MADRL 算法的一个显著优势是包含基于灵敏度矩阵的安全层,可保证训练和执行过程中的电压安全性。

SMS-MADRL 算法的流程如图 3 所示,主动配电网环境在 1.1 小节已经介绍过,每一个智能体包含一组评价网络和一组动作网络,两者均由深度神经网络构成。动作网络的输入为第 n 个智能体自身观察 $\mathbf{o}_n$,输出为第 n 个智能体的原始动作 $\mathbf{a}_n$ 。评价网络通过输入所有智能体的观察 $[\mathbf{o}_1^T, \dots, \mathbf{o}_N^T]^T$ 和动作 $[\mathbf{a}_1^T, \dots, \mathbf{a}_N^T]^T$ 估计折现回报 $J_{r,n}$,其估计值记为 Q_n 。

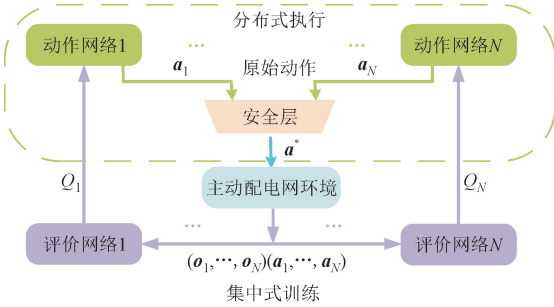


图 3 SMS-MADRL 算法原理图

Fig. 3 Block diagram of the SMS-MADRL algorithm

在集中式训练阶段,各个智能体将不断衰减的高斯噪声添加给动作网络所输出的动作,由此实现环境探索。安全层保证了探索阶段配电网的各个节点电压不会超出约束范围,在该阶段,智能体通过与主动配电网环境的交互获得数据,并不断更新动作网络和评价网络的参数。

在分布式执行阶段,评价网络将被移除。多个智能体共享一个集中式的安全层,由其收集所有智能体的原始动作 $\mathbf{a} = [\mathbf{a}_1^T, \dots, \mathbf{a}_N^T]^T$ 。安全层在不违反电压约束的情况下,对 $\mathbf{a}$ 进行最小程度地修改,得到动作 $\mathbf{a}^*$ 并在主动配电网中实际执行。

2.2 SMS-MADRL 算法训练过程

图 4 更加详细地展示了智能体的训练过程。可见,训练过程由若干个循环构成,每个循环又包含若干控制周期。图 4 的下方为时间轴,每个控制周期包含智能体与主动配电网环境交互,以及动作网络和评价网络更新两个阶段。在智能体与主动配电网

环境交互阶段,所有智能体的动作网络 μ 根据各自的观察 $\mathbf{o}$ 输出动作 $\mathbf{a}$,在对 $\mathbf{a}$ 添加高斯噪声 $\boldsymbol{\varepsilon}_1$ 后,经过安全层处理得到在主动配电网环境中执行的动作 $\mathbf{a}^*$ 。安全层可以保证训练过程满足各节点电压约束。配电网进入下一个状态后,智能体的观察更新为 $\mathbf{o}'$ 。接下来,智能体的观察 $\mathbf{x} = [\mathbf{o}_1^T, \dots, \mathbf{o}_N^T]^T$ 、动作 $\mathbf{a}$ 、奖励 $\mathbf{r} = [r_1, \dots, r_N]^T$ 、下一时刻的观察 $\mathbf{x}'$,将合成一组数据 $(\mathbf{x}, \mathbf{a}, \mathbf{r}, \mathbf{x}')$ 被存储在记忆回放 $\mathcal{B}$ 当中,用于评价网络和动作网络的参数更新。在动作网络和评价网络更新阶段,每一次与环境交互后,算法都会从记忆回放 $\mathcal{B}$ 中随机抽取 M 组数据 $(\mathbf{x}^j, \mathbf{a}^j, \mathbf{r}^j, \mathbf{x}'^j)$, $j=1, 2, \dots, M$,用于智能体参数的更新。

