

基于多智能体深度强化学习的实时配电网电压优化控制方法研究

周 荃

(国网江苏省电力有限公司丹阳市供电分公司)

摘要：本研究致力于配电网电压优化控制，特别是在动态和不确定性环境中。研究采用多智能体深度强化学习 (MADRL) 方法，并引入改进型双延迟深度确定性策略梯度 (TD3) 算法，克服了深度 Q 网络 (DQN) 和深度确定性策略梯度 (DDPG) 在高维问题处理和 Q 值估计中的局限性。通过仿真测试，研究展示了该方法在电压控制和降低网损方面的显著优势，特别在多变电网环境中表现出高效的适应性和鲁棒性。本研究不仅推进了 MADRL 技术在电力系统中的应用，也为实际电网运行提供了有效的控制策略，增强了电力系统的可靠性和稳定性。

关键词：配电网电压控制；多智能体深度强化学习；双延迟深度确定性策略梯度算法；电网鲁棒性

1 光伏系统的动态电压调控机制

1.1 光伏系统在电压管理中的作用原理

光伏系统在电压管理中起到关键作用，其调控机制主要依赖于可变无功功率。光伏系统的无功功率 $Q_{PV}^{\max}$ 与系统容量的关系遵循方程：

$$Q_{PV}^{\max} = \pm \sqrt{S_{inv}^2 - P_{PV}^2} \quad (1)$$

式中， S_{inv} 表示光伏逆变器的总容量； P_{PV}^2 表示光伏逆变器当前的有功功率输出。

在光伏系统达到标定有功功率 $P_{PV}^{\max}$ 时，其无功调节范围介于 $[-Q_{PV}^{\max}, Q_{PV}^{\max}]$ 。如需增加无功支持，有功功率可降至 P_{PV}' ，则无功能力增至 $[-Q_{PV}^{\max'}, Q_{PV}^{\max'}]$ 。光伏逆变器在日间通过最大功率点跟踪 (MPPT) 模式运行，夜间则切换至静止同步补偿器 (STATCOM) 模式，持续提供无功支撑，以在双碳战略和整县屋

顶光伏政策指导下最大化有功功率发电，同时调节无功输出^[1]。

1.2 电网内集成分布式能源的电压管理与优化模型

在配电网中，日内实时无功电压控制旨在通过协调无功调节装置，在日前调度计划的基础上执行更精细的短期调整。这种调整的目标是提升电压稳定性并减少网络损耗：

$$\min \sum_{i=0}^n \left(C_{Loss} P_{Loss,i} + C_a \sum_{j=1}^m \Delta U_{i,j} \right) \quad (2)$$

式中， n 表示节点的总数； C_{Loss} 表示单位电力损失的成本； $P_{Loss,i}$ 表示第 i 个节点或时间步的电力损失； C_a 表示电压偏差的成本系数； m 表示在每个节点或时间步考虑的电压测量点的总数； $\Delta U_{i,j}$ 表示第 i 个节点或时间步和第 j 个电压测量点之间的电压

偏差。

此模型的约束条件包括：

1) 电压限制

$$U^{\min} \leq U \leq U^{\max} \quad (3)$$

式中， $U^{\min}$ 表示电压的最小允许值或下限； U 表示当前或任一特定节点的电压水平； $U^{\max}$ 表示电压的最大允许值或上限。

2) 光伏功率范围

$$P_{PV}^{\min} \leq P_{PV} \leq P_{PV}^{\max} \quad (4)$$

式中， $P_{PV}^{\min}$ 表示光伏系统的最小有功功率输出； P_{PV} 表示当前的光伏系统有功功率输出； $P_{PV}^{\max}$ 表示光伏系统的最大有功功率输出。

3) 光伏系统容量和输出功率的关系

$$Q_{i,PV}^2 \leq S_{i,PV}^2 - P_{i,PV}^2 \quad (5)$$

式中， $Q_{i,PV}$ 表示第 i 个光伏逆变器的无功功率； $S_{i,PV}^2$ 表示第 i 个光伏逆变器的额定容量； $P_{i,PV}$ 表示第 i 个光伏逆变器的有功功率。

