DOI: 10.19783/j.cnki.pspc.231477

# 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

张梓枭[1]，崔明建[1]，张程彬[1]，张 剑[2]，蔡木良[3]，周求宽[3]

(1. 天津大学电气自动化与信息工程学院，天津 300072；2. 合肥工业大学电气与自动化工程学院，安徽 合肥 230009；3. 国网江西省电力有限公司电力科学研究院，江西 南昌 330096)

**摘要：**大量分布式电源的接入使配电网的结构与控制方式发生改变。针对分布式电源间歇性和波动性引起的电压越限问题，通过调节系统中无功潮流与有功潮流的分布来维持配电网的电压稳定。提出了一种基于反事实多智能体策略梯度(counterfactual multi-agent policy gradients, COMA)算法的配电网电压协同优化方法，通过反事实基线解决了多智能体强化学习中的“信度分配”问题，实现有功出力设备和无功补偿设备的联合优化调度。智能体通过局部观测值选定动作，减轻系统的通信压力，且不依赖精确的潮流模型，以实现配电网的实时优化控制。通过改进的 IEEE33 节点系统和 141 节点系统验证了所提算法的可行性与有效性。并与经典算法的控制效果进行比较，进一步证明所提算法在配电网电压优化控制方面的性能优势。

**关键词：**配电网；有功无功协同优化；多智能体深度强化学习；分布式电源

## Active and reactive power coordinated optimal voltage control of a distribution network based on counterfactual multi-agent reinforcement learning

ZHANG Zixiao[1], CUI Mingjian[1], ZHANG Chengbin[1], ZHANG Jian[2], CAI Muliang[3], ZHOU Qiukuan[3]

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; 2. School of Electrical and Automation Engineering, Hefei University of Technology, Hefei 230009, China; 3. State Grid Jiangxi Electric Power Research Institute, Nanchang 330096, China)

**Abstract:** The integration of a significant number of distributed generators has altered the structure and control methods in distribution networks. To address the voltage stability issues caused by the intermittency and fluctuation of distributed generators, this paper proposes the stabilization of the distribution network voltage by adjusting the distribution of reactive and active power flows within the system. A distribution network voltage coordinated optimization method is proposed based on the counterfactual multi-agent policy gradients (COMA) algorithm. The proposed method can use a counterfactual baseline to resolve the “credit assignment” challenge in multi-agent reinforcement learning, enabling the joint optimization scheduling of active power generation and reactive power compensation devices. Agents select actions based on local observations, thereby reducing the system’s communication load and eliminating the dependency on precise flow models, to achieve real-time optimization control of distribution networks. The feasibility and effectiveness of the proposed algorithm are demonstrated by using the improved IEEE33-node system and 141-node system. Compared with the classic control algorithms, the proposed method has further performance advantages in the voltage optimization and control problems for distribution networks.

This work is supported by the National Natural Science Foundation of China (No. 52207130).

**Key words:** distribution network; active and reactive power coordinated optimization; multi-agent deep reinforcement learning; distributed generator

**基金项目：**国家自然科学基金项目资助(52207130)；江西省重点研发计划项目资助(20223BBE51013)

## 0　引言

为应对化石能源短缺、全球气候变化等问题，实现能源安全与可持续发展，我国大力发展光伏、风电、地热能等分布式新能源，电力系统的电源结构和负荷特性也随之发生了深刻变革[1]。由于分布式电源(distributed generator, DG)具有安全、可靠、灵活、绿色等特点，各国都加快了发展 DG 的步伐。2022 年，我国新增光伏装机容量 87.41 GW，同比增长 28.1%。根据国家能源局发布的数据显示，截至 2023 年 4 月底，我国风电光伏发电总装机突破 8 亿 kW，达到 8.2 亿 kW，占全国发电装机的 30.9%，其中风电占比 14.3%，光伏发电占比 16.6%。

此外，随着我国经济发展与产业升级，越来越多的高新技术企业对配电网的供电可靠性与电能质量提出了更高要求[2]。而当自然资源(如光照、风速)较为丰富时，DG 注入的有功功率增大将使得节点电压上升并超出其上限值，甚至导致配电网潮流反向。在某些极端情况下，传统的电压调节方法无法将系统所有节点电压调整至稳定运行范围内[3]。

随着分布式电源的大规模并网，配网电压控制问题亟待解决。配电网中有功、无功解耦性已不存在，有功对电压影响显著。在低压配电网或大量使用电缆的配电网中，由于线路电阻远大于电抗，有功对电压的影响远大于无功。对于含高渗透率光伏、风电的配电网，单纯依靠无功补偿装置、调压器、DG 无功无法解决电压越限问题[4]。本文利用光伏逆变器协同多种调压设备(如静止无功补偿器(static var compensator, SVC))、储能系统(energy storage system, ESS)等深度参与电压调节，解决光伏大量接入后配电网电压越限、弃光等问题，为提升配电网对新能源的接纳能力及电压控制精细化水平提供技术支撑。

根据对通信系统的依赖程度，国内外主流的电压控制方法总体分为集中式控制、分布式控制和本地控制。集中式控制主要基于最优化方法[5]、模型预测控制[6]和灵敏度分析[7]等；分布式控制主要基于多智能体技术[8]、一致性算法[9-10]等；本地控制主要基于梯度投影法[11-12]、下垂控制[13-15]和模糊数学理论[16]等。其中，本地控制不能保证全局最优性，容易陷入局部最优的困境，其迭代步长的选取与网络结构、线路参数强相关，当网络结构发生变化时，不适用的迭代步长将导致优化算法不收敛，最终导致系统电压振荡。集中式控制方法过于依赖精确的配电网模型、强大的计算能力与健全、快速、可靠的通信系统，上述条件在实际问题中难以满足，且集中式控制方法难以处理大规模非线性约束、离散变量与不确定性问题，难以满足在线运行要求[17]。

以多智能体技术为代表的分布式控制方式，对通信系统的依赖程度较低，可以满足配电网电压协同优化的在线运行要求。为了弥补传统的集中式控制方法在通信、计算等方面存在的不足，本文将分布式控制方法应用到配电网电压协调优化控制领域，通过多智能体技术，实现配电网电压有功无功协同优化。

近年来，基于深度强化学习(deep reinforcement learning, DRL)的优化控制方法受到了国内外学者的广泛关注。DRL 利用深度神经网络强大的感知能力和特征提取能力处理复杂的环境特征，利用强化学习强大的决策能力控制智能体与环境交互，优化决策过程[18]。DRL 已广泛应用于解决多智能体协调控制问题，多智能体协调控制是指多个智能体通过相互合作，在集体层面呈现出有序的协同动作，完成既定目标得到联合奖励。但目前应用 DRL 实现多智能体协作存在一系列问题[19]，如状态-动作维数过大、多智能体交互、部分可观测、环境非静态等，因此相对单智能体问题而言，多智能体深度强化学习(multi-agent deep reinforcement learning, MADRL)的理论研究有待进一步完善和拓展。

多智能体深度策略梯度(multi-agent deep deterministic policy gradient, MADDPG)算法以深度策略梯度(deep deterministic policy gradient, DDPG)算法为基础，作为一种集中式训练、分布式执行的 MADRL 算法，提供了一个既可以适应于通信信道的协作场景，同时适用于智能体之间存在物理交互的合作性或竞争性场景[20]。文献[21]统筹考虑了无功补偿设备的差异化调节特性和不同深度强化学习算法的特点，分别选取 DQN(deep Q network)与 DDPG 算法，实现了双时间尺度无功电压优化；文献[22]提出了基于 MADDPG 算法的分布式电源就地电压控制框架，通过构建 DRL 智能体以实时感知配电网状态，制定分布式电源运行策略，自适应地应对电压波动；文献[23]提出了一种基于优化数学模型与 MADDPG 算法相结合的配电网多时间尺度电压调节策略；文献[24]基于 DDPG 算法实现了主动配电网的优化运行，使主动配电网的节点电压总偏差和线损最小。