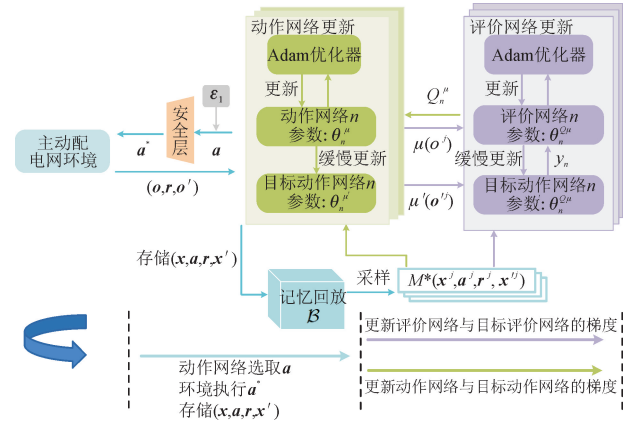


图 4 智能体的训练过程示意图

Fig. 4 Block diagram of the training process for agents

将每个智能体的动作网络 μ_n 和评价网络 Q_n^μ 的参数分别记为 θ_n^μ 和 $\theta_n^{Q^\mu}$,根据下式更新

$$\begin{aligned} \nabla_{\theta_n^\mu} J_{r,n}(\mu_n) &= E_{(\mathbf{x}, \mathbf{a}) \sim \mathcal{B}} \cdot \\ &[\nabla_{\theta_n^\mu} \mu_n(\mathbf{a}_n | \mathbf{o}_n) \nabla_{\theta_n^{Q^\mu}} Q_n^\mu(\mathbf{x}, \mathbf{a}_1, \dots, \mathbf{a}_N) |_{\mathbf{a}_n = \mu_n(\mathbf{o}_n)}] \end{aligned} \quad (18)$$

式中: $\nabla_{\theta_n^\mu} J_{r,n}(\mu_n)$ 表示动作网络为 μ_n 时获得的回报对于动作网络参数 θ_n^μ 的梯度; $\nabla_{\theta_n^\mu} \mu_n(\mathbf{a}_n | \mathbf{o}_n)$ 为动作网络的输出 $\mu_n(\mathbf{a}_n | \mathbf{o}_n)$ 对于动作网络参数 θ_n^μ 的梯度; $\nabla_{\theta_n^{Q^\mu}} Q_n^\mu(\mathbf{x}, \mathbf{a}_1, \dots, \mathbf{a}_N)$ 为评价网络的输出 $Q_n^\mu(\mathbf{x}, \mathbf{a}_1, \dots, \mathbf{a}_N)$ 对于第 n 个智能体的动作 $\mathbf{a}_n$ 的梯度; $E_{(\mathbf{x}, \mathbf{a}) \sim \mathcal{B}}$ 表示从记忆回放 $\mathcal{B}$ 中采样数据 $(\mathbf{x}, \mathbf{a})$ 中得到的数学期望。

评价网络通过最小化损失函数进行训练,损失函数 L_n 如下所示

$$L_n = E_{\mathbf{x}, \mathbf{a}, \mathbf{r}, \mathbf{x}'} [(y_n - Q_n^\mu(\mathbf{x}, \mathbf{a}_1, \dots, \mathbf{a}_N))^2] \quad (19)$$

$$y_n = r_n + \gamma Q_n^\mu(\mathbf{x}', \mathbf{a}'_1, \dots, \mathbf{a}'_N) |_{\mathbf{a}'_n = \mu_n(\mathbf{o}'_n)} \quad (20)$$

式中: y_n 为第 n 个智能体对抽取的 M 组数据回报

的估计值; μ'_n 与 Q'_n 为一组以较慢的速度更新的目标网络,其参数分别记为 θ'_n 和 $\theta^{Q'}_n$; a'^j_n 为第 n 个智能体的目标动作网络在接收到观察 o'^j_n 后所给出的动作。

目标网络 μ'_n 与 Q'_n 的参数按照如下表达式进行缓慢更新

$$\theta'_n \leftarrow \tau \theta'_n + (1 - \tau) \theta_n \quad (21)$$

$$\theta^{Q'}_n \leftarrow \tau \theta^{Q'}_n + (1 - \tau) \theta^{Q_n} \quad (22)$$

式中: τ 为目标网络更新率。

2.3 处理约束的安全层方法

虽然 MADDPG 算法可以通过与主动配电网环境的复杂交互自适应环境的不确定性,但在处理约束条件时却面临困境^[14-19]。在应对约束时,现有研究中惯用的奖励整定法是将电压偏移作为附加惩罚项加入到奖励当中。然而,惩罚系数的整定极其依赖于经验,若惩罚系数过小,则难以保证电压约束的满足;反之若惩罚系数过大,则会增大控制代价并恶化配网损耗优化效果^[20-23]。

针对此问题,本文提出在 MADDPG 算法中加入基于灵敏度矩阵的安全层来处理约束条件,算法的具体结构如图 5 所示。MADDPG 算法的动作网络在接收到智能体的观察后,所输出的原始动作 a 并不直接用于主动配电网控制,而是先经过一个安全层收集并进行最小程度的校正,之后再传递给各无功调节设备具体执行。

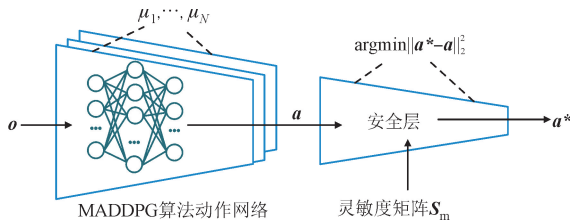


图5 处理约束的安全层示意图

Fig. 5 Block diagram of a safety layer for handling constraints

考虑式(17)中的约束 $c_j(s, a)$, 将其对于智能体的动作线性化^[28], 得到

$$c_j(s, a) = c_j(s) + g_j^T(\omega_j; a) \Delta a, \quad j = 1, 2, \dots, K \quad (23)$$

