

基于改进 TD3 算法的微电网群经济 优化调度研究

滕新亮, 田培文*, 徐建军, 韩司南

(东北石油大学 提高油气采收率教育部重点实验室, 黑龙江 大庆 163318)

Economic Optimal Scheduling of Microgrid Cluster Using Improved TD3 Algorithm

Teng Xinliang, Tian Peiwen*, Xu Jianjun, Han Sinan

(Laboratory of Enhanced Oil Recovery of Education Ministry, Northeast Petroleum University,
Daqing 163318, China)

Abstract: [Objective] To address the strong uncertainty in renewable energy output and load demand, complex operational constraints, and continuous decision variables in the cooperative scheduling of multi-microgrid systems, this paper proposes an economic optimization method based on an improved twin delayed deep deterministic policy gradient (TD3) algorithm. [Methods] Firstly, a three-microgrid cluster model was constructed, incorporating wind power, photovoltaic generation, micro-turbines, energy storage systems, EV charging loads, and main grid interactions. Its economic optimal dispatch process was modeled as a Markov decision process. The state space encompassed load demand, renewable generation, time-of-use electricity prices, and battery state of charge, while the action space comprised energy storage charge/discharge power, micro-turbine output, power exchange between microgrids, and interaction power with the main grid. Next, a reward function integrating normalized operating costs, environmental costs, and constraint penalty terms was designed to guide the agent in learning economical and feasible scheduling strategies. Finally, building on the original TD3 algorithm, a dual experience pool (DEP) mechanism and an adaptive exponentially decaying Gaussian exploration noise (AED-GEN) mechanism were introduced. The DEP mechanism classified experience samples into feasible solution samples and boundary exploration samples, improving sample utilization efficiency and enhancing the learning capability for

constraint boundaries. Meanwhile, the AED-GEN mechanism balanced exploration during early training and convergence stability in later stages. [Results] Case studies were conducted on a three-microgrid cluster. The results demonstrated that the improved TD3 algorithm outperformed the deep deterministic policy gradient (DDPG), original TD3, and SAC algorithms in terms of convergence stability, operational economy, and scheduling security. The standard deviation of reward fluctuation in the last 100 training episodes for the improved TD3 was 1.51, representing reductions of 35.74%, 45.09%, and 54.24% compared to SAC, original TD3, and DDPG, respectively. Its total daily operating cost was 4,186.94 yuan, which was 1.38%, 6.31%, and 8.71% lower than that of SAC, original TD3, and DDPG, respectively. Meanwhile, the action limit violation rate dropped rapidly during training and gradually approached zero in the later stages. Further comparisons of the improved mechanisms verified that both the adaptive exploration noise mechanism and the dual experience pool mechanism enhanced algorithm performance. [Conclusion] By satisfying the physical operational constraints, the proposed improved TD3 method leverages the mutual support capability among multiple entities within the microgrid cluster, thereby enhancing the economic efficiency and security of the system's collaborative scheduling.

Key words: microgrid cluster; improved twin delayed deep deterministic policy gradient algorithm; economic optimal dispatch; dual experience pool mechanism; adaptive exploration noise mechanism

基金项目: 黑龙江省重点研发计划(2024ZXJ01A04)

Provincial Key Research and Development Program of
Heilongjiang Province of China (2024ZXJ01A04)

摘要: [目的] 针对微电网群协同调度中可再生能源出

力和负荷需求不确定性强、运行约束复杂以及决策变量连续等问题,提出一种基于改进双延迟深度确定性策略梯度(TD3)算法的微电网群经济优化方法。【方法】首先,构建包含风电、光伏、微型燃气轮机、储能系统、电动汽车充电负荷及主网交互的三微电网群模型,并将其经济优化调度过程建模为马尔可夫决策过程。其中,状态空间包括负荷需求、可再生能源出力、分时电价和储能荷电状态等信息,动作空间包括储能充放电功率、微型燃气轮机出力、微网间交互功率及主网交互功率。其次,综合考虑归一化运行成本、环境成本和约束惩罚项构造奖励函数,引导智能体学习经济且可行的调度策略。最后,在原始 TD3 算法基础上引入双经验池(DEP)机制和自适应指数衰减高斯探索噪声(AED-GEN)机制。DEP 机制将经验样本划分为可行解样本和边界探索样本,以提高样本利用率并增强约束边界学习能力;AED-GEN 机制用于平衡训练前期探索能力和后期收敛稳定性。【结果】以三微电网群为对象进行算例分析,结果表明,改进 TD3 算法在收敛稳定性、运行经济性和调度安全性方面均优于深度确定性策略梯度(DDPG)、原始 TD3 和 SAC 算法。改进 TD3 算法最后 100 个训练回合的奖励波动标准差为 1.51,较 SAC、原始 TD3 和 DDPG 分别降低 35.74%、45.09% 和 54.24%;其日运行总成本为 4 186.94 元,较 SAC、原始 TD3 和 DDPG 分别降低 1.38%、6.31% 和 8.71%。同时,动作越限率在训练过程中快速下降,并在训练后期逐渐接近 0。改进机制对比结果进一步表明,自适应探索噪声机制和双经验池机制均能提升算法性能。【结论】本文所提改进 TD3 方法能够在满足物理运行约束的前提下,发挥微电网群内多主体的互济能力,提升了系统协同调度的经济效益与安全性。

关键词: 微电网群;改进双延迟深度确定性策略梯度算法;经济优化调度;双经验池机制;自适应探索噪声机制

0 引言

如今,风力、光伏(Photovoltaic, PV)发电为代表的可再生能源技术得到了快速发展,有效缓解了化石能源危机与环境污染问题^[1]。然而,分布式能源出力的间歇性和波动性又给传统电力系统的安全稳定带来了巨大挑战^[2-3]。微电网(Microgrid, MG)能够实现源网荷储的就地平衡,是解决分布式能源并网消纳问题的有效技术手段^[4]。多个微电网通过联络线互联形成的微电网群,能够通过区域内的功率支援与优势互补,提高可再生能源的利用率,增强供电可靠性,并降低整体运行成本^[5-6]。然而,作为典型的高维非线性强耦合系统,多微电网

面临着风光出力的随机波动、负荷动态变化及多主体间复杂能量交互等多重挑战^[7-8]。如何在源荷不确定性环境下,制定兼顾经济性与实时性的调度策略,已成为当前待攻克的难题。

针对微电网优化调度问题,传统研究多采用混合整数规划、模型预测控制和启发式优化等方法。文献[9]建立了适用于微电网日内调度的混合整数规划模型,通过对储能充放电过程、潮流约束进行线性化处理,提高了调度模型的求解稳定性与工程适用性。文献[10]提出基于模型预测控制的微电网多时间尺度协调优化调度方法,在日前阶段进行经济调度,在日内阶段通过滚动优化和反馈校正减小可再生能源预测误差对调度计划的影响。文献[11]以投资成本、电网电量不足期望及网损为目标,通过混合编码和罚函数处理约束,实现混合供电系统经济性与可靠性的综合最优。上述方法具有物理意义明确、约束表达直观等优点,但通常依赖精确模型和预测信息。当调度对象扩展至多微电网群时,系统状态和决策变量维度显著增加,传统优化方法在在线决策能力和不确定性适应性方面仍存在不足。

近年来,深度强化学习(Deep Reinforcement Learning, DRL)为不确定环境下的微电网在线调度提供了新的技术路径。DRL 将深度神经网络的非线性特征提取能力与强化学习的序贯决策能力相结合,能够通过智能体与环境的交互学习调度策略,减少对精确数学模型的依赖^[12-13]。从算法架构看,早期相关研究多采用值函数型方法。文献[14]利用深度 Q 网络(Deep Q-network, DQN)学习负荷、可再生能源出力和分时电价等状态信息,并通过经验回放和目标网络机制提升训练效果,实现了微能源网的经济运行。文献[15]进一步针对 Q 值估计偏差问题改进价值函数学习过程,提高了微电网储能控制策略的稳定性。然而,DQN 及其改进算法主要适用于离散动作空间,而微电网调度中的储能充放电功率、燃气轮机出力、微网间交易功率和主网交互功率均为连续变量。若将连续动作离散化,不仅会降低调度精度,还会导致动作组合规模迅速增加。

为适应连续动作调度问题,基于 Actor-Critic 架构的策略梯度方法逐渐应用于微电网能量管理。文献[16]采用深度确定性策略梯度(Deep

Deterministic Policy Gradient, DDPG) 算法, 通过差分状态、增量化动作定义及感知学习比例调整, 提升了调度策略对风、光、荷不确定性的自适应能力。DDPG 能够直接输出连续控制动作, 但其单 Critic 结构容易产生 Q 值过估计, 导致策略更新震荡甚至收敛困难。为缓解该问题, 双延迟深度确定性策略梯度 (Twin Delayed Deep Deterministic Policy Gradient, TD3) 算法通过双 Critic 网络、延迟策略更新和目标策略平滑机制提高了连续控制任务中的训练稳定性。文献[17]基于 TD3 算法开展微电网运行优化研究, 并通过在奖励函数中引入错误动作惩罚项限制设备出力, 从而降低越限风险。文献[18]基于柔性策略-评价网络提出微电网源储协同调度策略, 通过引入部分模型信息指导安全策略学习, 并结合神经网络对环境进行建模, 提高了算法训练效率和调度安全性。

