

基于强化学习 PER-MATD3 模型的智能配电网无功电压优化控制

周建华¹, 张睿凝², 陶锴², 徐俊俊²

(1. 国网江苏省电力有限公司电力科学研究院, 江苏 南京 211103;

2. 南京邮电大学 自动化学院, 江苏 南京 210023)

摘要:随着分布式新能源的高比例接入与新型电力电子器件的大规模应用,电压越限、波动等事故风险激增。因此,采用智能方法优化控制系统无功电压对保障智能配电网可靠性十分重要。提出一种基于优先经验回放(PER)-双重延迟确定性策略梯度(MATD3)的多智能体深度强化学习算法以优化无功电压。首先,将基于光伏逆变器的电压控制问题建模为分区马尔科夫决策过程。然后,提出PER采样方法,基于TD误差优化强化学习的采样操作。最后,利用MATD3策略开展模型求解,实现基于光伏逆变器的无功电压优化。基于开源光伏发电和负荷数据开展优化验证,对典型日场景开展电压优化实验。结果显示,相比于传统MATD3算法和MADDPG算法,所提算法平均电压偏差分别减小了原始结果的4.7%和15.5%,网损相较传统MATD3模型减小了1.6%,此结果验证了所提方法的控制效果和经济性。

关键词:智能配电网;光伏逆变器;无功电压优化;强化学习

中图分类号:TM28 **文献标识码:**A **DOI:**10.19457/j.1001-2095.dqed26578

Reactive Voltage Optimization and Control for Smart Distribution Networks Based on Reinforcement Learning PER-MATD3 Model

ZHOU Jianhua¹, ZHANG Ruining², TAO Kai², XU Junjun²

(1. Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 211103,

Jiangsu, China; 2. College of Automation, Nanjing University of Posts and

Telecommunications, Nanjing 210023, Jiangsu, China)

Abstract: The risk of voltage violations and fluctuations has surged due to the high penetration of distributed renewable energy and the large-scale application of new power electronic devices. Therefore, it is essential to optimize the control system of reactive voltage through intelligent methods for the reliability of smart distribution networks. A multi-agent deep reinforcement learning algorithm based on prioritized experience replay (PER) and multi-agent twin delayed deep deterministic policy gradient (MATD3) was proposed. First, the voltage control problem based on photovoltaic inverters was modeled as a partitioned Markov decision process. Then, the PER sampling method was used to optimize the reinforcement learning sampling operation based on TD error. Finally, the MATD3 policy was used to solve the model. The reactive voltage was optimized by means of photovoltaic inverter. Optimization validation was conducted using open-source photovoltaic generation and load data. The voltage optimization experiments were carried out using the data from typical days. Compared with the traditional MATD3 and MADDPG algorithms, the average voltage deviation decreases by 4.7% and 15.5%, respectively. Network losses are reduced by 1.6% compared to the traditional MATD3 model. These results validate the effectiveness and economic benefits of the proposed method.

Key words: smart distribution grid; PV inverter; reactive voltage optimization; reinforcement learning

基金项目:国家电网有限公司科技项目(J2024062)

作者简介:周建华(1983—),男,博士,高级工程师,主要研究方向为电力系统运行,Email: zhoujianhua83@126.com

通讯作者:陶锴(1993—),男,博士,副教授,主要研究方向为电网优化,Email: kai.tao@njupt.edu.cn

在“碳达峰、碳中和”战略目标的推动下,风、光等可再生能源的开发利用对能源结构转型十分重要^[1-2]。分布式光伏发电系统在智能配电网中的渗透率高^[3]。分布式光伏出力具有显著的随机性和波动性特征,易引发电压越限、波动等电能质量问题^[4-6]。这种不确定性源荷的大规模接入给配电网的安全稳定带来了严峻挑战^[7]。面对复杂的配电网拓扑结构和多变的运行工况,传统的电压控制方法难以满足实时控制需求^[8]。光伏逆变器电压控制中传统控制方法——下垂控制方法是分布式的,它通常只考虑局部电压变化与功率输出的关系,并不能全面考虑整个配电网的运行状态,难以实现智能配电网系统整体的无功电压优化。

近年来,人工智能技术在电力系统优化领域应用广泛,在负荷预测、电压控制以及故障诊断等方面都有好的应用效果^[9-10]。多智能体深度强化学习(multi-agent deep reinforcement learning, MADRL)在分布式决策和自适应学习方面具有优势。因此,在智能配电网无功电压优化领域有较好的应用。文献[11]利用MADDPG(multi-agent deep deterministic policy gradient)算法实现了基于光伏逆变器的电压实时控制。文献[12]设计了一种强化学习近端策略算法以调节各类环境下多能源微电网最优调度策略。文献[13]提出了一种基于混合整数优化方法和多智能体强化学习的双时间尺度混合电压控制策略。文献[14]将参数共享机制集成到延迟深度确定性策略梯度(TD3)算法中,用于克服多智能体环境的非平稳性。文献[15]提出了一种稳定性约束的强化学习方法,用于实时电压控制。文献[16]提出了一种多任务深度强化学习方法,将拓扑编码为深度强化学习的附加状态,实现了配电系统电压调节。

上述研究基于强化学习方法实现了电力系统电压优化与控制。然而,由于传统强化学习中的均匀采样策略将经验池中的所有样本视为同等重要,故智能体从中随机选择训练样本进行学习。在电力系统无功电压优化中,某些对于加速学习过程和提高控制性能具有重要作用的关键经验样本难以被该策略有效识别和充分利用,因此,传统强化学习方法的效率低。综上所述,传统强化学习方法在经验回放机制方面可被改进。

