

数据-知识融合驱动下光伏高渗透配电网电压/无功优化策略

吴明贺¹, 洪芦诚¹, 高远¹, 周爱华^{1,2}, 王栋³, 刘丰瑞⁴

(1. 东南大学电气工程学院, 江苏省南京市 210096; 2. 中国电力科学研究院有限公司, 北京市 100192;

3. 国网江苏省电力有限公司, 江苏省南京市 210009; 4. 智慧城市物联网全国重点实验室(澳门大学), 澳门 999078)

摘要: 海量分布式光伏出力的间歇性和波动性使配电网潮流呈强不确定性, 频繁电压越限与运行网损增大的问题愈发突出。为此, 文中提出一种基于数据-知识融合驱动的分布式光伏高渗透配电网电压/无功优化策略。首先, 针对配电网潮流状态的强不确定性和高动态特性, 提出一种改进的分布式软 Actor-Critic(DSAC)算法, 通过直接学习状态-动作回报的分布函数来探索累积折扣回报的不确定性, 缓解了状态-动作价值的高估问题, 并显著提升了电压/无功优化策略的性能; 其次, 为解决电压约束容易违反的问题, 在数据驱动的流程中嵌入基于 DistFlow 方程的专家知识安全层, 通过对 DSAC 智能体输出动作进行物理知识投影与矫正, 实现数据学习与知识约束的深度融合, 提高策略的可行性与安全性; 然后, 提出一种基于多元条件 t 分布 Copula 的数据增强模型, 通过生成高质量、多样化的训练数据, 显著提升了 DSAC 策略在复杂运行场景下的泛化能力与鲁棒性; 最后, 通过 IEEE 34 节点和 IEEE 118 节点系统的仿真结果验证了所提策略的有效性。

关键词: 配电网; 电压/无功优化; 光伏高渗透; 数据-知识融合驱动; 深度强化学习

0 引言

在双碳目标稳步推进及新型电力系统建设背景下, 配电网中分布式光伏(photovoltaic, PV)装机规模高速发展。截至 2024 年底, 中国分布式 PV 并网容量已达到 370 GW^[1]。配电网中高渗透分布式 PV 出力的随机波动特性导致潮流状态复杂多变, 使得配电网运行网损增大和频繁电压越限问题愈发严重, 成为影响新能源消纳和配电网安全经济运行的关键问题^[2-3]。同时, 分布式 PV 和多类型负荷发用电功率的间歇性和波动性使得配电网的电压/无功优化(voltage/VAR optimization, VVO)问题的模型和变量呈高维复杂特性, 导致传统优化方法难以高效求解。因此, 针对高比例 PV 接入下的 VVO 策略研究已成为当前配电网运行与控制领域的热点问题, 亟须探索兼具高效性、鲁棒性和实用性的解决方案。

PV 高渗透配电网的 VVO 任务是一个典型的随机优化问题^[4-5], 目前其求解方案可以被分为三大类: 基于模型的知识驱动方法、无模型的数据驱动方法和数据-知识融合驱动的方法。

1) 第 1 类方法^[6-9]构建了配电网的最优潮流模型

和不确定性量预测模型, 并基于解析方法进行求解, 求解过程可解释性强且解的最优性可以得到保障。但这类方法依赖对不确定性变量和系统动态演化模型的精准刻画, 且模型的求解规模会随着变量和场景的增加呈指数级增长^[10], 制约了其在高动态、随机场景中的推广应用。

2) 第 2 类方法以无模型深度强化学习(deep reinforcement learning, DRL)方法^[11-15]为主, 通过与配电网环境的持续交互来学习最佳策略, 并使用参数化的神经网络来表征配电网潮流转移的不确定性, 具备实时操作能力。然而, 这类方法主要从历史经验轨迹中捕获系统的动态特性, 在执行时缺乏对配电网实际运行知识的理解, 控制模型通常为黑盒模型, 控制策略的适用性不足^[16-17]。

3) 第 3 类数据-知识融合驱动的方法结合了前两类方法的优势, 通过在数据驱动方法中融入配电网运行的专家知识来实现精准安全的控制。文献[18-20]基于配电网功率-电压灵敏度矩阵构建了 DRL 策略的安全层以实现电压偏差的安全修正。文献[21]利用考虑专家知识的深度神经网络辅助 DRL 智能体将动作投影到安全区域。文献[22]提出基于最优潮流知识属性的 PV 调压数据集, 然后采用神经网络映射最优 VVO 策略。文献[23]基于最优潮流公式构建了一种称为“feasibility diamond”

收稿日期: 2025-05-06; 修回日期: 2025-08-04。

上网日期: 2025-11-18。

国家自然科学基金资助项目(52477180)。

的多维凸多面体,用于快速检查DRL策略的可行性。文献[24]提出一种基于专家指导信号的多智能体深度确定性策略梯度(deep deterministic policy gradient,DDPG)算法的电压优化方法。文献[25]计及配电网物理约束构建了二次规划模型,然后结合数据驱动算法提出一种配电网双时间尺度电压协调优化控制策略。文献[26]在DRL的在线执行阶段,根据配电网的运行机理知识,将神经网络选取动作的过程转化为混合整数规划问题来求解,以满足配电网的实时运行约束。

上述数据-知识融合驱动方法在新能源高渗透配电网的运行优化领域取得了显著进展,但是还存在一些问题有待进一步解决。首先,面对分布式PV高渗透配电网潮流状态的强不确定性和高动态特性,传统DRL智能体在进行系统状态-动作对的价值(也即Q值)评估时,因采用单一标量形式的Q值估计而缺乏对回报不确定性的有效感知,并且容易受到神经网络逼近误差的影响,出现Q值高估的现象。这种高估现象往往会导致智能体决策失准,严重影响了DRL策略的性能^[27],具体表现为决策动作引发节点电压频繁越限、无功调节策略偏离最优。其次,DRL策略的性能也高度依赖于状态-动作对的多样性以及电网时序数据的完整性^[28],然而全面、准确地获取具有代表性的功率数据极为困难。现有方法尚未考虑配电网小样本新型源荷数据对VVO性能的影响和解决方案。

为此,本文针对分布式PV高渗透配电网提出了一种新型数据-知识融合驱动的VVO策略。具体贡献为:

1)提出一种改进的分布式软Actor-Critic(distributional soft Actor-Critic,DSAC)算法。与传统DRL方法不同,所提算法通过直接学习状态-动作回报的分布函数,探索累积折扣回报的不确定性并有效缓解了Q值高估问题,从而提升了电压无功优化策略的稳定性。

2)基于配电网DistFlow方程构建DRL策略的专家知识安全层,该安全层显式刻画DSAC策略与节点电压幅值的物理耦合关系,相比于传统纯数据驱动的DRL方法,嵌入配电网物理知识后可以提升策略在实际运行中的安全性和可行性。

3)提出一种基于多元条件 t 分布Copula(multivariate conditional t -Copula,MCTC)的配电网数据增强模型。考虑到多类型负荷和PV数据在时序依赖性、重尾性及波动性方面的复杂特性,所提出的MCTC数据增强模型能有效扩充训练数据集中

的状态-动作对多样性,实现了在配电网小样本新型源荷数据下DRL策略性能的显著提升。

1 配电网VVO问题描述及MDP构建

本文将配电网建模为一个无环图 $G=(\mathcal{N},\mathcal{E})$,其中, \mathcal{N} 为节点集合, \mathcal{E} 为支路集合。对于每个节点 $i\in\mathcal{N}$,其节点电压幅值和相角分别为 v_i 和 δ_i ,其注入的复功率为 $s_i=p_i+jq_i$, p_i 和 q_i 分别为节点注入的有功功率和无功功率。所提PV高渗透配电网的VVO决策变量为PV逆变器和静止无功补偿器(static var compensator,SVC)的无功输出,节点的有功和无功功率平衡方程如下:

$$p_i^{\text{PV}} - p_i^{\text{L}} = v_i^2 \sum_{j\in\mathcal{N}_i} g_{ij} - v_i \sum_{j\in\mathcal{N}_i} v_j (g_{ij} \cos \delta_{ij} + b_{ij} \sin \delta_{ij}) \quad (1)$$

$$q_i^{\text{PV}} - q_i^{\text{SVC}} - q_i^{\text{L}} = -v_i^2 \sum_{j\in\mathcal{N}_i} b_{ij} + v_i \sum_{j\in\mathcal{N}_i} v_j (g_{ij} \sin \delta_{ij} + b_{ij} \cos \delta_{ij}) \quad (2)$$