但 MADDPG 算法存在一定的“信度分配”问题，信度分配主要指合作式环境下各智能体对全局奖励信号的分配，而在理论上 MADDPG 分配给每个智能体的信度是相同的[25]。基于此，文献[26]提出了反事实多智能体策略梯度(counterfactual multi-

agent policy gradients, COMA)算法，COMA 算法基于“集中训练、分布执行”的计算框架，使用反事实基线解决了多智能体强化学习中的“信度分配”问题，可以评价单一智能体对整体最终回报的贡献度大小，从而更好地优化整体性能[26]。但传统的 COMA 算法只能处理智能体的离散动作，本文对传统 COMA 算法进行改进，使其能够处理连续的动作空间，通过协同控制多种调压设备的有功、无功出力，实现配电网电压优化，最后在改进的 IEEE33 节点和 Matpower 中 141 节点算例系统中验证了所提方法的可行性与有效性。

## 1 配电网电压优化模型

### 1.1 目标函数

本文通过对多种设备的调度控制，缓解电压波动与电压越限问题，维持配电网电压稳定，构造目标函数为

$$\min f=\sum_{i=1}^{N}\left|v_i-v_{\text{ref}}\right| \tag{1}$$

式中：$f$ 表示目标函数；$v_i$ 表示节点 $i$ ($i=1,2,\cdots,N$) 的电压幅值；$v_{\text{ref}}$ 表示配电网额定电压；$N$ 表示配电网节点数。

### 1.2 约束条件

1) 潮流约束

$$P_i^{\text{G}}-P_i^{\text{L}}=v_i^2\sum_{i=1}^{N}G_{ij}-v_i\sum_{i=1}^{N}v_j(G_{ij}\cos\theta_{ij}+B_{ij}\sin\theta_{ij}) \tag{2}$$

$$Q_i^{\text{G}}-Q_i^{\text{L}}=v_i\sum_{i=1}^{N}v_j(B_{ij}\cos\theta_{ij}-G_{ij}\sin\theta_{ij})-v_i^2\sum_{i=1}^{N}B_{ij} \tag{3}$$

式中：$P_i^{\text{G}}$ 和 $Q_i^{\text{G}}$ 分别为母线 $i$ 处注入节点的有功功率和无功功率；$P_i^{\text{L}}$ 和 $Q_i^{\text{L}}$ 分别为母线 $i$ 处有功负荷和无功负荷；$v_i$ 和 $v_j$ 分别为母线 $i$ 与母线 $j$ 的电压；$G_{ij}$ 和 $B_{ij}$ 分别为母线 $i$ 与母线 $j$ 间的电导与电纳；$\theta_{ij}$ 为母线 $i$ 与母线 $j$ 的相位差。

2) 光伏出力约束

为提高经济性，规定光伏系统以实际功率发电，无“弃光”现象，仅无功功率可调。

$$-Q_{p,\max}^{\text{PV}}\leqslant Q_p^{\text{PV}}\leqslant Q_{p,\max}^{\text{PV}} \tag{4}$$

$$Q_{p,\max}^{\text{PV}}=\sqrt{s_p^2-(P_{p,\max}^{\text{PV}})^2} \tag{5}$$

式中：$Q_p^{\text{PV}}$ 为第 $p$ 个光伏逆变器的无功功率；$P_{p,\max}^{\text{PV}}$、$Q_{p,\max}^{\text{PV}}$ 分别为该光伏的最大有功、无功功率；$s_p$ 为第 $p$ 个光伏最大视在功率。

3) SVC 容量约束

$$Q_{y,\min}^{\text{SVC}}\leqslant Q_y^{\text{SVC}}\leqslant Q_{y,\max}^{\text{SVC}} \tag{6}$$

式中：$Q_y^{\text{SVC}}$ 为第 $y$ 个 SVC 的无功出力；$Q_{y,\max}^{\text{SVC}}$、$Q_{y,\min}^{\text{SVC}}$ 分别为第 $y$ 个 SVC 无功出力的上、下限。

4) 燃气机组出力约束

$$P_{c,\min}^{\text{DG}}\leqslant P_c^{\text{DG}}\leqslant P_{c,\max}^{\text{DG}} \tag{7}$$

$$Q_{c,\min}^{\text{DG}}\leqslant Q_c^{\text{DG}}\leqslant Q_{c,\max}^{\text{DG}} \tag{8}$$

式中：$P_c^{\text{DG}}$ 为第 $c$ 个机组的有功出力；$P_{c,\max}^{\text{DG}}$、$P_{c,\min}^{\text{DG}}$ 分别为第 $c$ 个机组有功出力的上、下限；$Q_c^{\text{DG}}$ 为第 $c$ 个机组的无功出力；$Q_{c,\max}^{\text{DG}}$、$Q_{c,\min}^{\text{DG}}$ 分别为第 $c$ 个机组无功出力的上、下限。

5) 储能系统约束

充电状态下：

$$C_{l,t}^{\text{SOC}}=C_{l,t-1}^{\text{SOC}}+\frac{\eta_{l,\text{ch}}P_{l,t}^{\text{ESS}}}{E_l}\Delta t \tag{9}$$

放电状态下：

$$C_{l,t}^{\text{SOC}}=C_{l,t-1}^{\text{SOC}}+\frac{P_{l,t}^{\text{ESS}}}{\eta_{l,\text{dis}}E_l}\Delta t \tag{10}$$

式中：$C_{l,t}^{\text{SOC}}$ 为第 $l$ 个储能设备在 $t$ 时刻的荷电状态；$E_l$ 为第 $l$ 个储能设备的额定容量；$\eta_{l,\text{ch}}$ 和 $\eta_{l,\text{dis}}$ 分别为第 $l$ 个储能设备的充电效率和放电效率；$P_{l,t}^{\text{ESS}}$ 为第 $l$ 个储能设备的有功功率；$\Delta t$ 为时间间隔。

$$C_{l,\min}^{\text{SOC}}\leqslant C_l^{\text{SOC}}\leqslant C_{l,\max}^{\text{SOC}} \tag{11}$$

$$-P_{l,\text{dis}}^{\text{ESS}}\leqslant P_l^{\text{ESS}}\leqslant P_{l,\text{ch}}^{\text{ESS}} \tag{12}$$

式中，$C_{l,\max}^{\text{SOC}}$、$C_{l,\min}^{\text{SOC}}$ 和 $P_{l,\text{dis}}^{\text{ESS}}$、$P_{l,\text{ch}}^{\text{ESS}}$ 分别为第 $l$ 个储能系统运行时荷电状态的上、下限和充电功率最大值、放电功率最大值。

## 2 基于 COMA 算法的协同电压优化

### 2.1 配电网电压优化的马尔可夫博弈模型

在 DRL 的框架下，智能体通过不断与环境交互，以“试错”的方式完成学习。每轮交互均从环境初始状态开始，各个智能体通过观察局部环境状态，根据观测状态选取并执行动作，与环境交互改变环境状态，然后获得环境给予的奖励，从而对自身的动作策略进行优化。重复交互过程中，智能体会不断优化自身的动作策略[27]。完成离线训练过程后，通过上述过程得到的动作策略满足在线运行的要求。且 MADRL 算法基于离线集中训练和在线分散执行的框架，智能体只需通过观测本地状态给出决策，无需复杂的通信设备。强化学习离线训练过程如图 1 所示。