本文提出了一种基于优先经验回放机制(prioritized experience replay, PER)的改进型

MATD3算法。该算法通过构建优先经验采样策略,在采样优先级上考虑时间差分误差(TD-error)实现了经验回放过程的优化。首先,PER机制引入优先经验采样策略,可有效地捕捉关键状态转移信息,提升学习速度与效率。然后,将优先经验采样机制与MATD3算法结合,增强了策略的稳定性和控制精度。

1 分布式配电网电压优化问题建模

1.1 基于光伏逆变器的电压控制策略

光伏逆变器是分布式智能配电网中的关键设备。它不仅能实现光伏发电的并网运行,还可通过调节无功功率参与配电网的电压调控。为了保证光伏消纳率,根据不同的运行时段,光伏逆变器被设置为最大功率点跟踪(maximum power point tracking, MPPT)模式和静止同步无功补偿器(STATCOM)两种模式的自动切换。

1) 日间最大功率点模式。在光照条件下,即光伏发电不为0,逆变器将优先保证光伏系统工作在最大功率点。同时,利用剩余容量调节无功功率,参与电压调控。此时逆变器的视在功率约束为

$$S_{\max}^2 = P_{\text{pv}}^2 + Q_{\text{pv}}^2 \quad (1)$$

式中: $S_{\max}$ 为逆变器额定容量; P_{pv} 为光伏实际输出有功功率; Q_{pv} 为逆变器的无功功率。

2) 夜间静止无功补偿器模式。在无光照条件下,逆变器将自动切换至STATCOM模式。此时逆变器全容量会用于无功功率调节^[17],即

$$|Q| = S_{\max} \quad (2)$$

在配电网中,节点电压与功率之间的近似关系为

$$\Delta U = \frac{RP + XQ}{U_n} \quad (3)$$

式中: ΔU 为节点电压偏差; U_n 为额定电压; R, X 分别为等效电阻和电抗; P, Q 分别为注入节点的有功功率和无功功率。

由于配电网普遍具有大的 R/X 比值特性,节点电压同时受有功功率和无功功率的影响。光伏逆变器通过调节无功功率输出,可有效改变节点电压。增大容性无功功率输出可使节点电压上升,增大感性无功功率输出则使节点电压下降。为充分利用逆变器容量并确保系统安全稳定运行,需要合理分配无功功率调节容量。在MPPT模式下,最大无功功率为

$$Q_{\max} = \pm \sqrt{S_{\max}^2 - P_{\text{pv}}^2} \quad (4)$$

实际可调节无功功率范围为

$$Q_{\text{available1}} = \min \{Q_{\max}, Q_{\text{required1}}\} \quad (5)$$

式中: $Q_{\max}$ 为逆变器额定容量; $Q_{\text{required1}}$ 为所需要的无功功率。

在STATCOM模式下,电压调节所需的无功功率为

$$Q_{\text{available2}} = \pm S_{\max} \quad (6)$$

调节所需要的无功功率根据节点电压偏差确定,即

$$Q_{\text{required2}} = K_q (U_{\text{ref}} - U) \quad (7)$$

式中: K_q 为比例系数; U_{ref} 为电压参考值; U 为实际电压值。

通过上述控制策略,光伏逆变器可在不同工作模式下灵活调节无功功率输出,有效参与配电网的电压调控并提高配电网的电压质量。同时,该控制策略考虑了逆变器容量约束,可确保系统安全稳定运行。

1.2 分布式配电网无功电压控制模型

分布式配电网在日内利用逆变器进行无功功率调节,可快速控制电压,并尽量减少电网损耗,即

$$\text{Min} f = \text{Min} (\beta_1 \Delta U + \beta_2 P_{\text{loss}}) \quad (8)$$

$$\text{其中} \quad \Delta U = \sum_{t=0}^T \sum_{j=1}^M |V_{j,t} - U_n| \quad (9)$$

式中: P_{loss} 为网络损耗; β_1 为电压偏差权重系数; β_2 为网络损耗权重系数; $V_{j,t}$ 为时步 t 的 j 节点电压; M 为总节点数; T 为总指令周期数。

网络损耗具体公式为

$$P_{\text{loss}} = \sum_{t=0}^T \sum_{k=1}^W g_{k,t} \sigma \quad (10)$$

$$\text{其中} \quad \sigma = V_{a,t}^2 + V_{b,t}^2 - 2V_{a,t}V_{b,t}\cos\theta_{ab,t} \quad (11)$$

式中: $g_{k,t}$ 为时步 t 的支路 k 的电导; $\cos\theta_{ab,t}$ 为时步 t 的节点电压相角余弦。

以下约束共同定义了式(8)~式(11)的可行域,用于保证电气可实现性、运行合规性与设备能力边界:

1) 常规潮流约束:

$$\eta_1 = G_{ab} \cos \theta_{ab,t} + B_{ab} \sin \theta_{ab,t} \quad (12)$$

$$\eta_2 = G_{ab} \sin \theta_{ab,t} + B_{ab} \cos \theta_{ab,t} \quad (13)$$

$$P_{G,a,t} - P_{L,a,t} = U_{a,t} \sum_{j=1}^n U_{b,t} \eta_1 \quad (14)$$

$$Q_{G,a,t} - Q_{L,a,t} = U_{a,t} \sum_{j=1}^n U_{b,t} \eta_2 \quad (15)$$

式中: $P_{G,a,t}$, $Q_{G,a,t}$ 分别为节点 a 在时步 t 的有功功率与无功功率; $P_{L,a,t}$, $Q_{L,a,t}$ 分别为节点 a 的有功负荷与无功负荷; $U_{a,t}$, $U_{b,t}$ 分别为节点 a 和节点 b 在时步 t 的电压幅值; G_{ab} , B_{ab} 分别为节点 a 和节点 b 之间支路的电导和电纳; $\theta_{ab,t}$ 为节点 a 和节点 b 在时步 t 的电压相位差。

2) 电压幅值约束:

$$U_{\min} \leq U \leq U_{\max} \quad (16)$$

式中: $U_{\min}$, $U_{\max}$ 为节点电压的下限和上限。

3) 逆变器机组出力约束(限制光伏逆变器的有功功率和无功功率):

$$P_{\text{pv}}^{\min} \leq P_{\text{pv}} \leq P_{\text{pv}}^{\max} \quad (17)$$

$$Q_{\text{pv}}^2 \leq S_{\text{pv}}^2 - P_{\text{pv}}^2 \quad (18)$$

式中: $P_{\text{pv}}^{\min}$, $P_{\text{pv}}^{\max}$ 分别为逆变器有功功率的下限和上限; S_{pv} 为逆变器视在功率。

1.3 分区马尔科夫决策建模

将分布式配电网中每个逆变器都看作智能体,将控制模型建模为多智能体深度强化学习(MADRL)模型。按智能体将电网划分成多个区域,每个智能体只能观测到所在区域的环境状态。采用中心化训练分散式执行(centralized training decentralized execution, CTDE)的框架。在离线训练阶段,智能体之间可以实现信息交换,每个智能体都能获取全局信息。在线执行阶段,智能体仅需自身的局部观测信息,利用训练好的策略网络决策,实时读取电网当前状态,获得分布式实时调度策略。将数学模型改为马尔科夫决策模型用于强化学习^[17]。

1) 状态空间 S_t 。 S_t 为所有智能体在 t 时刻观测的集合,包含6个要素:

$$\{V_{j,t-1}, \theta_{j,t-1}, Q_{\text{PVi},t-1}, P_{\text{PVi},t-1}, P_{\text{LOAD},t}, Q_{\text{LOAD},t}\}$$

式中: $V_{j,t-1}$ 为在 $t-1$ 时刻第 j 个节点的电压幅值; $\theta_{j,t-1}$ 为电压的相角; $Q_{\text{PVi},t-1}$ 为在 $t-1$ 时刻第 i 个逆变器的无功调节量; $P_{\text{PVi},t-1}$ 为在 $t-1$ 时刻第 i 个逆变器的有功调节量; $P_{\text{LOAD},t}$, $Q_{\text{LOAD},t}$ 分别为系统在 t 时刻的有功负荷和无功负荷。

2) 动作空间 A 。动作空间中的动作 a_t 表示光伏逆变器的无功功率调节动作。对于每个时步 t ,智能体可以选择调节逆变器无功功率 $Q_{\text{PVi},t}$,动作空间定义为每个逆变器的无功功率调节范围应满足所有约束条件。

3) 联合动作空间 A_t (i 个逆变器):

$$\{a_{1,t}, \dots, a_{I,t}\}$$

式中: $a_{i,t}$ 为逆变器 i 在时步 t 的动作空间。

4) 奖励函数 $R(s_t, a_t)$ 。奖励函数用于最小化电压偏差和网络损耗,同时用于确保系统遵守潮流约束。

电压偏差是电压与目标值 1.0(标么值)的差。控制目标是将节点电压控制在 0.95(标么值)~1.05(标么值)范围内。网络损耗是指逆变器的无功功率输出的平方。损耗与逆变器无功调节量的平方成正比。奖励函数定义为

$$U_c = \sum_{t=0}^T \sum_{j=1}^M |V_{j,t,pu} - 1| \quad (19)$$

$$R(s_t, a_t) = -(\beta_1 U_c + \beta_2 W_{\text{loss}}) \quad (20)$$

式中: U_c 为奖励函数的电压偏差部分; $V_{j,t,pu}$ 为时步 t 的 j 节点电压的标么值; W_{loss} 为奖励函数的网损部分。

2 PER-MATD3 算法

PER-MATD3 算法框架如图 1 所示。

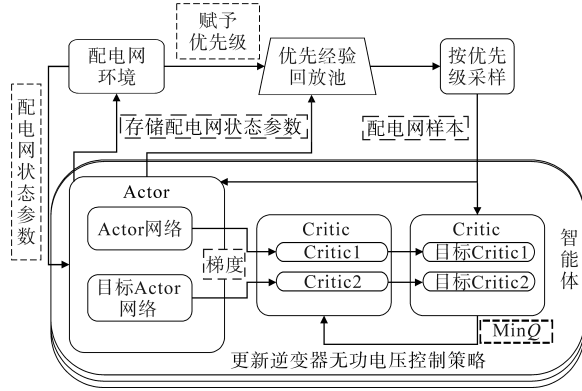


图1 PER-MATD3 算法框架图

Fig.1 The algorithm framework diagram of PER-MATD3

2.1 优先经验回放机制(PER)