式中: p_i^{PV} 和 q_i^{PV} 分别为节点 i 光伏的有功功率和无功功率; p_i^{L} 和 q_i^{L} 分别为节点 i 的负荷有功功率和无功功率; q_i^{SVC} 为SVC输出的无功功率; g_{ij} 和 b_{ij} 分别为支路 $ij\in\mathcal{E}$ 的电导和电纳; δ_{ij} 为相邻节点之间的相角差; \mathcal{N}_i 为与 i 直接相连的节点集合。

上述潮流方程描述了配电网的功率平衡关系,并可用于构建最优潮流模型来求解VVO问题。为了应对本文所提VVO问题的强不确定性变量和策略的实时性与安全性需求,本文采用DRL方法进行求解,其核心是将VVO问题转化为一个马尔可夫决策过程(Markov decision process,MDP)。在MDP的一个决策步中,DRL智能体通过策略 π_φ 并根据当前的状态 s_t 选择设备的操作动作 a_t ,其目标是找到一个使折扣累积奖励最大化的策略 π^* ,表示为:

$$\max J_R(\pi_\varphi) = E \left[\sum_{t=0}^T \gamma^t R_t(s_t, a_t) \right] \quad (3)$$

式中: $J_R(\cdot)$ 为目标函数; $E[\cdot]$ 为在给定策略下所有可能状态-动作对的期望回报; T 为控制视界的长度(VVO的优化周期); γ 为折扣因子; $R_t(s_t, a_t)$ 为 t 时刻在状态 s_t 下做出动作 a_t 所获取的即时奖励函数。配电网VVO更详细的MDP描述如下。

1)状态空间 \mathbf{S} :状态 s_t 反映了DRL智能体对 t 时刻配电网运行状态的感知,本文中, $s_t=(P_t^{\text{L}}, Q_t^{\text{L}}, V_t, P_t^{\text{PV}}, p_t^{\text{Loss}}, t)$, $s_t\in\mathbf{S}$ 。其中, $P_t^{\text{L}}/Q_t^{\text{L}}$ 和 V_t 分别为所有节点负荷的有功/无功功率向量和节点电压幅值向量, P_t^{PV} 为PV的有功功率向量, p_t^{Loss} 为配

电网当前时刻的总网损值, t 为当前决策步的时间戳。 V_i 和 p_i^{Loss} 根据配电网环境中的潮流计算工具计算得到。

2) 动作空间 A : 智能体的动作包含了配电网中所有 PV 和 SVC 输出的无功功率向量 Q_i^{PV} 和 Q_i^{SVC} , $a_i = (Q_i^{\text{PV}}, Q_i^{\text{SVC}})$, $a_i \in A$, t 时刻节点 i 的决策变量 $Q_{i,t}^{\text{PV}}$ 和 $Q_{i,t}^{\text{SVC}}$ 分别满足以下约束。

$$|Q_{i,t}^{\text{PV}}| \leq \sqrt{(S_i^{\text{PV}})^2 - (P_{i,t}^{\text{PV}})^2} \quad (4)$$

$$Q_{i,t,\min}^{\text{SVC}} \leq Q_{i,t}^{\text{SVC}} \leq Q_{i,t,\max}^{\text{SVC}} \quad (5)$$

式中: $P_{i,t}^{\text{PV}}$ 为 t 时刻节点 i 处 PV 输出的有功功率值; S_i^{PV} 为节点 i 处额定容量; $Q_{i,t,\max}^{\text{SVC}}$ 和 $Q_{i,t,\min}^{\text{SVC}}$ 分别为 SVC 设备上、下限。

3) 奖励函数 R : 本文的优化目标是在满足电压约束的情况下最小化配电网运行过程中的网络损耗, t 时刻的奖励函数为

$$R_t(s_t, a_t) = r_t = -p_t^{\text{Loss}} - \eta \sum_{i \in \mathcal{N}} C_{i,t}^V \quad (6)$$

式中: r_t 为 t 时刻的奖励值; $\eta \sum_{i \in \mathcal{N}} C_{i,t}^V$ 为节点电压违规的惩罚项, η 为惩罚因子。本文将 $C_{i,t}^V$ 定义为:

$$C_{i,t}^V = \max \left\{ 0, |v_n - v_{i,t}| - \frac{\bar{v} - \underline{v}}{2} \right\} \quad \forall i \in \mathcal{N} \quad (7)$$

式中: v_n 为额定电压; \bar{v} 和 \underline{v} 分别为电压幅值的上、下限。在安全与优化运行条件下, 通常允许 5% 的电压偏移, 即 $0.95 \text{ p.u.} \leq v_{i,t} \leq 1.05 \text{ p.u.}$ 。式(7)表示 t 时刻节点 i 的电压偏离额定电压超过一个范围时接受一个正的惩罚项, 否则为 0。

4) 状态转移概率分布 P : 状态转移概率函数通常被定义为 $P: \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$, $P_t(s_{t+1} | \delta(s_t, a_t))$ 代表 t 时刻代理采取动作 a_t 时状态从 s_t 转移到 s_{t+1} 的概率。在本文中, $\delta(s_t, a_t) \rightarrow s_{t+1}$ 代表配电网在输入新的变量后经过潮流计算转移到新的状态。DRL 求解的核心是通过历史运行数据(变量)和环境交互生成大量状态-动作序列 $\{s_0, a_0, s_1, a_1, \dots, s_T, a_T\}$, 并通过最大化状态转移概率分布下的预期累计奖励之和来学习最佳控制策略, 智能体则从这个过程中捕获 PV 和多类型负荷的不确定性和配电网潮流的动态特性。

5) 配电网环境: 本文根据真实节点和线路参数(电阻、电抗和拓扑连接关系)并采用三相平衡等效建模的方式, 构建可供 DRL 智能体交互的配电网环境, 同时使用 Pandapower 工具包进行潮流计算。在每个决策步 t , 配电网都会注入变量 P_t^L 、 Q_t^L 和 P_t^{PV} , 这些电气量均可通过数据采集与监控(SCADA)系统获取, 并可通过所提数据增强方法进行扩充。值得说明的是, 本文方法在设计上可兼容含有载调压开关与电容器组的实际配电网, 尽管未直接控制这

类离散设备, 其调节效应已通过功率分布动态隐含于状态输入中, 使策略具备隐式感知与适应能力, 从而在不干扰原有控制逻辑的前提下实现连续型资源的优化调节。

2 分布 DRL 框架及改进的 DSAC 算法原理

2.1 基于最大熵的 DRL 分布软策略迭代框架

传统 DRL 的目标是找到一个能够最大化期望累计回报的策略, 即式(3)。为了促进对随机性的探索并防止策略过早收敛, 最大熵强化学习额外引入了策略熵项, 从而使得目标函数变为:

$$J_R(\pi_\varphi) = \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t (r_t + \alpha \mathcal{H}(\pi_\varphi(\cdot | s_t))) \right] \quad (8)$$

式中: α 为温度系数, 用来平衡奖励与熵之间的权重; $\pi_\varphi(\cdot | s_t)$ 为状态 s_t 下的策略函数; $\mathcal{H}(\pi_\varphi(\cdot | s_t))$ 为熵函数。熵的存在可以使智能体在追求高回报的同时保持足够的随机性, 可被定义为:

$$\mathcal{H}(\pi_\varphi(\cdot | s_t)) = \mathbb{E}[-\log \pi_\varphi(a_t | s_t)] \quad (9)$$

传统的 DRL 方法仅估计状态-动作对的期望回报值 $Q(s, a)$, 为了在最大化累计期望回报时更充分地兼顾策略的探索性(通过熵项), 文献[27]提出了一种分布软策略迭代框架, 将 Q 值视为一个随机变量。具体地, 定义了随机变量 $Z^{\pi_\varphi}(s_t, a_t) := r_t + G_{t+1}$, 将其称为软状态-动作收益, G_{t+1} 为从时刻 $t+1$ 开始的累积折扣回报。因此, 可以令 $Q^\pi(s, a) = \mathbb{E}[Z^\pi(s, a)]$ 。

分布软策略迭代框架中引入了一个分布化的软一致性算子 T^{π_φ} , 定义为:

$$T^{\pi_\varphi} Z^{\pi_\varphi}(s, a) = r + \gamma (Z^{\pi_\varphi}(s_{t+1}, a_{t+1}) - \alpha \log \pi_\varphi(a_{t+1} | s_{t+1})) \quad (10)$$