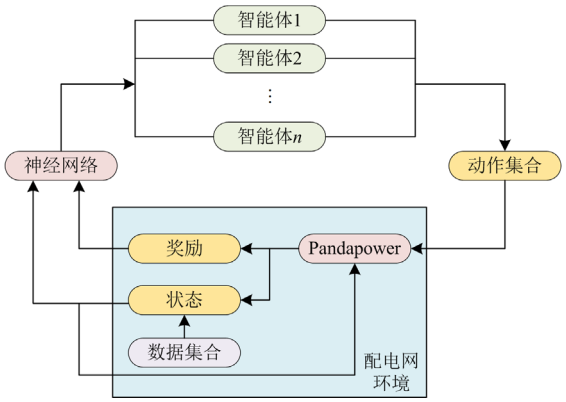



图 1 强化学习离线训练过程

Fig. 1 Offline training process of reinforcement learning

上述 MADRL 算法通常使用马尔可夫博弈过程(markov game process, MGP)进行定量描述，其使用元组 $<n,S,A,O,P,r,\Omega,\gamma>$ 表示。元组中：$n$ 表示智能体个数；$S$ 表示系统状态集，$t$ 时刻系统状态为 $s_t \in S$；$A$ 表示联合动作集且 $A_m \in A$，其中 $A_m$ 表示智能体 $m$ 的动作集，$t$ 时刻智能体 $m$ 的动作为 $a_{m,t} \in A_m$；$O$ 表示联合观察集且 $O_m \in O$，其中 $O_m$ 表示智能体 $m$ 的观察集，$t$ 时刻智能体 $m$ 的观察为 $o_{m,t} \in O_m$；$P:S\times A\times S\to[0,1]$ 表示状态转移概率函数，表示系统在状态 $s_t$ 下执行动作 $a_t$ 进入下一状态 $s_{t+1}$ 的概率；$r:S\times A\to R$ 为奖励函数，表示智能体在状态 $s_t$ 下执行联合动作 $a_t$ 并与环境交互，环境的反馈奖励，$t$ 时刻的奖励为 $r_t$；$\Omega:S\times A\times O\to[0,1]$ 表示智能体在执行动作 $a_t$ 后对用于联合观测的观测器(或传感器)的扰动；$\gamma\in(0,1)$ 为折扣因子。本文中 MADRL 的目标是找到一个最佳的联合策略 $\pi$ 来最大化折扣回报($\pi_m \in \pi$，其中 $\pi_m$ 表示智能体 $m$ 的动作策略)，用期望的形式表示为

$$\max \mathbb{E}_{\pi}\left[\sum_{t=0}^{T}\gamma^t r_t\right] \tag{13}$$

式中，$\mathbb{E}_{\pi}$ 表示在联合策略 $\pi$ 下的数学期望。

1) 观测空间

智能体 $m$ 的观测空间是智能体对本地环境的感知信息。

$$O_m=[P_b^{\mathrm{L}},Q_b^{\mathrm{L}},V_b,P_b^{\mathrm{PV}},C_b^{\mathrm{SOC}}] \tag{14}$$

式中：$b$ 表示与智能体 $m$ 在同一分区的可观测节点；$P_b^{\mathrm{L}}$、$Q_b^{\mathrm{L}}$、$V_b$、$P_b^{\mathrm{PV}}$ 和 $C_b^{\mathrm{SOC}}$ 分别表示可观测节点的负荷损耗有功、无功功率、电压幅值、光伏的有功功率注入和储能装置的荷电状态。

2) 动作空间

智能体 $m$ 的动作空间 $A_m=\{a_m:a_{m,\min}\leqslant a_m\leqslant a_{m,\max}\}$，其中，$a_m=\{Q_m^{\mathrm{PV}},Q_m^{\mathrm{SVC}},P_m^{\mathrm{DG}},P_m^{\mathrm{ESS}}\}$，包括光伏逆变器、SVC 的无功出力及 DG、ESS 的有功出力，$a_{m,\max}$、$a_{m,\min}$ 分别为第 $m$ 个智能体动作空间的上、下限，其取值满足式(4)—式(12)的约束条件。

3) 奖励函数

MADRL 训练过程中，奖励函数 $r$ 为

$$r=-\frac{k_{\mathrm{r}}}{N}\sum_{i=1}^{N}\left|v_i-v_{\mathrm{ref}}\right| \tag{15}$$

式中，$k_{\mathrm{r}}$ 为常量，旨在调整奖励函数的值使之位于一个合适区间内。奖励函数 $r$ 的目的是将配电网系统的节点电压控制在额定电压 $v_{\mathrm{ref}}$ 附近。

## 2.2 COMA 强化学习算法

传统的 MADRL 算法面临“多智能体回报分配”的问题：在合作环境中，多智能体联合行动通常只产生全局奖励，这使得智能体难以判断自身动作对全局奖励的贡献大小。如果只是为每个智能体设置单独的奖励函数，在合作的环境下通常训练效果不佳，而且难以令单个智能体为全局奖励做出牺牲。

COMA 是一种基于策略梯度的多智能体算法，其引入了“反事实基线”的概念，当智能体执行某一动作后，通过反事实基线判断动作的价值，从而使每个智能体都能有独立的奖励，这样能够更清楚地区分全局行为决策中单个智能体动作的贡献大小，解决 MADRL 中的“信度分配”问题。

传统 COMA 算法只能作用于离散动作空间，如果用其来控制调压设备，只能将调压设备的功率离散化，不利于发挥调压设备控制的灵活性。所以本文针对 COMA 算法做出一定修改，使其能够处理连续动作空间。即对 COMA 算法每个智能体的优势函数 $AF_m$ 进行调整，传统 COMA 算法智能体 $m$ 的优势函数 $AF_m$ 表示为

$$AF_m(\boldsymbol{s},\boldsymbol{a})=Q(\boldsymbol{s},\boldsymbol{a})-\sum_{a'_m\in A_m}\pi(a'_m\mid\tau_m)Q(\boldsymbol{s},\boldsymbol{a}_{-m},a'_m) \tag{16}$$

式中：$\boldsymbol{s}$ 表示系统环境的当前状态；$\boldsymbol{a}$ 表示智能体的联合动作；$\boldsymbol{a}_{-m}$ 表示除智能体 $m$ 外，其他智能体的联合动作；$a'_m$ 表示算法假设智能体执行的另一动作；$Q$ 函数用于评估动作的价值，即智能体 $m$ 执行某一动作 $a_m$ 后，能够获得奖励的期望值；$\tau_m$ 表示智能体 $m$ 的历史观测–动作序列；$Q(\boldsymbol{s},\boldsymbol{a})$ 表示集中式 critic 网络中输出的全局 $Q$ 值；$\sum_{a'_m\in A_m}\pi(a'_m\mid\tau_m)Q(\boldsymbol{s},\boldsymbol{a}_{-m},a'_m)$ 表示在智能体 $m$ 的所有可能动作下能够获得的全局 $Q$ 值的期望。

由式(16)可得，当智能体 $m$ 在执行完动作 $a_m$ 并获得一个奖励期望值后，在保证其他智能体动作

$\boldsymbol{a}_{-m}$ 不变的前提下，假设智能体 $m$ 执行的是另一个动作 $a'_m$，并计算出可能获得的奖励，通过比较这两个奖励期望值来判断智能体 $m$ 执行的动作 $a_m$ 对全局奖励的贡献大小。

