

# 采用决策空间与策略模型动态迭代的线路过载紧急控制混合学习

张寿志<sup>1</sup>, 陈戈<sup>2</sup>, 张俊勃<sup>1</sup>, 彭颖<sup>1</sup>

(1. 华南理工大学电力学院, 广东省广州市 510641; 2. 南方电网能源发展研究院有限责任公司, 广东省广州市 510663)

**摘要:** 在新型电力系统中, 新能源可快速调节功率, 具有参与线路过载紧急控制的潜力。然而, 引入该措施后, 基于深度强化学习的紧急控制策略生成方法面临决策空间过大、求解复杂度高的挑战。为此, 提出一种决策空间与策略模型动态迭代的紧急控制混合学习方法。首先, 构建包含控制地点网络和控制量网络的双网络模型, 设计针对两个网络的迭代学习框架; 其次, 提出控制地点网络及学习要点, 设计基于灵敏度的样本生成方法, 学习控制地点网络; 然后, 提出控制量网络深度强化学习方法, 设计分段探索策略, 高效学习控制量网络; 接着, 提出控制量网络和控制地点网络的动态迭代实施流程; 最后, 在 IEEE 39 节点、IEEE 300 节点系统以及中国某省级电网中验证了所提方法的有效性。

**关键词:** 紧急控制; 新能源; 深度强化学习; 线路过载; 混合学习; 动态迭代; 决策空间; 样本生成

## 0 引言

线路过载是电力系统过载型连锁故障发展初期的关键特征<sup>[1-2]</sup>, 由初始事件引发的潮流转移可能导致部分线路功率越限, 若不能在此阶段通过紧急控制快速消除过载, 则可能引发更多设备相继退出, 最终导致系统失稳。因此, 对线路过载问题实施快速、有效的紧急控制, 是阻断连锁故障演化的关键。

线路过载紧急控制分为事件驱动型和响应驱动型<sup>[2]</sup>。前者由调度中心集中计算后下发控制指令; 后者通过安稳装置实现本地即时控制<sup>[3-4]</sup>。由于事件驱动型控制的计算速度决定线路过载的阻断效果, 实际电网采用“离线预决策”模式, 在离线阶段针对典型运行方式的预想事故集生成基于映射策略的策略表, 在线直接匹配形成方案<sup>[5]</sup>。

随着双碳目标的深入推进, 电力系统正经历可再生能源及电力电子设备大规模接入的深刻变革。新能源强随机性和波动性导致系统运行方式多变且难以精准预测, 叠加故障的随机性, 导致各种运行方式下的风险场景难以穷举, 传统离线形成的策略表可能失效<sup>[6]</sup>, 系统失稳风险增加。为应对上述问题,

一方面许多研究采用具有泛化能力的神经网络表征策略<sup>[7-8]</sup>; 另一方面, 紧急控制策略生成也逐步由“离线预决策”模式转向日前制定甚至小时级刷新的“在线预决策”模式, 对策略生成效率提出了要求。

策略生成效率与其计算方法相关, 采用神经网络表征的策略通常由机器学习方法来学习, 通过学习系统运行方式与控制策略间的映射关系, 获得具有泛化能力的映射策略模型(policy model, PM), 目前已在紧急控制策略生成领域受到广泛关注。

深度学习和深度强化学习(deep reinforcement learning, DRL)是常见的机器学习方法。前者基于大量样本驱动学习, 常用于预测和辨识等应用<sup>[9-11]</sup>; 后者是求解序贯决策问题的方法, 不依赖于大量样本, 通过与电网仿真环境交互, 渐进学习反映电网运作机理的紧急控制策略<sup>[12]</sup>。目前, 已有许多基于 DRL 的紧急控制方法。例如, 采用深度 Q 网络(deep Q network, DQN)<sup>[13]</sup>、双深度 Q 网络(double DQN, DDQN)<sup>[14]</sup>、深度确定性策略梯度(deep deterministic policy gradient, DDPG)<sup>[7]</sup>等学习功角稳定问题紧急控制策略; 采用 DDPG 方法<sup>[6]</sup>、基于图卷积的 DQN 方法<sup>[15]</sup>等学习频率稳定问题紧急控制策略; 采用 DDPG 方法<sup>[8]</sup>、DQN 方法<sup>[16]</sup>、基于图卷积的 DQN 方法<sup>[17]</sup>等学习电压紧急控制策略。