式中: a_{t+1} 服从目标策略 $\pi_\varphi(\cdot | s_{t+1})$, 其中, φ 为目标策略的参数。软一致性算子将当前的价值分布转换为目标分布, 从而使得更新目标不再是一个标量而是整个分布。为使更新过程更稳定, 采用最小化目标分布与当前分布之间的 KL(Kullback-Leibler) 散度, 即更新过程可以表示为:

$$Z_{\text{new}} = \arg \min_{Z(\cdot | s, a)} D_{\text{KL}}(T^{\pi_\varphi} Z_{\text{old}}(\cdot | s, a) \| Z(\cdot | s, a)) \quad (11)$$

式中: Z_{new} 为更新后的软状态-动作收益值; $D_{\text{KL}}(\cdot \| \cdot)$ 为 KL 散度计算; $Z_{\text{old}}(\cdot | s, a)$ 为更新前的软状态-动作收益值; $Z(\cdot | s, a)$ 为 $Z^{\pi_\varphi}(s, a)$ 的分布。

2.2 改进的 DSAC 算法

与大多数 DRL 算法相似, DSAC 使用参数化的神经网络来近似状态-动作收益分布函数(Critic 网络)和随机策略函数(Actor 网络)。令 θ 表示 Critic

网络的参数, φ 表示 Actor 网络的参数。DSAC 算法通过策略评估 (Critic 更新) 和策略改进 (Actor 更新) 这两个步骤不断循环迭代, 从而逼近最优策略。

DSAC 算法使用了参数化的分布 $Z_\theta(s, a)$ 来表示状态-动作对 (s, a) 的回报分布, θ 为分布的参数。假设该分布为高斯分布 $G(\mu_\theta(s, a), \sigma_\theta(s, a))$, 其中 μ_θ 和 σ_θ 分别为均值和方差的网络输出。为了学习这个分布, DSAC 将目标分布 $T^{\pi_\varphi} Z_{\bar{\theta}}$ ($\bar{\theta}$ 为目标值网络的参数) 与当前分布 Z_θ 之间的差异通过 KL 散度进行最小化, 从而得到 Critic 网络的梯度更新式如下^[27]:

$$\nabla_{\theta} J_Z(\theta) \approx \mathbb{E}_{(s,a) \sim \mathcal{B}} \left[-\frac{(y_z - Q_\theta(s, a))}{\sigma_\theta^2(s, a)} \nabla_{\theta} Q_\theta(s, a) - \frac{(C(y_z; b) - Q_\theta(s, a))^2 - \sigma_\theta^2(s, a)}{\sigma_\theta^3(s, a)} \nabla_{\theta} \sigma_\theta(s, a) \right] \quad (12)$$

式中: ∇_{θ} 为梯度算子; 第 1 项为均值相关梯度, 第 2 项为方差相关梯度; \mathcal{B} 为经验回放池, 存储状态-动作对 (s, a) 所形成的经验轨迹; $C(y_z; b)$ 为裁剪函数, 用于确保平方误差保持在当前值分布的平均值附近, 以防止梯度爆炸; b 为裁剪边界; y_z 为目标随机回报, 具体计算见式 (13)。

$$y_z = r_t + \gamma (Z_{\bar{\theta}}(s_{t+1}, a_{t+1}) - \alpha \log \pi_{\varphi}(a_{t+1}|s_{t+1})) \quad (13)$$

DSAC 的策略更新目标函数可表示为:

$$J_{\pi}(\varphi) = \mathbb{E}_{s \sim \mathcal{B}, a \sim \pi_{\varphi}(a|s)} [Z_{\bar{\theta}}(s, a) - \alpha \log \pi_{\varphi}(a|s)] \quad (14)$$

温度系数 α 控制了高回报与高熵之间的权衡, 其自适应更新方式为:

$$\alpha \leftarrow \alpha - \beta_{\alpha} \mathbb{E}_{s \sim \mathcal{B}, a \sim \pi_{\varphi}(a|s)} [-\log \pi_{\varphi}(a|s) - \bar{\mathcal{H}}] \quad (15)$$

式中: $\bar{\mathcal{H}}$ 为目标熵; β_{α} 为学习率。

进一步地, 本文在 DSAC 的基础上, 为了解决训练中存在的不稳定性、Q 值高估以及对奖励尺度敏感等问题, 引入了 3 个关键的改进: 期望值替换、双值分布学习和基于方差的梯度调整。图 1 展示了所提改进 DSAC 算法的框架。

首先, 将随机采样得到的目标回报 y_z 用更平稳的期望值 y_q 替代, 降低训练过程的高方差问题:

$$y_q = r_t + \gamma (Q_{\bar{\theta}}(s_{t+1}, a_{t+1}) - \alpha \log \pi_{\varphi}(a_{t+1}|s_{t+1})) \quad (16)$$

式中: $Q_{\bar{\theta}}(s_{t+1}, a_{t+1})$ 为确定状态-动作值分布。由于 y_q 比 y_z 更加平稳, 修正后的 Critic 梯度显著降低了价值分布学习过程中的随机性。这一策略有效地减少了均值相关梯度的不稳定性, 从而增强了学习过程的稳定性和效率^[29]。

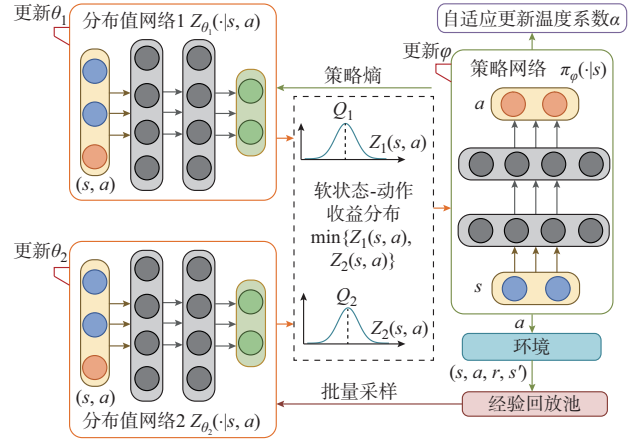


图 1 改进的 DSAC 算法框架
Fig. 1 Framework of improved DSAC algorithm

其次, 同时训练两个独立的 Q 值分布 Z_{θ_1} 和 Z_{θ_2} , 并在目标计算和策略更新时选择较小的均值, 以减少 Q 值高估偏差。因此, 新的 Q 值分布 Q_{\min} 和式 (14) 中的策略更新目标函数 $J_{\pi}(\varphi)$ 分别表示为:

$$Q_{\min} = \min \{ Z_{\theta_1}(s_{t+1}, a_{t+1}), Z_{\theta_2}(s_{t+1}, a_{t+1}) \} \quad (17)$$

$$J_{\pi}(\varphi) = \mathbb{E}_{s \sim \mathcal{B}, a \sim \pi_{\varphi}(a|s)} \left[\min_{i=1,2} Z_{\theta_i}(s, a) - \alpha \log \pi_{\varphi}(a|s) \right] \quad (18)$$

最后, 为了进一步缓解对奖励尺度的依赖和防止梯度爆炸, 改进的 DSAC 引入了自适应的裁剪边界和梯度缩放因子。针对式 (12) 中的 $C(y_z; b)$, 不再采用固定的裁剪边界, 而是通过分布的标准差 $\sigma_{\theta}(s, a)$ 自动确定裁剪范围, 有

$$b = \xi \mathbb{E}_{(s,a) \sim \mathcal{B}} [\sigma_{\theta}(s, a)] \quad (19)$$

式中: ξ 为三倍标准差参数。裁剪边界 b 会随着训练过程中 $\sigma_{\theta}(s, a)$ 的变化而自适应调整, 从而适应不同奖励尺度的任务需求。

为了防止当方差较大时梯度爆炸, 本文引入了一个梯度缩放因子 ω :

$$\omega = \mathbb{E}_{(s,a) \sim \mathcal{B}} [\sigma_{\theta}^2(s, a)] \quad (20)$$

在 Critic 网络的梯度更新中, 损失函数中会乘以 ω 使得梯度更新量能够根据当前 Q 值分布的方差自动调整, 改进后的 Critic 网络梯度更新式如下:

$$\nabla_{\theta_i} J_Z(\theta_i) \approx (\omega_i + \epsilon_{\omega}) \mathbb{E} \left[-\frac{(y_q - Z_{\theta_i}(s, a))}{\sigma_{\theta_i}^2(s, a) + \epsilon} \nabla_{\theta_i} Z_{\theta_i}(s, a) - \frac{(C(y_z^{\min}, b_i) - Z_{\theta_i}(s, a))^2 - \sigma_{\theta_i}^2(s, a)}{\sigma_{\theta_i}^3(s, a) + \epsilon} \nabla_{\theta_i} \sigma_{\theta_i}(s, a) \right] \quad (21)$$