在上述理论基础上，为适应连续动作空间，本文对 COMA 算法的优势函数部分作如下修改[28]，将智能体 $m$ 的优势函数 $AF_m$ 修改为

$$AF_m(\boldsymbol{s},\boldsymbol{a})=Q(\boldsymbol{s},\boldsymbol{a})-\int_{a'_m\in A_m}Q(\boldsymbol{s},\boldsymbol{a}_{-m},a'_m)\,\mathrm{d}\,\pi_m(a'_m\mid\tau_m) \tag{17}$$

式中，$\pi_m(a'_m\mid\tau_m)$ 为 $a'_m$ 的一个高斯分布，在实际计算过程中，$\int_{a'_m\in A_m}Q(\boldsymbol{s},\boldsymbol{a}_{-m},a'_m)\,\mathrm{d}\,\pi_m(a'_m\mid\tau_m)$ 可以通过 Monte Carlo 抽样近似，故在计算过程中智能体 $m$ 的优势函数 $AF_m$ 可表示为

$$AF_m(\boldsymbol{s},\boldsymbol{a})=Q(\boldsymbol{s},\boldsymbol{a})-\frac{1}{M}\sum_{k_m=1}^{M}Q(\boldsymbol{s},\boldsymbol{a}_{-m},(a'_m)_{k_m}) \tag{18}$$

式中：$(a'_m)_{k_m}$ 表示抽样样本中假设的智能体动作，且 $(a'_m)_{k_m}\sim\pi_m(a'_m\mid\tau_m)$；$M$ 表示使用 Monte Carlo 抽样近似的样本数量。

由式(18)可知，调整后的优势函数可以计算智能体 $m$ 的动作 $a_m$ 在连续动作空间下相对于全局的贡献，且不需要额外增加交互数据。此时 COMA 的策略梯度计算公式表示为

$$g_k=\mathbb{E}_\pi\left[\sum_{m=1}^{n}\nabla_{\theta_k}\log\pi_m(a_m\mid\boldsymbol{s})AF_m(\boldsymbol{s},\boldsymbol{a})\right] \tag{19}$$

式中：$g_k$ 表示第 $k$ 次迭代的策略梯度；$\theta_k$ 表示第 $k$ 次迭代时策略网络的参数；$\pi_m(a_m\mid\boldsymbol{s})$ 表示智能体 $m$ 根据当前系统的状态 $\boldsymbol{s}$ 给出的动作策略。

## 3　算例分析

### 3.1　算例系统介绍

本文 MADRL 算法均由 Python 3.11.4 基于深度学习框架 PyTorch 2.0.1 实现。软件系统为 Windows 11 操作系统，硬件系统为 NVIDIA GTX 3060，AMD Ryzen 7 6800H CPU，使用改进的 IEEE33 节点系统算例进行仿真测试，验证所提算法的有效性。改进 33 节点系统的拓扑结构及分区情况如图 2 所示。为了便于操作与计算，按照耦合母线和终端母线间的最短路径，将 33 节点系统划分为 4 个区域。在 33 节点配电网中，网络额定电压为 12.66 kV，配电网设备参数如表 1 所示。

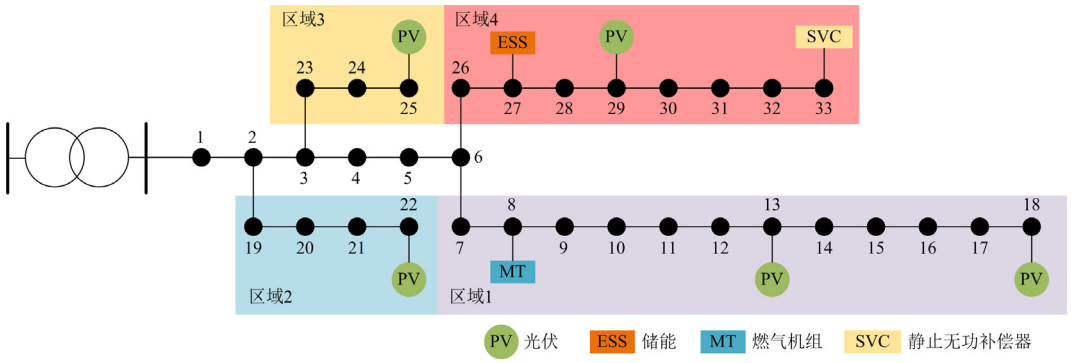



图 2　33 节点配电网系统

Fig. 2 A 33-node distribution network system

表 1　配电网可控设备参数

Table 1 Parameters of controllable devices in distribution network

| 节点 | 设备 | 出力范围 |
|---|---|---|
| 8 | 燃气机组 | 0.1~0.9 MW |
| 13, 18, 22 | PV 逆变器 | −2.0~2.0 Mvar |
| 25 | PV 逆变器 | −3.2~3.2 Mvar |
| 27 | ESS | −0.9~0.9 MW |
| 29 | PV 逆变器 | −0.3~0.3 Mvar |
| 33 | SVC | −0.9~0.9 Mvar |

### 3.2　训练过程

本节使用 MADDPG 和 COMA 两种多智能体强化学习算法实现配电网电压协同优化控制，将配电网运行的大量历史数据作为输入，在训练过程中不断优化多种调压设备的出力策略，达到最小化电压偏差的目的。本文选用训练过程中配电网的电压偏差 $\delta$、奖励函数 $r$ 和电压控制率 $\eta$ 来表示两种 MADRL 算法的训练效果。电压偏差 $\delta$ 定义为

$$\delta=\frac{1}{N_{\mathrm{T}}N}\sum_{t=1}^{N_{\mathrm{T}}}\sum_{i=1}^{N}|v_{i,t}-v_{\mathrm{ref}}| \tag{20}$$

式中：$v_{i,t}$ 表示节点 $i$ 在 $t$ 时刻的电压幅值的标幺值；$N_{\mathrm{T}}$ 表示测试日测量时刻的数量。电压偏差能在一定程度反映系统的电能质量，通常情况下，电压偏差越小，配电网电能质量越高，稳定性越强。

电压控制率表示在训练的一个步长下，电压处

于稳定状态的节点数与节点总数的比值，其数学表达式为

$$\eta = \frac{N_C}{N} \tag{21}$$

式中，$N_C$ 表示电压处于稳定状态的节点数。设节点系统的额定电压为 $v_{ref}$，电压安全运行的上限为 $1.05v_{ref}$，电压安全运行的下限为 $0.95v_{ref}$。对于节点 $i$ 电压 $v_i$，当 $0.95v_{ref} \leqslant v_i \leqslant 1.05v_{ref}$ 时，认为该节点电压处于稳定状态。在训练过程中，两种 MADRL 算法的超参数设置如表 2 所示。

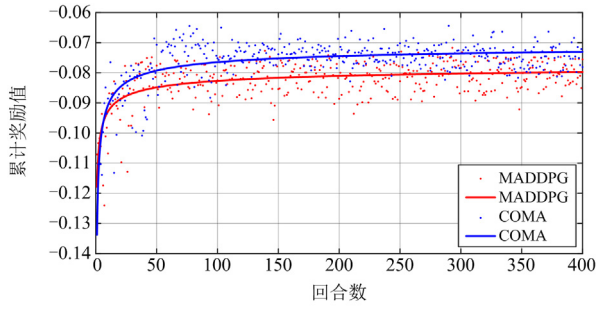
**表 2 超参数设置**

Table 2 Hyperparameter setting

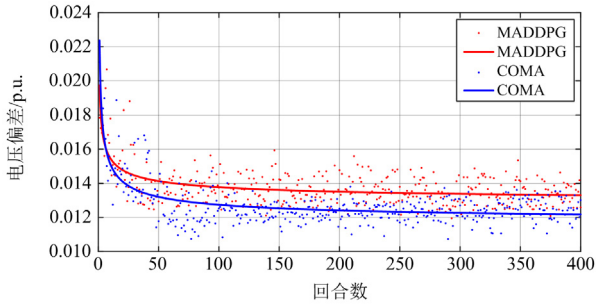
| 算法 | MADDPG | COMA |
|---|---|---|
| 学习率 | 0.001 | 0.001 |
| 批处理规模 | 32 | 32 |
| 训练回合数 | 400 | 400 |
| 训练总步长 | 19 200 | 19 200 |
| 折扣因子 | 0.99 | 0.99 |
| 经验池大小 | 5000 | Null |

