

基于结构感知深度强化学习的配电网动态重构

王睿谦, 刘鑫蕊[†], 富若涵, 张浩晨, 孙秋野

(东北大学 信息科学与工程学院, 沈阳 110819)

摘要: 随着分布式能源的大规模接入, 配电网运行的不确定性显著增加, 配电网动态重构 (DDNR) 成为保障系统经济性与安全性的关键技术. 深度强化学习 (DRL) 因其对不确定性的自适应能力和数据驱动特性而应用于 DDNR 求解. 针对离散动作空间的 DRL 方法求解重构时存在的维数灾难和动作盲区问题, 提出一种分阶段的结构感知 (SA)DRL 动态重构求解框架. 首先, 提出一种性能与结构驱动的重构决策优化方法, 通过启发式生成与结构化聚类, 构建一个高质量且规模可控的拓扑集合, 从维度、性能和结构 3 方面优化 DDNR 的决策空间; 然后, 改进 DRL 架构, 结合图卷积网络和动作嵌入, 设计 SA-Q 网络, 利用图卷积网络 (GCN) 将拓扑编码为图嵌入向量, 实现对配电网运行状态与开关动作的结构化感知和并行处理, 使智能体能够利用拓扑间的结构关系进行高效泛化, 并将模型集成到改进的 Rainbow DQN 算法框架中, 实现离线训练和在线求解; 最后, 通过仿真结果验证所提出方法可以有效降低配电网运行损耗, 改善电压分布并减少开关不必要动作, 收敛速度快且稳定, 能够显著改善配电网的运行水平.

关键词: 动态重构; 深度强化学习; 动作嵌入; 数据驱动控制; 运行优化; 结构感知

中图分类号: TM732 文献标志码: A

DOI: 10.13195/j.kzyjc.2025.1009

引用格式: 王睿谦, 刘鑫蕊, 富若涵, 等. 基于结构感知深度强化学习的配电网动态重构 [J]. 控制与决策, xxxx, x(x): xxxx-xxxx.

Dynamic distribution network reconfiguration based on structure-aware deep reinforcement learning

WANG Rui-qian, LIU Xin-rui[†], FU Ruo-han, ZHANG Hao-chen, SUN Qiu-ye

(College of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Abstract: With large-scale integration of distributed energy resources, operational uncertainty in distribution networks has increased significantly, making dynamic distribution network reconfiguration (DDNR) essential for secure and economical operation. Deep reinforcement learning (DRL) is applied to the DDNR due to its adaptability to uncertainty and data-driven nature. To address the curse of dimensionality and action blind zones in discrete-action DRL-based reconfiguration, this paper proposes a phased structure-aware (SA) DRL framework. A performance- and structure-driven topology screening strategy is first developed to construct a compact and high-quality action space. Then, an SA-Q network integrating graph convolutional networks and action embedding is designed to encode topologies into graph embeddings, enabling structural awareness and improved generalization. The model is incorporated into an enhanced Rainbow DQN framework for offline training and online decision-making. Simulation results validate that the proposed method effectively reduces network losses, improves voltage profiles, and minimizes unnecessary switching actions, achieving fast and stable convergence while significantly enhancing distribution network operational performance.

Keywords: dynamic reconfiguration; deep reinforcement learning; action embedding; data-driven control; operational optimization; structure-aware

0 引言

随着风电、光伏等分布式能源 (DER) 在配电网中大量接入, 网络运行的不确定性显著增加, 传统静态运行方式难以兼顾经济性与安全性^[1-3]. 配电网动

态重构 (DDNR) 通过调整分段开关和联络开关状态改变网络拓扑, 能够有效降低网损、改善电压分布并增强系统运行灵活性, 对主动配电网运行管理起到重要作用^[4-6].

收稿日期: 2025-09-23; 录用日期: 2026-03-08.

基金项目: 国家自然科学基金 (52477075), 中央高校基本科研业务费专项资金 (N25GFZ004).

[†]通信作者. E-mail: liuxinrui@ise.neu.edu.cn.

本文附带电子附录文件, 可登录本刊官网该文“资源附件”区自行下载阅览.

DDNR本质上是一个具有时序耦合特性的混合整数非线性规划问题,其解空间随网络规模呈指数级增长,属于典型的NP-hard问题^[7-9].传统数学规划方法虽可获得较优解,但建模复杂、计算负担重,难以满足实时需求^[10];元启发式算法在求解灵活性方面具有优势,但对参数依赖较强,容易陷入局部最优^[11],在复杂场景下不稳定.近年来,深度强化学习(Deep Reinforcement Learning, DRL)因其数据驱动、适应不确定环境的特点广泛应用于电力系统优化调度领域^[12-14].然而,现有基于DRL的动态重构方法仍面临以下关键问题:

首先,离散动作空间维度过高.重构的核心在于对配电网中分段开关和联络开关进行操作,解空间一般定义为离散的开关组合.以标准的IEEE33节点配电网为例,其可能的开关组合有 2^{37} 种,即使只考虑符合运行约束的可行组合,数量依然有数万种,若直接进行求解会导致维数灾难,寻优效率极低,难以求解.尤其是在DRL算法中,由于策略对维度更加敏感,维数灾难的影响相比传统算法要更大.因此,现有研究大多需要对解空间进行降维压缩.文献[15]通过掩码层确保动作符合拓扑约束,引入批量约束机制,通过变分自编码器学习历史行为分布,确保了重构决策在安全区域内进行.文献[16]将简化后网络的生成树映射到原网络以达到削减动作空间的效果,并使用PER DQN算法求解.文献[17]采用了基本环编码的方式,将系统中每条构成回路的开关作为编码基本单元,并使用Noisy DQN算法进行求解.以上方法从全部可能的网络拓扑中筛选出部分可行解来压缩解空间规模,然而,这些筛选方法通常仅从是否符合运行约束出发,并未将其实际运行表现纳入筛选的考虑因素,很可能导致可行域内潜在的优解被剔除,而差解因符合运行约束被保留的情况,有可能会从根本上限制优化上限.

此外,算法对于配电网的结构信息利用不足.配电网具备明显的图属性,虽然基于多层感知机(MLP)的DRL算法具有数据驱动控制、适应不确定性等特点,但对于图的特征提取能力依然不足^[18-19].在配电网重构问题中,不同动作之间存在明确的物理层面联系,且由于需要考虑开关动作成本,一般不会大范围改变网络结构,而是用尽量少的开关动作达成优化目的,很多拓扑的结构差异并不大.如果仍采用扁平化的处理方式,将大量复杂的拓扑强行映射为离散且独立的动作,则算法无法有效利用网络信息^[20],每个动作都需要独立进行探索和评估,存在大量冗余计算,对探索效率和泛化能力有很大负面影响.