然而, 线路过载紧急控制涉及大范围潮流转移, 且新型电力系统中弃风弃光及储能充放电均可实现

收稿日期: 2025-06-25; 修回日期: 2026-01-26。

上网日期: 2026-04-03。

国家自然科学基金企业创新发展联合基金集成项目(U22B6007); 国家自然科学基金资助项目(52277101); 中央高校基本科研业务费专项资金资助项目(2024ZYGXZR109)。

秒级功率调节<sup>[18-20]</sup>,控制策略决策空间(decision space, DS)大,求解难度提升,虽可通过裁剪DS改善DRL收敛性<sup>[21]</sup>,但其裁剪规则缺乏灵活性,裁剪后的决策网络效果和泛化能力受到影响。

为此,本文针对线路过载紧急控制问题,提出DS与PM动态迭代的线路过载紧急控制混合学习(hybrid learning, HL)(以下简称DS-PM-HL)方法,通过构建包含控制地点网络和控制量网络的双网络模型,一方面由控制地点网络实现DS裁剪,另一方面设计基于分段探索策略的DRL方法以提高控制量网络学习效率,最终在迭代学习框架下实现控制地点网络与控制量网络的迭代收敛。

本文在IEEE 39节点、IEEE 300节点系统以及中国某省级电网中验证所提方法的有效性,其应用范围有望从线路过载紧急控制问题推广至其他电力系统安全稳定问题的紧急控制策略生成,为解决多类型稳定问题提供可行技术路径借鉴。

## 1 面向线路过载的紧急控制问题

### 1.1 线路过载紧急控制优化问题

新型电力系统中,线路过载紧急控制问题聚焦于稳态层面,其可表述为以最小化发电机功率调节、切机或切负荷的调节代价为目标,以电网安全为约束的数学问题。

1)控制变量 $a_k$ 为发电机和负荷的控制措施,包括调节发电机功率、切机、切负荷等,其中, $k$ 为节点编号。定义控制向量 $a$ 如下:

$$a = [a_{m,G}, a_{d,L}, a_{m,GC}] \quad (1)$$

$$m = 1, 2, \dots, N_M; d = 1, 2, \dots, N_D$$

式中:下标 $m$ 为发电机索引;下标 $d$ 为负荷索引; $N_M$ 、 $N_D$ 分别为发电机节点和负荷节点数; $a_{m,G}$ 、 $a_{m,GC}$ 、 $a_{d,L}$ 分别为调节发电机功率措施、切机措施和切负荷措施,其中,调节发电机功率措施与切机措施满足互斥关系。

2)约束条件包括潮流方程约束 $h_1(x_t, y_t, a_t) = 0$ 以及控制变量作用于系统后的系统代数变量越限约束 $h_2(x_t, y_t, a_t) > 0$ ,后者包括线路功率、节点电压等限制。其中, $x_t, y_t$ 分别为 $t$ 时刻电网运行状态的代数变量和系统参数,包括电压幅值、相角等代数变量,以及系统拓扑、给定节点功率等系统参数; $a_t$ 为 $t$ 时刻控制向量。

3)控制目标为最小化控制变量的动作代价。定义 $\rho_k$ 为调节 $a_t$ 中每个 $a_k$ 的代价,一次控制总代价为所有 $\rho_k$ 和 $a_k$ 乘积的总和。

基于上述定义,可建立优化模型如下:

$$\min F = \sum_k \rho_k a_k \quad (2)$$

$$\text{s.t. } h_1(x_t, y_t, a_t) = 0 \quad (3)$$

$$h_2(x_t, y_t, a_t) > 0 \quad (4)$$

$$a_t \in A_t \quad (5)$$

式中: $A_t$ 为 $t$ 时刻的DS,不仅体现了控制变量的可调范围,还体现了调节发电机功率措施和切机措施的互斥关系。

### 1.2 面向线路过载的紧急控制策略学习问题

面向线路过载的紧急控制策略学习问题采用“在线预决策”的应用模式,根据未来电网运行方式和预想事故形成潜在风险场景,针对风险场景定期学习并更新紧急控制策略,然后,在实际应用时,直接匹配紧急控制策略形成控制方案。其中,预想事故通常考虑线路 $N-1$ 、发电机 $N-1$ 、新能源波动等不确定事件<sup>[21-22]</sup>。在线刷新策略的周期通常由实际需求决定。例如,日前策略是日前阶段生成针对第2日的控制策略,实时策略是日内阶段生成针对分钟级或小时级内风险场景的策略。