式中: ϵ 和 ϵ_{ω} 为防止除零的小常数。为保证数值稳定性, 裁剪边界 b_i 与梯度缩放因子 ω_i 通常采用滑动

平均方式更新,即

$$\begin{cases} b_i \leftarrow \tau \xi \mathbb{E}_{(s,a) \sim \mathcal{B}}[\sigma_{\theta_i}(s,a)] + (1-\tau)b_i \\ \omega_i \leftarrow \tau \mathbb{E}_{(s,a) \sim \mathcal{B}}[\sigma_{\omega_i}^2(s,a)] + (1-\tau)\omega_i \end{cases} \quad (22)$$

式中: $\tau \in (0,1)$ 为平滑系数。

3 数据-知识融合驱动下的配电网 VVO 框架及实施流程

3.1 基于 DistFlow 方程的专家知识策略安全层构建

配电网的潮流可表示为如下 DistFlow 方程式:

$$P = TP + D(o)c^2 + TF^T P \quad (23)$$

$$Q = TQ + D(x)c^2 + TF^T Q \quad (24)$$

$$M(v^{(2)} - v_0^2 \mathbf{1}_{|\mathcal{N}|}) = 2(D(o)P + D(x)Q) - D(r^2 + x^2)c^2 \quad (25)$$

式中: P 和 Q 分别为配电网所有节点的有功功率和无功功率向量; F 和 T 分别为线路的起始节点和终止节点连接矩阵; M 为配电网的关联矩阵; $D(o)$ 和 $D(x)$ 分别为线路的电阻 o 和电抗 x 组成的对角阵; $v^{(2)}$ 为除平衡节点外所有节点的电压平方值的向量; $\mathbf{1}_{|\mathcal{N}|}$ 为全1向量; v_0^2 为平衡节点的电压平方值; c^2 为线路电流的平方,用于计算功率损耗,在配电网中可以将 c^2 忽略使得功率流方程变为线性的形式^[30]。

式(25)中的线性表达展示了功率向量和电压向量之间的关系。根据第1章的定义,智能体的策略对应的动作空间 A 为 $(Q_1^{PV}, Q_2^{PV}, \dots, Q_{N_{PV}}^{PV}, Q_1^{SVC}, Q_2^{SVC}, \dots, Q_{N_{SVC}}^{SVC})$,故动作 a 和 $v^{(2)}$ 存在如下关系:

$$Mv^{(2)} = Mv_0^2 \mathbf{1}_{|\mathcal{N}|} + 2(D(o)BT(P) + D(x)BT(Q - a)) \quad (26)$$

式中: $B = (I - TF^T)^{-1}$,其中, I 为单位矩阵,根据这个关系可以构建 DRL 智能体的专家知识策略安全层。该安全层的目的是根据动作向量 a 和电压平方 $v^{(2)}$ 间的耦合关系,最小化原始的不安全动作与目标安全动作之间的欧几里得距离,进而将 DRL 的不安全动作投影到安全区域内。具体的投影过程可通过解决一个优化问题来实现,优化目标是找到距离原始动作 a 最近的安全动作 \hat{a} ,其数学形式为:

$$\hat{a} = \underset{a}{\operatorname{argmin}} \left\{ \frac{1}{2} (\hat{a} - a)^2 \right\} \quad (27)$$

s.t.

$$\begin{cases} [v_0^2 \mathbf{1}_{|\mathcal{N}|} + 2M^{-1}(D(r)BT(P) + D(x)BT(Q - \hat{a}))] \leq \mathbf{1}_{|\mathcal{N}|}(\bar{v} - \kappa) \\ [v_0^2 \mathbf{1}_{|\mathcal{N}|} + 2M^{-1}(D(r)BT(P) + D(x)BT(Q - \hat{a}))] > \mathbf{1}_{|\mathcal{N}|}(\underline{v} + \kappa) \end{cases} \quad (28)$$

式中: \bar{v} 和 \underline{v} 分别为电压幅值的上、下限; κ 为松弛参数,被用于对电压幅值的限制条件进行管理。通过加入松弛参数,能够为操作约束提供一定的缓冲空间,以适应预测与实际电压幅值之间可能的偏差,从而确保生成的动作在安全操作范围内,同时尽可能地减小策略调整带来的偏差。

图2展示了数据-知识融合驱动的执行框架,本文在数据驱动的 DRL 框架中引入以 DistFlow 方程为基础的专家知识安全层,将配电网的物理知识约束显式映射到 DRL 的动作空间中,实现了数据-知识一体化融合驱动的策略求解。

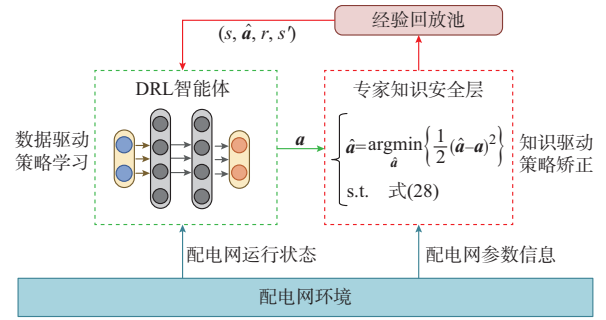


图2 数据-知识融合驱动框架示意图
Fig. 2 Schematic diagram of hybrid data-knowledge-driven framework

3.2 基于 MCTC 的数据增强模型

Copula 的核心思想是通过一个多元函数将每个随机变量的边际分布与它们之间的依赖关系解耦,从而提供灵活的依赖建模能力^[31]。在配电网环境下,多类型负荷及 PV 数据通常具有复杂的时序依赖性、明显的重尾特性以及较大的波动性,使得传统的数据建模方法难以准确刻画其真实分布特征。本文所提基于 MCTC 的数据增强模型通过构建符合真实数据分布的联合概率结构,能够有效刻画电网负荷的动态变化模式,并增强对极端值的鲁棒性,从而提升生成数据的真实性和多样性。

在 MCTC 中,给定每个边际的分布(如负荷和 PV 数据等),通过多元 t 分布 T_d 来描述这些变量之间的依赖关系,可表示为:

$$C(u_1, u_2, \dots, u_d; \Sigma, \nu) = T_d(T_d^{-1}(u_1), T_d^{-1}(u_2), \dots, T_d^{-1}(u_d); \Sigma, \nu) \quad (29)$$

式中: $C(\cdot)$ 为 Copula 函数; u_1, u_2, \dots, u_d 为每个边际分布的累积分布函数转换的伪观察值,通常位于区间 $[0,1]$ 内; Σ 为协方差矩阵; ν 为自由度; T_d 为 d 维 t 分布; T_d^{-1} 为 t 分布的逆累积分布函数。

MCTC的密度函数表示为:

$$C(u_1, \dots, u_d; \boldsymbol{\Sigma}, \nu) = t_d(z_1, \dots, z_d; \boldsymbol{\Sigma}, \nu) \prod_{i=1}^d t_v(z_i; \nu) \quad (30)$$

式中: $t_d(\cdot)$ 为标准 t 分布的密度函数; $t_v(\cdot)$ 为单维标准 t 分布的密度函数; $z_i = T_d^{-1}(u_i)$ 为每个边际变量的标准化观察值。密度函数反映了数据在各个边际分布下的联合分布结构。在 MCTC 模型中, 主要的估计参数包括协方差矩阵 $\boldsymbol{\Sigma}$ 和自由度 ν 。这些参数的估计方法通常使用最大似然估计, 然后使用贝叶斯信息准则来选择最佳的模型参数。具体过程为:

$$\boldsymbol{\Sigma} = \arg \max_{\boldsymbol{\Sigma}, \nu} \prod_{d=1}^{N_d} C(u_1, u_2, \dots, u_d; \boldsymbol{\Sigma}, \nu) \quad (31)$$

$$B_{\text{BIC}} = -2 \ln L_d + p_d \ln N_d \quad (32)$$

式中: N_d 为样本数量; L_d 似然函数; p_d 为参数的数量; B_{BIC} 为贝叶斯信息准则值, 其目标是平衡模型的拟合优度(通过 L_d 衡量)和模型复杂度(通过 p_d 衡量)。

为了更好地捕捉随机变量的时序特征, 所提模型通过引入外部条件变量来进行条件化, 条件化过程可以通过式(33)进行表示:

$$F(x_1|x_2) = C(u|u_2) = C(z_1|z_2; \chi_{12}) \quad (33)$$

式中: $F(x_1|x_2)$ 为条件分布, 表示给定 x_2 时, x_1 的分布; χ_{12} 为条件化的参数; z_1 和 z_2 为投影后的随机变量。