式中: $c_j(s)$ 为关于状态的约束函数; $g_j^T(\omega_j; a)$ 为线性映射函数组成的向量, 该向量与智能体动作 a 具有相同维度; Δa 为相邻时刻 HDT 变压器无功功率的变化量, 可写为

$$\Delta a = \Delta Q_H = [\Delta Q_{H,1}, \dots, \Delta Q_{H,N_H}]^T \quad (24)$$

式中: N_H 为 HDT 变压器的数量。一般来说, $g_j^T(\omega_j; a)$ 可通过独立的神经网络进行拟合, 但这意味着对于每一个约束都要单独设置一个神经网络^[29], 若用于求解含有几十、上百个电压约束的主动配电网电压优化控制问题, 存在计算量大、容易过拟合的缺陷。为此, 采用电压灵敏度矩阵 S_m 来表示约束函数 $c_j(s, a)$ 和 Δa 之间的线性关系, 可得

$$S_m = \begin{bmatrix} \frac{\partial V_1}{\partial Q_{H,1}} & \cdots & \frac{\partial V_1}{\partial Q_{H,N_H}} \\ \vdots & & \vdots \\ \frac{\partial V_{N_B}}{\partial Q_{H,1}} & \cdots & \frac{\partial V_{N_B}}{\partial Q_{H,N_H}} \end{bmatrix}_{N_B \times N_H} \quad (25)$$

式中: N_B 为主动配电网节点数量。

灵敏度矩阵 S_m 中的每一个元素, 都包含一个特定节点电压对于 HDT 变压器无功功率的偏导数。灵敏度矩阵的获取方式主要有两种, 一种是对潮流计算的雅可比矩阵取倒数来获得整个灵敏度矩阵^[28]; 另一种是在离线潮流计算过程中, 每次对 HDT 变压器输出的无功功率作微小修改, 通过分析节点电压的变化来逐一获得灵敏度矩阵的每一列^[4]。

在本文采用的修改后的 IEEE 33 节点算例中, HDT 变压器分别安装在 3、10、12、16、26 号节点, 节点电压对于 HDT 变压器无功功率 Q_H 的偏导等于节点电压对于 HDT 变压器安装节点 i 的无功功率 Q_i 的偏导, 故式(25)中灵敏度矩阵 S_m 的具体信息为

$$S_m = \begin{bmatrix} \frac{\partial V_1}{\partial Q_3} & \frac{\partial V_1}{\partial Q_{10}} & \frac{\partial V_1}{\partial Q_{12}} & \frac{\partial V_1}{\partial Q_{16}} & \frac{\partial V_1}{\partial Q_{26}} \\ \frac{\partial V_2}{\partial Q_3} & \frac{\partial V_2}{\partial Q_{10}} & \frac{\partial V_2}{\partial Q_{12}} & \frac{\partial V_2}{\partial Q_{16}} & \frac{\partial V_2}{\partial Q_{26}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial V_{33}}{\partial Q_3} & \frac{\partial V_{33}}{\partial Q_{10}} & \frac{\partial V_{33}}{\partial Q_{12}} & \frac{\partial V_{33}}{\partial Q_{16}} & \frac{\partial V_{33}}{\partial Q_{26}} \end{bmatrix}_{33 \times 5} \quad (26)$$

安全层在不违反电压约束的前提下, 对智能体所给出的动作 a 进行最小程度的校正, 得到修改后的动作 a^* 并在主动配电网中执行。因此, 安全层所求解的问题可以写为带线性约束的二次规划向量形式

$$\begin{cases} a^* = \underset{a^*}{\operatorname{argmin}} ||a^* - a||_2 \\ \text{s. t. } \mathbf{V} + S_m \Delta Q_H \leq (\mathbf{V}_{\max} - \xi_1) \\ \mathbf{V} + S_m \Delta Q_H \geq (\mathbf{V}_{\min} + \xi_2) \\ Q_{H,\min} \leq Q_H + \Delta Q_H \leq Q_{H,\max} \end{cases} \quad (27)$$

在式(27)所示的优化问题中,随着主动配电网工作状态的波动,电压灵敏度矩阵 $\mathbf{S}_m$ 存在变化的可能。对此,可通过设置紧缩变量 ξ_1 和 ξ_2 ,对节点电压约束进行紧缩来解决这一问题。含紧缩变量的问题仍是一个标准的二次规划问题,现有的求解方法已非常成熟。