针对上述问题,本文提出一种基于结构感知

(Structure Aware, SA)深度强化学习的分阶段配电网动态重构求解框架,通过离线高质量动作空间构建与在线结构化决策协同,系统性提升重构优化的效率与性能.分阶段重构求解框架如图1所示.本文的主要工作如下:

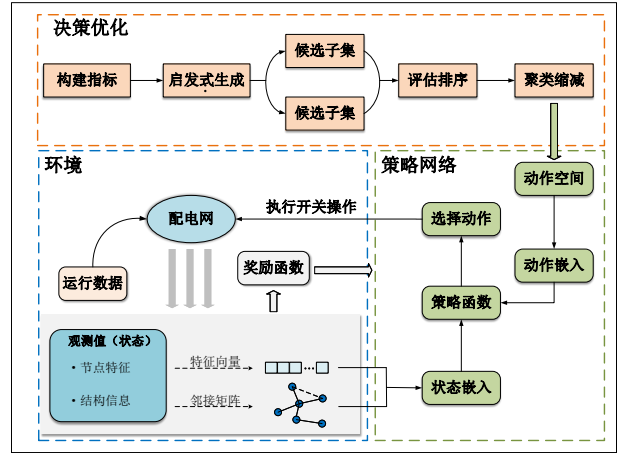


图1 分阶段重构求解框架示意图

1) 针对DDNR的维数灾难问题与现有降维方法的局限性,提出一种性能-结构驱动的重构决策优化方案,通过启发式生成与结构聚类,从数量、性能和结构多样性三个层面优化重构的候选拓扑,构建规模可控且表征性良好的高质量解空间.

2) 针对DRL难以利用配电网图信息的问题,构建含状态-动作双图嵌入的SA-Q网络,将拓扑结构编码为图嵌入向量,实现价值计算在结构层面的泛化,从而提升学习效率与稳定性.

3) 基于上述模型构建SA-Rainbow DQN算法框架,实现离线训练与在线快速求解的协同,提高动态重构的实时决策能力.

1 基于启发式生成及筛选的重构决策优化

1.1 动态重构模型

1.1.1 目标函数

动态重构的主要目标包括最小化网络损耗和开关动作次数,公式如下:

$$F = \min(\alpha \sum_l P_{\text{loss}}^l(t) + \beta \|b_t - b_{t-1}\|_0). \quad (1)$$

其中 b_t 为当前拓扑对应的开关向量, $P_{\text{loss}}^l(t)$ 为线路 l 上的功率损耗, α 和 β 为对应的权重系数.

1.1.2 拓扑约束

将配电网视为一个无向图 $G = (N, E)$,其中 N 表示节点集合, E 表示线路集合,则符合连通辐射约束的拓扑 T 应为此图的一个生成树(ST),因此拓扑约束表示为:

$$T \subseteq G_{ST}. \quad (2)$$

1.1.3 潮流约束

系统运行要满足潮流约束, 包括每个节点的有功无功平衡, 潮流约束表达式如下:

$$P_i(t) = \sum_{j: (i,j) \in E} V_i(t)V_j(t)(G_{ij} \cos \theta_{ij}(t) + B_{ij} \sin \theta_{ij}(t)) = P_{DG,i}(t) - P_{L,i}(t). \quad (3)$$

$$Q_i(t) = \sum_{j: (i,j) \in E} V_i(t)V_j(t)(G_{ij} \sin \theta_{ij}(t) - B_{ij} \cos \theta_{ij}(t)) = Q_{DG,i}(t) - Q_{L,i}(t). \quad (4)$$

其中 j 代表所有和节点 i 相邻的节点集合, $P_i(t)$ 和 $Q_i(t)$ 代表 t 时刻节点 i 注入的有功和无功功率; $V_i(t)$ 和 $V_j(t)$ 代表 t 时刻线路首尾节点的电压; $\theta_{ij}(t)$ 代表 t 时刻线路首尾的电压相角差; $P_{DG,i}(t)$ 和 $Q_{DG,i}(t)$ 代表 t 时刻节点 i 分布式能源注入的有功和无功功率 $P_{L,i}(t)$ 和 $Q_{L,i}(t)$ 代表 t 时刻节点 i 负荷的有功和无功功率.

1.1.4 安全运行约束

$$V_{\min} \leq V_i(t) \leq V_{\max}, \quad (5)$$

$$\begin{cases} P_{DG,i}^{\min} \leq P_{DG,i} \leq P_{DG,i}^{\max} \\ Q_{DG,i}^{\min} \leq Q_{DG,i} \leq Q_{DG,i}^{\max} \end{cases}, \quad (6)$$

$$I_{ij}(t) \leq I_{ij}^{\max}. \quad (7)$$

其中 V_{\max} 和 V_{\min} 分别为系统节点电压的幅值上下限; $P_{DG,i}^{\min}$ 、 $P_{DG,i}^{\max}$ 、 $Q_{DG,i}^{\min}$ 、 $Q_{DG,i}^{\max}$ 代表节点 i 分布式能源的有功无功出力上下限. I_{ij}^{\max} 代表线路 ij 上可通过的最大电流.

1.2 启发式拓扑生成策略

1.2.1 损耗导向的拓扑生成

在线路阻抗基础上引入到电源的电气距离进行加权作为边权重, 调节系数得到配电网无向图 G 的生成树集合 S_1 :

$$G_{MST} = \arg \min_{T \subseteq G_{ST}} \sum_{(i,j) \in E} R_{ij}(1 + \beta \frac{d_{ij}}{d_{\max}}). \quad (8)$$

其中 R_{ij} 为支路 ij 的电阻, d_{ij} 为支路两端节点到最近电源的平均电气距离, d_{\max} 为最大电气距离, β 为调节系数, 范围取 $[0.5, 1]$.

1.2.2 负荷重心导向的拓扑生成

采用负荷重心导向的改进 Prim 算法, 构建对重负荷区域提供较短电气路径的网络结构. 计算配电网的负荷重心:

$$C = \frac{\sum_{i \in N} (P_i - P_i^{DG}) \cdot r_i}{\sum_{i \in N} (P_i - P_i^{DG})}. \quad (9)$$

其中 P_i 为节点 i 的有功负荷, P_i^{DG} 为节点 i 的 DG 出力, r_i 为节点 i 的位置矢量. 定义边 ij 的扩展成本函数为:

$$E(i, j) = \omega \cdot \lambda_{ij} + (1 - \omega) \cdot \frac{d(i, C) + d(j, C)}{2}. \quad (10)$$

其中 λ_{ij} 为边权重, $d(i, C)$ 和 $d(j, C)$ 为节点 i 和 j 到负荷重心的距离, $\omega \in [0, 1]$ 为权重因子. 从电源节点逐步扩展, 每次从支路中选取扩展成本最小的边并加入生成树, 直至获得可行拓扑. 基于参数组合得到拓扑子集 S_2 . 将 S_1 和 S_2 合并, 得到初选集 S . 相比于全部可行拓扑的集合, S 规模极大缩减, 同时并未损失表征性.