进一步地, 引入外部条件变量并进行数据反投影的过程如下:

$$\hat{z}_w = T^{-1}(F_T(\hat{w}; t); \nu) \quad (34)$$

$$\hat{x}_i = F_T^{-1}(T(\hat{z}_i; \nu; t); \boldsymbol{\Sigma}) \quad i \in \{1, 2\} \quad (35)$$

式中: \hat{z}_w 为变量经过条件化变换后的估计值; \hat{w} 为容量或年能耗等外部特征; t 为季节性时间变量; F_T 为条件化后的累积分布函数; \hat{x}_i 为经过反投影最终生成的模拟数据。

3.3 数据-知识融合驱动下的 VVO 实施流程

所提数据-知识融合驱动下的 VVO 方法的实施流程可分为以下 4 个阶段:

1) 配电网数据增强阶段: 使用 3.2 节提出的 MCTC 模型对从 SCADA 系统获取的 PV 和多类型负荷数据进行增强处理。数据的时间分辨率为 15 min, 同时将 PV 容量及配电台区类型、时间戳、季节和工作日等外部信息作为 MCTC 的条件, 以增强对数据复杂耦合关系和时序特征的感知。

2) DRL 智能体离线探索阶段: 该阶段 DRL 智能体通过多样化数据与配电网环境交互, 收集不同状态下的操作经验轨迹并存储到经验回放池。设置每回合的交互步数为 96, 对应一天中数据的分辨率。同时, 该阶段还通过求解式(27)、式(28)中的策略矫正机制将不满足约束的动作投影至安全域。

3) DRL 智能体离线训练阶段: 该阶段从经验回放池中随机采样来训练 Actor 和 Critic 网络的参数 φ 、 θ_1 、 θ_2 , 以及对应目标网络的参数 $\bar{\varphi}$ 、 $\bar{\theta}_1$ 、 $\bar{\theta}_2$ 和温度系数 α 。根据式(21), 参数 θ 的优化可被定义为:

$$\theta_i \leftarrow \theta_i - \beta_z \nabla_{\theta_i} J_z(\theta_i) \quad i \in \{1, 2\} \quad (36)$$

式中: β_z 为 Critic 网络的学习率。

根据式(18), 参数 φ 的优化可被定义为:

$$\varphi \leftarrow \varphi + \beta_\pi \nabla_{\varphi} J_\pi(\varphi) \quad (37)$$

式中: β_π 为 Actor 网络的学习率。温度系数 α 根据式(15)来优化。

目标网络的参数 $\bar{\theta}_i$ 和 $\bar{\varphi}$ 通过以下方式更新:

$$\begin{cases} \bar{\theta}_i \leftarrow \lambda \bar{\theta}_i + (1 - \lambda) \bar{\theta}_i \\ \bar{\varphi} \leftarrow \lambda \bar{\varphi} + (1 - \lambda) \bar{\varphi} \end{cases} \quad (38)$$

式中: $\lambda \ll 1$ 为目标网络的软更新系数。

4) VVO 任务在线执行阶段: 离线训练完成后保存具有最佳参数 φ 的策略网络, 并部署到配电调度中心执行 VVO 任务。所提数据-知识融合驱动方法的伪代码如附录 A 表 A1 的算法 1 所示。

4 仿真算例分析

4.1 仿真系统、数据及算法参数设置

为了验证所提 VVO 方法的有效性和先进性, 本文使用了 IEEE 34 节点测试系统进行仿真验证, 如附录 A 图 A1 所示。在该系统中, 节点 1 为变压器节点(即平衡节点), 节点 5、9、12、16、18、23、27、29 和 34 为分布式 PV 所在节点, 节点 20 和 29 分别连接了两个 SVC。除平衡节点外, 每个节点上均聚合了一定的电力负荷, 包括居民、商业和工业用户, 负荷的无功功率根据功率因数和随机噪声进行生成。本文中使用的数据均来自江苏某地的真实运行场景, PV 和负荷数据的时间分辨率为 15 min。本文假设节点 12、16、18 和 29 上的 PV 为电网侧资源, 故将其作为可控设备进行无功输出。两台 SVC 的容量均为 $-600 \sim 600$ kvar, 根据 IEEE 1547 系列电网标准, PV 逆变器的容量均设置为 PV 阵列额定容量的 110%。

本文使用Python的PyTorch工具包实现模型训练,计算和仿真在搭载Intel i9-13900KF CPU和NVIDIA RTX 3080Ti GPU的服务器上完成,所提算法超参数设置如附录A表A2所示。3.2节中提到的安全层优化问题使用GUROBI求解器进行求解。

4.2 基于MCTC模型的数据增强结果分析

本节分别选取居民、商业和工业3个典型的台区和一个PV场站一年的数据对所提方案进行验证,分别选取一个月的运行数据来生成全年的数据。为验证所提MCTC模型的优势,在居民住宅台区的数据样本上与现有混合高斯Copula(GMC)模型进行对比。

图3(a)展示了两种模型生成数据和原始数据的分布,GMC在捕捉主要负荷集中区时具备一定的准确度,但对尾部区域的拟合存在低估或偏差。而MCTC通过 t 分布的重尾特性和条件化策略,能够在极端负荷场景下也与原始数据保持较高的吻合度,从而更好地满足电网运行和规划中对极端负荷的仿真需求。图3(b)展示了两种模型生成数据和原始数据四个季节的分布箱线图。在冬季和夏季,负荷分布整体较高,MCTC在这两个季节下的上须和离群值与原始数据更为接近,显示了更好的极端值捕捉能力,而GMC模型在高负荷场景下存在低估现象。在春季和秋季,MCTC在尾部的离群值上仍能与原始数据保持相似的范围,而GMC的离群点分布则相对偏小。

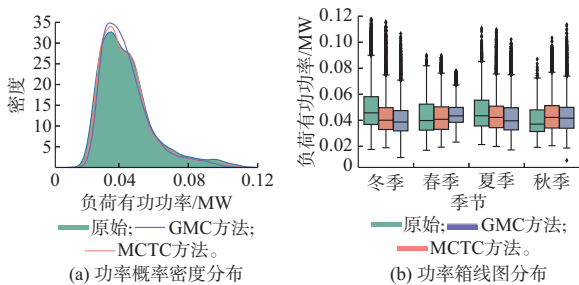


图3 生成数据和原始数据分布对比
Fig. 3 Comparison of distribution between generated data and original data

附录B图B1展示了3个不同类型的负荷台区和PV场站一年中MCTC模型生成数据及原始数据的日发用电曲线。值得注意的是,MCTC模型只使用4个月数据生成的年运行数据,生成的数据在整体曲线形态、峰谷时段以及日间功率波动范围上,与原始数据表现出高度相似性,充分体现了其对原始分布和时间相关性的良好保留能力。

4.3 所提VVO方法有效性验证及对比

本节首先验证了所提数据-知识融合驱动的VVO方法训练时的性能,并与DDPG、SAC(Soft Actor-critic)算法以及基于灵敏度矩阵的安全强化学习(safe reinforcement learning, SRL)方法^[19]进行了对比。图4展示了4种方法训练时的回合平均奖励值、回合平均电压惩罚值和回合平均网损改善情况的演化过程。使用了5个不同的随机种子进行了5次实验,图中实线表示5次实验的均值,阴影部分表示方差。从图中可以看出,训练达到稳定后,所提方法的VVO性能明显优于其他3种方法,并得益于专家知识安全层对动作的矫正,表现出最少的电压违规情况。

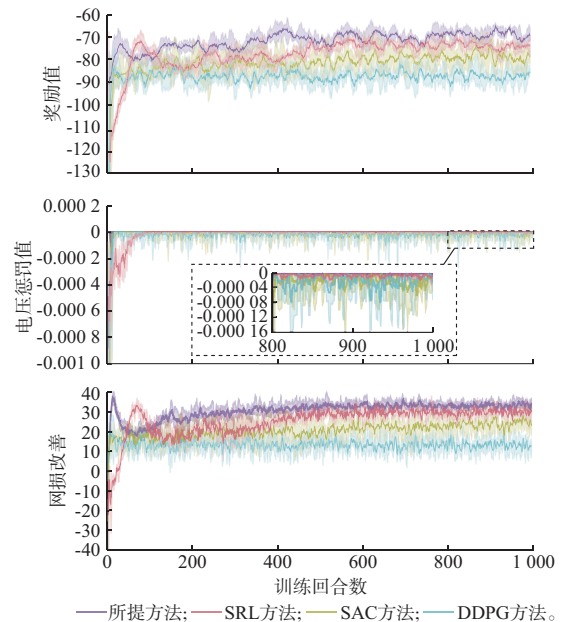


图4 DRL方法训练过程奖励演化情况
Fig. 4 Evolution of rewards during DRL training process

表1的第3至5列展示了训练过程中最后200回合的平均回报对比,所提方法在这3个指标中均取得了最好的效果。尽管所提方法引入了安全层来矫正动作,在面对一些极端场景时仍可能出现少量电压越限情况,但安全层的矫正优化也会尽可能使策略往安全域靠拢。同时,本文对比了4种方法的Q值估计值与真实值的偏差。为了近似真实的Q值,本文在最后200回合的每个回合内随机选取5个状态-动作对,从每个采样到的状态-动作对出发,利用当前策略执行并累加奖励。表1最后一列给出了4种方法的相对Q值估计偏差,所提方法的分布值函数与双值分布技术,使其在Q值估计时能更好地抑制过度高估或严重低估,显著优于其他方法。