综上所述,在主动配电网电压优化控制问题背景下,SMS-MADRL 的具体实现可归纳为算法 1。

算法 1:主动配电网的 SMS-MADRL 算法电压优化控制策略

输入:电压灵敏度矩阵 $\mathbf{S}_m$, 算法超参数

输出:智能体的目标动作网络参数 θ^i

1. 初始化 N 个动作网络 μ_n 和评价网络 Q_n^{π}
2. 初始化 N 个延迟更新的目标网络 μ'_n 与 $Q_n^{\pi'}$
3. 初始化记忆库 $\mathcal{B}$,初始化交互环境
4. for $i \leftarrow 1$ to N_f do: // N_f 为总训练次数
5. 动作网络输出并添加噪声 $\mathbf{a} + \boldsymbol{\varepsilon}_1$
6. 安全层根据式(27)修改动作,得到无功补偿方案 $\mathbf{a}^*$
7. 执行该无功补偿方案,并获得主动配电网环境的奖励 r 和环境新的状态 $\mathbf{s}'$
8. 将交互数据 $(\mathbf{x}, \mathbf{a}, r, \mathbf{x}')$ 储存在记忆库 $\mathcal{B}$ 中
9. for $n \leftarrow 1$ to N do:
10. 从记忆库中随机抽取 M 组交互数据 $(\mathbf{x}^j, \mathbf{a}^j, r^j, \mathbf{x}'^j)$, $j = 1, 2, \dots, M$
11. 根据式(18)更新动作网络
12. 根据式(19)~(20)更新评价网络
13. 按照式(21)~(22)更新目标网络
14. end for
15. if $t = T$ then:
16. 重新初始化交互环境
17. end if
18. end for

3 算例仿真

以修改后的 IEEE 33 节点系统为主动配电网案例,对所提出的 SMS-MADRL 算法进行仿真验证。为检验 SMS-MADRL 算法的优越性,选择基于机理模型的 MPC 策略,以及文献 [21] 中采用奖励整定法处理约束的 MADDPG 算法作为对比算法。算法参数如表 1 所示,控制器参数设置如表 2 所示。

表 1 SMS-MADRL 算法超参数	
Table 1 The SMS-MADRL algorithm hyperparameters	
参数	数值
动作网络隐藏层尺寸	[128,128,128]
动作网络学习率	1×10^{-5}
评价网络隐藏层尺寸	[256,256,256]
评价网络学习率	2×10^{-5}
目标网络更新率	0.001
记忆回放容量/组	2×10^5
采样容量/组	256
折现系数	0.95
探索噪声/dB	$-20 \sim -26$

表 2 控制器参数设置	
Table 2 The controller parameter settings	
参数	数值
无功功率约束范围/Mvar	± 0.4
电压约束	1 ± 0.05
总约束数量 K	76
电压约束紧缩项 ξ	0.002
控制周期 $\Delta t/s$	10
仿真总时长 T/h	24
平均网络损耗归一化因子 $\Phi_{\text{loss}}^*/\text{kW}$	35
平均控制代价归一化因子 $\Phi_{\text{H}}^*/\text{Mvar}$	0.25
加权项 β	0.9

3.1 算例介绍与分区

主动配电网案例即修改后的 IEEE 33 节点系统分区如前文中的图 1 所示,依据分布式电源和 HDT 变压器的安装位置分为 3 个子网络,每个子网络的 HDT 变压器各由一个智能体控制。

图 6 给出了 IEEE 33 节点系统中 24 h 有功功率和无功功率的负荷曲线,其数据来源于陕西电网某区域。为了使每个节点的负荷不同,各节点的负荷依照 IEEE 33 节点系统的原始数据乘以相应的系数。电压、有功负荷和无功负荷同样被转换为了标么值,基准功率为 1 MW,基准电压为 12.66 kV。其中,1 号节点为平衡节点,不施加有功和无功负荷。

各分布式电源为 2 MW 永磁同步风电机组,均运行于最大功率跟踪方式,其 24 h 功率曲线如图 7 所示,功率因数取值为 0.9,风速数据来源于美国国家可再生能源实验室^[30]。