1.3 基于 K-Medoids 的结构聚类缩减

虽然通过启发式生成得到了初选集, 但依然存在着以下问题: 若两个树状拓扑仅在末端支路存在一定差异, 则它们的运行表现也大概率相近, 因此 1.2 获得的初选集中包含大量结构雷同、性能表现相近的拓扑. 直接将该集合作为重构的候选决策会导致冗余, 计算效率低下且容易陷入局部最优. 因此, 本文采用基于 K-Medoids 的分段式结构聚类缩减策略, 对初选集内的拓扑执行基于结构相似度的聚类, 步骤如下:

(1) 将拓扑 T_i 表示为长度为 L 的结构向量 $b_i = [b_i^{(1)}, \dots, b_i^{(L)}]$, 其中分量为二进制变量, 表示可操作开关的通断状态, 0 表示断开, 1 表示闭合.

(2) 使用汉明距离 (Hamming Distance) 作为指标衡量两个拓扑的结构相似度:

$$d_H(\mathbf{b}_j, \mathbf{b}_k) = \sum_{i=1}^L |b_j^{(i)} - b_k^{(i)}|. \quad (11)$$

(3) 设定聚类簇数为 M , 执行 K-Medoids 聚类算法, 目标为使得簇内所有非质心样本点到其所属质心的距离最小化. 其优化目标可表示为:

$$\min \sum_{j=1}^M \sum_{b_T \in C_j} d_H(b_T, m_j). \quad (12)$$

其中 C_j 为质心为 m_j 的簇, d_H 为汉明距离, b_T 为该簇中所有非质心的结构向量. 最优聚类总数可根据轮廓系数法确定. 聚类缩减后的集合既保留了运行表现良好的拓扑, 又确保了聚类簇间的结构多样性.

2 基于 SA-DRL 的重构求解模型

2.1 DDNR 的 MDP 形式建模

2.1.1 动作空间

将对配电网中开关的操作映射为具体的网络拓扑, 并采用 1.3 中获得的维度为 M 的高质量拓扑集

合作为动作空间:

$$A = \{a_1, a_2, \dots, a_M\}. \quad (13)$$

在执行动作 a_i 时,从当前拓扑以最少的开关动作切换至拓扑 T_i .在每个时间步,智能体根据策略选择动作并执行,不同动作之间仅通过动作编号来进行区分.但在配电网重构场景下,动作(即不同拓扑)之间包含显式的物理关系,这种动作空间的建模方式忽略了这点,会影响寻优和泛化能力.因此,本文采用动作嵌入的方式,将拓扑结构转换为图对象,经过GCN提取层后,作为深度神经网络的核心输入.

采用动作图嵌入后,每个动作 a_i 都由其对应的图对象 $G(a_i)$ 表示.动作的图对象定义为:

$$G(a_i) = (N_i, E_i). \quad (14)$$

其中 N_i 表示图的节点集合,每个节点对应配电网的一个节点, E_i 表示图的边集合,由 a_i 代表的开关状态组合决定.

2.1.2 状态空间

在实际运行时,负荷需求和分布式电源出力会受到气象波动、用户行为、预测误差等多种因素影响,呈现出很强的不确定性.本文采用高斯噪声模拟不确定性,其概率密度函数为:

$$\varepsilon_i = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma_i^2}\right). \quad (15)$$

其中 i 代表施加噪声的变量,总数为系统内PQ节点数与分布式电源数之和, ε_i 代表第 i 个变量的高斯噪声, σ_i 代表噪声分布服从的标准差.考虑不确定性后的变量为:

$$\tilde{x}_i = x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma_i^2). \quad (16)$$

状态是对环境(即配电网)的观测,需要能够反映配电网的运行状况以及执行动作后系统的变化.考虑到图属性的信息经过向量化展平后会被模糊,将状态空间分解为向量形式的电气信息 s_1 与图形式的网络结构信息 s_2 .网络结构信息设计与动作图对象一致,并采用共享的GCN编码器进行特征提取.电气信息设计如下:

$$s_1 = (\tilde{P}_1^i, \tilde{Q}_1^i, V^i, P_{WT}^k, P_{PV}^l, h). \quad (17)$$

其中 \tilde{P}_1^i 和 \tilde{Q}_1^i 为第 i 个节点的有功和无功负荷; V^i 为第 i 个节点的电压; \tilde{P}_{WT}^k 和 \tilde{P}_{PV}^l 为第 k 个风电机组和第 l 个光伏机组的有功功率;考虑到分布式能源出力与负荷的时序特性,状态空间内加入了当前时间步 h .为保证数据的尺度一致,数据皆采用最大-最小缩放法进行标准化处理.

2.1.3 奖励函数

奖励函数包含网络损耗和开关动作成本,并引

入电压违约作为惩罚项,奖励函数 r 设计如下:

$$r(t) = \omega_1 \sum_{l \in T} P_{loss}^l(t) + \omega_2 \|b_t - b_{t-1}\|_0 + \omega_3 \sum_{i \in N} V_{viol}^i(t). \quad (18)$$

其中 P_{loss}^l 表示线路 l 上的有功损耗; b_t 和 b_{t-1} 表示当前时刻状态和上一时刻网络结构对应的结构向量,0代表断开,1代表闭合; V_{viol}^i 代表节点 i 电压的非线性惩罚项,当潮流不收敛或越限值过大时, V_{viol}^i 将被设置为极大惩罚项.

2.2 结构感知Q网络

在标准Q网络架构中,动作并未作为显式输入特征参与价值函数的构建,智能体难以有效学习动作之间的差异,仅能通过状态价值来间接进行区分.该特性使得网络在高维动作空间和结构化决策任务中容易出现价值估计偏差、泛化能力不足等问题.