综上, 现有 DRL 方法已在微电网能量管理中取得一定成效, 但在微电网群协同调度场景下仍存在连续动作处理、训练稳定性和约束安全性等

问题^[19]。因此, 本文在 TD3 框架下引入约束感知双经验池与自适应指数衰减探索噪声机制, 前者依据物理约束违背量对交互样本进行分流, 并通过动态混合采样兼顾可行高价值样本利用与边界风险识别; 后者根据训练进程自适应调整探索强度, 实现前期充分搜索和后期稳定寻优的平衡。本文所提改进 TD3 算法从样本组织和探索调节两个层面增强了智能体对微电网群可行运行区域的学习能力, 有助于提高调度策略的经济性、收敛稳定性和约束安全性。

1 微电网群结构

三个互联的微电网构成的微电网群, 如图 1 所示, 每个微电网主要由风力发电 (Wind Turbine, WT) 机组、PV 系统、微型燃气轮机 (Microturbine, MT)、日常负荷、储能以及电动汽车 (Electric Vehicle, EV) 组成。在微电网群系统中, 每个微电网可以向其余微电网进行购电或售电。配电网与微电网之间以及微电网之间通过通信线和电力线进行信号与电力传输。

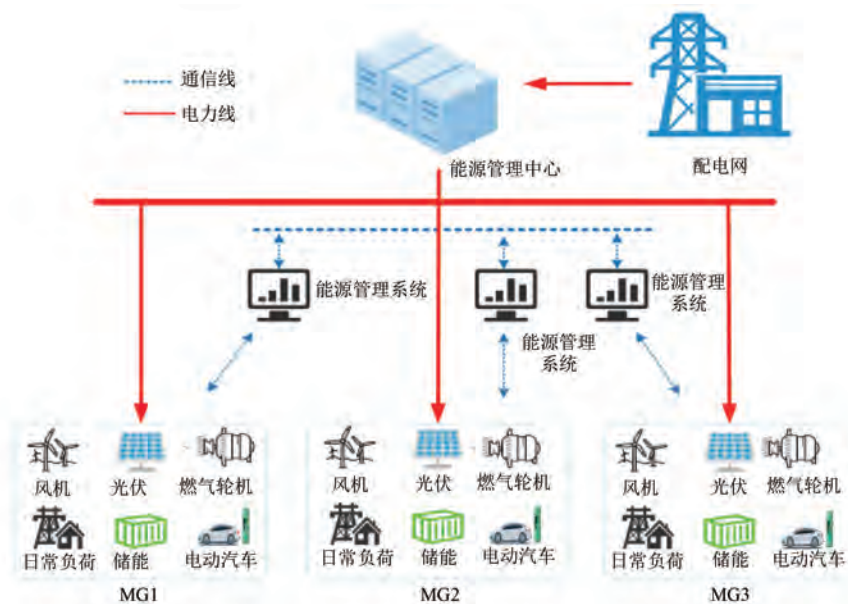


图 1 多微电网系统结构

Fig. 1 Structure of multi-microgrid system

2 微电网群系统内部单元模型

2.1 分布式发电设备模型

(1) PV 发电模型

PV 发电系统的输出功率主要取决于光照和

温度两个参数, 其关系式为

$$P_i^{PV}(t) = f_{PV} P_{STC} \frac{S_c}{S_{STC}} [1 + \eta_{PV}(T_c - T_{STA})] \quad (1)$$

式中: P_i^{PV} 为 t 时刻下第 i 个 MG 的 PV 发电机的输出功率; f_{PV} 为 PV 衰减系数; S_c 为辐射强度;

S_{STC} 为 PV 系统标准环境下的辐射强度; T_c 为当前光照强度; T_{STA} 为光照强度的基准值; η_{PV} 为温度系数。

(2) WT 模型

WT 系统的输出功率主要取决于当前时刻的风速,其关系式为

$$P_i^{\text{WT}}(t) = \begin{cases} 0, & 0 \leq v_t \leq v_{\text{in}} \\ P_r \times \frac{v_t^3 - v_{\text{in}}^3}{v_r^3 - v_{\text{in}}^3}, & v_{\text{in}} \leq v_t \leq v_r \\ P_r, & v_r \leq v_t \leq v_{\text{out}} \\ 0, & v_{\text{out}} \leq v_t \end{cases} \quad (2)$$

式中: P_i^{WT} 为 t 时刻下第 i 个 MG 风力发电机组的输出功率; v_t 为 t 时刻下发电机组的风速; P_r 为机组的额定输出功率; v_{in} 、 v_{out} 、 v_r 分别为机组的切入额定风速、切出风速、额定风速。

(3) MT 模型

当某时刻新能源出力不足以满足需求时, MT 出力来满足负荷用电需求,其数学模型为

$$f_m(t) = \begin{cases} \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left[-\frac{(t - \mu_m)^2}{2\sigma_m^2}\right], & \mu_m - 12 < t < 24 \\ \frac{1}{\sigma_m \sqrt{2\pi}} \exp\left[-\frac{(t + 24 - \mu_m)^2}{2\sigma_m^2}\right], & 0 < t < \mu_m - 12 \end{cases} \quad (5)$$

式中: t 为起始充电时间; μ_m 为起始充电时间的均值; σ_m 为起始充电时间的方差。

日行里程概率密度函数为

$$f_x = \frac{1}{s} \frac{1}{\sigma_n \sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu_n)^2}{2\sigma_n^2}\right] \quad (6)$$

式中: s 为里程; μ_n 为日行里程的均值; σ_n 为日行里程的方差。

EV 充电时长为

$$T_{\text{ch}} = \frac{s W_{100}}{100 P_{\text{ch}} \eta} \quad (7)$$

式中: T_{ch} 为充电时长; W_{100} 为百公里耗电量; P_{ch} 为充电功率; η 为充电效率。

2.4 目标函数

本文综合考虑系统的经济效益与环境效益,以微电网一天内的总运行成本最小化为目标,构建了微电网群经济优化调度模型。其中运行成本包括分布式发电单元的发电成本、储能运行成本、设备维护成本和电能交易成本;环境成

$$C_i^{\text{MT}}(t) = \frac{C_{\text{nl}}}{L} \times \frac{P_i^{\text{MT}}}{\eta_i^{\text{MT}}} \quad (3)$$

式中: C_{nl} 为天然气价格; L 为天然气低热值; P_i^{MT} 为 t 时刻下第 i 个 MG 的 MT 机组输出功率; η_i^{MT} 为第 i 个 MG 的 MT 机组发电效率。

2.2 储能设备模型

微电网中在用电需求低时储能将多余能量进行存储,在用电需求高时将能量释放,能有效地解决微电网中供需平衡问题。其数学模型为

$$C_i^{\text{BT}}(t) = \frac{C_{\text{cost}}}{P_R T} \times \frac{d(1+d)}{(1+d)^l - 1} P_{\text{BT}}(t) \quad (4)$$

式中: C_i^{BT} 为储能设备运行成本; C_{cost} 为储能设备总投资成本; P_R 为储能设备额定功率; T 为年运行小时数; d 为储能设备折旧率; l 为储能设备使用年限; P_{BT} 为充放电功率。

2.3 EV 负荷模型

基于蒙特卡洛法模拟 EV 负荷,其概率密度函数为^[20]

本为 CO_2 、 SO_2 、 NO_x 污染物气体的排放成本。目标函数为

$$\min F = F_1 + F_2 \quad (8)$$

2.4.1 运行成本

$$F_1 =$$

$$\sum_{t=1}^{24} \sum_{i=1}^3 [C_i^{\text{MT}}(t) + C_i^{\text{BT}}(t) + C_i^{\text{OM}}(t) + C_i^{\text{Grid}}(t)] \quad (9)$$

式中: C_i^{MT} 为 MT 发电成本; C_i^{BT} 为蓄电池(Battery, BT)的运行成本; C_i^{OM} 为发电设备维护成本; C_i^{Grid} 为电能交易成本。

(1) 设备维护成本

$$C_i^{\text{OM}}(t) = \sum_{j \in J} K_{\text{OM},j} P_i^j(t) \quad (10)$$

式中: K_{OM} 为各微电网设备的运行维护成本系数; $J = \{\text{PV}, \text{WT}, \text{MT}\}$ 。

(2) 电能交易成本

由于本文以微电网群整体运行成本最小化为目标,微电网之间的交易费用属于群内主体之间

的内部转移支付,在系统总成本统计中不重复计入,仅作为能量互济程度的分析指标。故单个微电网的交易结算成本如式(11)所示:

$$C_i^{\text{Grid}}(t) = P_i^{\text{Grid}}(t)\rho_{\text{Grid}}(t) + P_i^{\text{MG}}(t)\rho_{\text{MG}}(t) \quad (11)$$

式中: P_i^{Grid} 、 P_i^{MG} 分别为微电网与主电网、微电网与其他微网间的交互功率; ρ_{Grid} 、 ρ_{MG} 分别为主电网交互的等效电价、微网间内部交易电价。

2.4.2 环境成本

$$F_2 = 10^{-3} \sum_{i=1}^3 \sum_{t=1}^{24} \sum_{p=1}^3 [C_p \lambda_p^{\text{MT}} P_i^{\text{MT}}(t) + C_p \lambda_p^{\text{Grid}} P_i^{\text{Buy}}(t)] \quad (12)$$

式中: C_p 为污染物 p 的单位处理成本; λ_p^{MT} 、 λ_p^{Grid} 分别为 MT、配电网供电时产生的 p 类污染物的排放系数; P_i^{Buy} 为微电网从主电网购买的功率。

2.5 约束条件

2.5.1 功率平衡约束

$$P_i^{\text{PV}} + P_i^{\text{WT}} + P_i^{\text{MT}} + P_i^{\text{MG}} + P_i^{\text{disch}} + P_i^{\text{Grid}} - P_i^{\text{Load}} - P_i^{\text{ch}} = 0 \quad (13)$$