上述紧急控制策略可用强化学习获取,具体学习框架如图1所示。为了增强紧急控制策略的泛化能力,策略采用神经网络的参数化表达方式,记为 $\pi$ ,使其经过数据驱动学习后具备从局部样本向全局状态空间泛化的潜力。策略网络 $\pi$ 由控制地点网络 $\phi$ 和控制量网络 $\eta$ 组成。前者以电网运行状态 $s$ 为输入,控制地点选择0-1向量 $c$ 为输出,用于裁剪低贡献或无贡献控制地点,减少控制变量数量;后者以电网运行状态 $s$ 为输入,以各种紧急控制措施的调节量向量 $a$ 为输出,确定给定控制地点后的具体调节量。 $\phi$ 和 $\eta$ 共同作用,由 $a$ 和 $c$ 相乘得到裁剪DS后的调节量向量 $a^*$ 。

考虑到控制地点的选择与电力专业知识密切相关,可以利用领域知识指导 $\phi$ 深度学习,再对 $\eta$ 进行强化学习,以降低学习难度。工作流程概括如下:给定 $\psi^{(k)}, \phi^{(k)}$ 与 $\eta^{(i)}$ 基于电网状态 $s$ 产生控制动作 $a^*$ ;  $a^*$ 作用到电网后,电网状态转变为 $s'$ ,并产生回报 $r$ ;基于 $r, s', s, a^*$ ,用DRL更新 $\eta^{(i)}$ 为 $\eta^{(i+1)}$ ,其中, $k$ 和 $i$ 为迭代次数;如果一直学不到有效 $\eta$ ,考虑DS过度裁剪,动态增加控制地点,扩大DS,用深度学习更新 $\psi^{(k)}$ 为 $\psi^{(k+1)}$ ;然后,在新的 $\psi^{(k+1)}$ 下继续学习 $\eta^{(i+1)}$ 。

## 2 控制地点网络及学习要点

### 2.1 控制地点网络模型

定义控制地点选择向量 $c$ 如下:

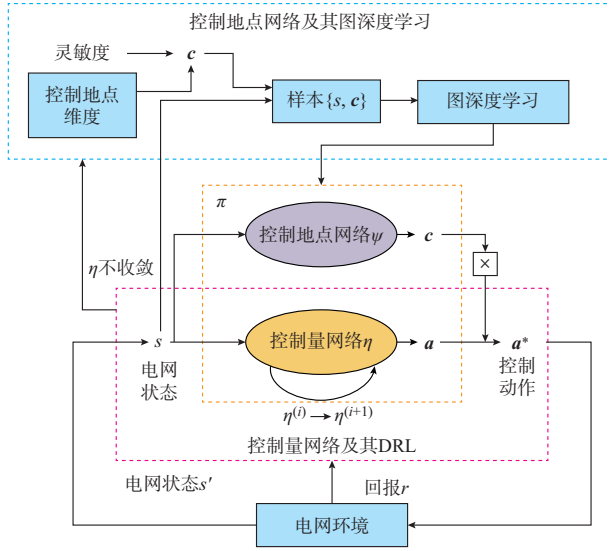


图1 基于DS-PM-HL的紧急控制策略学习框架  
Fig.1 Emergency control strategy learning framework based on DS-PM-HL

$$c = [c_{1,G}, c_{2,G}, \dots, c_{N_m,G}, c_{1,L}, c_{2,L}, \dots, c_{N_d,L}, c_{1,GC}, c_{2,GC}, \dots, c_{N_m,GC}] \quad (6)$$