优先经验回放机制是一种改进的经验回放策略,可以提高深度强化学习的样本效率和收敛速度。与传统的均匀采样经验回放方法不同,它计算了样本时间差分误差。每个经验样本赋予不同的优先级权重。PER 采用 TD-error 指标来评估优先级,该指标能反映当前 Q 值估计与目标 Q 值之间的差异。误差越大,表明该经验对模型更新越重要。TD 误差计算公式如下:

$$\delta_t = r_t + \gamma \cdot Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t) \quad (21)$$

式中: δ_t 为第 t 步的 TD 误差,表示当前 Q 值估计与目标 Q 值之间的差异; r_t 为在第 t 步获得的即时奖励; $Q(s_t, a_t)$ 为在状态 s_t 下采取动作 a_t 时的 Q 值估计(当前估计); $Q(s_{t+1}, a_{t+1})$ 为在下一状态 s_{t+1} 下采取动作 a_{t+1} 时的 Q 值估计(目标估计); γ 为折

扣因子,取值范围为 $0 \leq \gamma \leq 1$,用于衡量未来奖励的重要性。

为了在电网优化这一应用环境中充分考虑未来长期效果,确保强化学习模型能够充分考虑到未来的奖励或潜在的风险,优化系统的长期运行稳定性、效率和可持续性,本文算例中,经过实验迭代调整,设定折扣因子 γ 为 0.99。

通过这种优先级评估机制,PER 能够更有效地利用重要的经验样本。此外,可以加快学习速度,减少样本效率损失。在多智能体系统中,这种机制能够更好地捕捉智能体之间的复杂交互关系。

2.2 多智能体双重延迟确定性策略梯度算法(MATD3)

多智能体双重延迟确定性策略梯度算法是一种 Actor-Critic 结构算法。MATD3 通过引入 3 种关键技术来解决多智能体系统中的高方差问题:

1) 裁剪双 Q 学习:采用两个独立的 Critic 网络来评估每个智能体的动作。在计算目标 Q 值时取两个网络的最小值,从而有效降低 Q 值的高估问题,在 MADDPG 算法的基础上进一步增强稳定性^[18-19]。

2) 延迟策略更新:将 Actor 网络的更新频率设置为低于 Critic 网络,确保 Critic 网络在 Actor 更新时保持相对稳定,提高策略更新的可靠性。

3) 目标策略平滑正则化:在目标策略的输出动作中加入高斯噪声。通过平滑目标 Q 值来减小方差,提高算法的稳定性。

该方法通过目标 Critic 网络计算目标 Q 值,利用下一个状态和动作来计算下一步的 Q 值,再结合当前奖励和折扣因子迭代 Q 值,即

$$Q_t = r_t + \gamma \cdot \min(Q_{n1}, Q_{n2}) \quad (22)$$

式中: Q_{n1}, Q_{n2} 分别为两个目标 Critic 网络计算的下一状态 s_{t+1} 和下一动作 a_{t+1} 的 Q 值估计。

取两者的最小值是为了减小策略的高方差^[20]。而后,通过当前 Critic 网络计算当前的 Q 值,并与目标 Q 值对比,得到 TD 误差。

在计算出 TD 误差之后, Critic 网络的损失如下式所示:

$$Loss_{\text{Critic}} = (\delta_t)^2 \quad (23)$$

损失反映了当前 Q 值估计和目标 Q 值之间的差距。通过反向传播优化 Critic 网络参数以最小化损失,可以让 Critic 网络更准确地评估状态-动

作对的价值。在Critic网络优化后,更新Actor网络。Actor网络的目标是最小化 Q 函数^[21]。Actor网络的损失通过最大化Critic网络的 Q 值来实现。

2.3 PER-MATD3算法

PER-MATD3算法将优先经验回放机制与多智能体双重延迟确定性策略梯度算法相结合,网络核心结构包含以下3个组件:

1) Actor网络:负责根据当前状态输出动作,其策略更新基于Critic网络提供的 Q 值反馈。

2) Critic网络:用于评估状态-动作对的价值,采用双重网络结构以降低 Q 值估计的方差。

3) 优先经验回放池:存储智能体与环境交互产生的经验元组(状态、动作、奖励、下一个状态),并根据优先级计算公式对经验样本进行加权采样。

PER-MATD3的训练过程可以分为4个阶段:

1) 初始化阶段。在训练开始时,目标网络(target network)会被初始化为当前Actor和Critic网络的参数副本,并将用于目标 Q 值的计算。而后设置经验回放池的容量和优先级存储方式,为优先经验回放机制的实现做好准备,其中经验回放池用于存储智能体与电网的交互数据。在经验回放池中,相对重要的经验可以以更高优先级存储,确保在后续训练中被更多地使用,从而加快学习速度。

2) 经验收集阶段。强化学习的智能体会与配电网环境持续交互,不断收集经验。首先,智能体在每个时间步根据当前状态选择一个动作即调节逆变器的无功功率。接着,智能体执行该动作,配电网环境根据智能体执行的动作,返回一个新的状态和对应的奖励值,新的状态则反映了执行动作后的电网配置或状态变化。最后,将这一交互过程中的状态、动作、奖励、新状态以及“done”标志存入优先经验回放池。