4) 电力平衡方程，确保每个时刻发电量与负荷量相匹配

$$\begin{cases} P_{G,i} - P_{L,i} = U_i \sum_{j=1}^n U_j (G_{ij} \cos \delta_{ij} + B_{ij} \sin \delta_{ij}) \\ Q_{G,i} - Q_{L,i} = U_i \sum_{j=1}^n U_j (G_{ij} \sin \delta_{ij} - B_{ij} \cos \delta_{ij}) \end{cases} \quad (6)$$

式中， $P_{G,i}$ 、 $Q_{G,i}$ 分别表示第 i 节点的有功发电和无功发电； $P_{L,i}$ 、 $Q_{L,i}$ 分别表示第 i 节点的有功负荷和无功负荷； U_i 、 U_j 分别表示第 i 和 j 节点的电压幅值； G_{ij} 、 B_{ij} 表示第 i 节点和 j 节点之间的电导和电纳； δ_{ij} 表示第 i 和 j 节点电压相角的差值。

2 配电网电压控制的多智能体深度强化学习方法

2.1 多智能体深度学习与强化学习的结合

在配电网电压控制领域，MADRL 方法的融合展现出显著的优势。深度强化学习（DRL）结合深度学习的数据识别与状态感知能力和强化学习的决策制定

优势，实现从策略空间的直接控制决策提取^[2]。特别是在 Actor-Critic（AC）框架下，深度策略梯度方法以其优化策略的高效性在电网调度等领域受到重视。

多智能体系统（MAS）中，合作型智能体通过协调互动形成协同控制策略，提升系统整体效益。MADRL 在离线阶段通过数据训练智能体，以应对在线执行阶段的实时决策挑战，有效处理配电网中分布式电源和负荷的不确定性问题^[3]。

2.2 配电网电压控制中的双延迟深度确定性策略梯度算法

为解决 DQN 在高维问题处理上的局限性及 DDPG 算法的过高估计问题，本研究采用了改进型 TD3 算法。TD3 适用于连续动作空间，采用 Actor-Critic 框架，高效解决无功电压控制问题。它通过最小化损失函数 L_θ 来调整策略，有效优化电网电压控制策略。

$$L_\theta = \mathbb{E} \left\{ \left[r_t + \gamma Q_\theta(s_{t+1}, \pi_\phi(s_{t+1})) - Q_\theta(s_t, a_t) \right]^2 \right\} \quad (7)$$

式中， $\mathbb{E}$ 表示计算损失函数的预期值； r_t 表示在时间 t 的即时奖励； γ 表示折扣因子； Q_θ 表示价值函数； s_{t+1} 、 s_t 分别表示在时间 $t+1$ 和 t 的状态。

Actor 则根据状态 s_t 输出动作 $\pi_\phi(s_t)$ ，并依据 critic 的 Q 值更新策略 ϕ 。actor 更新参数的策略梯度计算方法：

$$\nabla_\phi J(\phi) = \mathbb{E}_{s \sim p_\pi} \left[\nabla_a Q_\theta(s_t, a_t) \Big|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s_t) \right] \quad (8)$$

式中， $\nabla_\phi J(\phi)$ 表示关于策略参数 ϕ 的策略性能函数 J 的梯度； $\mathbb{E}_{s \sim p_\pi}$ 表示在策略 π 下的状态分布 p_π 中对状态 s 的期望值； $\nabla_a Q_\theta(s_t, a_t)$ 表示 Q 函数（价值函数）关于动作 a 的梯度，评估在当前状态 s_t 和动作 a_t 下，动作的微小变化如何影响预期的回报； $a=\pi(s)$ 表示根据策略 π 在状态 s 下选择的动作； $\nabla_\phi \pi_\phi(s_t)$ 表示策略函数 π 关于其参数 ϕ 的梯度。