式中: P_i^{Load} 为总负荷大小; P_i^{ch} 、 P_i^{disch} 分别为储能的充、放电功率。

2.5.2 MT 爬坡约束

$$-R_{i,\text{down}}^{\text{MT}} \Delta t \leq P_i^{\text{MT}}(t) - P_i^{\text{MT}}(t-1) \leq R_{i,\text{up}}^{\text{MT}} \Delta t \quad (14)$$

式中: $R_{i,\text{up}}^{\text{MT}}$ 、 $R_{i,\text{down}}^{\text{MT}}$ 分别为 MT 的上、下爬坡速率。

2.5.3 PV 出力约束

$$P_{\min}^{\text{PV}} \leq P_i^{\text{PV}} \leq P_{\max}^{\text{PV}} \quad (15)$$

式中: $P_{\max}^{\text{PV}}$ 、 $P_{\min}^{\text{PV}}$ 分别为 PV 出力的上、下限。

2.5.4 WT 出力约束

$$P_{\min}^{\text{WT}} \leq P_i^{\text{WT}} \leq P_{\max}^{\text{WT}} \quad (16)$$

式中: $P_{\max}^{\text{WT}}$ 、 $P_{\min}^{\text{WT}}$ 分别为 WT 出力的上、下限。

2.5.5 储能系统约束

(1) 荷电状态(State of Charge, SOC):

$$\text{SOC}_{\min} < \text{SOC}_i < \text{SOC}_{\max} \quad (17)$$

(2) 充放电功率约束:

$$\begin{cases} 0 \leq P_i^{\text{ch}}(t) \leq P_i^{\text{max}} \\ 0 \leq P_i^{\text{disch}}(t) \leq P_i^{\text{max}} \\ P_i^{\text{ch}}(t) \times P_i^{\text{disch}}(t) = 0 \end{cases} \quad (18)$$

2.5.6 微电网之间交互功率约束

$$P_{\min}^{\text{MG}} \leq P_i^{\text{MG}} \leq P_{\max}^{\text{MG}} \quad (19)$$

式中: $P_{\min}^{\text{MG}}$ 、 $P_{\max}^{\text{MG}}$ 分别为 MG 之间交互功率的最小

值、最大值。

2.5.7 微电网与配电网之间交互功率约束

$$P_{\min}^{\text{Grid}} \leq P_i^{\text{Grid}} \leq P_{\max}^{\text{Grid}} \quad (20)$$

式中: $P_{\min}^{\text{Grid}}$ 、 $P_{\max}^{\text{Grid}}$ 分别为 MG 与配电网之间交互功率的最小值、最大值。

3 基于改进 TD3 算法的微电网群优化调度模型

针对多微电网系统中源荷不确定性强、调度维度高及动作空间连续等特点,传统的优化算法往往难以兼顾计算效率与策略精度。虽然常规的 DRL 算法能够应对高维连续问题,但在复杂环境中普遍面临着价值函数过估计及训练不稳定的瓶颈^[21-22]。因此,本文提出一种基于改进 TD3 算法的优化调度框架。通过改进经验回放机制与探索策略,提升算法在多微电网经济调度任务中的样本利用率与收敛性能,并将其构建为马尔可夫决策过程进行求解。

3.1 TD3 算法基本框架

TD3 算法是在 DDPG 算法的基础上改进的连续动作空间 DRL 算法,采用 Actor-Critic 架构,由策略网络 Actor 和价值网络 Critic 共同构成^[23-24]。Actor 根据当前状态输出连续控制动作, Critic 对动作进行价值评估,并指导策略网络更新。为缓解 DDPG 中存在的 Q 值过估计和训练不稳定问题,TD3 引入双 Q 网络、延迟策略更新和目标策略平滑正则化三项机制^[25-26],提高了连续控制任务中的学习稳定性和策略精度。

3.1.1 双 Q 网络

TD3 采用了两套单独 Critic 网络 Q1 和 Q2,分别计算下一个状态的动作值,并选取两个目标网络输出中的较小值解决 Q 值过估计问题^[27],如式(21)所示:

$$y = r + \gamma \min_{i=1,2} Q_{\theta_i}(s', a') \quad (21)$$

式中: r 为奖励; γ 为奖励折扣率; a' 为目标策略网络针对下一状态 s' 输出的动作; $Q_{\theta_i}(s', a')$ 为目标 Q 网络针对下一状态 s' 与目标动作 a' 的动作价值估计。

Critic 网络 Q1 和 Q2 均通过最小化预测值与目标值之间的均方误差来更新参数,设当前估计 Q 值为 (s, a) , Q 网络的损失函数为

$$L(\theta_i) = E_{(s,a,r,s') \sim D} \{ \{ Q(s,a) - [r + \gamma \min_{i=1,2} Q_{\theta'_i}(s',a')] \}^2 \} \quad (22)$$

式中: $L(\theta_i)$ 为第 i 个 Critic 网络的损失; $E_{(s,a,r,s') \sim D}$ 为采样批次数据的期望。

3.1.2 延迟更新

为解决 Actor-Critic 架构中策略网络 (Actor) 因频繁更新而导致的震荡问题, 算法降低了策略网络的更新频率, 当 Critic 网络更新一定步数之后 ($n=2$), 再更新 Actor 网络, 从而保证 Actor 网络的训练更加稳定^[28]。

Actor 的参数 ϕ 按照确定性策略梯度方向进行更新, 如式 (23) 所示:

$$\nabla_{\phi} J(\phi) \approx \frac{1}{N} \sum \nabla_a Q_{\theta_1}(s,a) \big|_{a=\pi_{\phi}(s)} \nabla_{\phi} \pi_{\phi}(s) \quad (23)$$

式中: $\nabla_{\phi} J(\phi)$ 为目标函数对 Actor 参数 ϕ 的梯度; N 为采样批次; $\pi_{\phi}(s)$ 为主 Actor 网络在状态 s 下输出的确定性动作。

与 Actor 更新同步目标网络参数 (θ', ϕ') 执行软更新:

$$\begin{cases} \theta'_i \leftarrow \tau \theta_i + (1 - \tau) \theta'_i \\ \phi' \leftarrow \tau \phi + (1 - \tau) \phi' \end{cases} \quad (24)$$

式中: τ 为软更新系数, 一般取 0.005。

3.1.3 目标策略平滑正则化

进一步地, 为了减少策略评估过程中的方差, TD3 算法在目标策略中引入噪声 ε ^[29], 如式 (25) 所示, 从而防止 Critic 网络对特定的动作值产生过拟合。

$$\varepsilon \sim \text{clip}[N(0, \sigma), -n, n] \quad (25)$$

式中: clip 为截断函数; n 为噪声截断阈值。

3.2 改进策略设计

传统 TD3 算法虽然在处理连续动作空间问题上表现优异, 但仍面临着样本利用率低和探索与利用难以平衡的问题^[30]。针对上述局限性, 本文在传统 TD3 算法的基础上对 TD3 算法进行改进, 以满足多微电网经济优化调度这一高维、强约束的复杂场景。

3.2.1 经验回放机制改进

传统 TD3 采用单一经验池进行均匀随机采样, 训练初期大量低质量或越限样本会降低样本

利用效率。为此, 本文构建双经验回放池, 将交互样本划分为可行解经验池 r_f 和边界探索经验池 r_b 。与优先经验回放主要依据 TD 误差确定采样优先级、分层经验回放主要按照任务阶段或样本类别组织经验不同, 本文双经验池机制面向微电网群调度中的物理约束特征, 以约束违背情况作为样本分流依据。可行解经验池存储满足主要运行约束且奖励较高的样本, 用于提高高质量样本利用率; 边界探索经验池存储越限或近边界样本, 用于保留动作越过储能 SOC、设备出力及交互功率等安全边界后的惩罚反馈信息。该类样本虽然奖励较低, 但能够帮助 Critic 网络识别不可行动作区域, 从而引导策略网络减少后续越限动作。因此, 双经验池机制能够兼顾可行样本利用和安全边界学习, 具体判定标准为

$$\sum_{k=1}^K \xi_k(s_i, a_i) = 0 \quad (26)$$

式中: $\xi_k(s_i, a_i)$ 为第 k 项物理约束的违背量, 若满足约束, 则 $\xi_k = 0$; 若违反约束, 则 $\xi_k > 0$ 。

在训练过程中, 算法不再进行简单的随机采样, 而是引入动态采样比例 η 。设总采样批次大小为 N , 则 r_f 采样的数量为 ηN , r_b 中采样的数量为 $(1-\eta)N$ 。为兼顾训练初期的边界探索能力和训练后期的高质量样本利用率, 将可行解经验池采样比例设计为随训练回合数逐渐增大的自适应参数 η , 其表达式为

$$\eta = \eta_{\min} + (\eta_{\max} - \eta_{\min})(1 - e^{-\lambda \eta^c}) \quad (27)$$

式中: $\eta_{\min}$ 为可行解经验池初始采样比例, 取 0.50; $\eta_{\max}$ 为可行解经验池采样比例上限取 0.75; λ_{η} 为采样比例增长系数, 取 0.003。

训练初期策略尚不稳定, η 接近 0.50, 使可行解经验池与边界探索经验池保持相对均衡采样, 有利于智能体学习约束边界; 随着训练回合数增加, η 逐渐趋近于 0.75, 以提高高质量可行样本的利用率。训练后期仍保留 25% 的边界探索样本, 避免完全削弱对约束边界的学习。

3.2.2 探索策略改进

传统的探索策略通常使用固定的方差 σ , 难以兼顾训练初期的全局搜索需求与后期的精度收敛需求。为此, 本文采用自适应指数衰减噪声机制, 使探索噪声随训练轮次逐渐减小。