3) 策略更新阶段。在经验数量达到一定阈值后,开始进行无功电压控制策略更新。当回放池中的经验数量达到一定阈值后,系统会开始执行策略更新过程。策略更新的核心步骤包括优先级采样、目标 Q 值计算、TD误差计算以及网络更新。首先,根据样本优先级从回放池中采样一批经验,得到对应配电网的状态、动作、奖励、下一个状态和“done”标志。优先级采样机制能够确保关键经验被优先利用,加速智能体的学习。然后更新目标网络参数,确保目标网络的权重缓

慢跟随当前网络。

4) 优先级更新。根据优先级评估机制计算优先级,每次策略更新后,系统会计算所有样本的TD误差并更新经验的优先级。该经验的权重越大,意味着该经验对于当前策略的影响越大,从而具有更高的优先级。对于配电网,TD误差较大的情况是智能体选择的动作没有有效提升电网稳定性或降低损耗,反之,TD误差小意味着智能体的动作对电网的控制已经比较精确。这一优先级更新过程确保经验回放池中的优先级能够动态反映智能体的学习状态。在后续的采样中,重要性高的经验被更频繁地利用,有助于智能体在短时间内更有效地调整电压控制策略。

PER-MATD3算法流程图如图2所示。

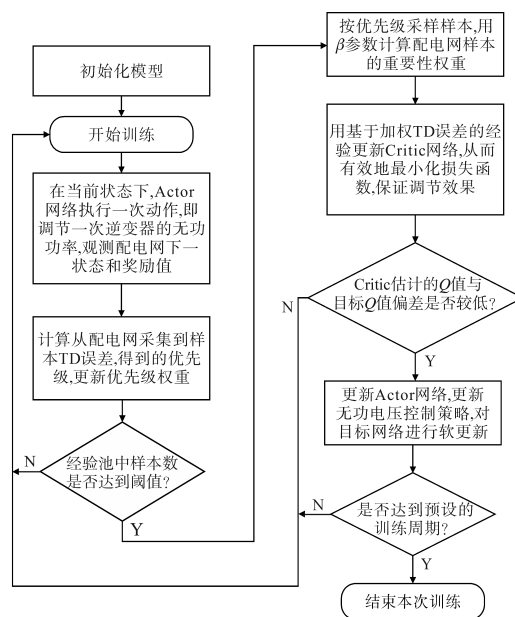


图2 PER-MATD3算法训练流程图

Fig.2 The training flowchart of the PER-MATD3 algorithm

3 算例分析

3.1 数据介绍与参数设置

本研究的训练和测试过程基于搭载 AMD Ryzen97945HX CPU, NVIDIA RTX4060 GPU 的硬件平台实现。算例使用修改后的 IEEE 33 节点案例, 节点系统的基准电压设定为 12.66 kV, 根节点的电压值为 1.0(标么值), 电网允许的最大电压偏差为 ± 0.05 (标么值), 在节点 6、节点 11、节点 16、节点 23、节点 26、节点 33 上安装逆变器容量为 1.5 MVA 的分布式光伏装置。本算例采用的光伏发电数据来自比利时电网的开源数据集, 所采用的负荷数据来自葡萄牙的开源负荷数据。数据

时间跨度为1 a,选取其中330 d的数据用于训练,其余数据用于测试。本文使用的光伏发电数据和负荷数据均来源于西欧地区^[22],都遵循欧洲电力系统(ENTSO-E)共同的技术标准和市场规则,在天气条件和季节性变化上也存在共性,因此在电网稳定性、负荷管理和光伏发电接入等方面具有相似的特性。在原数据的基础上利用模糊采样方法减少数据集维度,并进行缩放和插值计算^[23-25]。IEEE 33节点系统拓扑结构如图3所示。

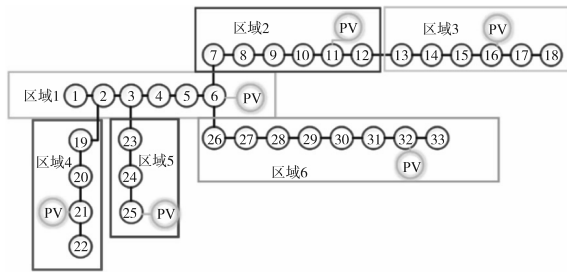


图3 修改后的IEEE 33节点系统拓扑结构

Fig.3 The modified IEEE 33-bus system topology

本次深度强化学习训练的总回合数设为600回合。每个回合随机选取半天的数据用于进行强化学习的训练。数据间隔为3 min。为了提高智能体决策的稳定性,将智能体动作时间间隔设置为15 min,即智能体每5个时步下发一个电压控制指令。为了模拟现实中分布式配电网的不确定性,系统在每回合开始前都会进行初始化。具体算法参数如表1所示。

表1 算法参数

Tab.1 Algorithm parameter

参数	取值
训练回合数	600
软更新系数	0.001
批尺寸	64
Actor网络学习率	0.000 1
Critic网络学习率	0.000 2
每回合步长	240
经验池容量	10 000
折扣因子	0.99
优先级程度 alpha	0.65

3.2 训练结果分析

训练过程中,每5回合计算奖励值曲线的平均值,结果如图4所示。由图4中可知,PRE-MATD3的模型收敛速度明显优于MATD3和MADDPG模型,训练结束时的奖励值也高于原模型。说明PER-MATD3在训练过程中的效果较好,相比传统的MATD3模型和MADDPG模型具有优越性。