TD3 算法通过引入裁剪双 Q 学习、延迟策略更新和目标策略平滑正则化三种技术手段，有效应对维

度灾难和次优解问题。具体而言，TD3 采用两套目标 critic 网络 Q_{θ_1} 和 Q_{θ_2} ，计算更新目标值时选取较小者作为时间差分（TD）目标，从而降低过高估计的风险。此外，在 TD 误差较小时更新 actor 网络参数，减缓学习速率，避免过度适应特定样本。引入的策略平滑正则化技术通过向动作增加高斯噪音，有效平均目标值，增强算法的泛化能力。

在中心化训练分散式执行（CTDE）框架下，TD3 算法通过离线训练中的信息交换和集中式学习优化多智能体决策。在线执行阶段，智能体依据局部观测信息作决策，降低计算复杂度。TD3 因此在处理电压控制等复杂问题中显示出卓越性能。

2.3 配电网无功电压控制的部分可观测马尔科夫决策模型

本研究将配电网无功电压控制建模为分布式部分可观测马尔科夫决策过程（Dec-POMDP），适用于多智能体深度强化学习（MADRL）。由于，光伏逆变器被视为独立的智能体，在通信限制下无法获取完整状态信息，因此 Dec-POMDP 比完全可观测的马尔科夫决策过程（MDP）更适用于 MADRL，Dec-POMDP 模型定义为元组 M 表示为：

$$M = \langle n, S, A, T, O, R, \gamma \rangle \quad (9)$$

式中， n 为参与决策的智能体数量； S 为状态集，包括所有智能体的状态信息如负荷的有功和无功、光伏逆变器的出力等； A 为联合动作集，包括所有智能体的动作； T 为状态转移概率函数，描述智能体的联合动作如何影响环境，从而确定下一状态的概率； O 为联合观测信息集，包含每个智能体的局部量测信息； O 为观测概率函数，描述智能体观测到特定信息的概率； R 为奖励函数，根据配电网无功电压控制目标函数设计的即时奖励函数，反映了网损和电压偏差。

配电网无功电压控制的部分可观测马尔科夫决策模

型中，智能体依据本地观测数据和动作结果做出决策，旨在最大化网损与电压偏差相关的奖励。该模型利用潮流计算模拟配电网运行，考虑光伏出力和负荷不确定性，指导智能体在每个时间步实现最优化决策^[5]。Dec-POMDP 模型的实施使智能体能够在部分可观测且动态变化的环境中有效学习，从而提高配电网无功电压控制的性能和效率。

3 算例仿真与分析

3.1 算例介绍和算法参数配置

本研究的仿真测试基于搭载 AMD Ryzen7-5800H CPU 和 NVIDIA RTX3050Ti GPU 的硬件平台，以改进版 IEEE 33 节点配电系统为测试案例。系统中第 13、18、22、25、29 和 33 号节点装有 1.5MW 容量的分布式光伏装置，如图 1 所示。基准电压等级为 12.66kV，根节点电压标么值设为 1.00pu，安全运行电压范围为 0.95pu 至 1.05pu。

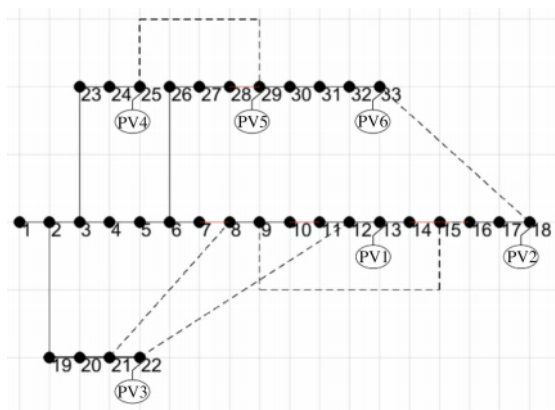


图 1 基于深度强化学习优化的 IEEE-33 节点配电系统结构图

在仿真环境设置中，本研究采用比利时电网的光伏数据和英国的负荷数据，通过加入标准差为 0.1 的高斯扰动，有效模拟了实际环境中的波动。该仿真环境包含一年数据，其中选取 30 天数据作为测试集，其余作为训练集。强化学习训练过程涵盖 400 回合，