表1 训练过程最后200回合的回报对比
Table 1 Comparison of rewards in the last 200 episodes during training process

方法	类型	奖励值	惩罚值	网损改善	Q值估计/%
所提方法	均值	-68.89	-4.52×10^{-7}	33.91	-13.98
	方差	2.68	8.89×10^{-7}	1.84	1.69
SRL	均值	-73.85	-5.62×10^{-6}	30.06	-51.92
	方差	2.69	5.64×10^{-6}	2.35	26.72
SAC	均值	-78.98	-1.91×10^{-5}	24.38	-33.26
	方差	4.21	2.43×10^{-5}	3.62	3.35
DDPG	均值	-87.60	-2.05×10^{-5}	13.37	-64.66
	方差	3.51	2.73×10^{-5}	3.31	31.18

进一步地,本节选取一个PV出力及波动较大运行日的测试数据,对所提出的方法进行详细分析,并与DDPG、SAC以及SRL三种DRL方法进行对比。为更加直观评估所提方法的优化性能,本文构建了基于二阶锥规划的最优电力潮流(OPF)求解模型^[32],并将该模型在理想条件(不确定性信息完全已知)下的求解结果作为网损优化的理论最优值。图5展示了所提的数据-知识融合驱动方法与其他3种DRL方法以及二阶锥规划方案的有功损耗对比。结果表明,所提方法的网损优化性能最为接近二阶锥规划的理论最优值,且相较于DDPG、SAC和SRL三种DRL方法,每个决策时刻的平均网损分别降低了55.7、64.5、35.1 kW。

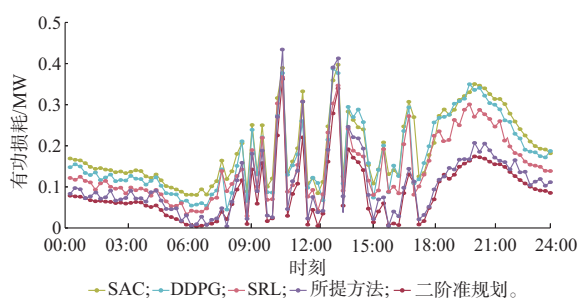


图5 测试日有功损耗对比
Fig. 5 Active power loss comparison on test day

图6展示了所提出方法控制下的6台无功补偿设备在典型运行日24 h内的无功出力情况。从图中可见,所提的智能体策略能够根据配电网中PV发电和负荷的波动特性,动态地调整各设备的无功输出。

附录B图B2展示了典型测试日中无任何控制、二阶锥规划方法、SRL方法以及本文所提方法的节点电压分布情况。可以看出,无控制方案下节点电压波动最为剧烈,多个节点在PV发电高峰期明显越过了上限阈值,并在负荷用电高峰期出现严重的

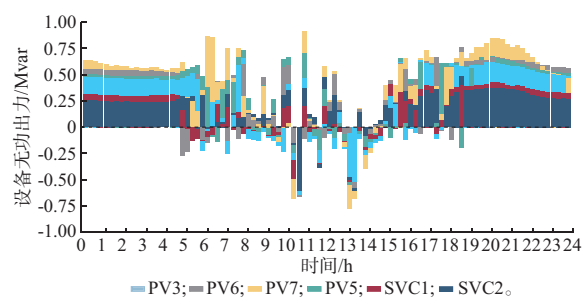


图6 所提方法测试日无功设备出力
Fig. 6 Reactive power equipment output on test day of proposed method

低电压现象。SRL方法由于动作保守,部分节点电压仍存在一定的偏差和局部过限趋势。相比而言,本文所提出方法不仅显著降低了节点电压越限的发生次数和幅值波动,还能有效保持节点电压更加稳定和均匀分布。

4.4 所提VVO方法鲁棒性分析

本节首先验证了所提基于MCTC的数据增强模型对DRL策略泛化性和鲁棒性的提升效果。利用MCTC模型对配电网1年的历史运行数据进行增强,分别生成1、3、5年的增强数据集。随后,基于原始1年数据、增强的1年数据、增强的3年数据以及增强的5年数据,对4种DRL方法在IEEE 34节点系统上进行训练,并评估其在最后100回合的性能表现。附录B表B1的结果表明,在一年数据集上,增强数据与原始数据表现出相近的性能,验证了MCTC模型生成数据的可靠性,未导致DRL性能失真。与未增强数据集相比,使用增强的3年数据集训练的DRL方法性能提升了3%~9%。

为验证所提方法在极端条件下的鲁棒性,本文设计了3组对比实验,分别将负荷与PV功率放大至1.1、1.3、1.5倍。附录B表B2展示了各个方法在3种场景下一天内所有节点的电压越限次数总和。实验结果表明,在功率显著增加的极端条件下,所提方法在电压优化方面表现出优于其他方法的鲁棒性。

此外,本节进一步探讨了所提专家知识安全层的鲁棒性,在IEEE 34节点、69节点和118节点系统上分别评估了DistFlow线性化设计的电压近似性能,并计算了这3个系统中DistFlow近似值与一年实际电压值之间的偏差。附录B图B3展示了不同规模配电网下DistFlow模型的电压幅值近似误差分布。结果表明,大部分电压近似误差在0.001 0至0.001 6之间,尽管规模最大的配电网表现出稍宽的误差分布,但其最大误差未超过0.002。这表明,在

安全层优化过程中,将松弛参数设置为 0.002 能够有效抵消线性化近似引入的不准确性。

4.5 所提 VVO 方法在大规模配电网的性能分析

为进一步验证所提 VVO 方法在大规模配电网中的适用性与可扩展性,本节将仿真场景扩展至 IEEE 118 节点配电系统^[33],其拓扑与资源布置如附录 C 图 C1 所示。不同于传统的 IEEE 118 节点大电网模型,该系统具有典型的配电网结构特征,并已被 MATPOWER 收录为标准测试案例(case118ZH)。在该系统中,共接入 21 个分布式 PV 及 4 台 SVC 装置,假设电网可控资源分别分布在节点 15、25、46、48、54、58、61、70、74、85、108 和 112 上。

图 7 给出了在 IEEE 118 节点系统中将训练回合数扩展至 3 000 的全过程对比曲线,附录 C 表 C1 汇总了最后 200 回合的统计指标。从实验结果中可以看出,训练达到稳定后,所提方法在大规模系统上的网损和电压的优化性能表现出明显的优势,网损改善幅度是 SRL 方法的 2 倍,平均电压惩罚绝对值维持在 10^{-6} 量级,而其他 3 种基线方法在高动态负荷区段仍出现明显的越限尖峰。值得注意的是,分布式回报估计配合双值分布技术显著抑制了 Q 值偏差,同时从训练曲线的波动幅度看,所提方法在整个 3 000 回合内始终呈现最小方差。

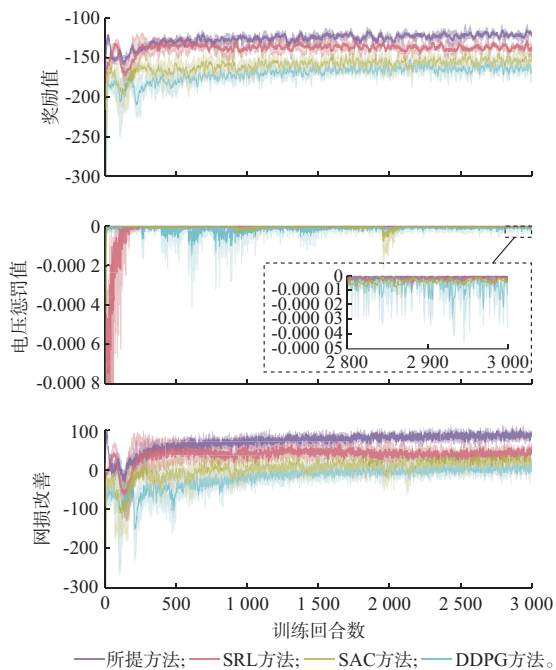


图 7 DRL 方法在大规模配电网的训练奖励演化
Fig. 7 Training reward evolution of DRL methods in large-scale distribution networks