因此,本文对Q网络模型进行改进,构建SA-Q网络,通过基于图卷积网络(GCN)的状态-动作双图嵌入,实现对状态和动作的结构感知,显著提升了网络泛化能力.SA-Q网络的结构如图2所示,其核心在于双流并行的信息处理、价值分离计算与信息融合机制:

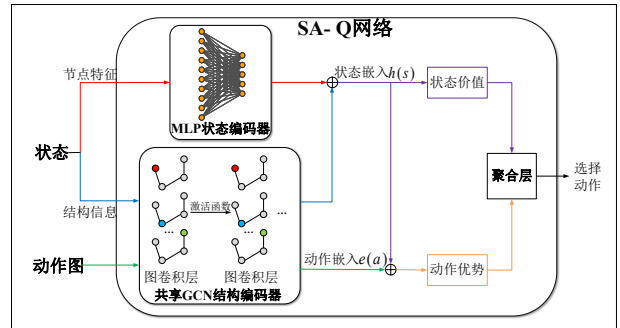


图2 SA-Q网络结构图

(1) 并行处理:网络包含两个并行独立的特征提取流用以处理异构输入.状态电气向量 $s_1(t)$ 通过MLP状态编码器进行处理,提取状态电气特征 $h_1(s)$.状态图和动作图则通过共享的GCN编码器,提取拓扑结构的深层特征,得到状态结构特征 $h_2(s)$ 与动作嵌入 $e(a)$.

(2) 价值分离:基于对偶网络架构,将Q值分解为状态流 $V(s)$ 和动作优势流 $A(s, a)$ 两个对象,分别计算再重新聚合.对偶网络架构下的Q值计算公式如下:

$$Q(s, a) = V(s) + (A(s, a) - \frac{1}{|A|} \sum_{a' \in A} A(s, a')). \quad (19)$$

式中第二项为去中心化操作,目的是确保所有动作

的优势值均值为零. $V(s)$ 评价状态本身, 与具体动作无关, $A(s, a)$ 评价在当前状态下各动作相对平均水平的优势.

(3) 信息融合: 将 $h_1(s)$ 和 $h_2(s)$ 拼接为状态嵌入向量 $h(s)$. 基于价值分离计算 Q 值, 状态价值流 $V(s)$ 的输入为状态嵌入 $h(s)$, 动作优势流 $A(s, a)$ 的输入则为 $e(a)$ 与 $h(s)$ 的融合向量. 在标准的对偶网络架构中, $A(s, a)$ 的实际输入仅为状态特征 $h(s)$, 并没有输入具体动作 a , 而是直接输出所有动作的优势函数. 改进后, 动作优势流的输入直接包含了配电网的结构信息, 智能体能够学习结构特征和价值收益之间的映射关系. 这种设计实现了价值估计在结构层面的泛化, 在动作对价值影响差异不大的重构场景下效果显著.

2.3 SA-Rainbow DQN 求解算法

本文在 Rainbow DQN^[21] 基础上, 将 SA-Q 网络作为核心价值函数, 改进形成 SA-Rainbow DQN 算法. Rainbow DQN 集成了 DQN 的一系列改进, 包括双 Q 值计算、对偶网络、优先经验回放、多步学习、Noisy Nets 以及值分布, 在训练稳定性、收敛速度和策略表现上都有显著提升.

SA-Rainbow DQN 算法流程见表 1, 强化学习训练流程见图 3. 交互过程中经验样本以单步形式进行采集, 并基于多步 TD 回放和优先级采样, 获取多步经验元组. 多步 TD 目标增强了远期收益的传播效果, 提高了长期回报相关样本的采样优先级, 从而加速价值收敛, 达到正反馈的效果; 引入值分布后网络不再预测 Q 值的期望, 而是学习回报的概率分布, 结合价值分离和动作嵌入, 能够更加有效地应对配电网环境中的不确定性. 更新时目标网络根据 s_{j+n} 计算得到目标 Q 分布, 在线网络则根据当前状态 s_j 和动作 a_j 计算预测的 Q 分布. 由于输出为分布, 损失函数改为计算分布间的 KL 散度. 随后优化器根据损失函数进行梯度反向传播并更新在线网络参数; 噪

声探索相比 ϵ -贪婪探索策略随机性更强, 不易陷入局部最优, 但也更加不稳定, 为保证训练过程的稳定性, 目标网络参数采用软更新的方式.

表1 SA-Rainbow DQN 算法流程

算法流程	
1	初始化在线网络参数 θ , 目标网络参数 θ^- , 样本 TD 误差 δ
2	for 回合数 = 1, 2, 3, ..., N
3	智能体接收状态观测量 s_0 和动作嵌入矩阵 $E(A)$
4	for 时间步 = 1, 2, 3, ..., H
5	基于在线 Q 网络, 对当前状态的每个动作计算回报的预测分布, 并基于噪声探索策略选择动作 a_t
6	从环境获得奖励 r_t 和状态 s_{t+1} , 将经验元组 $(s_t, a_t, s_{t+1}, r_t, \delta_t)$ 存入经验池, 并赋予最大优先级
7	经验池中数据量足够后, 从中根据优先级采样 n 步的经验样本, 计算累计回报, 获得 n 步经验元组 $(s_t, a_t, s_{t+n}, R_t^{(n)}, \delta_t)$
8	在状态 s_{t+n} 基于在线网络选择最优动作, 并用目标网络计算该动作的回报分布
9	将 n 步回报与回报分布进行贝尔曼投影, 构建目标分布
10	计算预测分布和目标分布的 KL 散度, 根据损失函数更新在线网络参数和经验优先级
11	软更新目标网络参数
12	end for
13	end for

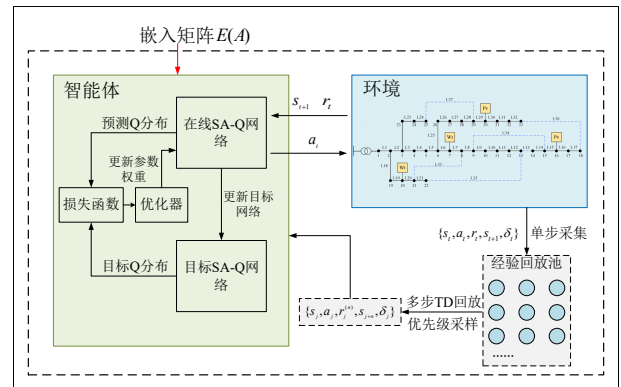


图3 训练交互框图

3 算例分析

在如图 4 所示的 IEEE 33 节点配电网中验证本文方法的有效性, 假设所有 DG 以恒功率因数模式发电, 功率因数为 0.9, 基准电压为 12.66 kV, 基准容

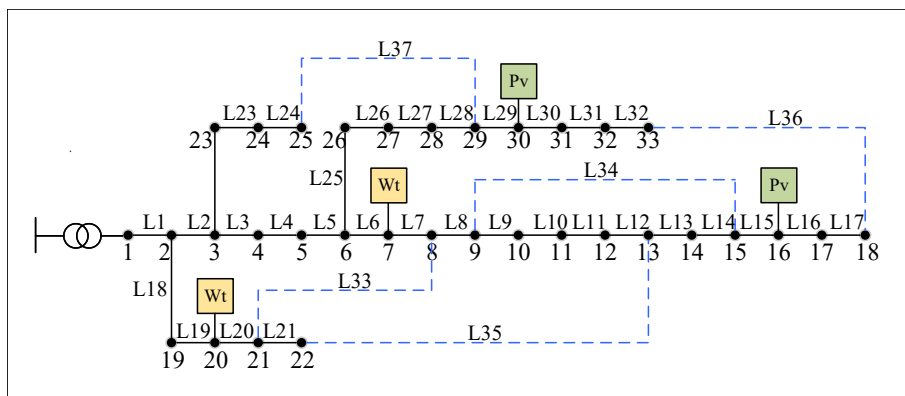


图4 IEEE 33 节点配电网结构图

量为 100 MVA. 模型训练和求解在 Python 3.9.20 版本下进行, 环境基于 OpenDSS 进行构建, 所有实验均在配备 Intel i7-13700K 处理器以及 32 GB 内存的计算机上完成.