每回合使用半天数据，智能体每 3min 做出一次动作决策。算法参数配置为：批尺寸 64，经验池容量 5000，折扣因子 0.99，步长 240，学习率 0.0001。该设置为模型训练提供了均衡的数据分布，有助于智能体更好地适应并优化电网运行状态。

此仿真环境旨在测试深度强化学习方法在实际电网电压控制中的应用效果，考虑光伏发电和负荷波动性，模拟电网运行的不确定性，为控制策略评估提供测试平台。

3.2 训练效果分析

在训练效果分析中，实验选取光伏出力最大的典型日进行验证，如图 2 所示，揭示了在 09:00~15:00 的高峰时间段内，光伏出力显著高于负荷需求。此现象导致功率倒送和电压越限问题，反映了模型中需进一步考虑这种波动性因素以确保电网的稳定运行。

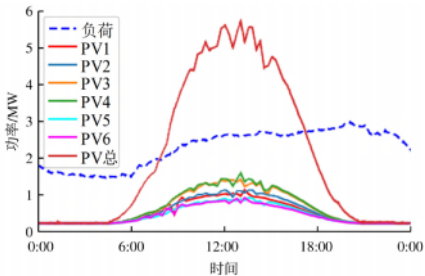


图 2 典型日配电网光伏发电与负荷变化曲线

经 MATD3 策略优化后，典型日测试显示各节点电压分布均处于安全运行范围内。曲线图进一步证实 MATD3 在电压控制方面的有效性，展示了优化后策略的显著控制效果，如图 3 所示。

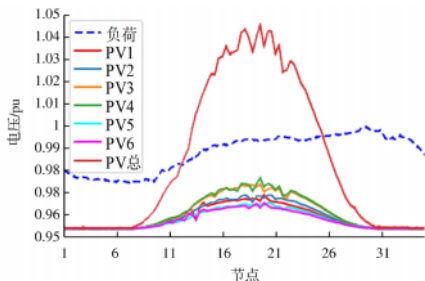


图 3 MATD3 策略优化下典型日配电网电压波动分析

3.3 对比分析

在本研究中，采用的 MATD3 策略与传统 Q-V 下垂控制策略、无控制策略及 MADDPG 策略进行对比。结果显示，MATD3 在电压控制和减少网损方面具有显著优势。例如，MATD3 策略在电压控制方面实现了最小的平均电压偏差 0.0069pu，计算效率为 36ms，优于其他策略。结果见表 1。

表 1 深度强化学习在配电网电压控制中的应用效能比较

控制策略	平均电压偏差 (pu)	电压范围 (pu)	计算效率 (ms)
原始方法	0.0196	0.9358~1.0659	—
下垂控制	0.0077	0.9737~1.019	69
MADDPG	0.0071	0.9706~1.0277	40
MATD3	0.0069	0.9753~1.0219	36

此外，MATD3 策略在网损减少方面也展现出最佳性能，相较于下垂控制和 MADDPG 方法分别减小了 5.29% 和 10.61% 的网损，证明了其在配电网无功电压控制领域的应用效能和经济性（见图 4）。通过这些对比分析，MATD3 策略被证实有效提升配电网的整体性能和稳定性。