在训练过程中，以 12 h 为一个回合，每个回合包括 240 个步长，步长为 3 min。训练过程中，智能体每轮训练的累计奖励、电压偏差和电压控制率如图 3 所示。

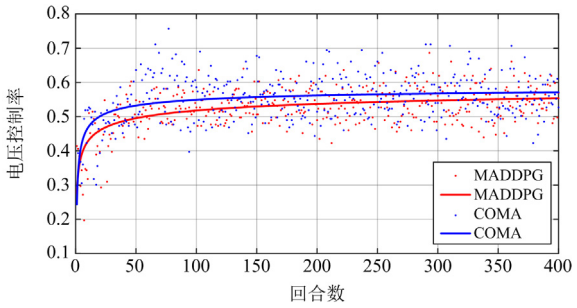
图 3 中散点表示训练过程中的实际数据，曲线表示散点数据的拟合结果。由图 3 可知，随着训练回合数的增加，奖励不断上升，电压偏差逐渐下降，




(a) 奖励值收敛过程




(b) 电压偏差收敛过程




(c) 电压控制率收敛过程

**图 3 两种深度强化学习算法的训练过程**

Fig. 3 Training process of two deep reinforcement learning algorithms

系统的电压控制率逐步上升，智能体能够采用更优的动作策略提高系统的电压稳定性。在训练约 100 个回合后，两种算法均收敛，COMA 算法训练过程的累计奖励、电压偏差和电压控制率依次趋近于–0.073、0.012、0.57。MADDPG 算法的训练结果依次趋近于–0.08、0.013、0.55。本文所提方法在训练过程中的奖励值、电压偏差和电压控制率均优于 MADDPG 算法，具有更好的训练效果。

为了证明 COMA 算法可以有效解决传统 MADRL 算法中存在的“信度分配”问题，定义迭代过程中各个智能体梯度 $g_{m,k}$ 与梯度标准差，如式(22)—式(24)所示。

$$g_{\text{std}} = \frac{1}{n_s}\sum_{k=1}^{n_s}\sqrt{\frac{\sum_{m=1}^{n}(g_{m,k}-g_m)^2}{n}} \tag{22}$$

$$g_{m,k}^{\text{COMA}} = \mathbb{E}_{\pi}\left[\nabla_{\theta_k}\log\pi_m(a_m\,|\,\boldsymbol{s})AF_m(\boldsymbol{s},\boldsymbol{a})\right] \tag{23}$$

$$g_{m,k}^{\text{MADDPG}} = \mathbb{E}_{\pi}\left[\nabla_{\theta_k}\log\pi_m(a_m\,|\,\boldsymbol{s})Q_m^{\pi}(\boldsymbol{s},\boldsymbol{a})\right] \tag{24}$$

式中：$g_{\text{std}}$ 为策略梯度标准差；$g_{m,k}$ 表示智能体 $m$ 在第 $k$ 次迭代的策略梯度；$n_s$ 为迭代总数；$Q_m^{\pi}(\boldsymbol{s},\boldsymbol{a})$ 为 MADDPG 算法中智能体 $m$ 的价值函数。

策略梯度标准差如表 3 所示。在训练过程中，MADDPG 算法每个智能体的策略梯度相差很小，而 COMA 算法不同智能体的策略梯度相差较大，因为 COMA 算法通过反事实基线判断动作的价值，从而使每个智能体都能有相对独立的策略梯度，这样能够更清楚地区分全局行为决策中单个智能体动作的贡献大小，解决 MADRL 中的“信度分配”问题。

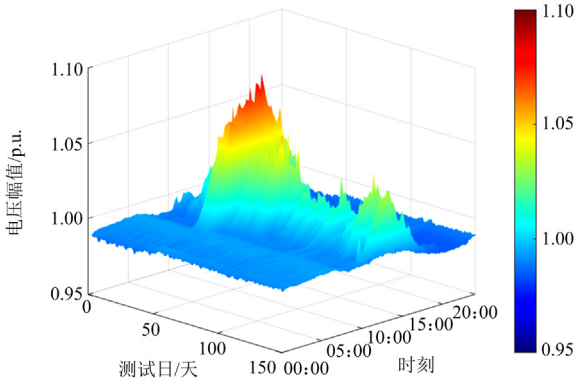
**表 3 策略梯度标准差**

Table 3 Standard deviation of policy gradient

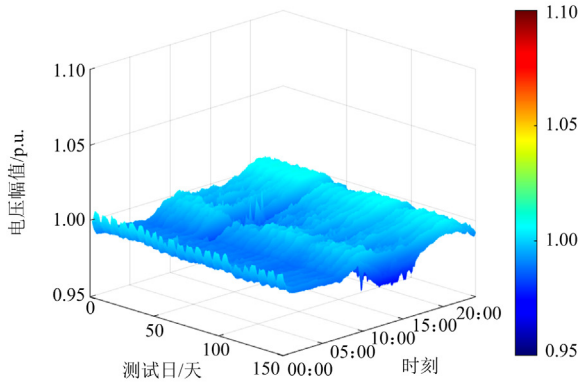
| 算法 | MADDPG | COMA |
|---|---|---|
| 平均标准差 | 0.045 | 0.811 |

### 3.3 仿真结果分析

为直观地体现 COMA 算法的电压控制效果，以 33 节点系统中配置光伏的 18 号节点为例，比较优化调度前后系统不同测试日的电压幅值情况，如图 4 所示。同一测试日中各节点各时刻的电压幅值情况如图 5 所示。单个测试日优化步长为 3 min。



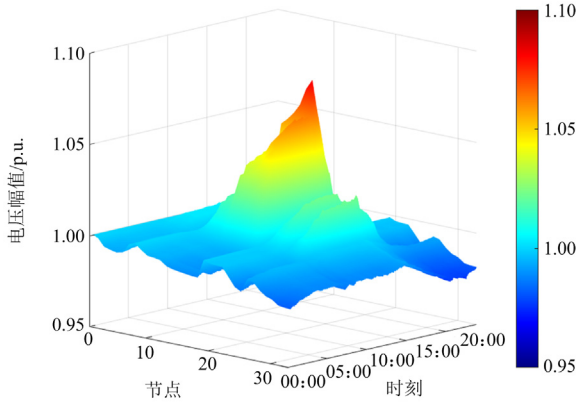

(a) 优化前电压幅值



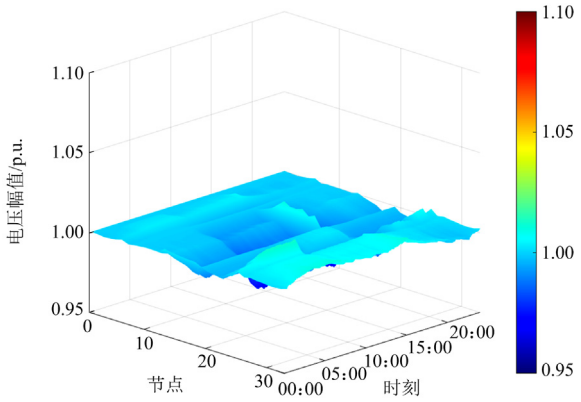

(b) 优化后电压幅值

**图 4 优化前后不同测试日节点电压幅值**

Fig. 4 Node voltage amplitudes on different test days before and after optimization