3.1 决策优化方法验证

设置以下 5 种不同方案, 在测试系统中采用 Dueling DQN 算法进行训练. 方案 1: 采用本文方法, 筛选出数量为 100 的离散动作; 方案 2: 采用本文方法, 筛选出数量为 732 的离散动作; 方案 3: 采用开关交换法^[22] 筛选出数量为 100 的离散动作; 方案 4: 采用基本环编码法^[23] 筛选出数量为 732 的离散动作; 方案 5: 动作空间内包含全部符合运行约束的拓扑, 维度为 50751. 所有方案除采用的动作空间不同外, 其余设置均完全相同.

由于奖励值波动较大, 采用每 100 回合取滑动平均值的方式以更直观地展示训练过程, 滑动平均奖励曲线如图 5 所示. 观察奖励曲线可以发现, 方案 3、4、5 奖励跌落现象严重, 方案 5 未对动作集进行任何处理, 网络输出层维度过大, 无法进行有效训练; 方案 4 的奖励在 10000 回合后并未完全收敛, 还有一定上升空间, 但依然低于方案 2 训练初期的水平. 方案 3 动作空间维度小, 探索难度较低, 因此可以观察到已经收敛, 但由于动作集质量低, 限制了优化效果, 奖励水平甚至要低于方案 4. 方案 1、2 在训练初期奖励有一定差异, 奖励曲线趋势和收敛后奖励水平基本相同, 且奖励水平明显高于其他方案.

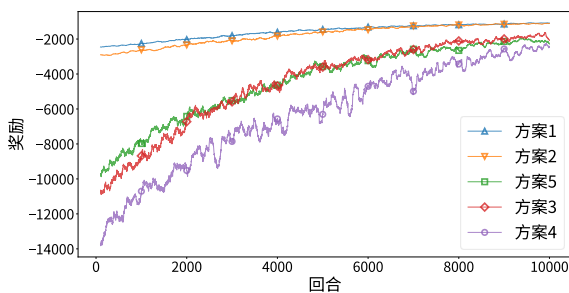


图5 不同动作集奖励曲线对比

综合对比各方案奖励曲线, 可以验证本文所采用的解空间筛选优化方法相比于基本环编码、开关编码等方法更加有效. 因在决策优化阶段已经筛选掉了不符合约束和运行表现不佳的拓扑, 动作集整体质量更高, 有效避免了训练初期的奖励波动, 使得智能体能够进一步学习最优策略. 进一步对比方案 1 和 2, 可以看到依据本文方法筛选得到的动作空间并未因压缩维度而损失性能, 同时压缩维度对探索和收敛有着正面影响. 需要注意的是, 为适应配电网

复杂的运行环境, 动作空间规模不能过小, 否则会与方案 3 类似, 影响最终的优化效果.

3.2 DRL 算法对比

为验证 SA-DRL 架构的有效性, 分别采用 SA-Rainbow DQN、Rainbow DQN、SA-Dueling DQN、Dueling DQN 算法进行训练, 主要超参数取值见表 2, 其中动作空间维度为 100; GCN 层数设置为 2; 激活函数采用 ReLU 函数.

表2 超参数设置

超参数	取值
学习率	0.0001
状态编码器维度	{256,128}
价值流维度	64
优势流维度	64
GCN层数	2
嵌入维度	64
批次大小	128
网络更新率	0.005
折扣因子	0.99
优先级指数	0.6
重要性权重	0.4
学习步数	3

不同算法的奖励曲线如图 6 所示. Dueling DQN 和 SA-Dueling DQN 都基于 ϵ -贪婪策略进行探索, 可以看到二者的奖励曲线都稳步上升并逐渐平缓, 这符合 ϵ -贪婪探索的特征. 相比前两者, Rainbow DQN 和 SA-Rainbow DQN 的表现明显更好. 由于采用噪声探索而非 ϵ -贪婪探索, 奖励曲线趋势有明显变化, 收敛速度更快, 但在训练初期存在较大波动, 这是因为网络的权重还未得到充分的优化, 智能体对环境的理解有限, 会在策略的引导下选择极端或是不合理的动作. 可以看到二者在训练初期都存在小幅的暂时奖励跌落现象, 而 SA-Rainbow DQN 的稳定性更强.

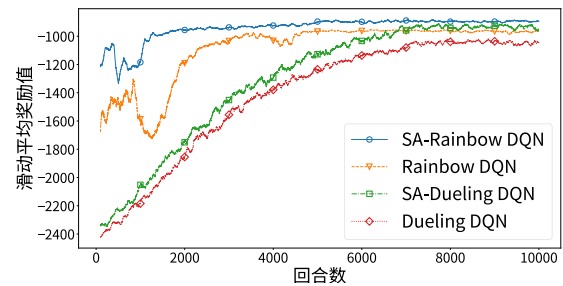


图6 不同算法奖励曲线对比

综合对比各算法的曲线, 可以看到采用 SA-DRL 的算法性能有明显提升, Rainbow DQN 在 4000 回合左右达到收敛, 而 SA-Rainbow DQN 在 2000 回合

左右就已收敛, 且奖励水平更高. SA-Dueling DQN 算法的表现相比 Dueling DQN 算法也有全面提升, 收敛后的奖励达到了略高于 Rainbow DQN 算法的水平. 由此可见, SA-DRL 在寻优能力、收敛速度及泛化性能方面均表现出显著提升, 能够有效优化配电网运行.

3.3 重构结果比较分析

设置以下 4 种方案, 随机抽取非训练集的数据样本作为测试场景, 将求解出的重构决策进行对比分析. 方案 1: 不进行重构; 方案 2: 对一日时间进行分段, 使用改进二进制粒子群 (BPSO) 算法^[24] 求解分时段动态重构; 方案 3: 采用 Rainbow DQN 算法求解全时段动态重构; 方案 4: 采用 SA-Rainbow DQN 算法求解全时段动态重构.