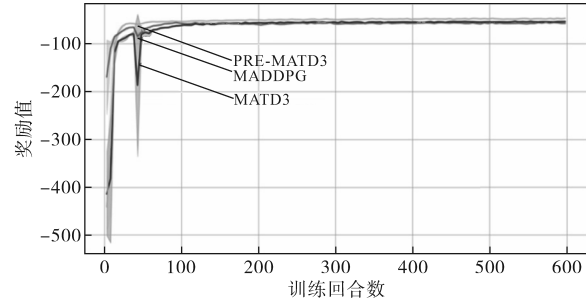


图4 训练过程奖励曲线

Fig.4 The reward curve of the training process

3.3 电压控制结果与讨论

取测试集中一典型日进行策略的对比分析。典型日光伏出力 and 负荷曲线如图5所示。

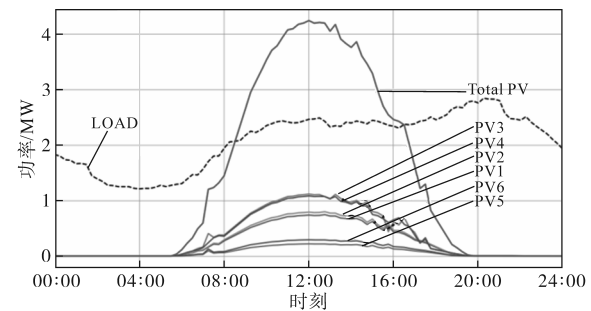


图5 典型日光伏出力和负荷曲线

Fig.5 Typical daily photovoltaic output and load curve

对比实验方案有:1)未经控制方案,即不对光伏逆变器施加任何无功控制策略。2)MADRL中的典型方法MADDPG^[26]。3)MATD3方法^[27]。4)本文所提的基于MATD3进行优化得到的PER-MATD3方法。当日系统原始数据最高电压出现在节点17,最低电压在节点18。在PER-MATD3调节后最高电压仍出现在节点17,最低电压在节点25。共选取节点17、节点18、节点25进行分析。不同电压控制策略下各个节点的电压标幺值如图6所示。

$Q-V$ 下垂控制算法作为传统的电压调节方法,广泛应用于配电网特别是微电网和分布式发电系统中。在光伏发电等可再生能源系统接入的情况下,下垂控制被认为是电力系统中非常典型且有效的无功电压调节方法。为了与传统电压控制方法开展对比,本研究采用 $Q-V$ 下垂控制算法进行电压控制,其节点17、节点18、节点25电压标幺值调控结果如图7所示。

由图6、图7可知,在接入逆变器后,电压波动较大。不加任何电压控制策略时,由于光伏发电过多,负荷无法消纳,从而发生潮流倒送导致部分节点出现电压越上限的情况。节点17和节

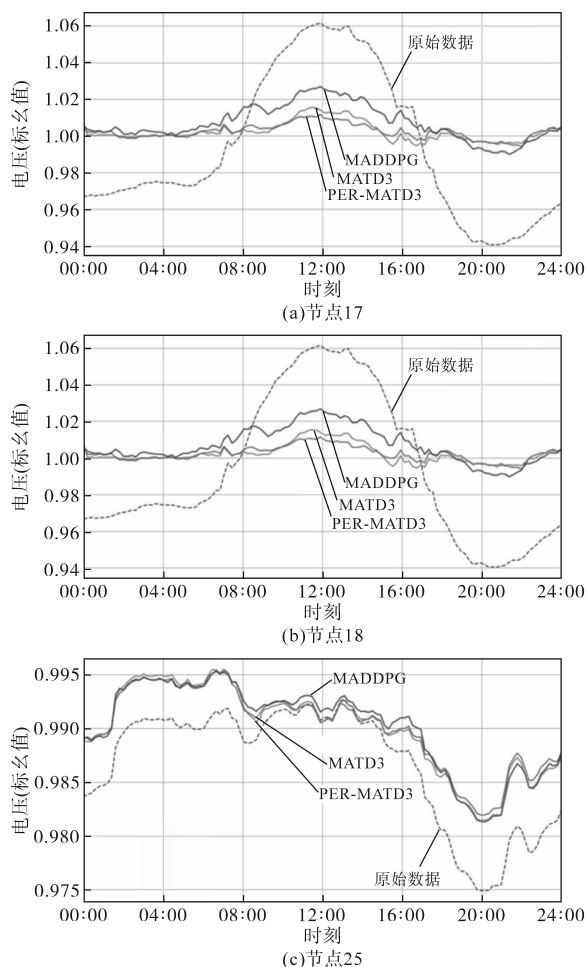


图6 典型日不同控制策略下部分节点电压

Fig.6 Voltage at some nodes under different control strategies on a typical day

点18在中午12:00时附近由于光伏出力超过负荷的需求,一直处于电压越上限状态。且节点17和节点18由于夜间的负荷需求量过大的原因发生了电压越下限。这给系统的安全稳定运行造成了危害。3种深度强化学习策略及传统 $Q-V$ 下垂控制策略都能较好地解决上述问题,将电压曲线维持在更加稳定的范围内。在所展示的典型日4个节点电压分布曲线图中,除了节点25由于调节前未出现电压越限现象,3种强化学习方法调节幅度较小且调节后的电压偏差几乎一致,下垂控制调节能力略差于3种强化学习算法,在其余两个节点上PER-MATD3算法的电压调节能力均优于其他3种算法。