(a) 优化前电压幅值




(b) 优化后电压幅值

**图 5 优化前后各节点电压幅值**

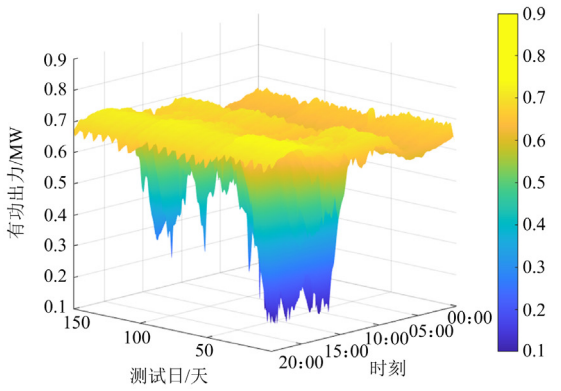
Fig. 5 Voltage amplitude of each node before and after optimization

图 4 表示 18 号节点在多个测试日的电压变化情况。由图 4 可知，在光照较强、光伏有功出力过高的时段(如 10：00—16：00)，本文所提方法能有效降低电压峰值，避免电压越限。而在光照较弱、光伏有功出力较少的时段，COMA 算法能适当抬高节点电压，降低电压偏差。为了更全面地反映有功无功协同优化调度的电压控制效果，图 5 展示了同一测试日中各节点的电压幅值随时间的变化情况。由图 5 可知，COMA 算法可以通过协调设备的有功出力和无功出力，使不同时段、不同种类节点的电压值尽可能地接近标幺值，从而减少配网的电压偏差，改善电压分布，提高配网的供电质量。

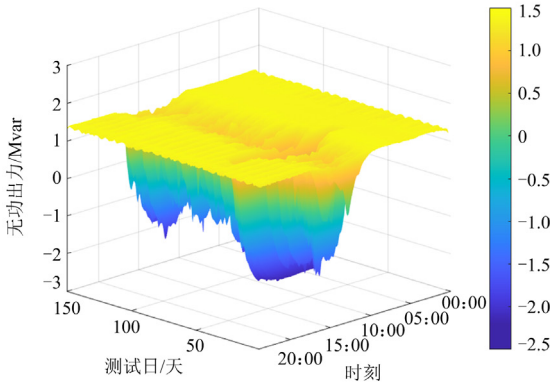
以 8 号节点的燃气机组和 25 号节点的光伏逆变器为例，分析优化过程的设备出力情况。

测试日系统的主要调压装置出力如图 6 所示。燃气机组的有功出力如图 6(a)所示，在 10：00—17：00 光伏出力较高时，燃气机组的有功出力维持在较低水平，在光伏出力较低时，燃气机组的有功出力较高，以支撑负荷、抬高电压。光伏逆变器无功出力如图 6(b)所示，在光伏出力较高时逆变器吸收无功功率，防止电压越限；大部分时间光伏逆变器发出无功功率实现电压支撑，进而平抑电压波动。因此，合理调控调压设备在保证电压不越限的情况下实现了配电网稳定运行。

为了进一步验证 COMA 算法的可行性和有效性，选用 MADDPG 算法与本文方法进行对比分析，选取了 150 个测试日的数据，从电压偏差、电压方差和电压极值 3 个指标对不同优化策略下的系统电压质量进行对比分析。




(a) 8号节点燃气机组测试日有功出力




(b) 25节点光伏逆变器测试日无功出力

**图 6　主要设备的出力情况**

Fig. 6 Output of main equipment

为了检验优化后电压的波动情况，衡量节点电压幅值的波动性，更全面地评价配电网的电能质量，本文定义电压方差指标 $S^2$ 为

$$S^2=\frac{1}{N_{\mathrm{T}}N}\sum_{t=1}^{N_{\mathrm{T}}}\sum_{i=1}^{N}(v_{i,t}-\overline{v}_i)^2 \tag{25}$$

式中，$\overline{v}_i$ 表示节点 $i$ 在测试日的平均电压。

两种优化策略下多测试日的电压偏差、电压方差和电压极大值情况如图 7—图 9 所示。

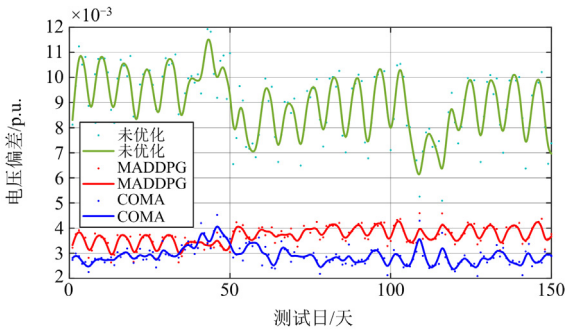



**图 7　多测试日电压偏差**

Fig. 7 Voltage deviation on multiple test days

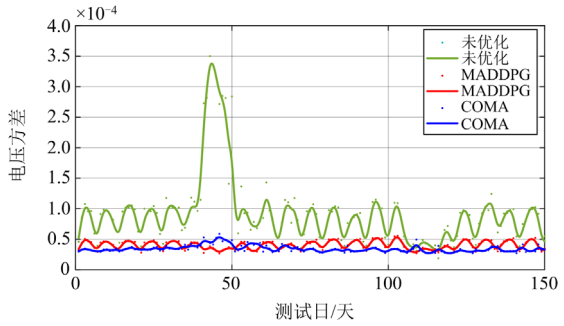



**图 8　多测试日电压方差**

Fig. 8 Voltage variance on multiple test days

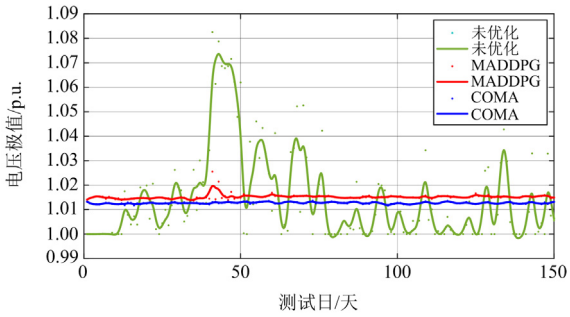



**图 9　多测试日电压极大值**

Fig. 9 Maximum voltage on multiple test days

电压偏差均值与电压方差均值反映 150 个测试日的总体电能质量，该指标避免了单一测试日结果的偶然性，能更准确地反映算法性能。测试日中各算法的综合电压控制结果具体数据如表 4 所示。

**表 4　测试日不同算法电压控制效果对比**

Table 4 Comparison of voltage control performance by various algorithms on test days

| 算法 | 电压偏差(均值)/p.u. | 电压方差(均值)/p.u.$^2$ | 电压极大值/p.u. | 电压极小值/p.u. |
|---|---|---|---|---|
| 未优化 | 0.008 97 | $9.1855\times10^{-5}$ | 1.0826 | 0.9555 |
| MADDPG | 0.003 69 | $4.0286\times10^{-5}$ | 1.0255 | 0.9827 |
| COMA | 0.002 90 | $3.4267\times10^{-5}$ | 1.0136 | 0.9620 |