式中: $c$ 维度以及节点位置与 $a$ 一致,其元素为0和1,分别表示裁剪和保留该元素对应的控制地点。

考虑到电网运行状态与拓扑强相关,选用对电网拓扑特征提取能力较强的基于Transformer的图神经网络(Transformer-based graph neural network, T-GNN)表征控制地点网络 $\psi$ ,如图2所示,包括3层Transformer卷积(Transformer convolution, TC)层、4层批归一化(batch normalization, BN)层和1个全连接网络(fully connected network, FCN)(记为FCN<sub>1</sub>),网络参数设为 $\theta_\psi$ ,激活函数为修正线性单元(ReLU)。 $\psi$ 以状态 $s_L = [X_L, M]$ 为输入,其中, $X_L = [V, \varphi, P, Q]$ 为节点特征, $V, \varphi, P, Q$ 分别为节点电压、相位、注入有功和无功功率, $M$ 为考虑线路和变压器通断的电网拓扑邻接矩阵。 $\psi$ 通过BN和TC进行交替特征提取。然后,经过FCN<sub>1</sub>进行概率拟合,并通过Sigmoid函数变换输出各控制地点的选择概率。当输出概率大于0.5时,选择该地点;否则,裁剪该地点。最后,输出向量 $c$ 。

## 2.2 基于灵敏度的样本生成方法

为训练 $\psi$ ,需生成样本 $\{s, c\}$ 。本节提出基于灵敏度的DS裁剪模型样本生成流程。

步骤1:裁剪对过载线路功率控制效果相反或无效的控制地点。

首先,针对过载线路 $L_{ij}$ ,计算线路功率 $P_{ij}$ 与节点 $k$ 注入功率 $P_k$ 的灵敏度 $S_{ij,k}$ 如下:

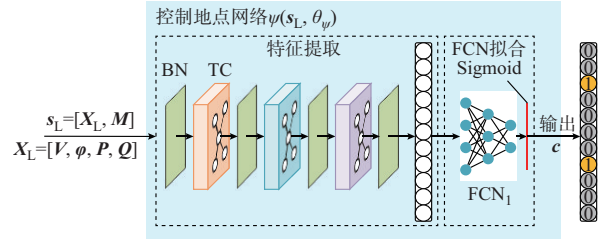


图2  $\psi$ 网络结构  
Fig.2 Structure of  $\psi$  network

$$S_{ij,k} = \frac{\partial P_{ij}}{\partial P_k} = \left( \frac{\partial P_{ij}}{\partial \varphi} \right)^T \frac{\partial \varphi}{\partial P_k} = [\dots B_{ij} \dots -B_{ij} \dots] (B')^{-1} \quad (7)$$

式中: $B_{ij}$ 为线路 $L_{ij}$ 的电纳; $B'$ 为电纳矩阵。

其次,根据 $S_{ij,k}$ 的符号,结合过载线路功率 $P_{ij}$ 、过载线路功率的调节方向 $\Delta P_{ij}$ ,通过表1各变量的符号关系确定节点功率调节方向 $\Delta P_k$ 和控制变量类型。

表1  $P_{ij}, \Delta P_{ij}, S_{ij,k}$ 与 $\Delta P_k$ 的符号关系  
Table 1 Symbolic relationships between  $P_{ij}, \Delta P_{ij}, S_{ij,k}$  and  $\Delta P_k$

$P_{ij}$	$\Delta P_{ij}$	$S_{ij,k}$	$\Delta P_k$	可选控制变量类型
$P_{ij} > 0$	$\Delta P_{ij} < 0$	$S_{ij,k} > 0$	$\Delta P_k < 0$	减小发电机出力/切机
$P_{ij} > 0$	$\Delta P_{ij} < 0$	$S_{ij,k} < 0$	$\Delta P_k > 0$	增加发电机出力/切负荷
$P_{ij} < 0$	$\Delta P_{ij} > 0$	$S_{ij,k} > 0$	$\Delta P_k > 0$	增加发电机出力/切负荷
$P_{ij} < 0$	$\Delta P_{ij} > 0$	$S_{ij,k} < 0$	$\Delta P_k < 0$	减小发电机出力/切机