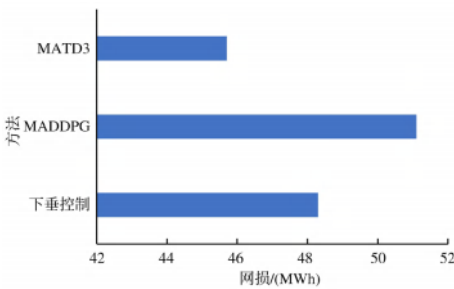


图 4 典型日内各控制策略对配电网网损影响分析

3.4 扩展性验证

通过在 141 个节点的配电系统进行仿真测试，成功验证了所提方法的扩展性（见图 5）。使用 Matpower 7.1 参数库和典型日的 PV、负荷曲线数据，确保了测试的准确性和一致性。

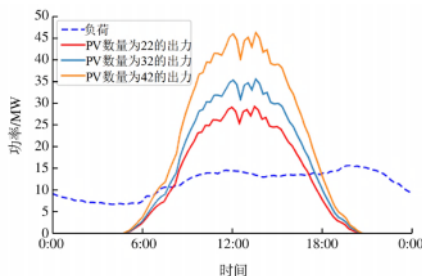


图5 141节点配电系统在典型日的光伏发电与负荷动态曲线

测试结果表明,如表2所示,在增加智能体数量的情况下,模型成功适应了更大的状态和动作空间。采用CTDE框架有效避免维度灾难问题,确保算法在复杂环境下的有效收敛,同时维持优秀的计算性能和电压控制效果。

表2 智能体数量对配电网电压控制性能影响的测试结果

智能体数目	电压合格率 (%)	平均电压偏差 (pu)	计算时间 (ms)
22	100	0.0323	56
32	100	0.0149	64
42	100	0.0273	66

3.5 鲁棒性验证

在IEEE 33节点系统的测试中,模型面对50%光伏渗透率场景展现出强鲁棒性。通过调整年度光伏最大出力与最大负荷之比,模型适应了3种不同高渗透率场景。测试结果证实模型能有效应对不同PV容量,显示出卓越的适应力和控制效果,如表3所示。

表3 光伏渗透率对电网电压稳定性和网损影响的测试分析

光伏渗透率 (%)	平均电压偏差 (pu)	电压合格率 (%)	系统网损 (MWh)
75	0.0203	100	20.21
100	0.0175	100	18.87
125	0.0161	100	19.64
150	0.0159	100	22.43

结果表明,随着渗透率增加,逆变器调节容量的提升显著增强了无功调节能力和系统电压质量。研究方法展现出对配电网环境的强泛化能力和鲁棒性,即在光伏出力显著增加的场景下,依然能有效实施高质量的电压控制策略。

4 结束语

本研究展示了MADRL在配电网电压优化控制中的有效性。通过实施改进型TD3算法,解决了DQN和DDPG算法在高维问题处理和Q值估计中的局限性。仿真测试显示,该方法能显著优化电压控制和降低网损,尤其在多变的配电网环境下表现出高效的适应性和鲁棒性。此研究不仅在理论上推进了MADRL技术在电力系统中的应用,也为实际电网运行提供了有效的电压控制策略,进而支持电力系统的可靠性和稳定性,贡献于电力工程领域的持续发展。

参考文献

- [1] 郑弘奇,江岳文,戴锦山.不确定环境下基于多智能体Q学习的海上风电输电工程电压调整降损优化[J/OL].中国电机工程学报:1-16[2024-04-10].<http://kns.cnki.net/kcms/detail/11.2107.tm.20230726.1702.006.html>.
- [2] 梅铭洋,寇鹏,张智豪,等.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[J].西安交通大学学报,2023,57(12):157-167.
- [3] 朱振山,张新炳,陈豪.基于深度强化学习的含智能软开关配电网电压控制方法[J].高电压技术,2023,50(3).
- [4] 齐韵英,许潇,殷科,等.基于深度强化学习的含储能有源配电网电压联合调控技术[J].电力建设,2023,44(11):64-74.
- [5] 廖思阳,皮山泉,徐箭,等.新型电力系统直控式负荷多层级协同调控关键技术综述[J].高电压技术,2023,49(9):3669-3685.

(收稿日期:2024-01-24)