附录 C 表 C2 给出了在 IEEE 118 节点系统中,分别采用原始 1 年数据与基于 MCTC 模型增强生成的 3 年数据对 4 种 DRL 方法进行训练后,在最后 100 回合的性能对比结果。结果表明,采用增强数据集训练的各类 DRL 策略在奖励值、电压惩罚及网损优化等指标上较原始数据集实现了 2%~6% 的性能提升,验证了所提 MCTC 数据增强方法在大规模系统上提升策略鲁棒性的有效性。

进一步地,本节选取了一天包含“中午极端 PV 高注入及夜间尖顶负荷”特征的典型日作为测试集,对训练完成的智能体在 IEEE 118 节点系统上进行在线执行。附录 B 图 B2 给出了优化前后测试日全天各节点的电压幅值,可以看出无控制方案下中午和晚上分别出现显著的电压越限问题,而本文所提 VVO 方法可以将所有节点的电压控制在合理的范围内。图 8 展示了 8 台可控 PV 逆变器与 4 台 SVC 在典型测试日 96 个决策步内的无功调节轨迹,可清晰看出方法对中午过压场景吸收无功及夜间欠压场景补偿无功的动态响应机理。同时,所提方法在测试日的每个决策步上平均减少了 63.38 kW 有功损耗,且所有决策步总耗时仅为 27.36 s,相比基于模型的 SOCP 方法的策略计算效率提升了近 42 倍。

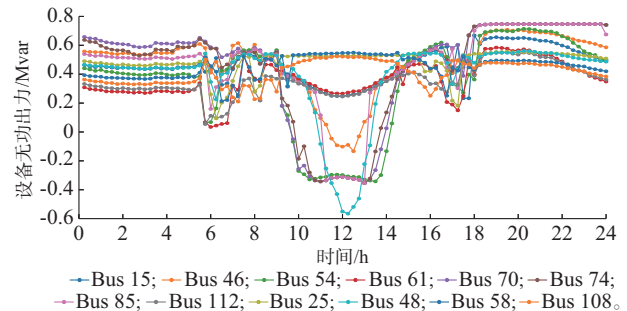


图 8 所提方法测试日上可控 PV 和 SVC 的无功输出
Fig. 8 Reactive power outputs of controllable PVs and SVCs on test day of proposed method

5 结语

本文提出了一种基于数据-知识融合驱动的分布式 PV 高渗透配电网 VVO 策略。首先,通过改进的 DSAC 算法直接学习状态-动作回报的分布函数,显著缓解了传统 DRL 方法中因 Q 值高估导致的策略性能下降问题,提升了算法的收敛性和稳定性。其次,基于配电网 DistFlow 方程构建的专家知识安全层,通过动作矫正机制确保了 DRL 策略在实际配电网运行中的安全性和可靠性,避免了因策略探索

导致的电压越限等安全问题。最后,提出的MCTC数据增强模型有效扩充了训练数据集的多样性,弥补了现实数据获取不足的问题,进一步提升了DRL策略的泛化能力和鲁棒性。

在真实的源荷数据集及IEEE 34节点系统和IEEE 118节点系统上的数值实验表明,本文所提方法在优化效果、鲁棒性和可扩展性方面均优于其他VVO策略。特别是在极端条件和更大规模系统中,所提方法展现了更强的适应性和稳定性。同时,所提数据增强技术的引入使VVO策略的性能提升了2%~9%,进一步验证了其在实际应用中的潜力。

本文提出的优化策略在标准节点系统和典型数据集上取得了良好效果。但当前研究主要针对单相配电网的集中式优化场景,未充分考虑实际三相不平衡配电网的复杂性及分布式多智能体协同控制问题。下一步研究将拓展至三相不平衡配电网环境,结合分布式多智能体强化学习方法,深入探讨多智能体协同下的所提VVO策略的应用潜力。

附录见本刊网络版,点击<http://www.aeps-info.com/aeps/article/abstract/20250506005>,或扫描英文摘要后二维码,可阅读全文。

参 考 文 献

- [1] 国家能源局.2024年光伏发电建设情况[EB/OL].[2025-01-27].<https://www.nea.gov.cn/20250221/f04452701c914d51a89d0c0ea6f4acd1/c.html>.
National Energy Administration. Photovoltaic construction in 2024 [EB/OL]. [2025-01-27]. <https://www.nea.gov.cn/20250221/f04452701c914d51a89d0c0ea6f4acd1/c.html>.
- [2] 张波,张永康,孙英钧,等.考虑光伏度电成本的配电网数据-知识驱动优化调控策略[J].电力系统自动化,2025,49(18):74-82.
ZHANG Bo, ZHANG Yongkang, SUN Yingjun, et al. Data-knowledge-driven optimal regulation strategy of distribution network considering photovoltaic power cost[J]. Automation of Electric Power Systems, 2025, 49(18): 74-82.
- [3] 祝泳琪,刘友波,唐志远,等.基于数据驱动预测控制的有源配电网电压控制策略[J].电力系统自动化,2024,48(20):100-108.
ZHU Yongqi, LIU Youbo, TANG Zhiyuan, et al. Voltage control strategy for active distribution network based on data-enabled predictive control [J]. Automation of Electric Power System, 2024, 48(20): 100-108.
- [4] SUN X Z, QIU J, ZHAO J H. Real-time volt/VAR control in active distribution networks with data-driven partition method[J]. IEEE Transactions on Power Systems, 2020, 36(3): 2448-2461.
- [5] 朱涛,海迪,李文云,等.基于图神经网络与强化学习的配电网电压与无功功率优化方法[J].南方电网技术,2024,18(11):67-78.
ZHU Tao, HAI Di, LI Wenyun, et al. Voltage and reactive power optimization method for distribution networks based on graph neural network and reinforcement learning [J]. Southern Power System Technology, 2024, 18(11): 67-78.
- [6] 巨云涛,康晓帆,刘文武,等.基于分布鲁棒优化的分布式配-微协同日前调度方法[J].电力系统自动化,2024,48(20):48-58.
JU Yuntao, KANG Xiaofan, LIU Wenwu, et al. Distributed coordinated day-ahead scheduling method for distribution network and microgrid based on distributionally robust optimization [J]. Automation of Electric Power System, 2024, 48(20): 48-58.
- [7] GAO H J, WANG L F, LIU J Y, et al. Integrated day-ahead scheduling considering active management in future smart distribution system [J]. IEEE Transactions on Power Systems, 2018, 33(6): 6049-6061.
- [8] XU Y, DONG Z Y, ZHANG R, et al. Multi-timescale coordinated voltage/VAR control of high renewable-penetrated distribution systems [J]. IEEE Transactions on Power Systems, 2017, 32(6): 4398-4408.
- [9] 杨炜晨,苗世洪,刘志伟,等.面向分布式电源功率波动平抑的变频空凋集群多时间尺度模型预测控制策略[J].电工技术学报,2022,37(19):4848-4861.
YANG Weichen, MIAO Shihong, LIU Zhiwei, et al. Multi-time-scale model predictive control of inverter air conditioner cluster for distributed power fluctuation stabilization [J]. Transactions of China Electrotechnical Society, 2022, 37(19): 4848-4861.
- [10] CUI H, YE Y J, HU J X, et al. Online preventive control for transmission overload relief using safe reinforcement learning with enhanced spatial-temporal awareness [J]. IEEE Transactions on Power Systems, 2024, 39(1): 517-532.
- [11] 胡维昊,曹迪,黄琦,等.深度强化学习在配电网优化运行中的应用[J].电力系统自动化,2023,47(14):174-191.
HU Weihao, CAO Di, HUANG Qi, et al. Application of deep reinforcement learning in optimal operation of distribution network [J]. Automation of Electric Power System, 2023, 47(14): 174-191.
- [12] HONG L C, WU M H, WANG Y F, et al. MADRL-based DSO-customer coordinated bi-level volt/VAR optimization method for power distribution networks [J]. IEEE Transactions on Sustainable Energy, 2024, 15(3): 1834-1846.
- [13] CAO D, ZHAO J B, HU J X, et al. Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control [J]. IEEE Transactions on Smart Grid, 2024, 15(1): 233-246.
- [14] ZHANG T J, YU L, YUE D, et al. Two-timescale coordinated voltage regulation for high renewable-penetrated active distribution networks considering hybrid devices [J]. IEEE Transactions on Industrial Informatics, 2017, 20(3): 3456-3467.
- [15] 李鹏,钟瀚明,马红伟,等.基于深度强化学习的有源配电网多时间尺度源荷储协同优化调控[J].电工技术学报,2025,40(5):1487-1502.
LI Peng, ZHONG Hanming, MA Hongwei, et al. Multi-timescale optimal dispatch of source-load-storage coordination in active distribution network based on deep reinforcement learning