该机制在训练初期增强动作探索能力, 在训

练后期减小扰动幅度,从而提升调度策略的稳定性和精度。其自适应探索噪声曲线如图 2 所示,将固定的方差 σ^2 改为随训练回合数 t 变化的动态参数 $\sigma_{\text{exp}}^2(t)$,改进后的动作生成表达式为

$$a_t = \text{clip} \{ \pi_{\phi}(s_t) + N[0, \sigma_{\text{exp}}^2(t)], a_{\min}, a_{\max} \} \quad (28)$$

式中: $\sigma_{\text{exp}}(t)$ 为标准差,其服从指数衰减规律,如式(29)所示:

$$\sigma_{\text{exp}}(t) = \sigma_{\min} + (\sigma_0 - \sigma_{\min}) \cdot e^{-\mu \cdot t} \quad (29)$$

式中: $a_{\max}$ 、 $a_{\min}$ 分别为动作空间的上、下边界; σ_0 为初始探索强度; $\sigma_{\min}$ 为最小维持噪声; μ 为衰减率。

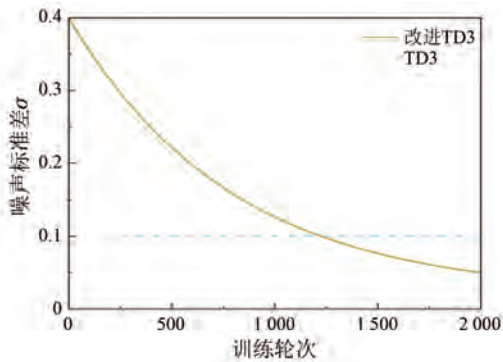


图 2 自适应探索噪声曲线

Fig. 2 Adaptive exploration noise curves

3.2.3 改进 TD3 算法基本框架

综上所述,基于上述算法改进策略,本文构建的基于改进 TD3 算法的多微电网群优化调度总体框架如图 3 所示。

图 3 展示了改进 TD3 算法在微电网群优化

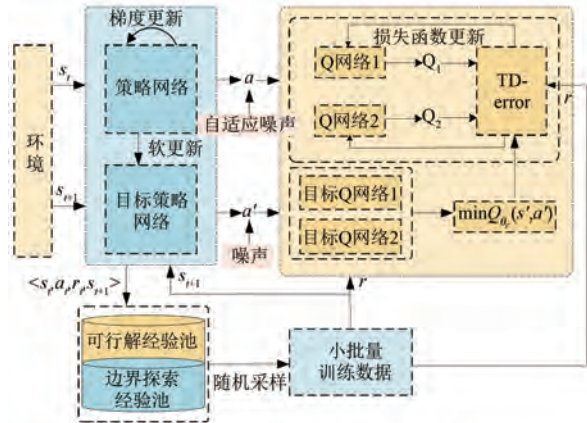


图 3 改进 TD3 算法微电网优化调度框架

Fig. 3 Framework of improved TD3 algorithm for microgrid optimal scheduling

调度中的整体流程。智能体根据当前状态 s_t 通过策略网络输出调度动作 a_t ,并在动作中加入自适应探索噪声,以增强训练初期的搜索能力;环境执行动作后反馈奖励 r_t 和下一状态 s_{t+1} ,形成经验样本并根据约束满足情况分别存入可行解经验池和边界探索经验池。训练时,算法按照动态采样比例从两个经验池中抽取小批量数据,用于更新双 Q 网络和策略网络,其中双 Q 网络用于抑制 Q 值过估计,目标网络软更新用于提高训练稳定性。通过上述流程,算法能够同时提升高质量样本利用率和约束边界学习能力,从而获得更稳定、可行的调度策略。

3.3 DRL 模型构建

基于上述微电网物理模型与改进 TD3 算法框架,将微电网经济调度问题建模为马尔可夫决策过程,智能体多次交互后可以产生如图 4 的序列^[31]。

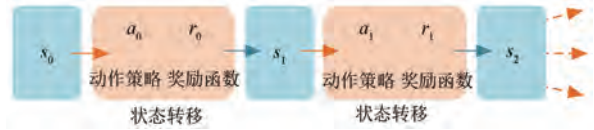


图 4 马尔可夫决策过程序列

Fig. 4 Markov decision process sequence

具体的状态空间、动作空间及奖励函数设计如下。

3.3.1 状态空间

t 时刻智能体观察到的环境状态变量主要包括负荷功率、可再生能源出力、储能 SOC、市场电价信息以及时间,即:

$$s_t = \{P_i^{\text{Load}}, P_i^{\text{WT}}(t), P_i^{\text{PV}}(t), \text{SOC}_i(t), \rho_i^{\text{MG}}, \rho_i^{\text{Grid}}, t\} \quad (30)$$

3.3.2 动作空间

多微电网系统选择的动作空间中主要包括储能充放电功率、MT 出力、微网之间交易功率、微网与配电网交易功率,即:

$$a_t = \{P_i^{\text{BT}}(t), P_i^{\text{MT}}(t), P_i^{\text{MG}}(t), P_i^{\text{Grid}}(t)\} \quad (31)$$

3.3.3 奖励函数

本文构建的奖励函数是在满足微电网物理约束的前提下,实现系统运行成本最小化。奖励函数由基础经济奖励与约束惩罚项两部分构成。为避免不同成本项数值尺度差异影响强化学习训练

稳定性,本文分别对运行成本和环境成本进行归一化处理,并采用等权重方式构造基础奖励函数。

(1)系统运行成本为目标函数时,奖励函数为

$$r_1(t) = -F_1(t) \quad (32)$$

环境成本为目标函数时,奖励函数为

$$r_2(t) = -F_2(t) \quad (33)$$

考虑综合效益时,奖励函数为运行成本和环境成本之和,奖励函数为

$$r(t) = - \left[\omega_1 \frac{F_1(t)}{F_{1,base}} + \omega_2 \frac{F_2(t)}{F_{2,base}} \right] \quad (34)$$

式中: ω_1 、 ω_2 分别为不同目标的权重, $\omega_1 = \omega_2 = 0.5$; $F_{1,base}$ 、 $F_{2,base}$ 分别为运行成本、环境成本的归一化基准,仅用于奖励函数数值尺度调整,不参与最终日运行总成本统计。

本文根据典型日运行成本量级,分别取值 5 000,1 000。最终算例中的日运行总成本仍按各项实际货币成本直接求和得到。

(2)约束惩罚机制

本文采用外点罚函数法设置约束惩罚项,并结合奖励函数归一化后的数值尺度,通过预试验确定惩罚因子 M 数值。当 M 取值过小时,越限动作受到的惩罚不足,智能体可能频繁输出违反 SOC 或交互功率约束的动作;当 M 取值过大时,奖励函数会过度受惩罚项主导,使策略过于保守,降低经济寻优能力。综合考虑动作越限率、奖励收敛稳定性和日运行成本,本文取 $M = 10$ 。

为防止储能电池过充或过放,当储能超出安全区间时需要施加惩罚项: $M \cdot \{ \max[g(x), 0] \}^2$, $g(x)$ 为约束违背函数,用于描述当前调度动作相对于安全运行边界的越限程度。当动作满足约束时,惩罚项为 0;当动作违反约束时,惩罚项随越限程度增大而增大,从而引导智能体减少不可行动作。储能惩罚函数由式(35)来表示,同时微电网与配电网的交互功率不超过联络线传输限值,需要施加惩罚项,如式(36)所示:

$$r_p^{SOC}(t) = \sum_{i=1}^3 M \cdot \{ \{ \max[SOC_i(t) - SOC_{max}, 0] \}^2 + \{ \min[SOC_{min} - SOC_i(t), 0] \}^2 \} \quad (35)$$

$$r_p^{Grid}(t) = \sum_{i=1}^3 M \cdot \{ \{ \max[P_i^{Grid}(t) - P_{max}^{Grid}, 0] \}^2 + \{ \min[P_{min}^{Grid} - P_i^{Grid}(t), 0] \}^2 \} \quad (36)$$

则总惩罚项为

$$r_p = r_p^{SOC}(t) + r_p^{Grid}(t) \quad (37)$$

总奖励为式(34)与式(37)之和,则:

$$R(t) = r(t) - r_p(t) \quad (38)$$

3.4 算法训练流程

结合上述改进策略,本文提出的改进 TD3 算法在多微网系统中的整体运行流程如图 5 所示。

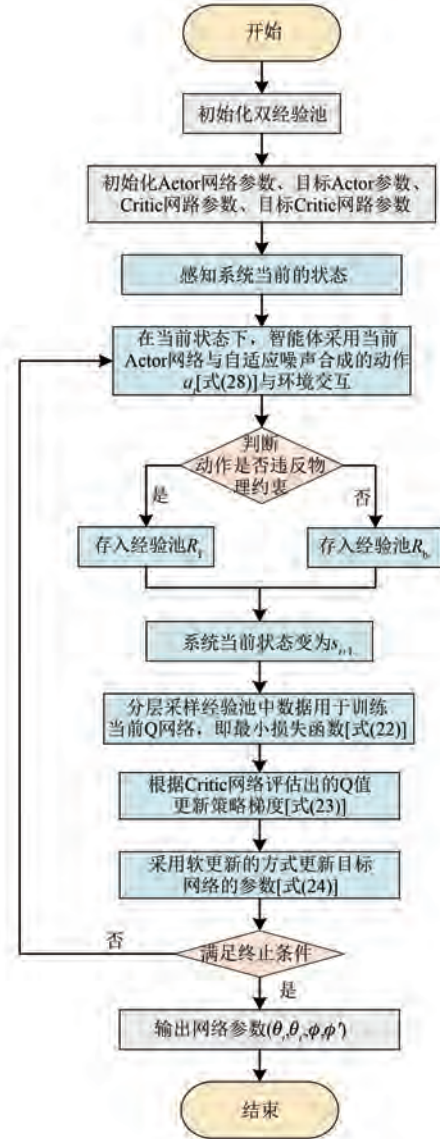


图5 算法流程图

Fig. 5 Algorithm flowchart

4 算例分析

本文选取包含 3 个互联微电网(MG1、MG2、MG3)的微电网群系统作为仿真对象,仿真采用的三微网群整体接线示意图如图 6 所示。三个微网均包含 WT、PV、MT、储能、EV 负荷和日常负荷,并