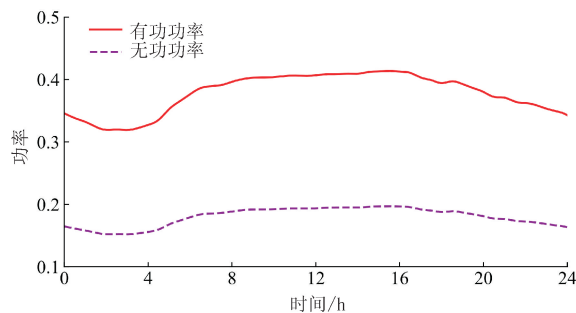


图 6 节点的 24 h 负荷曲线
Fig. 6 24-hour load profile of a node

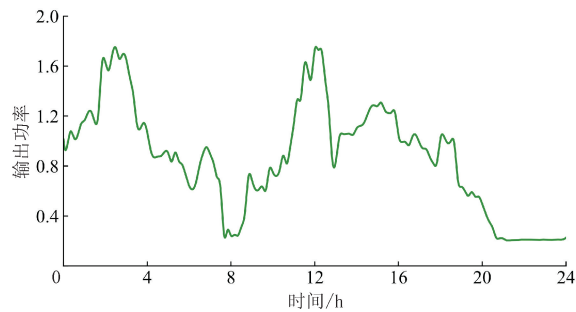


图 7 单台分布式电源 24 h 输出功率曲线
Fig. 7 24-hour output power profile of a DG unit

在未投入任何 HDT 变压器的情况下,对主动配电网电压进行 24 h 仿真,仿真步长为 10 s,结果如图 8 所示,图中,每条曲线对应一个节点电压的变化。可以看到,电压峰值出现在两个时段,分别为凌晨 2:00 和正午 12:00 左右,此时风电机组的输出功率较大,主动配电网向上级输电网馈送功率。在离风机最近的 9~18 号节点,电压上升尤为明显。这一现象表明,在 2:00 和 12:00 时段附近,1 号子网络的无功过剩最严重。在上午 8:00 和晚上 21:00 时段附近,负荷需求很高而风机出力较小,配电网从上级输电网吸收功率,因此出现了电压低于电压下

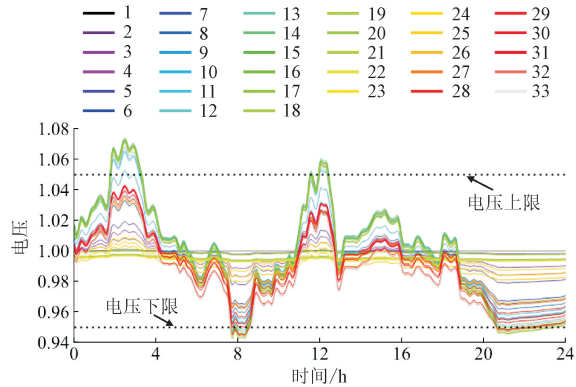


图 8 不施加无功补偿下各节点 24 h 电压曲线
Fig. 8 Daily voltage profiles without reactive power compensation

限的现象。

3.2 仿真结果分析

本节对比分析了所提出的 SMS-MADRL 算法,以及作为对照的 MPC 算法、采用奖励整定法处理电压约束的 MADDPG 算法的控制效果。

主动配电网电压优化控制的首要目标是维持所有节点电压在约束范围内,并在此基础上尽量减少网络损耗和控制代价。因此,在本节中,控制效果的对比包含是否满足电压约束、24 h 平均网络损耗以及 24 h 平均控制代价 3 个方面。

图 9 展示了 SMS-MADRL 算法控制下的 24 h 电压仿真结果,可以看出,在安全层的作用下,电压被控制在了 0.95~1.05 范围内,没有出现电压越限情况。

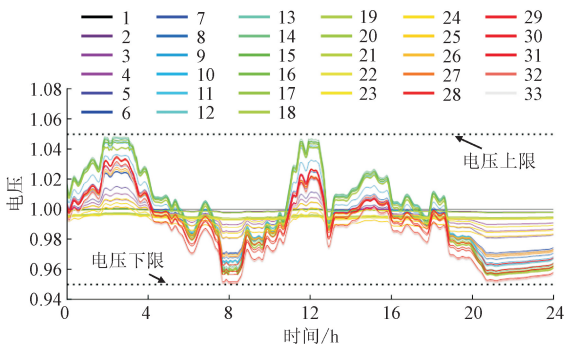


图 9 SMS-MADRL 算法控制下各节点 24 h 电压变化曲线
Fig. 9 Voltage profiles under the control of SMS-MADRL

图 10 展示了 MADDPG 算法控制下的 24 h 电压变化曲线,可见当电压惩罚系数较小时,即凌晨 2:00 左右,节点 10~18 的电压出现了违反电压约束上限的情况。这是由于惩罚系数过小,智能体在训练过程中更加关注网络损耗和控制代价,而忽视了对电压的控制。

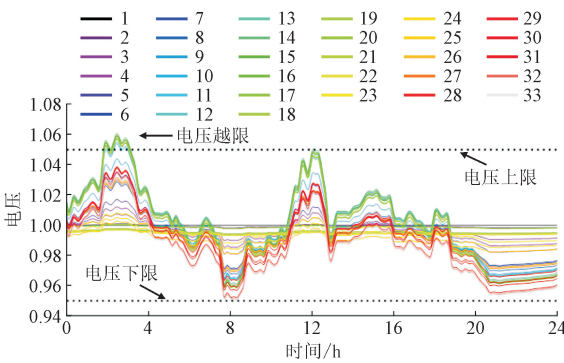


图 10 MADDPG(惩罚系数较小)算法控制下各节点 24 h 电压变化曲线
Fig. 10 Voltage profiles under the control of MADDPG (small penalty coefficient)