然后,考虑到不同电源的功率可调方向不同,结合控制地点 $k$ 的电源类型,裁剪控制效果相反或无效的控制地点,保留具有一定可调裕度的控制地点,如表2所示。其中, $P'_{hy}, P'_{pv}, P'_{st}$ 分别为调节后的水电、风光、储能有功出力; $P_{hy, max}$ 为水电最大可调出力; $P_{pv}$ 为当前风光有功出力; $P_{st, max}$ 和 $P_{st, min}$ 分别为储能可调上、下限, $P_{st, min}$ 为负时表示充电状态。

表2 各种形式发电机的调节范围  
Table 2 Adjustment range of various kinds of generators

类型	调节范围
火电	
风光	$0 \leq P'_{pv} \leq P_{pv}$
独立储能	$P_{st, min} \leq P'_{st} \leq P_{st, max}$
水电	$0 \leq P'_{hy} \leq P_{hy, max}$
风光+储	$P_{st, min} \leq P'_{pv} \leq P_{pv} + P_{st, max}$

步骤2:确定初始控制地点选择向量 $c$ ,形成用于学习 $\psi$ 的样本集 $\{s, c\}$ 。

对步骤1中保留的控制地点按照 $|S_{ij,k}|$ 的大小

由高到低排序,依次选择控制地点,且所选发电机和负荷节点的可用控制量应相对平衡以维持源荷功率平衡。

选择控制地点后,在  $c$  中相应位置置 1,直到通过式(8)计算的  $I'_{ij}$  满足式(9)约束,由此确定  $c$ 。

$$I'_{ij} = \frac{P'_{ij}}{\sqrt{3} V_{ij} \cos \varphi} \quad (8)$$

$$I'_{ij} \leq \beta I_{ij, \text{limit}} \quad (9)$$

式中:  $I'_{ij}$  和  $P'_{ij}$  分别为各节点功率变化后线路  $L_{ij}$  的电流和有功功率;  $V_{ij}$  为线路电压;  $\varphi$  为线路功率因数角;  $I_{ij, \text{limit}}$  为线路电流限值;  $\beta$  为过载线路的电流控制裕度。

步骤 3: 检查  $c$  中元素,即调节发电机功率、切机和切负荷的可用控制量是否平衡,防止只有发电机或负荷控制措施,打破系统功率平衡。如果  $c$  中仅有发电机控制措施,则根据灵敏度  $S_{ij,k}$  额外选择控制效益最高的负荷节点;如果  $c$  中仅有负荷控制措施,则根据灵敏度  $S_{ij,k}$  额外选择控制效益最高的发电机节点。

针对所有过载场景,通过上述规则确定控制地点向量  $c$  后,形成控制地点网络的学习样本集。然后,采用常规图深度学习训练方法训练  $\psi$ 。

### 3 控制量网络 DRL 方法

给定  $\psi$  后,可通过 DRL 学习控制量网络  $\eta$ 。将式(2)一式(4)转化成强化学习回报  $r$ 。其中,式(2)中目标函数越小,回报越大;约束式(3)和式(4)不能满足时,负回报较大。为此,设计考虑分段探索策略的控制量网络 DRL 方法。

#### 3.1 控制量网络 DRL 模型

控制量网络的 DRL 模型要素如下:

1) 状态。选取电网节点特征  $X_V$  和邻接矩阵  $M$  作为状态向量  $s$ , 记为  $s = [X_V, M]$ 。考虑到控制地点网络作用下 DS 已被裁剪变小,为提升控制量网络的学习效率,  $X_V$  仅考虑节点电压  $V$  和相位  $\varphi$ ,  $\varphi$  为弧度制,反映电网运行方式;  $M$  随着线路或主变压器的通断而变化,当线路和主变压器为连通状态时,  $M$  中相应位置的元素置 1, 否则置 0。在一个  $n$  节点电网中,节点特征  $X_V$  可表示为:

$$X_V = [V_1, V_2, \dots, V_n, \varphi_1, \varphi_2, \dots, \varphi_n] \quad (10)$$