3.3.1 网损及开关动作对比分析

各方案求解得到的重构策略见表 3, 网络损耗见图 7. 方案 1 不进行重构, 开关动作数为 0, 日内总网损为 1.280 MWh. 方案 2 手动划分 5 个时段, 总网损为 0.903 MWh, 降损效果为 29.45%; 方案 3 总网损为 0.814 MWh, 降损达到 36.41%; 方案 4 总网损为 0.746 MWh, 降损达到 41.72%. 方案 2 采用手动划分时段的方式, 实施繁琐且开关动作次数多, 对长期维护产生不利影响, 且在划分更多的时段的前提下, 降损效果依然不及方案 3 和 4. 比较方案 3 和 4, 可以看到 SA-Rainbow DQN 对网络损耗和开关动作优化效果都要强于 Rainbow DQN, 说明本文提出方法在求解应用层面更加有效.

表3 各方案动态重构结果

方案	时间段	断开线路编号	动作次数
无重构	0:00-23:00	33, 34, 35, 36, 37	0
	0:00-5:00	7, 12, 35, 36, 37	
	6:00-8:00	7, 11, 34, 36, 37	
	9:00-16:00	8, 31, 33, 34, 37	
	17:00-20:00	7, 10, 34, 36, 37	
BPSO	21:00-23:00	7, 13, 35, 36, 37	20
	0:00-12:00	7, 9, 14, 31, 37	
	13:00-17:00	7, 9, 14, 28, 31	
RDQN	18:00-23:00	6, 11, 14, 31, 37	8
	0:00-9:00	7, 9, 14, 31, 37	
SA-RDQN	10:00-18:00	7, 10, 13, 31, 37	6
	19:00-23:00	6, 10, 13, 31, 37	

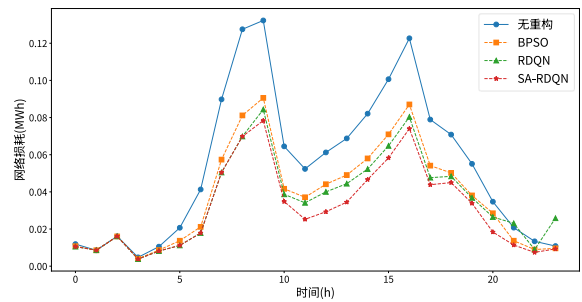


图7 日内网络损耗对比

3.3.2 电压对比分析

图 8 中 (a)(b)(c)(d) 分别绘制了 4 种方案的电压三维曲面图, 从时间尺度上分析, 夜间时段系统中负荷水平很低, 而由于夜间风速大, 风电输出较大功率, 虽然光伏夜间不发电, 电压仍有小幅抬升; 在早晚时段, 负荷水平高, 而光伏受日照强度制约, 发电量较

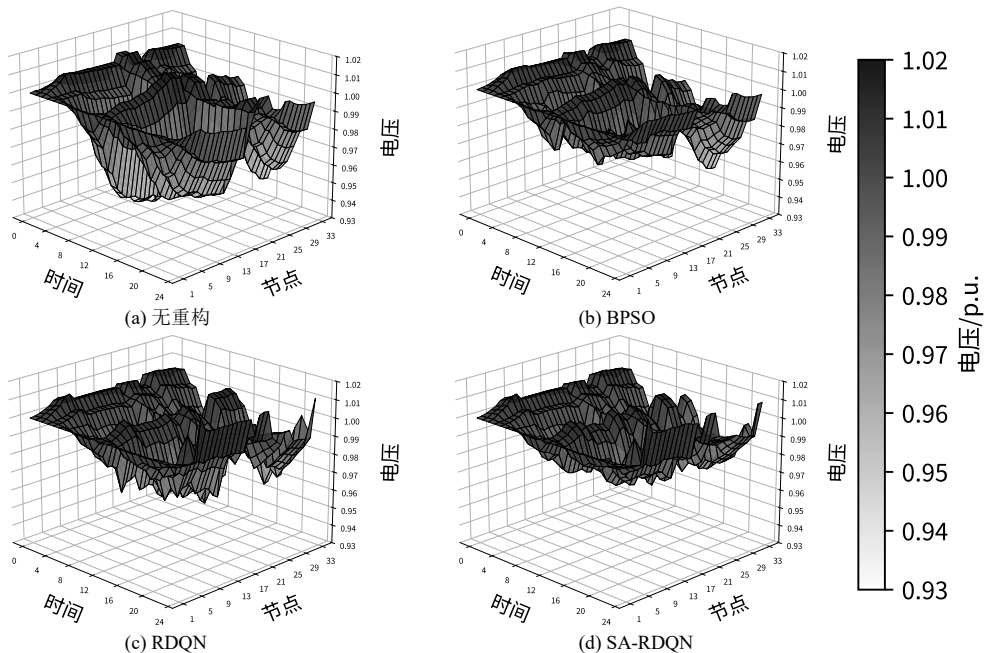


图8 各方案电压三维散点图

少或为0,系统功率依赖外部输入,损耗较大,因此在这两个时间段系统电压明显跌落,电压越限主要也发生在此时段;在中午时段,风电与光伏同时发电,且光伏发电量达到高峰,而负荷水平略低于早晚高峰时段,因此电压水平更高。

表4展示了电压数据的详细情况,方案1不进行重构,最低电压达到0.934,经常出现电压越限情况,无法满足运行需要.方案2相比方案1电压水平有所改善,但依然存在电压越限的情况.方案3和4没有发生电压越限情况,相比方案3,方案4的电压水平更高且更加稳定。

表4 电压详细数据

方案	最低电压	标准差	越限场景数
无重构	0.934	0.019	43
BPSO	0.946	0.013	9
RDQN	0.951	0.011	0
SA-RDQN	0.962	0.010	0

3.3.3 计算速度对比

表5展示了各方案求解花费的时间,方案3和4采用离线训练的模型,决策优化和训练花费时间较长,但离线过程不会影响实时调度,同时仅需1.3秒左右便可以完成求解,速度优势明显.而方案2的求解时间达到了1126秒,求解速度难以满足实时调度需求。

表5 各方案求解时间

方案	决策优化/s	训练时间/s	求解时间/s
BPSO	无	无	1126.75
RDQN	337.27	4449.62	1.36
SA-RDQN	337.27	4953.35	1.31