典型日具体电压偏差及网络损耗等详细数据如表2所示。

平均电压偏差为一天所有时段各节点电压偏差的平均值。相比MADDPG算法以及优化前的MATD3方法,PER-MATD3算法除了克服了

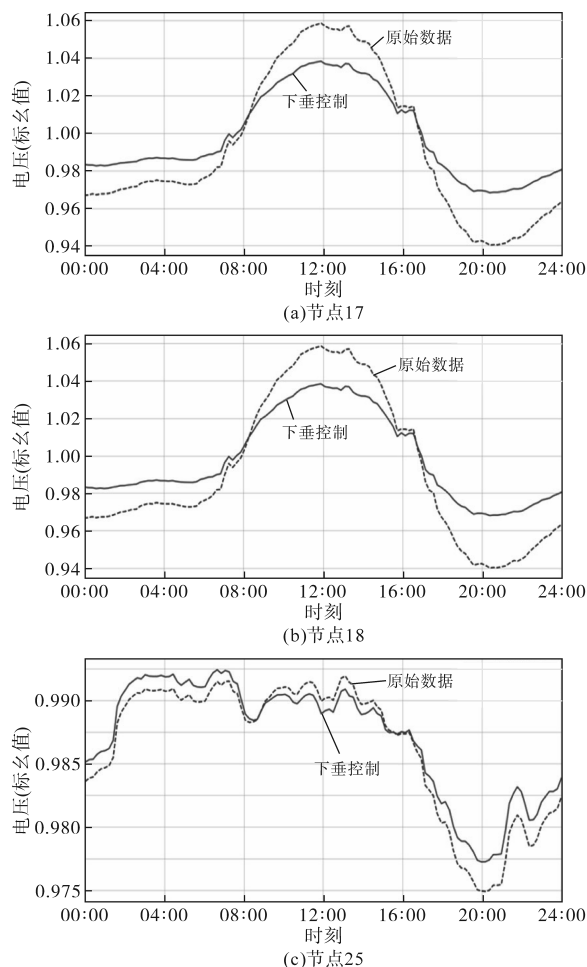


图7 典型日下垂控制策略下部分节点电压

Fig.7 Voltage at some nodes under droop control strategy on a typical day

表2 典型日平均电压偏差及网络损耗表

Tab.2 Table of average voltage deviation and network loss on a typical day

方法	平均电压偏差(标么值)	电压最大值(标么值)	电压最小值(标么值)	网络损耗/MW
原始数据	0.019 3	1.059 4	0.939 7	—
MADDPG	0.006 3	1.027 2	0.981 1	54.93
MATD3	0.004 2	1.016 7	0.974 8	58.21
PER-MATD3	0.003 3	1.013 0	0.983 0	57.28
下垂控制	0.013 3	1.040 8	0.959 0	44.11

MADDPG过估计的问题外,在原本的MATD3的基础上进一步提升了算法的控制效果。典型日中,对比MATD3算法和MADDPG算法,PER-MATD3方法的平均电压偏差分别降低了原始数据的4.7%和15.5%。对比传统的 $Q-V$ 下垂控制方法,PER-MATD3方法的平均电压偏差降低了原始数据的51.8%,传统 $Q-V$ 下垂控制方法的电压优化能力大幅度劣于PER-MATD3方法。

这说明本研究方法可有效地应对分布式电

源出力的波动,提高配电网的安全稳定性。

MADDPG算法和传统下垂控制方法在控制电压偏差方面的劣势大。PER-MATD3控制下的网损相比优化前MATD3方法降低了1.6%,略高于MADDPG算法和下垂控制方法。因此,PER-MATD3算法在综合降低电压偏差和网络损耗方面上相比其他方法有一定优越性。

4 结论

本文提出一种基于PER-MATD3方法的分布式配电网无功电压控制策略。在此方法中,应用优先经验回放机制结合MATD3算法。在训练过程中考虑TD误差给经验赋予不同的优先级。这使得模型在学习过程中更倾向于复用那些重要的经验,提高了模型的学习效率并优化了控制效果。在原模型的基础上,可以更好地实现对光伏逆变器的协同控制。通过对比实验可知,本方法在降低电压偏差方面具有较好的性能,在降低网络损耗这一次要目标方面仍有优化空间。