通过互联变换器接入公共联络线路,实现微网间能量互济及与配电网的功率交互。仿真步长为 1 h,调度周期为 24 h,以获得日调度结果和日运行成本。仿真试验在 Windows11 操作系统下完成,采用

Python 3.9 编程语言,并基于 PyTorch 2.8.0 深度学习框架搭建 Actor-Critic 神经网络模型。硬件平台配置为 NVIDIA GeForce RTX 5060 Laptop GPU,利用 CUDA12.8 实现神经网络训练加速。

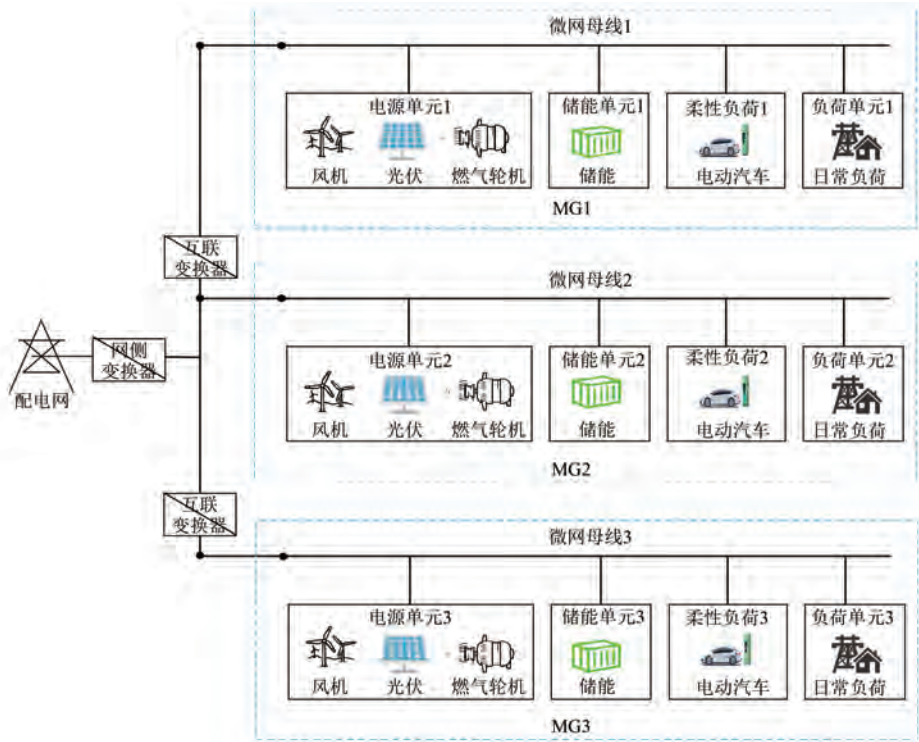


图 6 三微网群整体接线示意图

Fig. 6 Overall configuration diagram of the three-microgrid cluster

4.1 参数设置

改进 TD3 算法相关超参数如表 1 所示,污染物种类与治理费用以及排放系数如表 2 所示,微电网群内各设备的出力设定如表 3 所示,储能设备运行维护系数及参数如表 4 所示,分时电价如表 5 所示。

表 1 超参数设置

Tab.1 Hyperparameter settings	
参数名称	参数值
折扣率 γ	0.98
Actor 网络学习率 α	0.000 1
Critic 网络学习率 β	0.000 2
经验回放池容量 b	500 000
最小样本容量 K	256
混合采样比例 $\eta_{\min}, \eta_{\max}$	0.50, 0.75
采样比例增长系数 λ_{η}	0.003
探索噪声标准差 σ	0.20/0.05
延迟更新步数 n	2
软更新参数 τ	0.005
训练轮次 c	2 000

表 2 环境参数

Tab. 2 Environmental parameters			
污染物种类	治理费用/ (元·kg ⁻¹)	污染物排放系数/(g·kWh ⁻¹)	
		配电网	MT
CO ₂	0.22	870	730
SO ₂	15.32	1.87	0.038
NO _x	67.56	1.67	0.24

表 3 设备参数

Tab.3 Device parameters		
类型	参数名称	MG1/MG2/MG3
WT	出力上限/kW	225/210/280
	出力下限/kW	0/0/0
	运行维护系数	0.042/0.075/0.095
PV	出力上限/kW	150/190/230
	出力下限/kW	0/0/0
	运行维护系数	0.031/0.048/0.072
MT	出力上限/kW	110/105/85
	出力下限/kW	0/0/0
	爬坡速率/(kW·h ⁻¹)	40/40/30
	运行维护系数	0.081/0.072/0.065

表 4 储能参数

Tab. 4 Energy storage parameters

参数名称	参数值
额定容量/kWh	150
最大充放电功率/kW	50
充放电效率	0.9
折旧率	0.005
最大/最小电荷容量/kWh	140/30
使用年限 a	10
初始容量/kWh	75

表 5 分时电价

Tab. 5 Time-of-use electricity pricing

类别	售电/(元·kWh ⁻¹)			购电/ (元·kWh ⁻¹)
	峰段 9:00~11:00 18:00~22:00	平段 7:00~8:00 12:00~17:00	谷段 1:00~6:00 23:00~24:00	全段 0~24:00
配电网	1.45	1.13	0.55	
MG1	0.52	0.46	0.41	0.40
MG2	0.62	0.55	0.47	
MG3	0.57	0.52	0.43	

上述参数用于构建算法训练环境和经济调度目标,为后续算例分析提供统一仿真条件,确保不同算法在相同系统规模、设备约束及电价条件下进行比较。

4.2 源荷预测分析

图 7 为基于蒙特卡洛法得到的 EV 负荷预测曲线。由图 7 可知, EV 负荷整体呈双峰特征,分别对应日间出行后补电和傍晚集中充电时段。图 8 给出了典型日各微电网日常负荷及风光预测出力。由图 8 可知, PV 出力主要集中在白天,并在中午附近达到峰值; WT 全天波动,夜间出力相对较高;各微电网负荷呈多峰变化特征。

4.3 出力分析

图 9 为改进 TD3 算法下三个微电网的典型日优化调度结果。系统能够根据风光出力、负荷变化和分时电价,协调燃气轮机、储能、微网互济及主网购售电,各时段功率基本保持平衡。由图 9 可知, 1:00~6:00 为谷段, PV 出力为 0, WT 出力相对较高。此时系统优先利用 WT 盈余和低价主网电对储能充电,为后续高负荷时

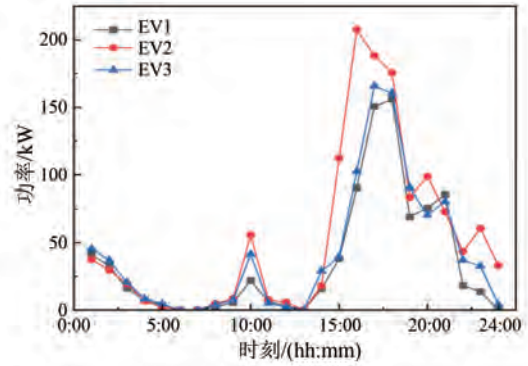
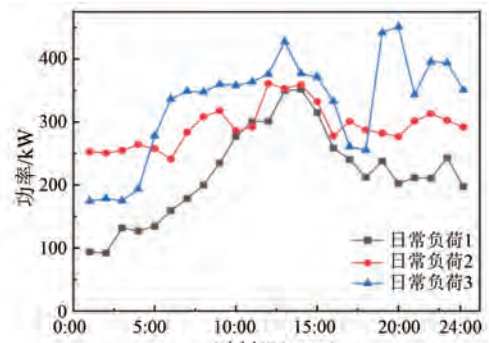
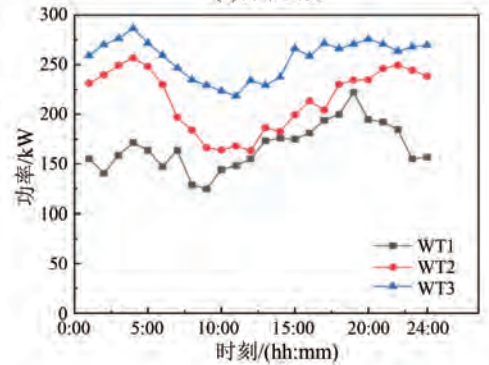