3 种控制方式下,各节点所安装 HDT 变压器无功功率的变化情况如图 11 所示。可以看出,尽管采用较大惩罚系数的 MADDPG 算法可以满足电压约束,但在相同时刻相较于 SMS-MADRL 算法需要更大的无功功率,经济性不佳。

表 3 给出了不同算法控制下平均网络损耗、电压约束和平均控制代价的对比结果。可以看出,主动配电网在不施加任何无功补偿的状态下平均网络损耗最大,采用较小惩罚系数的 MADDPG 虽然取得了最低的网络损耗,但由于惩罚系数过小,出现了电压越限现象。所提出的 SMS-MADRL 算法相较于惩罚系数较大的 MADDPG 算法,网络损耗减少了 4.18%,控制代价减少了 70.5%。

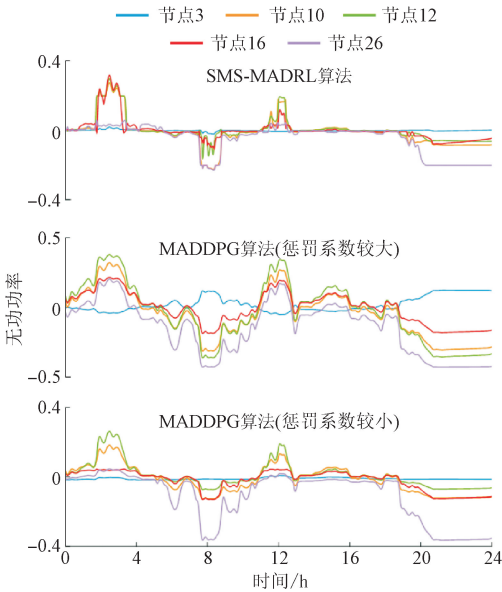


图 11 不同算法控制下 HDT 变压器无功功率 24 h 变化曲线
Fig. 11 HDT reactive power profiles under the control of different algorithms

表 4 给出了优化目标中不同加权系数 β 对 SMS-MADRL 算法电压优化控制结果的影响。由表 4 可见,其总体趋势为 β 越大,平均网络损耗越小,平均控制代价越大。由此可知,网络损耗受 β 变化的影响较小,而控制代价受 β 变化的影响更加显著。

表 4 SMS-MADRL 算法控制下不同加权系数 β 对电压优化结果的影响

Table 4 Results under the control of SMS-MADRL with different weighting factors β		
β	平均网络损耗/kW	平均控制代价/Mvar
0.10	35.01	0.143 7
0.30	35.06	0.150 0
0.50	34.76	0.173 4
0.70	34.81	0.181 8
0.90	33.89	0.188 6
0.95	33.79	0.391 0

综上所述,相较于采用奖励整定法处理约束的 MADDPG 算法,所提出的 SMS-MADRL 算法无论在网络损耗还是控制代价方面均具有明显的优越性。

3.3 HDT 变压器无功容量缩减的情况

在 3.2 节的仿真设置中,HDT 变压器的无功容量较为充裕,然而,随着电力电子设备的老化,HDT 变压器可能出现无功容量缩减的情况,其对电压优化控制问题的影响主要体现在无功约束的收紧。假设其它参数不变,仅将安装在节点 12 的 HDT 变压器无功容量从 ± 0.4 Mvar 下降至 ± 0.2 Mvar。图 12 给出了节点 12 和 16 的 HDT 变压器在节点 12 的 HDT 变压器无功功率受限情况下功率的变化情况。

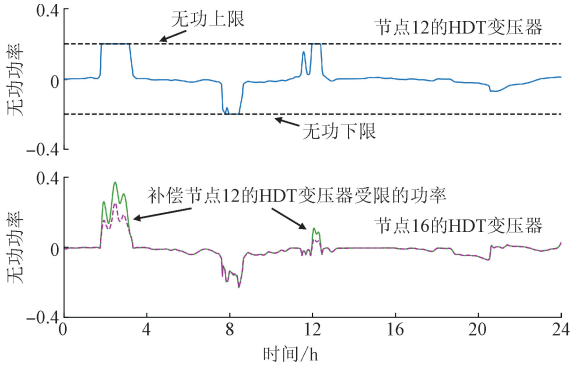


图 12 节点 12 无功功率受限时节点 12 和 16 的 HDT 变压器的无功功率
Fig. 12 HDT reactive power profiles at node 12 and 16 when reactive power is limited at node 12