3.3.4 鲁棒性验证

为验证本文提出方法对不同场景的适应能力,在非训练集的数据中抽取50个不同场景部署智能体并运行动态重构,各场景重构优化的降损效果以

及动作次数如图9所示,其中降损效果以无重构为基准.结果表明,大多数场景下动态重构的降损效果维持在35%到45%之间;开关动作次数基本稳定在6到12次之间,有助于延长设备寿命并稳定运行;此外所有测试中都没有电压越限情况发生.实验结果表明,训练后智能体在面对多种不确定场景依然表现出高度的稳定性和一致性,验证了本文方法具有良好的鲁棒性和泛化能力,能够有效应对配电网复杂的运行环境。

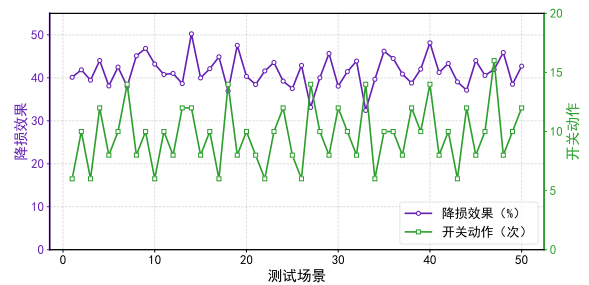


图9 多场景下优化效果展示

综合以上分析,本文提出的重构策略求解方法在降低网损、开关动作等方面优化效果显著,并且具有计算速度快、无需对时段手动划分等优点,能够有效提高配电网运行的经济性和可靠性。

3.4 复杂系统测试

为验证本文方法对于不同配电系统的适用性,在如图10所示的IEEE 69节点配电网中进行复杂系统测试,测试方案设计同3.3,动作空间维度为300.各方案优化结果如表6所示,不执行重构操作时,系统总网损达到2.187 MWh,经方案4降损后总网损为1.447 MWh,开关动作次数仅为14次,且没有电压越限情况,验证了本文方法针对复杂网络的扩展性良好。

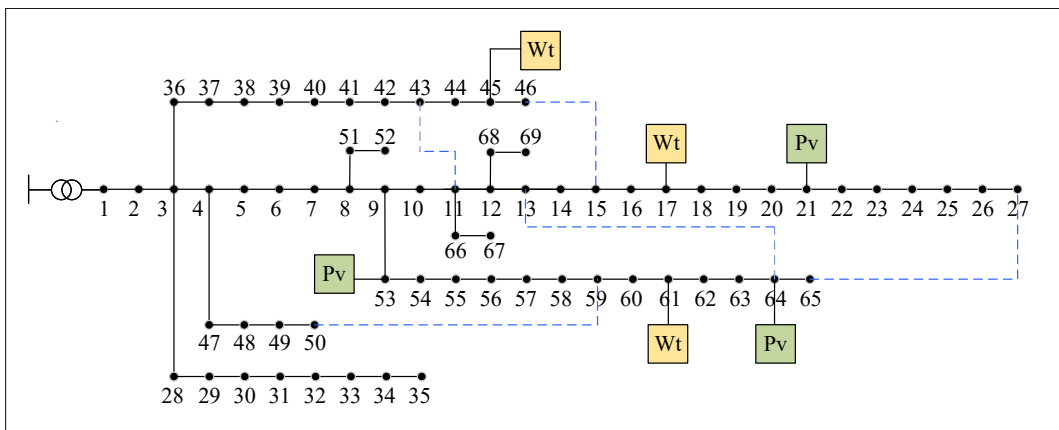


图10 IEEE 69节点配电网结构图

表6 69节点系统优化效果对比

方案	降损效果	越限场景数	动作次数
无重构	0.00%	92	0
BPSO	22.47%	30	28
RDQN	27.13%	7	22
SA-RDQN	33.84%	0	14

4 结论

本文提出一种基于结构感知深度强化学习的配电网动态重构决策求解模型, 通过性能-结构双重驱动对配电网拓扑进行筛选, 有效优化了重构的决策空间, 在保证优化效果的同时精简了动作维度; 同时, 本文所设计的 SA-Q 网络将配电网结构信息融入到价值评估过程中, 使得智能体能够感知并利用网络结构进行学习, 对动作空间高维且复杂的动态重构问题适配度极高. 实验结果表明, 本文方法在收敛速度和优化效果层面均展现出了明显优势, 并且具备良好的可拓展性, 能够有效改善配电网运行水平.

参考文献 (References)

- [1] 刘鑫蕊, 李新宇, 郭亮亮, 等. 考虑自治-互济的多园区端网协同弹性控制[J]. *控制与决策*, 2025, 40(9): 2693-2700.
(Liu X R, Li X Y, Guo L L, et al. Resilience cooperative control for MMG considering autonomy-mutual aid of edge-network[J]. *Control and Decision*, 2025, 40(9): 2693-2700.)
- [2] 姜思远, 高红均, 马望, 等. 基于深度学习的城市配电网多级动态重构决策方法[J]. *高电压技术*, 2024, 50(4): 1468-1477.
(Jiang S Y, Gao H J, Ma W, et al. Multi-level dynamic reconfiguration method for urban distribution networks based on deep learning algorithm[J]. *High Voltage Engineering*, 2024, 50(4): 1468-1477.)
- [3] 王丹璐, 孙秋野, 苏涵光. 多微网系统端网协同分布式实时智能优化[J]. *控制与决策*, 2024, 39(11): 3801-3809.
(Wang D L, Sun Q Y, Su H G. Collaborative distributed real-time intelligent optimization of multimicrogrid system[J]. *Control and Decision*, 2024, 39(11): 3801-3809.)
- [4] Lotfi H, Hajiabadi M E, Parsadust H. Power distribution network reconfiguration techniques: A thorough review[J]. *Sustainability*, 2024, 16(23): 10307.
- [5] 李扬, 韦钢, 马钰, 等. 含电动汽车和分布式电源的主动配电网动态重构[J]. *电力系统自动化*, 2018, 42(5): 102-110.
(Li Y, Wei G, Ma Y, et al. Dynamic reconfiguration of active distribution network considering electric vehicles and distributed generations[J]. *Automation of Electric Power Systems*, 2018, 42(5): 102-110.)
- [6] 马大中, 邢茗淇, 胡旭光, 等. 新型电力系统分布式协同优化调度研究综述[J]. *控制与决策*, 2025, 40(8): 2337-2360.
(Ma D Z, Xing M Q, Hu X G, et al. Distributed collaborative optimization dispatch in new power systems: A survey[J]. *Control and Decision*, 2025, 40(8): 2337-2360.)
- [7] 丁涛, 杨岳洋, 汪顺其, 等. 配电网拓扑重构下辐射状约束建模综述[J]. *电网技术*, 2025, 49(10): 4228-4244.
(Ding T, Yang Y Y, Wang S Q, et al. A review of radiality constraint modeling in distribution network reconfiguration[J]. *Power System Technology*, 2025, 49(10): 4228-4244.)
- [8] 汤一达, 吴志, 顾伟, 等. 主动配电网故障恢复的重构与孤岛划分统一模型[J]. *电网技术*, 2020, 44(7): 2731-2737.
(Tang Y D, Wu Z, Gu W, et al. Research on active distribution network fault recovery strategy based on unified model considering reconstruction and island partition[J]. *Power System Technology*, 2020, 44(7): 2731-2737.)
- [9] 傅长熠, 杨镜非, 胡继匀. 基于改进环网编码法的配电网辐射性判别方法[J]. *电测与仪表*, 2023, 60(9): 102-108.
(Fu C Y, Yang J F, Hu J Y. Discrimination method of radiation for distribution network based on improved ring network coding method[J]. *Electrical Measurement & Instrumentation*, 2023, 60(9): 102-108.)
- [10] Bahrami S, Chen Y C, Wong V W S. Dynamic distribution network reconfiguration with generation and load uncertainty[J]. *IEEE Transactions on Smart Grid*, 2024, 15(6): 5472-5484.
- [11] 高冠中, 杨胜春, 郭晓蕊, 等. 深度强化学习在含分布式柔性资源的电网优化调度中的应用研究综述[J]. *中国电机工程学报*, 2024, 44(16): 6385-6403.
(Gao G Z, Yang S C, Guo X R, et al. A review of research on the application of deep reinforcement learning in optimization dispatch of power grids with distributed flexible resources[J]. *Proceedings of the CSEE*, 2024, 44(16): 6385-6403.)
- [12] 王子晗, 高红均, 高艺文, 等. 基于深度强化学习的城市配电网多级动态重构优化运行方法[J]. *电力系统保护与控制*, 2022, 50(24): 60-70.
(Wang Z H, Gao H J, Gao Y W, et al. Multi-level dynamic reconfiguration and operation optimization method for an urban distribution network based on deep reinforcement learning[J]. *Power System Protection and Control*, 2022, 50(24): 60-70.)
- [13] 高放, 姚浩天, 高庆, 等. 考虑参数共享的深度强化学习双阶段分布式电源优化[J]. *中国电机工程学报*, 2025, 45(19): 7493-7509.
(Gao F, Yao H T, Gao Q, et al. Two-stage distributed generators optimization based on deep reinforcement learning with parameter sharing[J]. *Proceedings of the Chinese Society for Electrical Engineering*, 2025, 45(19): 7493-7509.)
- [14] 季颖, 王建辉. 基于深度强化学习的微电网在线优化调度[J]. *控制与决策*, 2022, 37(7): 1675-1684.