图 7 EV 预测曲线

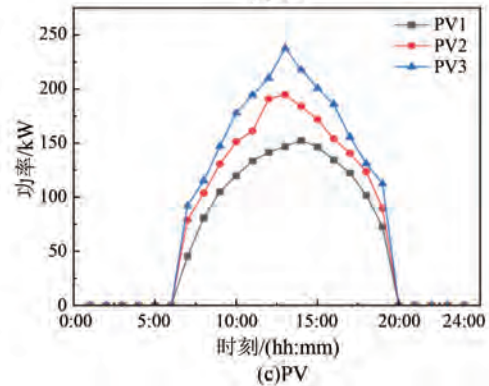
Fig. 7 The forecast curves of EV



(a) 日常负荷



(b) WT



(c) PV

图 8 微电网出力预测曲线

Fig. 8 Forecast curves of microgrid output

段提供备用能量。7:00~14:00 时 PV 逐渐升高,风光出力较充足,系统在满足本地负荷和储能充电需求后,将剩余电量通过微网互济或向主网售出;当系统整体富余时,MT 出力较低,避免了不必要的燃料消耗。15:00~17:00 负荷逐步升高,PV 出力开始下降,MT 开始爬坡并与储能共同补足系统缺额。18:00~22:00 为晚高峰阶段,PV 逐渐降为 0,负荷处于高位,MT 和储能成为主要调节资源。该阶段储能深度放电,配合 MT 降低峰段主网购电量,体现了削峰作用。23:00~24:00 重新进入谷段,储能利用低价电进行充电,最终各微电网 SOC 恢复至初始水平,保证次日调度连续性。

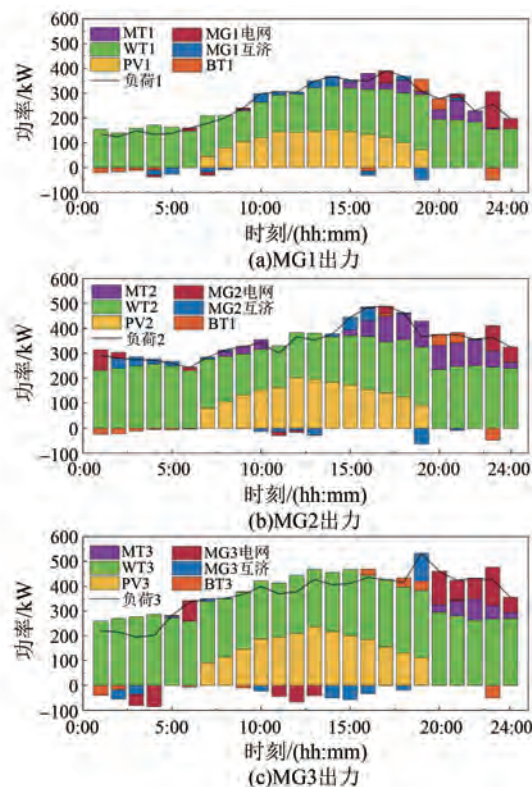


图 9 各微电网出力曲线

Fig. 9 Output curves of each microgrid

各微电网储能充放电功率以及 SOC 状态如图 10 所示。由图 10 可知,各 MG 的储能设备均具备初始储能。整体来看,各微电网储能 SOC 均保持在安全范围内,且调度末端均恢复至初始值,满足储能运行约束和日调度连续性要求。由图 10 可知,在 1:00~6:00 谷段及 WT 相对充足时段,MG1 和 MG3 储能优先吸收本地富余电能,SOC 逐步升高,为后续高负荷时段提供备用容量。

7:00~14:00 随着 PV 出力增加,系统新能源供应较充足,储能整体维持在较高 SOC 水平。15:00~22:00 负荷逐渐升高并进入晚高峰,储能开始放电参与削峰,其中 MG2 和 MG3 储能放电,在该阶段承担了较强的负荷支撑作用。23:00 后系统重新进入谷段,储能利用低价电进行充电,使各微电网 SOC 在调度周期末恢复至初始状态。由此可见,储能系统能够根据分时电价和源荷变化实现低谷充电、高峰放电的运行策略,提高了微电网群运行的经济性和灵活性。

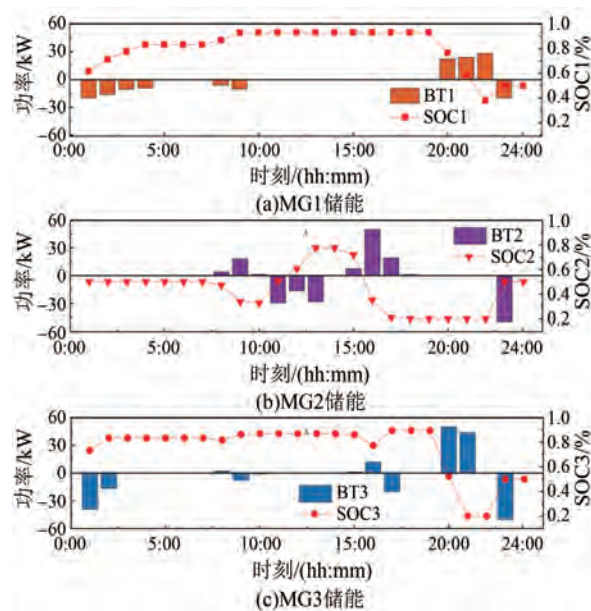


图 10 储能充放电功率以及 SOC

Fig. 10 Charging/discharging power and SOC of energy storage systems

4.4 算例对比分析

为验证本文所提改进 TD3 算法在微电网群能量管理中的有效性,将其与 DDPG、原始 TD3 和软演员-评论家 (Soft Actor-Critic, SAC) 算法进行对比。各算法均采用相同的微电网群调度环境、状态空间、动作空间、奖励函数和主要训练超参数。折扣率、学习率、经验池容量、批量大小、软更新系数和训练轮次等公共参数按照表 1 进行统一设置;对于不同算法特有参数,则按其原始机制设置。其中,DDPG 采用单 Actor-单 Critic 结构和固定高斯探索噪声;原始 TD3 采用单 Actor-双 Critic 结构,引入目标策略平滑和延迟策略更新机制,延迟更新步数为 2;SAC 采用最大熵随机策略,熵温度系数设置为 0.20;改进 TD3 在原始 TD3 基础上

进一步引入自适应指数衰减高斯噪声和双经验池动态采样机制。

4.4.1 训练结果对比分析

为降低随机初始化和探索噪声对训练结果的影响,本文对上述 4 种算法分别设置 5 组随机种子进行重复训练,随机种子取值为 0、1、2、3、4。图 11 给出了 4 种算法在某一典型随机种子下的奖励收敛曲线。由图 11 可知,训练初期各算法均处于探索阶段,奖励值波动较大;随着训练进行,奖励值逐渐上升并趋于稳定。其中,改进 TD3 算法收敛速度较快,后期奖励波动较小,说明所提双经验池机制和自适应探索噪声机制有助于提高样本利用效率和训练稳定性。

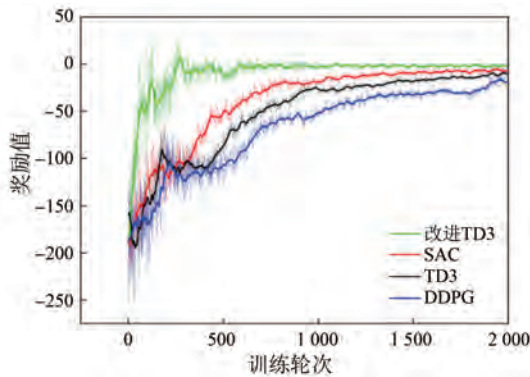


图 11 算法奖励收敛曲线

Fig. 11 Statistics of reward in the last 100 episodes for different algorithms

为进一步定量评价不同算法的收敛性能,表 6 统计了 5 组随机种子下各算法最后 100 个训练回合的奖励均值、种子间标准差和奖励波动标准差。

表 6 不同算法最后 100 回合奖励统计结果

Tab. 6 Reward statistics of different algorithms in the last 100 episodes

算法	均值	种子间标准差	奖励波动标准差
改进 TD3	-1.88	0.24	1.51
SAC	-6.51	0.38	2.35
TD3	-10.24	0.52	2.75
DDPG	-18.20	0.69	3.30

由表 6 可知,改进 TD3 算法的奖励均值最高,种子间标准差为 0.24,说明其在不同随机初始化条件下具有较好的鲁棒性;同时,其奖励波动标准差为 1.51,较 SAC、原始 TD3 和 DDPG 分别降低 35.74%、45.09% 和 54.24%,表明改进 TD3 在收

敛阶段具有更小的奖励波动和更好的训练稳定性。

4.4.2 算法越限率对比分析

为验证双经验池机制对调度安全性的提升,图 12 对比了两种算法的动作越限率。由图 12 可知,传统 TD3 算法采用单经验池均匀采样,优质经验易被海量试错数据淹没,导致越限率下降缓慢且剧烈震荡,训练末期仍高达 10% 左右,无法满足 5% 的安全容忍阈值。相比之下,改进 TD3 算法凭借双经验池与动态采样机制,越限率在 350 轮时便呈断崖式下降至 5% 左右,中后期稳定接近于 0%。这直观证明了改进机制能有效引导智能体快速规避违约动作,表明该机制能够有效提升微电网群调度策略的安全性与可行性。由于动作越限会引入约束惩罚项,越限率的降低有助于减小训练过程中的惩罚损失,从而与图 11 中改进 TD3 算法奖励值更快提升、后期波动更小的现象相对应。