由图 7 和图 8 可知，在多数的测试日中，COMA 算法的电压偏差和电压方差小于 MADDPG 算法。图 9 表示测试日中 33 节点系统出现的节点电压极大值。根据图 9 可知，测试日中 COMA 算法的电压极大值低于 MADDPG 算法，其电压越限的风险更低。由表 4 可知，测试结果中，COMA 算法的电压极大值与极小值均在稳定运行范围内，无电压越限问题。相较 MADDPG 算法，COMA 算法的测试结果平均电压偏差与平均电压方差分别降低了 21.41%和 14.94%，表明本文所提方法的优越性，其电压的偏差值更小，

且波动性更小。该结果避免了单一测试日中测试数据存在的偶然性对实验结果的影响，进一步说明COMA 算法能够更好地协调调压设备的有功出力和无功出力，实现对配电网节点电压的有效控制。

不同优化方法的实时决策能力如表 5 所示。由表 5 可知，COMA 算法拥有良好的在线决策能力，可以满足配电网运行优化的实时性需求。深度强化学习算法在决策速度上具有优势，因为其在线优化过程只需计算神经网络结果而无需进行迭代，大大缩短了决策时间。其决策时间远小于粒子群优化(particle swarm optimization, PSO)算法[23]和混合整数动态优化方法[29]，且相比于这两种算法分别降低了 95.06%和 99.98%。

表 5 不同优化方法的实时决策能力比较

Table 5 Comparison of real-time decision-making capabilities of different optimization methods

| 算法 | 平均决策时间/ms |
|---|---|
| 混合整数动态优化 | 17 629 |
| PSO | 79 |
| MADDPG | 4.0 |
| COMA | 3.9 |

### 3.4 算例系统补充

为了证明COMA 算法在大规模配电网系统的可拓展性，选取 141 节点系统进行补充验证。141 节点系统在 Matpower 的 case141[30]基础上进行改进，网络额定电压为 12.5 kV，设备分布及配电网拓扑如表 6 和图 10 所示。

表 6 141 节点配电网可控设备分布

Table 6 Distribution of controllable devices in a 141-node distribution network

| 节点 | 可控设备类型 |
|---|---|
| 5, 39, 45, 54, 80, 141 | 燃气机组 |
| 36, 59, 62, 68, 69, 75, 82, 87, 100, 106, 111, 116, 117, 130, 133, 137, 139, 141 | PV 逆变器 |
| 83, 88, 112, 118 | ESS |
| 53, 77, 110, 138 | SVC |

使用 MADDPG 和 COMA 两种多智能体强化学习算法在 141 节点配电网系统仿真环境下进行电压优化训练，训练过程如图 11 所示。

由图 11 可知，在完成 400 回合训练后，MADDPG 算法的累计奖励趋近于–0.916，COMA 算法训练过程的累计奖励趋近于–0.828。因此，在 141 节点系统中，本文所提方法在训练过程中的表现优于 MADDPG 算法，验证了在较大规模配电网中 COMA 算法的电压控制性能。

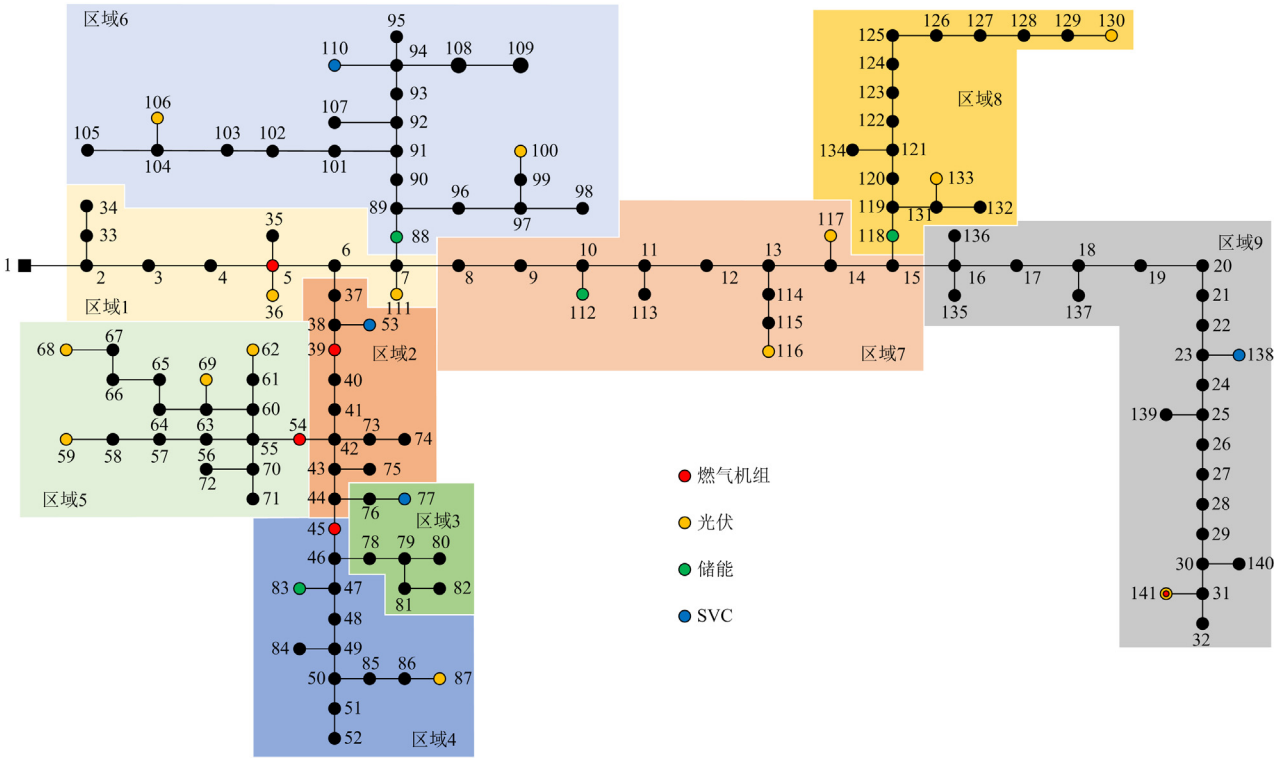



图 10 141 节点配电网系统

Fig. 10 A 141-node distribution network system

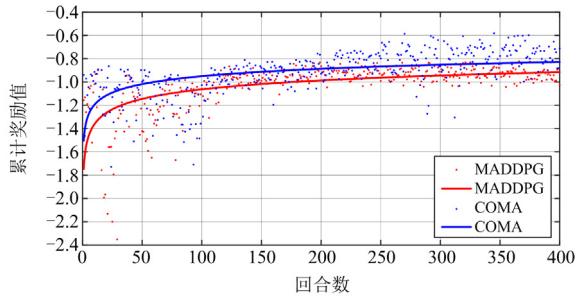



图 11 两种深度强化学习算法的训练过程

Fig. 11 Training process of two deep reinforcement learning algorithms

## 4 结论

针对大规模分布式电源并网下的配电网电压稳定控制问题，本文提出了一种基于 COMA 算法的配电网电压有功无功协同优化控制策略，采用集中学习、分散执行的框架，引入反事实基线作为奖励对照，有效解决传统 MADRL 算法中存在的“信度分配”问题。此外本文针对 COMA 算法作出改进，使其能够控制连续调压设备。算法以 3 min 作为调度周期的时间尺度，实现了配电网协同电压优化的离线训练与在线控制。最后通过改进 IEEE33 节点系统和 141 节点系统，验证了本文所提算法的可行性和有效性。

后续工作将进一步拓展 COMA 算法的应用场景，考虑不同设备调节和响应速度不同的问题，实现配电网多时间尺度的有功无功协同电压优化。

## 参考文献

[1] 赵东元, 胡楠, 傅靖, 等. 提升新能源电力系统灵活性的中国实践及发展路径研究[J]. 电力系统保护与控制, 2020, 48(24): 1-8.
ZHAO Dongyuan, HU Nan, FU Jing, et al. Research on the practice and road map of enhancing the flexibility of a new generation power system in China[J]. Power System Protection and Control, 2020, 48(24): 1-8.