- (Ji Y, Wang J H. Online optimal scheduling of a microgrid based on deep reinforcement learning[J]. *Control and Decision*, 2022, 37(7): 1675-1684.)
- [15] Gao Y Q, Wang W, Shi J, et al. Batch-constrained reinforcement learning for dynamic distribution network reconfiguration[J]. *IEEE Transactions on Smart Grid*, 2020, 11(6): 5357-5369.
- [16] 董雷, 吴怡, 张涛, 等. 基于强化学习的含智能软开关主动配电网双层优化方法[J]. *电力系统自动化*, 2023, 47(6): 59-68.
(Dong L, Wu Y, Zhang T, et al. Reinforcement learning based double-layer optimization method for active distribution network with soft open point[J]. *Automation of Electric Power Systems*, 2023, 47(6): 59-68.)
- [17] Wang B B, Zhu H, Xu H H, et al. Distribution network reconfiguration based on NoisyNet deep Q -learning network[J]. *IEEE Access*, 2021, 9: 90358-90365.
- [18] 江昌旭, 郭辰, 刘晨曦, 等. 基于深度强化学习的主动配电网动态重构综述[J]. *高电压技术*, 2025, 51(4): 1801-1816.
(Jiang C X, Guo C, Liu C X, et al. Review of active distribution network dynamic reconfiguration based on deep reinforcement learning[J]. *High Voltage Engineering*, 2025, 51(4): 1801-1816.)
- [19] Wang R H, Bi X W, Bu S Q. Real-time coordination of dynamic network reconfiguration and volt-VAR control in active distribution network: A graph-aware deep reinforcement learning approach[J]. *IEEE Transactions on Smart Grid*, 2024, 15(3): 3288-3302.
- [20] 江昌旭, 郭辰, 林俊杰, 等. 图强化学习驱动的主动配电网动态重构与无功补偿协同优化方法[J/OL]. *中国电机工程学报*, 1-13.
(Jiang C X, Guo C, Lin J J, et al. A graph reinforcement learning-driven cooperative optimization method for dynamic reconfiguration and reactive power compensation in active distribution networks[J/OL]. *Proceedings of the Chinese Society for Electrical Engineering*, 1-13.)
- [21] Hessel M, Modayil J, Van Hasselt H, et al. Rainbow: Combining improvements in deep reinforcement learning[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): ■-■.
- [22] Zhan J P, Liu W J, Chung C Y, et al. Switch opening and exchange method for stochastic distribution network reconfiguration[J]. *IEEE Transactions on Smart Grid*, 2020, 11(4): 2995-3007.
- [23] Kim H W, Ahn S J, Yun S Y, et al. Loop-based encoding and decoding algorithms for distribution network reconfiguration[J]. *IEEE Transactions on Power Delivery*, 2023, 38(4): 2573-2584.
- [24] 李豪, 马驰, 孙菊, 等. 基于改进BPSO算法的含微网的配电网故障恢复方法研究[J]. *电测与仪表*, 2022, 59(11): 105-110.
(Li H, Ma C, Sun J, et al. Research on distribution network fault recovery method with micro-grid based on improved BPSO algorithm[J]. *Electrical Measurement & Instrumentation*, 2022, 59(11): 105-110.)

作者简介

王睿谦 (2000-), 男, 硕士生, 主要研究方向为电力系统优化调度, E-mail: seek4thefire@qq.com;

刘鑫蕊 (1981-), 女, 教授, 博士生导师, 主要研究方向为交直流微电网控制、配电网故障分析及自愈控制、综合能源系统优化, E-mail: liuxinrui@ise.neu.edu.cn;

富若涵 (1996-), 女, 博士生, 主要研究方向为电力系统优化调度、多智能体协调控制, E-mail: fh77701@163.com;

张浩晨 (2000-), 男, 博士生, 主要研究方向为电力系统优化调度、多智能体协调控制, E-mail: zhanghaochen@mails.neu.edu.cn;

孙秋野 (1977-), 男, 教授, 博士生导师, 主要研究方向为能源互联网的建模与优化运行、新能源并网运行、电力电子化的电力系统、多智能体协调控制、人工智能及其在能源系统中的应用, E-mail: sunqiuye@ise.neu.edu.cn.