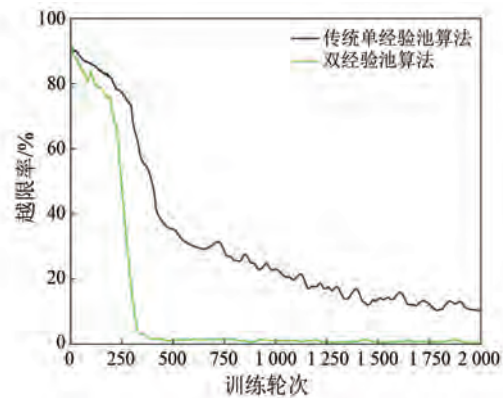


图 12 动作越限率对比

Fig. 12 Comparison of action violation rates

4.4.3 成本与计算时长对比分析

为验证改进 TD3 算法在多微电网协同调度中的可行性,图 13 给出了该算法与其他基准算法在系统日运行总成本及训练时长上的对比数据。

在系统运行经济性方面,改进 TD3 算法的经济表现得到了进一步优化。由图 13 可知,本文改进 TD3 算法的系统日运行总成本最低,为 4 186.94 元。具体构成如表 7 所示。相较于 SAC (4 245.67 元)、原始 TD3 (4 468.75 元) 以及 DDPG (4 586.32 元) 算法,改进 TD3 算法的日运行成本分别显著降低了 1.38%、6.31% 与 8.71%。这表明自适应衰减噪声与双经验池机制能够提高

策略搜索效率和约束满足能力,使算法在分时电价和多设备协同作用下获得更优的经济调度结果,从而利用分时电价机制下的经济寻优空间。

在计算效率方面,算法的耗时主要取决于其底层网络架构与数据处理逻辑的复杂度。DDPG 采用单 Actor-Critic 网络架构,反向传播的梯度计算量最小,因此训练耗时最短(1 800 s);原始 TD3 引入双 Critic 网络以缓解过估计问题,导致耗时大幅增加(1 980 s)。改进 TD3 算法的训练时长(2 130 s)略长于前两者,原因在于其在环境交互过程中,增加了双经验池的越限判定与分流抽样逻辑,以及自适应噪声的指数运算。尽管如此,相较于包含复杂对数与自动熵温调节运算的 SAC 算法(2 550 s),改进 TD3 依然保持了较高的计算效率。总体而言,本文算法在略微增加计算时长的情况下,降低了日运行成本,并提升了调度策略的物理安全性,展现出较好的工程实用价值。

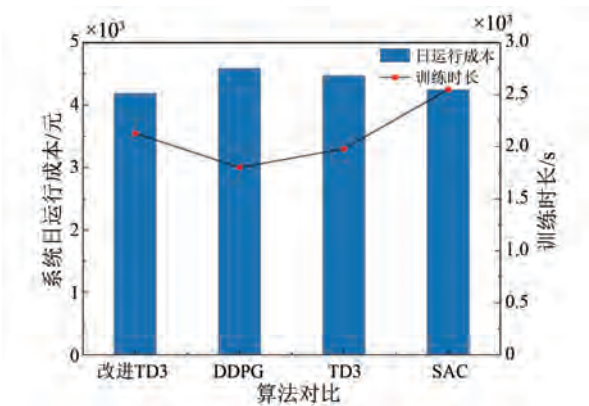


图 13 各算法成本与计算时长对比

Fig. 13 Comparison of cost and computation time of different algorithms

表 7 日运行总成本		
Tab. 7 Total daily operating cost		
成本项	数值/元	占比/%
主网购售电净成本	520.38	12.43
燃气轮机燃料成本	1 418.62	33.89
环境成本	624.30	14.91
储能充放电损耗成本	51.27	1.22
设备运行维护成本	1 572.37	37.55
日运行总成本	4 186.94	100.00

4.4.4 改进机制有效性分析

上述从整体上验证了改进 TD3 算法相较于

DDPG、原始 TD3 和 SAC 算法具有更好的收敛性能。为进一步分析自适应探索噪声机制和双经验池机制的作用,在原始 TD3 算法基础上设置不同改进机制下的对比试验。对比算法包括原始 TD3、仅引入自适应探索噪声机制的 TD3-AN、仅引入双经验池机制的 TD3-DER 以及同时引入两种机制的改进 TD3。各组算法采用与 4.4.1 节相同的训练环境、公共超参数和随机种子设置,以保证试验结果的可比性,具体结果如表 8 所示。

表 8 不同改进机制下的奖励统计结果

Tab. 8 Reward statistics under different improvement mechanisms

算法	均值	种子间标准差	奖励波动标准差
改进 TD3	-1.88	0.24	1.51
TD3-AN	-6.95	0.41	2.05
TD3-DER	-4.35	0.37	1.85
TD3	-10.24	0.52	2.75

由表 8 可知,TD3-AN 和 TD3-DER 相较原始 TD3 均表现出更高的奖励均值和更低的奖励标准差,说明自适应探索噪声机制和双经验池机制均对算法性能提升具有积极作用。其中,完整改进 TD3 算法的奖励均值最高、标准差最低,表明两种改进机制能够形成互补,提高算法的收敛稳定性和策略寻优能力。

图 14 给出了不同改进机制下的奖励收敛曲线。由图 14 可知,相比原始 TD3,TD3-AN 在训练后期的奖励波动有所减小,说明自适应探索噪声机制能够降低固定探索噪声带来的无效扰动,提高收敛稳定性;TD3-DER 的奖励收敛效果优于原

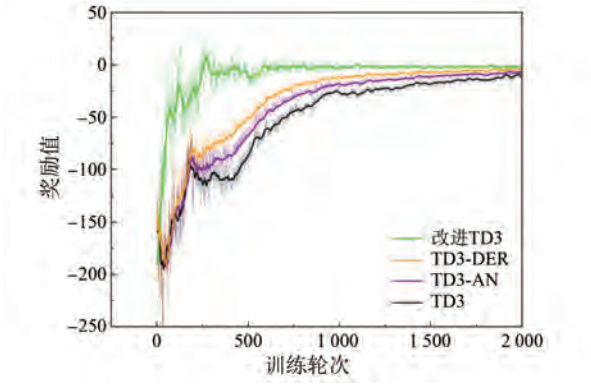


图 14 不同改进机制下的奖励收敛曲线

Fig. 14 Reward convergence curves under different improvement mechanisms

始 TD3,说明双经验池机制能够有效提高有效样本利用率,并增强智能体对约束边界的学习能力。完整改进 TD3 算法在收敛速度、最终奖励值和后期稳定性方面均表现最优,表明自适应探索噪声机制与双经验池机制具有协同作用。

5 结语

针对微电网群经济优化调度中源荷波动性强、运行约束复杂以及动作空间连续等问题,本文提出一种基于双经验池机制和自适应指数衰减高斯噪声机制的改进 TD3 算法,并将其应用于三微电网群经济调度。主要结论如下。

(1)改进 TD3 算法具有更好的收敛稳定性。多随机种子试验结果表明,改进 TD3 算法最后 100 个训练回合的奖励均值为-1.88,种子间标准差为 0.24,奖励波动标准差为 1.51,较 SAC、原始 TD3 和 DDPG 的奖励波动标准差分别降低 35.74%、45.09% 和 54.24%。

(2)两种改进机制均能提升算法性能。对比结果表明,TD3-AN 和 TD3-DER 的奖励波动标准差分别为 2.05 和 1.85,均低于原始 TD3 的 2.75,完整改进 TD3 的奖励波动标准差最低,说明自适应探索噪声机制和双经验池机制具有互补作用。

(3)改进 TD3 算法能够降低系统日运行总成本。算例结果表明,改进 TD3 算法下微电网群日运行总成本为 4 186.94 元,低于其他对比算法;训练时长低于 SAC 算法的 2 550 s,说明该算法在经济性和计算效率方面具有一定优势。

综上,本文所提改进 TD3 算法能够在提高训练稳定性和降低动作越限风险的基础上,实现微电网群运行成本的有效降低,具有一定的工程应用价值。

利益冲突声明

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

作者贡献

滕新亮、田培文进行了方案设计、内容总结与论文撰写,田培文、徐建军进行了试验研究,田培文、徐建军、韩司南参与了论文的审核与修改。所

有作者均阅读并同意了最终稿件的提交。

The scheme design, content summary and paper writing were carried out by Teng Xinliang and Tian Peiwen. The experiment was conducted by Tian Peiwen and Xu Jianjun. The manuscript was revised by Tian Peiwen, Xu Jianjun and Han Sinan. All authors have read and approved the final version of the paper for submission.

参考文献

- [1] 梁志峰,叶畅,刘子文,等. 分布式电源集群并网调控: 体系架构与关键技术[J]. 电网技术, 2021, 45(10): 3791-3802.
Liang Z F, Ye C, Liu Z W, et al. Grid-connected scheduling and control of distributed generations clusters: Architecture and key technologies [J]. Power System Technology, 2021, 45(10): 3791-3802.
- [2] 侯健生,陈新斌,王赢聪,等. 适用于微电网的 MGP 协调运行方法及装置研究[J]. 电机与控制应用, 2024, 51(8): 12-19.
Hou J S, Chen X B, Wang Y C, et al. Research on the coordinated operation method and devices of MGP suitable for microgrid [J]. Electric Machines & Control Application, 2024, 51(8): 12-19.
- [3] 李智,刘辉,刘明波,等. 虚拟同步发电机并网稳定性与控制技术研究综述[J]. 电机与控制应用, 2023, 50(11): 1-10.
Li Z, Liu H, Liu M B, et al. Research overview of grid-connected virtual synchronous generator stability and control technologies [J]. Electric Machines & Control Application, 2023, 50(11): 1-10.
- [4] 易文飞,朱卫平,郑明忠. 计及数据中心和风电不确定性的微电网经济调度[J]. 中国电力, 2024, 57(2): 19-26.
Yi W F, Zhu W P, Zheng M Z. Economic dispatch of microgrid considering data center and wind power uncertainty [J]. Electric Power, 2024, 57(2): 19-26.
- [5] 吴彦伟,姚刚,王海全,等. 基于能量共享的多微网协同优化调度[J]. 电机与控制应用, 2024, 51(8): 1-11.
Wu Y W, Yao G, Wang H Q, et al. Multi-microgrids cooperative optimal scheduling based on energy sharing [J]. Electric Machines & Control