参考文献

- [1] 张彼德,邱杰,姜广鑫,等. 基于CNN和Transformer的轻量化电能质量扰动识别模型[J]. 电力工程技术, 2025, 44(1): 69-78.
- [2] 丁皓月,吕干云,史明明,等. 基于SCG优化SSAE-FFNN的电能质量复合扰动深度特征提取与分类[J]. 电力工程技术, 2024, 43(3): 99-110.
- [3] REN G, ZHA X G, JIANG B, et al. Location of multiple types of faults in active distribution networks considering synchronization of power supply area data[J]. Applied Sciences, 2022, 12(19): 10024.
- [4] 王守相,程耀祥,赵倩宇,等. 考虑光储协调的配电网多阶段就地-分布式电压控制策略[J]. 电力自动化设备, 2024, 44(1): 1-9.
- [5] LIU H Y, ZHANG C, CHAI Q M, 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.
- [6] 谢黎龙,李勇汇,范培满,等. 基于深度强化学习的孤立多微电网系统频率和电压综合控制[J]. 电力自动化设备, 2024, 44(6): 118-126.
- [7] XIE Lilong, LI Yonghui, FAN Peixiao, et al. Deep reinforcement learning-based integrated frequency and voltage control for isolated multi-microgrid system[J]. Electric Power Automation Equipment, 2024, 44(6): 118-126.
- [8] 寇晓适,王德琳,柯佳颖,等. 基于雷达回波与LSTM的输电线路强对流灾害风险预警方法[J]. 电力工程技术, 2024, 43(5): 189-198.
- [9] KOU Xiaoshi, WANG Delin, KE Jiaying, et al. Risk early warning method of severe convective disasters for transmission lines based on radar echo and LSTM[J]. Electric Power Engineering Technology, 2024, 43(5): 189-198.
- [10] 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.
- [11] ZHANG Xinxiao, 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.
- [12] NING J, XI C, XUE Q, et al. Transformer fault identification based on GWO-optimized dual-channel M-A method[J]. PloS One, 2024, 19(10): e0312474.
- [13] CHEN X, JI N, QIN X, et al. Transformer fault diagnosis based on the improved sparrow search algorithm and random forest feature selection[C]//International Conference on Energy and Electrical Power Systems, IEEE, 2024: 1086-1091.
- [14] LI C F, CHEN Y A, JIN C R, et al. Online PV smart inverter coordination using deep deterministic policy gradient[J]. Electric Power Systems Research, 2022, 209: 107988.
- [15] 李子凯,杨波,周忠堂,等. 基于强化学习算法的微电网优化策略[J]. 山东电力技术, 2024, 51(6): 27-35.
- [16] LI Zikai, YANG Bo, ZHOU Zhongtang, et al. Optimization strategy for microgrid based on reinforcement learning algorithm[J]. Shandong Electric Power, 2024, 51(6): 27-35.
- [17] HU D, 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.
- [18] XU J, GAO H J, WANG R J, et al. Real-time operation optimization in active distribution networks based on multi-agent deep reinforcement learning[J]. Journal of Modern Power Systems and Clean Energy, 2024, 12(3): 886-899.
- [19] FENG J, SHI Y Y, QU G N, et al. Stability constrained reinforcement learning for decentralized real-time voltage control[J]. IEEE Transactions on Control of Network Systems, 2023, 11(3): 1370-1381.

- [16] PEI Y S, ZHAO J B, YAO Y Y, et al. Multi-task reinforcement learning for distribution system voltage control with topology changes[J]. IEEE Transactions on Smart Grid, 2023, 14(3): 2481–2484.
- [17] YAO Z K, LIANG X L, JIANG G P, et al. Model-based reinforcement learning control of electrohydraulic position servo systems[J]. IEEE/ASME Transactions on Mechatronics, 2023, 28(3): 1446–1455.
- [18] LI J W, YU T, YANG B. A data-driven output voltage control of solid oxide fuel cell using multi-agent deep reinforcement learning[J]. Applied Energy, 2021, 304: 117541.
- [19] 邱建, 朱煜昆, 张建新, 等. 基于安全强化学习的电网稳控策略智能生成方法[J]. 电力系统保护与控制, 2024, 52(10): 147–155.
- QIU Jian, ZHU Yukun, ZHANG Jianxin, et al. Intelligent generation method of power system stability control strategy based on safe reinforcement learning[J]. Power System Protection and Control, 2024, 52(10): 147–155.
- [20] 梁涛, 柴露露, 谭建鑫, 等. 基于深度强化学习算法的氢耦合电-热综合能源系统优化调度[J]. 电力自动化设备, 2025, 45(1): 59–66.
- LIANG Tao, CHAI Lulu, TAN Jianxin, et al. Optimal scheduling of hydrogen coupled electrothermal integrated energy system based on deep reinforcement learning algorithm[J]. Electric Power Automation Equipment, 2025, 45(1): 59–66.
- [21] YAO Z K, LIANG X L, WANG S P, et al. Model-data hybrid driven control of hydraulic Euler-Lagrange systems[J]. IEEE/ASME Transactions on Mechatronics, 2025, 30(1): 131–143.
- [22] WANG J H, XU W K, GU Y J, et al. Multi-agent reinforcement learning for active voltage control on power distribution networks[C]//the 35th International Conference on Neural Information Processing Systems (NIPS '21), Article 250, 2021: 3271–3284.
- [23] TAO K, WANG Q, YUE D. Data compression and damage evaluation of underground pipeline with musicalized sonar GMM[J]. IEEE Transactions on Industrial Electronics, 2023, 71(3): 3093–3102.
- [24] TAO K, LIU T J, WANG Q, et al. A structural monitoring data processing model based on signal musicalization[J]. Measurement, 2022, 199: 111563.
- [25] TAO K, CHEN G, WANG Q, et al. Ultrasonic curved coordinate transform-RAPID with bayesian method for the damage localization of pipeline[J]. IEEE Transactions on Industrial Electronics, 2024, 71(11): 15066–15076.
- [26] LI P, WEI M J, JI H R, et al. Deep reinforcement learning-based adaptive voltage control of active distribution networks with multi-terminal soft open point[J]. International Journal of Electrical Power and Energy Systems, 2022, 141: 108138.
- [27] 全欢, 彭显刚, 刘涵予, 等. 基于深度强化学习的配电网实时电压优化控制方法[J]. 电网技术, 2023, 47(5): 2029–2039.
- QUAN Huan, PENG Xiangang, LIU Hanyu, et al. Voltage optimal control of distribution network based on deep reinforcement learning[J]. Power System Technology, 2023, 47(5): 2029–2039.

收稿日期: 2025-04-17

修改稿日期: 2025-05-27