表 3 不同算法控制下的性能指标对比结果 Table 3 Results under the control of different algorithms			
算法	平均网络 损耗/kW	平均控制 代价/Mvar	电压 越限
MADDPG (惩罚系数较小)	32.62	0.286 9	出现
MADDPG (惩罚系数较大)	35.37	0.639 8	未出现
SMS-MADRL	33.89	0.188 6	未出现
MPC	35.25	0.094 6	未出现
不施加任何无功补偿	37.42		出现

况。由图可见, SMS-MADRL 算法可以有效地处理无功功率约束, 使得安装在节点 12 的 HDT 变压器无功功率被限制在约束范围内, 有效地阻止了 HDT 变压器的过载。同时, 为了弥补节点 12 的 HDT 变压器功率受限对电压的影响, 在安全层的作用下, 节点 16 的 HDT 变压器无功功率相应增大。

4 结论与展望

针对主动配电网电压优化控制中模型不确定性和通讯代价大的问题, 提出了 SMS-MADRL 算法, 该算法的优点在于给智能体的动作网络增添基于灵敏度矩阵的安全层, 从而显式地处理各节点的电压约束。通过理论分析和仿真计算, 得到以下结论。

(1) 所提出的主动配电网 SMS-MADRL 电压优化控制算法, 可有效处理电压约束和无功功率约束, 相较于采用奖励整定法的 MADDPG 算法, 网络损耗减少了 4.18%, 控制代价减少了 70.5%。

(2) 采用安全层处理约束的 SMS-MADRL 算法, 具有不依赖经验的优势。相比之下, 采用奖励整定法处理约束的 MADDPG 算法非常依赖于经验。若惩罚系数过小, 会出现电压越限; 若惩罚系数过大, 会增大控制代价, 并恶化配网损耗优化效果。

事实上, 在电力系统中应用深度强化学习的情景下, 安全层法所能处理的约束不仅仅局限于电压约束。在不久的将来, 输电线上的电流约束、储能的充电状态约束以及变流器功率约束等均可以通过安全层法进行处理, 而奖励整定法针对这种多类型的约束问题则显得束手无策。

参考文献:

- [1] SUN Hongbin, GUO Qinglai, QI Junjian, et al. Review of challenges and research opportunities for voltage control in smart grids [J]. IEEE Transactions on Power Systems, 2019, 34(4): 2790-2801.
- [2] GILL S, KOCKAR I, AULT G W. Dynamic optimal power flow for active distribution networks [J]. IEEE Transactions on Power Systems, 2014, 29(1): 121-131.
- [3] 王秀丽, 张择策, 侯雨伸. 主动配电网多维度静态安全评估 [J]. 西安交通大学学报, 2016, 50(8): 110-116.
WANG Xiuli, ZHANG Zece, HOU Yushen. Multidimensional static security assessment for active distribution network [J]. Journal of Xi'an Jiaotong University, 2016, 50(8): 110-116.

ty, 2016, 50(8): 110-116.

- [4] 高荣, 寇鹏, 梁得亮, 等. 含混合式配电变压器的主动配电网电压鲁棒模型预测控制 [J]. 中国电机工程学报, 2020, 40(7): 2081-2090.
GAO Rong, KOU Peng, LIANG Deliang, et al. Robust model predictive control for the voltage regulation in active distribution networks with hybrid distribution transformers [J]. Proceedings of the CSEE, 2020, 40(7): 2081-2090.
- [5] 徐志成, 赵波, 丁明, 等. 基于电压灵敏度的配电网光伏消纳能力随机场景模拟及逆变器控制参数优化整定 [J]. 中国电机工程学报, 2016, 36(6): 1578-1587, 1770.
XU Zhicheng, ZHAO Bo, DING Ming, et al. Photovoltaic hosting capacity evaluation of distribution networks and inverter parameters optimization based on node voltage sensitivity [J]. Proceedings of the CSEE, 2016, 36(6): 1578-1587, 1770.
- [6] 谷永刚, 肖凯, 夏经德, 等. 利用增量二次规划和启发式方法的电力系统动态无功优化 [J]. 西安交通大学学报, 2010, 44(8): 106-111.
GU Yonggang, XIAO Kai, XIA Jingde, et al. Incremental quadratic programming and heuristic combination algorithm for dynamic optimal reactive power in power system [J]. Journal of Xi'an Jiaotong University, 2010, 44(8): 106-111.
- [7] 刘一兵, 吴文传, 张伯明, 等. 基于混合整数二阶锥规划的三相有源配电网无功优化 [J]. 电力系统自动化, 2014, 38(15): 58-64.
LIU Yibing, WU Wenchuan, ZHANG Boming, et al. Reactive power optimization for three-phase distribution networks with distributed generators based on mixed integer second-order cone programming [J]. Automation of Electric Power Systems, 2014, 38(15): 58-64.
- [8] CHRISTAKOU K, PAOLONE M, ABUR A. Voltage control in active distribution networks under uncertainty in the system model: a robust optimization approach [J]. IEEE Transactions on Smart Grid, 2018, 9(6): 5631-5642.
- [9] 汤涌, 姚伟, 王宏志, 等. 电网仿真分析与决策的人工智能方法 [J]. 中国电机工程学报, 2022, 42(15): 5384-5405.
TANG Yong, YAO Wei, WANG Hongzhi, et al. Artificial intelligence techniques for power grid simulation analysis and decision making [J]. Proceedings of the CSEE, 2022, 42(15): 5384-5405.
- [10] 李琦, 乔颖, 张宇精. 配电网持续无功优化的深度强化学习方法 [J]. 电网技术, 2020, 44(4): 1473-1480.