- Application, 2024, 51(8): 1-11.
- [6] 桑博, 张涛, 刘亚杰, 等. 多微电网能量管理系统研究综述[J]. 中国电机工程学报, 2020, 40(10): 3077-3093.
- Sang B, Zhang T, Liu Y J, et al. Energy management system research of multi-microgrid: A review [J]. Proceedings of the CSEE, 2020, 40(10): 3077-3093.
- [7] Chen H P, Gao L, Zhang Z. Multi-objective optimal scheduling of a microgrid with uncertainties of renewable power generation considering user satisfaction [J]. International Journal of Electrical Power & Energy Systems, 2021, 131: 107142.
- [8] 葛磊蛟, 范延赫, 来金钢, 等. 面向低碳经济的人工智能赋能微电网优化运行技术[J]. 高电压技术, 2023, 49(6): 2219-2238.
- Ge L J, Fan Y H, Lai J G, et al. Artificial intelligence enabled microgrid optimization technology for low carbon economy [J]. High Voltage Engineering, 2023, 49(6): 2219-2238.
- [9] 孙浩, 张磊, 许海林, 等. 微电网日内调度计划的混合整数规划模型[J]. 电力系统自动化, 2015, 39(19): 21-27.
- Sun H, Zhang L, Xu H L, et al. Mixed integer programming model for microgrid intra-day scheduling [J]. Automation of Electric Power Systems, 2015, 39(19): 21-27.
- [10] 肖浩, 裴玮, 孔力. 基于模型预测控制的微电网多时间尺度协调优化调度[J]. 电力系统自动化, 2016, 40(18): 7-14+55.
- Xiao H, Pei W, Kong L. Multi-time scale coordinated optimal dispatch of microgrid based on model predictive control [J]. Automation of Electric Power Systems, 2016, 40(18): 7-14+55.
- [11] 盛万兴, 叶学顺, 刘科研, 等. 基于 NSGA-II 算法的分布式电源与微电网分组优化配置[J]. 中国电机工程学报, 2015, 35(18): 4655-4662.
- Sheng W X, Ye X S, Liu K Y, et al. Optimal allocation between distributed generations and microgrid based on NSGA-II algorithm [J]. Proceedings of the CSEE, 2015, 35(18): 4655-4662.
- [12] 范慧妍, 王爽. 基于深度强化学习的永磁同步电机控制算法研究[J]. 电机与控制应用, 2026, 53(3): 269-278.
- Fan H Y, Wang S. Research on control algorithm of permanent magnet synchronous motors based on deep reinforcement learning [J]. Electric Machines & Control Application, 2026, 53(3): 269-278.
- [13] 张有兵, 林一航, 黄冠弘, 等. 深度强化学习在微电网系统调控中的应用综述[J]. 电网技术, 2023, 47(7): 2774-2788.
- Zhang Y B, Lin Y H, Huang G H, et al. Review on applications of deep reinforcement learning in regulation of microgrid systems [J]. Power System Technology, 2023, 47(7): 2774-2788.
- [14] 刘俊峰, 陈剑龙, 王晓生, 等. 基于深度强化学习的微能源网能量管理与优化策略研究[J]. 电网技术, 2020, 44(10): 3794-3803.
- Liu J F, Chen J L, Wang X S, et al. Energy management and optimization of multi-energy grid based on deep reinforcement learning [J]. Power System Technology, 2020, 44(10): 3794-3803.
- [15] 梁宏, 李鸿鑫, 张华赢, 等. 基于深度强化学习的微网储能系统控制策略研究[J]. 电网技术, 2021, 45(10): 3869-3877.
- Liang H, Li H X, Zhang H Y, et al. Control strategy of microgrid energy storage system based on deep reinforcement learning [J]. Power System Technology, 2021, 45(10): 3869-3877.
- [16] 彭刘阳, 孙元章, 徐箭, 等. 基于深度强化学习的自适应不确定性经济调度[J]. 电力系统自动化, 2020, 44(9): 33-42.
- Peng L Y, Sun Y Z, Xu J, et al. Self-adaptive uncertainty economic dispatch based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2020, 44(9): 33-42.
- [17] 曾磊, 丁泉, 陈孝煜, 等. 基于深度强化学习的微电网运行优化方法[J]. 浙江电力, 2025, 44(6): 31-40.
- Zeng L, Ding Q, Chen X Y, et al. A DRL-based optimization method for microgrid operation [J]. Zhejiang Electric Power, 2025, 44(6): 31-40.
- [18] 刘林鹏, 朱建全, 陈嘉俊, 等. 基于柔性策略-评价网络的微电网源储协同优化调度策略[J]. 电力自动化设备, 2022, 42(1): 79-85.
- Liu L P, Zhu J Q, Chen J J, et al. Cooperative optimal scheduling strategy of source and storage in microgrid based on soft actor-critic [J]. Electric Power Automation Equipment, 2022, 42(1): 79-85.
- [19] 王亚东, 崔承刚, 钱申晟, 等. 基于深度强化学习的微电网储能调度策略研究[J]. 可再生能源,

- 2019, 37(8): 1220-1228.
- Wang Y D, Cui C G, Qian S S, et al. Study on micro-grid energy storage dispatching strategy based on deep Q-value reinforcement learning [J]. Renewable Energy Resources, 2019, 37(8): 1220-1228.
- [20] 肖仁鑫, 马四琳, 潘楠, 等. 考虑需求响应和 EV 有序接入的微电网双层优化调度[J]. 太阳能学报, 2026, 47(5): 123-134.
- Xiao R X, Ma S L, Pan N, et al. Bi-level optimization scheduling of microgrid considering demand response and ordered integration of EVs [J]. Acta Energiæ Solaris Sinica, 2026, 47(5): 123-134.
- [21] 李凯文, 张涛, 王锐, 等. 基于深度强化学习的组合优化研究进展[J]. 自动化学报, 2021, 47(11): 2521-2537.
- Li K W, Zhang T, Wang R, et al. Research reviews of combinatorial optimization methods based on deep reinforcement learning [J]. Acta Automatica Sinica, 2021, 47(11): 2521-2537.
- [22] 雷嘉明, 姜爱华, 吴新飞, 等. 计及源荷不确定性的综合能源系统近端策略优化调度[J]. 电力科学与技术学报, 2023, 38(5): 1-11.
- Lei J M, Jiang A H, Wu X F, et al. Proximal policy optimization dispatch of integrated energy system considering source-load uncertainty [J]. Journal of Electric Power Science and Technology, 2023, 38(5): 1-11.
- [23] 杨家令, 陈涛, 高赐威. 基于双延迟深度确定性策略梯度算法的微电网能源优化分配策略研究[J]. 电力需求侧管理, 2024, 26(4): 1-8.
- Yang J L, Chen T, Gao C W. Research on energy optimization allocation strategy for microgrids based on double delay deep deterministic strategy gradient algorithm [J]. Power Demand Side Management, 2024, 26(4): 1-8.
- [24] 宋金威, 何悠悠, 孙大帅, 等. 基于改进双延迟深度确定性策略梯度算法的微电网低碳经济优化调度策略[J]. 现代电力, 2026, 43(3): 455-463.
- Song J W, He Y Y, Sun D S, et al. A low-carbon economic optimization dispatching strategy for microgrids based on improved twin delayed deep deterministic policy gradient algorithm [J]. Modern Electric Power, 2026, 43(3): 455-463.
- [25] 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 J X, Liu Y M. Deterministic policy gradient algorithm coordinated optimization of active distribution network based on deep [J]. Automation of Electric Power Systems, 2020, 44(6): 113-120.
- [26] Liu J J, Meng X Y, Wu J T. Data-driven optimal scheduling for integrated electricity-heat-gas-hydrogen energy system considering demand-side management: A deep reinforcement learning approach [J]. International Journal of Hydrogen Energy, 2025, 103: 147-165.
- [27] Hasselt H. Double Q-learning [J]. Advances in Neural Information Processing Systems, 2010, 23.
- [28] 康朝海, 孙超, 荣垂霆, 等. 基于动态延迟策略更新的 TD3 算法[J]. 吉林大学学报(信息科学版), 2020, 38(4): 474-481.
- Kang C H, Sun C, Rong C T, et al. TD3 algorithm with dynamic delayed policy update [J]. Journal of Jilin University (Information Science Edition), 2020, 38(4): 474-481.
- [29] Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods [C]// International Conference on Machine Learning, PMLR, 2018: 1587-1596.
- [30] 邱革非, 何虹辉, 刘铠铭, 等. 基于改进 TD3 算法的综合能源系统低碳经济调度[J]. 电力科学与工程, 2023, 39(10): 52-62.
- Qiu G F, He H H, Liu K M, et al. Low-carbon economic dispatch of integrated energy system based on improved TD3 algorithm [J]. Electric Power Science and Engineering, 2023, 39(10): 52-62.
- [31] Domínguez-Barbero D, García-González J, Sanz-Bobi M Á. Twin-delayed deep deterministic policy gradient algorithm for the energy management of microgrids [J]. Engineering Applications of Artificial Intelligence, 2023, 125: 106693.

收稿日期:2026-04-30

收到修改稿日期:2026-06-02

作者简介:

滕新亮(1986—),男,博士研究生,研究方向为电力系统优化调度,408286173@qq.com;

* 通信作者:田培文(2000—),男,硕士研究生,研究方向为微电网优化调度,2351225035@qq.com。