[2] 王成山, 武震, 李鹏. 分布式电能存储技术的应用前景与挑战[J]. 电力系统自动化, 2014, 38(16): 1-8.
WANG Chengshan, WU Zhen, LI Peng. Prospects and challenges of distributed electricity storage technology[J]. Automation of Electric Power Systems, 2014, 38(16): 1-8.

[3] 王伟杰, 黄海宇, 徐远途, 等. 电动汽车参与主动配电网电压调控的策略研究[J]. 广东电力, 2023, 36(10): 93-104.
WANG Weijie, HUANG Haiyu, XU Yuantu, et al. Strategy research on electric vehicles participating in active distribution network voltage regulation[J]. Guangdong Electric Power, 2023, 36(10): 93-104.

[4] 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.

[5] 米阳, 申杰, 卢长坤, 等. 考虑含储能的三端智能软开关与需求侧响应的主动配电网有功无功协调优化[J]. 电力系统保护与控制, 2024, 52(3): 104-118.
MI Yang, SHEN Jie, LU Changkun, et al. Active and reactive power coordination optimization of an active distribution network considering a three-terminal soft open point with energy storage and demand response[J]. Power System Protection and Control, 2024, 52(3): 104-118.

[6] VALVERDE G, CUTSEM T V. Model predictive control of voltages in active distribution networks[J]. IEEE Transactions on Smart Grid, 2013, 4(4): 2152-2161.

[7] CHENG L, CHANG Y, HUANG R. Mitigating voltage problem in distribution system with distributed solar generation using electric vehicles[J]. IEEE Transactions on Sustainable Energy, 2013, 6(4): 1475-1484.

[8] 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.

[9] ZERAATI M, GOLSHAN M, GUERRERO J M. A consensus-based cooperative control of PEV battery and PV active power curtailment for voltage regulation in distribution networks[J]. IEEE Transactions on Smart Grid, 2019, 10(2): 670-680.

[10] KOUVELIOTIS-LYSIKATOS I N, KOUKOULA D, HATZIARGYRIOU N D. A double-layered fully distributed voltage control method for active distribution networks[J]. IEEE Transactions on Smart Grid, 2019, 10(2): 1465-1476.

[11] ZHU H, LIU H. Fast local voltage control under limited reactive power: optimality and stability analysis[J]. IEEE Transactions on Power Systems, 2016, 31(5): 3794-3803.

[12] LIU H, SHI W, ZHU H. Decentralized dynamic optimization for power network voltage control[J]. IEEE Transactions on Signal and Information Processing, 2017, 3(3): 568-579.

[13] JAHANGIRI P, ALIPRANTIS D C. Distributed volt/var control by PV inverters[J]. IEEE Transactions on Power Systems, 2013, 28(3): 3429-3438.

[14] 施家博, 苗虹, 曾成碧. 基于复合型虚拟阻抗与自适应下垂控制的并联逆变器功率均分策略[J]. 电测与仪表, 2022, 59(12): 77-82.
SHI Jiabo, MIAO Hong, ZENG Chengbi. Power sharing strategy of parallel inverter based on complex virtual

impedance and adaptive droop control[J]. Electrical Measurement & Instrumentation, 2022, 59(12): 77-82.

[15] 李华, 东琦, 李旭东. 考虑变工况的光伏下垂控制与二级电容响应的调频策略[J]. 电网与清洁能源, 2023, 39(10): 137-146.
LI Hua, DONG Qi, LI Xudong. The frequency regulation strategy for photovoltaic droop control and secondary capacitor response considering changing working conditions[J]. Power System and Clean Energy, 2023, 39(10): 137-146.

[16] 王守相, 陈思佳, 赵玮. 一种智能配电网三相调压器的模糊控制方法[J]. 电力系统自动化, 2016, 40(1): 72-80.
WANG Shouxiang, CHEN Sijia, ZHAO Wei. A fuzzy control method of voltage regulator in smart power distribution system[J]. Automation of Electric Power Systems, 2016, 40(1): 72-80.

[17] CAO D, HU W, ZHAO J, 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.

[18] 蔺伟山, 王小君, 孙庆凯, 等. 不确定性环境下基于深度强化学习的综合能源系统动态调度[J]. 电力系统保护与控制, 2022, 50(18): 50-60.
LIN Weishan, WANG Xiaojun, SUN Qingkai, et al. Dynamic dispatch of an integrated energy system based on deep reinforcement learning in an uncertain environment[J]. Power System Protection and Control, 2022, 50(18): 50-60.

[19] VINYAL O, BABUSCHKIN I, CZARNECKI W M. Grandmaster level in StarCraft II using multi-agent reinforcement learning[J]. Nature, 2019, 575: 350-354.

[20] LOWE R, WU Y, 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, February 4-9, 2017, San Francisco, California, USA: 6382-6393.

[21] 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.

[22] 习伟, 李鹏, 蔡田田, 等. 基于深度强化学习的分布式电源就地自适应电压控制方法[J]. 电力系统自动化, 2022, 46(22): 25-31.
XI Wei, LI Peng, CAI Tiantian, et al. Adaptive local voltage control method for distributed generator based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2022, 46(22): 25-31.

[23] 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.

[24] 龚锦霞, 刘艳敏. 基于深度确定策略梯度算法的主动配电网协调优化[J]. 电力系统自动化, 2020, 44(6): 113-120.
GONG Jinxia, LIU Yanmin. Coordinated optimization of active distribution network based on deep deterministic policy gradient algorithm[J]. Automation of Electric Power Systems, 2020, 44(6): 113-120.

[25] WANG J, ZHANG Y, KIM, et al. Shapley Q-value: a local reward approach to solve global reward games[C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2019: 7285-7292.

[26] FOERSTER J, FARQUHAR G, AFOURAS T, et al. Counterfactual multi-agent policy gradients[C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2017: 3054-3063.

[27] 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.

[28] WANG J, XU W, GU Y, et al. Multi-agent reinforcement learning for active voltage control on power distribution networks[C] // Advances in Neural Information Processing Systems, December 6-14, 2021: 3271-3284.

[29] 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.
JU Yuntao, CHEN Xi, LI Jiawei, et al Active and reactive power coordinated optimal dispatch of networked microgrids based on distributed deep reinforcement learning[J]. Automation of Electric Power Systems, 2023, 47(1): 115-125.

[30] KHODR H M, OLSINA F G, JESUS O D, et al. Maximum savings approach for location and sizing of capacitors in distribution systems[J]. Electric Power Systems Research, 2008, 78(7): 1192-1203.

**收稿日期：2023-12-28；　　修回日期：2024-04-18**

**作者简介：**

张梓枭(2001—)，男，硕士研究生，研究方向为配电网优化运行、人工智能等；E-mail: zhangzx2023@tju.edu.cn

崔明建(1987—)，男，通信作者，博士，教授，博士生导师，研究方向为电力系统优化运行等。E-mail: mj_cui@tju.edu.cn

(编辑　许　威)