- LI Qi, QIAO Ying, ZHANG Yujing. Continuous reactive power optimization of distribution network using deep reinforcement learning [J]. Power System Technology, 2020, 44(4): 1473-1480.
- [11] 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-time-scale online optimization for reactive power of distribution network based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2021, 45(10): 77-85.
- [12] CAO Di, HU Weihao, ZHAO Junbo, et al. A multi-agent deep reinforcement learning based voltage regulation using coordinated PV inverters [J]. IEEE Transactions on Power Systems, 2020, 35(5): 4120-4123.
- [13] LOWE R, WU Yi, TAMAR A, et al. Multi-agent actor-critic for mixed cooperative-competitive environments [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2017: 6382-6393.
- [14] AMODEI D, OLAH C, STEINHARDT J, et al. Concrete problems in AI safety [EB/OL]. (2016-07-25) [2022-10-20]. <https://arxiv.org/abs/1606.06565>.
- [15] DALAL G, DVIJOTHAM K, VECERIK M, et al. Safe exploration in continuous action spaces [EB/OL]. (2018-01-26) [2022-10-28]. <https://arxiv.org/abs/1801.08757>.
- [16] GU Shangding, KUBA J G, WEN M, et al. Multi-agent constrained policy optimisation [EB/OL]. (2022-02-10) [2022-11-25]. <https://arxiv.org/abs/2110.02793>.
- [17] LI Hepeng, HE Haibo. Learning to operate distribution networks with safe deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2022, 13(3): 1860-1872.
- [18] KOU Peng, LIANG Deliang, WANG Chen, et al. Safe deep reinforcement learning-based constrained optimal control scheme for active distribution networks [J]. Applied Energy, 2020, 264: 114772.
- [19] GAO Yuanqi, YU Nanpeng. Model-augmented safe reinforcement learning for volt-VAR control in power distribution networks [J]. Applied Energy, 2022, 313: 118762.
- [20] ZHANG Xi, LIU Youbo, DUAN Jiajun, et al. DDPG-based multi-agent framework for SVC tuning in urban power grid with renewable energy resources [J]. IEEE Transactions on Power Systems, 2021, 36(6): 5465-5475.
- [21] LIU Hangyue, ZHANG Cuo, CHAI Qingmian, et al. Robust regional coordination of inverter-based volt/var control via multi-agent deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2021, 12(6): 5420-5433.
- [22] CHEN Xin, QU Guannan, TANG Yujie, et al. Reinforcement learning for selective key applications in power systems: recent advances and future challenges [J]. IEEE Transactions on Smart Grid, 2022, 13(4): 2935-2958.
- [23] WANG Jianhong, XU Wangkun, GU Yunjie, et al. Multi-agent reinforcement learning for active voltage control on power distribution networks [C]//Advances in Neural Information Processing Systems 34 (NeurIPS 2021). San Francisco, CA, USA: Curran Associates, Inc., 2021: 3271-3284.
- [24] BARAN M E, WU F F. Network reconfiguration in distribution systems for loss reduction and load balancing [J]. IEEE Power Engineering Review, 1989, 9(4): 101-102.
- [25] LIU Yibin, LIANG Deliang, KOU Peng, et al. Compound control system of hybrid distribution transformer [J]. IEEE Transactions on Industry Applications, 2020, 56(6): 6360-6373.
- [26] 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.
- [27] LIU Haotian, WU Wenchuan. Online Multi-agent reinforcement learning for decentralized inverter-based volt-VAR control [J]. IEEE Transactions on Smart Grid, 2021, 12(4): 2980-2990.
- [28] SHEEBAELHAMD Z, ZISIS K, NISIOTI A, et al. Safe deep reinforcement learning for multi-agent systems with continuous action spaces [EB/OL]. (2021-08-11) [2022-12-11]. <https://arxiv.org/abs/2108.03952>.
- [29] BIDGOLI H S, VAN CUTSEM T. Combined local and centralized voltage control in active distribution networks [J]. IEEE Transactions on Power Systems, 2018, 33(2): 1374-1384.
- [30] NREL. Wind data and tools [EB/OL]. [2023-02-19]. <https://www.nrel.gov/wind/data-tools.html>.

(编辑 刘杨 李慧敏)