- [J]. Transactions of China Electrotechnical Society, 2025, 40(5): 1487-1502.
- [16] HOU S R, FU A H, MAURICIO SALAZAR DUQUE E, et al. DistFlow safe reinforcement learning algorithm for voltage magnitude regulation in distribution networks [J]. Journal of Modern Power Systems and Clean Energy, 2024, 13(1): 300-311.
- [17] WU Y Z, YE Y J, HU J X, et al. Chance constrained MDP formulation and Bayesian advantage policy optimization for stochastic dynamic optimal power flow [J]. IEEE Transactions on Power Systems, 2024, 39(5): 6788-6791.
- [18] SUN X Z, XU Z, QIU J, et al. Optimal volt/VAR control for unbalanced distribution networks with human-in-the-loop deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2024, 15(3): 2639-2651.
- [19] KOU P, LIANG D L, WANG C, et al. Safe deep reinforcement learning-based constrained optimal control scheme for active distribution networks [J]. Applied Energy, 2020, 264: 114772.
- [20] 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.
MEI Mingyang, KOU Peng, ZHANG Zhihao, et al. Safe multi-agent deep reinforcement learning for optimal voltage control in active distribution networks [J]. Journal of Xi'an Jiaotong University, 2023, 57(12): 157-167.
- [21] ZHANG M F, GUO G D, ZHAO T Y, et al. DNN assisted projection based deep reinforcement learning for safe control of distribution grids [J]. IEEE Transactions on Power Systems, 2024, 39(4): 5687-5698.
- [22] 杨浩, 王佳怡, 易文飞, 等. 知识-数据融合驱动的配电网光伏逆变器电压/无功优化自适应控制 [J]. 中国电机工程学报, 2025, 45(22): 8691-8706.
YANG Hao, WANG Jiayi, YI Wenfei, et al. Hybrid knowledge-data driven adaptive voltage/VAR optimization control of photovoltaic inverters in distribution networks [J]. Proceedings of the CSEE, 2025, 45(22): 8691-8706.
- [23] HOSSEINI M M, PARVANIA M. On the feasibility guarantees of deep reinforcement learning solutions for distribution system operation [J]. IEEE Transactions on Smart Grid, 2010, 14(2): 954-964.
- [24] 李士丹, 李航, 李国杰, 等. 考虑分区与模仿学习的深度强化学习配电网电压优化策略 [J]. 电力系统保护与控制, 2024, 52(22): 1-11.
LI Shidan, LI Hang, LI Guojie, et al. Voltage optimization strategy for a distribution network based on deep reinforcement learning considering regionalization and imitation learning [J]. Power System Protection and Control, 2024, 52(22): 1-11.
- [25] 张剑, 崔明建, 何怡刚. 结合数据驱动与物理模型的主动配电网双时间尺度电压协调优化控制 [J]. 电工技术学报, 2024, 39(5): 1327-1339.
ZHANG Jian, CUI Mingjian, HE Yigang. Dual-time-scale voltage coordinated optimal control of active distribution network based on data-driven and physical model [J]. Transactions of China Electrotechnical Society, 2024, 39(5): 1327-1339.
- [26] 洪芦诚, 吴明贺, 朱进, 等. 基于约束增强安全强化学习的光储-充高渗透配电网有功/无功优化决策方法 [J]. 中国电机工程学报, 2025, 45(22): 8764-8779.
HONG Lucheng, WU Minghe, ZHU Jin, et al. Active/reactive power optimization decision method for optical-storage-charging high permeability distribution network based on constraint-enhanced security reinforcement learning [J]. Proceedings of the CSEE, 2025, 45(22): 8764-8779.
- [27] DUAN J, GUAN Y, LI S E, et al. Distributional soft actor-critic: off-policy reinforcement learning for addressing value estimation errors [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(11): 6584-6598.
- [28] XIA W J, HUANG H Y, DUQUE E M S, et al. Comparative assessment of generative models for transformer- and consumer-level load profiles generation [J]. Sustainable Energy, Grids and Networks, 2024, 38: 101338.
- [29] DUAN J L, WANG W X, XIAO L M, et al. Distributional soft actor-critic with three refinements [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(5): 3935-3946.
- [30] SCHWEITZER E, SAHA S, SCAGLIONE A, et al. Lossy DistFlow formulation for single and multiphase radial feeders [J]. IEEE Transactions on Power Systems, 2020, 35(3): 1758-1768.
- [31] DUQUE E M S, VERGARA P P, NGUYEN P H, et al. Conditional multivariate elliptical copulas to model residential load profiles from smart meter data [J]. IEEE Transactions on Smart Grid, 2021, 12(5): 4280-4294.
- [32] 张波, 高远, 王磊, 等. 考虑光伏电源可靠性的配电网无功电压调控策略 [J]. 高电压技术, 2023, 49(7): 2775-2784.
ZHANG Bo, GAO Yuan, WANG Lei, et al. Reactive power and voltage regulation strategy of distribution network considering the reliability of photovoltaic power supply [J]. High Voltage Engineering, 2023, 49(7): 2775-2784.
- [33] ZHANG D, FU Z C, ZHANG L C. An improved TS algorithm for loss-minimum reconfiguration in large-scale distribution systems [J]. Electric Power Systems Research, 2007, 77(5/6): 685-694.

吴明贺(1997—), 男, 博士研究生, 主要研究方向: 人工智能在电力系统中的应用、分布式能源系统优化、新型源荷特性建模。E-mail: wu_mh@seu.edu.cn

洪芦诚(1985—), 男, 通信作者, 副教授, 博士生导师, 主要研究方向: 分布式储能与新能源、人工智能在电力系统中的应用、电力电子化配电网。E-mail: hlc3061@seu.edu.cn

高远(1999—), 男, 博士研究生, 主要研究方向: 智能电网优化、电力电子系统可靠性分析。E-mail: gaoy202404@163.com

(编辑 代长振)

Hybrid Data-Knowledge-driven Voltage/VAR Optimization Strategy for Distribution Networks with High Penetration of Photovoltaic

WU Minghe¹, HONG Lucheng¹, GAO Yuan¹, ZHOU Aihua^{1,2}, WANG Dong³, LIU Fengru⁴

(1. School of Electrical Engineering, Southeast University, Nanjing 210096, China;

2. China Electric Power Research Institute, Beijing 100192, China;

3. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210009, China;

4. State Key Laboratory of Internet of Things for Smart City (University of Macau), Macau 999078, China)

Abstract: The intermittency and volatility of massive distributed photovoltaic (PV) output introduce strong uncertainty into distribution network power flows, making frequent voltage violations and increased operation network losses increasingly prominent issues. To address this, this paper proposes a hybrid data-knowledge-driven voltage/VAR optimization (VVO) strategy for distribution networks with high penetration of distributed PV. First, to tackle the strong uncertainty and high dynamic characteristics of distribution network power flow states, an improved distributed soft actor-critic (DSAC) algorithm is proposed. By directly learning the distribution function of state-action returns, it explores the uncertainty of cumulative discounted returns, mitigating the overestimation problem of state-action values and significantly enhancing the performance of the VVO strategy. Second, to resolve the issue of frequent voltage constraint violations, an expert knowledge safety layer based on the DistFlow equations is embedded into the data-driven VVO process. By projecting and correcting the actions output by the DSAC agent through physical knowledge, a deep integration of data learning and knowledge constraints is achieved, improving the feasibility and safety of the strategy. Subsequently, a data augmentation model based on the multivariate conditional t -distribution Copula is proposed. By generating high-quality and diverse training data, it significantly enhances the generalization capability and robustness of the DSAC strategy in complex operation scenarios. Finally, the effectiveness of the proposed strategy is validated by the simulation results on the IEEE 34-bus and IEEE 118-bus systems.

This work is supported by National Natural Science Foundation of China (No. 52477180).

Key words: distribution network; voltage/VAR optimization; high penetration of photovoltaic; hybrid data-knowledge-driven; deep reinforcement learning