式中:  $V_n$  和  $\varphi_n$  分别为节点  $n$  的电压和相位。

2) 动作。动作  $a$  定义如式(1)所示,其中,  $a_{m,G}$ 、 $a_{d,L}$  和  $a_{m,GC}$  的取值均在  $[0, 1]$  间,分别表示各措施调节量的百分比。定义第  $m$  台发电机出力上、下限分别为  $P_{m, \text{up}}$  和  $P_{m, \text{low}}$ , 发电机出力  $P_{m,G}$  由式(11)计算。

其中,各种类型发电机的具体可调范围见表 2 中发电机调节范围约束。切负荷后的负荷有功、无功功率  $P'_{d,L}$ 、 $Q'_{d,L}$  分别如式(12)和式(13)所示。其中,负荷节点的有功功率  $P_{d,L}$  和无功功率  $Q_{d,L}$  以等比例  $a_{d,L}$  切除。切机措施属于二分类问题,通过阈值法定义  $a_{m,GC}$  超过阈值  $N_{\text{limit}} = 0.5$  时为切机,如式(14)所示。

$$P_{m,G} = a_{m,G}(P_{m, \text{up}} - P_{m, \text{low}}) \quad m = 1, 2, \dots, N_M \quad (11)$$

$$P'_{d,L} = a_{d,L} P_{d,L} \quad d = 1, 2, \dots, N_D \quad (12)$$

$$Q'_{d,L} = a_{d,L} Q_{d,L} \quad d = 1, 2, \dots, N_D \quad (13)$$

$$P_{m,G} = \begin{cases} 0 & a_{m,GC} \geq N_{\text{limit}} \\ P_{m,G} & a_{m,GC} < N_{\text{limit}} \end{cases} \quad (14)$$

3) 控制量网络  $\eta$ 。  $\eta$  为 T-GNN, 包含 3 层 TC 层、4 层 BN 层和 2 个 FCN(记为 FCN<sub>2</sub>、FCN<sub>3</sub>), 采用 ReLU 激活函数,如图 3(a)紫色方框所示。网络输入为故障后的潮流状态  $s = [X_V, M]$ , 经过特征提取和拟合后,输出动作  $a$ 。特征提取由 4 层 BN 层和 3 层 TC 层交替组成,分别用于调节特征间分布以及提取电网拓扑特征。然后,将提取特征后的图数据平摊成一维向量,作为 FCN<sub>2</sub> 和 FCN<sub>3</sub> 的输入,拟合动作向量  $a$ 。其中,FCN<sub>2</sub> 拟合连续变量  $a_G$  和  $a_L$ , FCN<sub>3</sub> 拟合离散变量  $a_{GC}$ , 合成输出  $a$ 。最终与  $\psi$  生成的向量  $c$  相乘得到考虑控制地点裁剪约束的向量  $a^*$ 。

4) 回报。回报由电网风险回报  $R_1$ 、节点电压约束回报  $R_2$ 、线路 80% 过载约束回报  $R_3$ 、控制成本回报  $R_4$  组成,如式(15)所示。

$$r = \begin{cases} \sum_{l=1}^4 \omega_l R_l & h_1(x_t, y_t, a_t) = 0 \\ -1 & h_1(x_t, y_t, a_t) \neq 0 \end{cases} \quad (15)$$

式中:  $\omega_l$  为权重。当潮流方程约束不满足,即  $h_1(x_t, y_t, a_t) \neq 0$  时,潮流不收敛,直接给定负回报 -1。

$R_1$  用于评估电网是否存在过载,如式(16)所示。

$$R_1 = \begin{cases} 1 & \forall L_{ij}, I_{ij} \leq I_{ij, \text{limit}} \\ -1 & \exists L_{ij}, I_{ij} > I_{ij, \text{limit}} \end{cases} \quad (16)$$

式中:  $I_{ij}$  为  $L_{ij}$  的电流。

$R_2$  用于评估电网节点电压的越限情况。当存在节点电压越限时,给予负回报,如式(17)所示。

$$R_2 = \begin{cases} 1 & \forall 1 \leq k \leq n, V_{k, \text{min}} \leq V_k \leq V_{k, \text{max}} \\ -1 & \exists 1 \leq k \leq n, V_k < V_{k, \text{min}} \text{ 或 } V_k > V_{k, \text{max}} \end{cases} \quad (17)$$

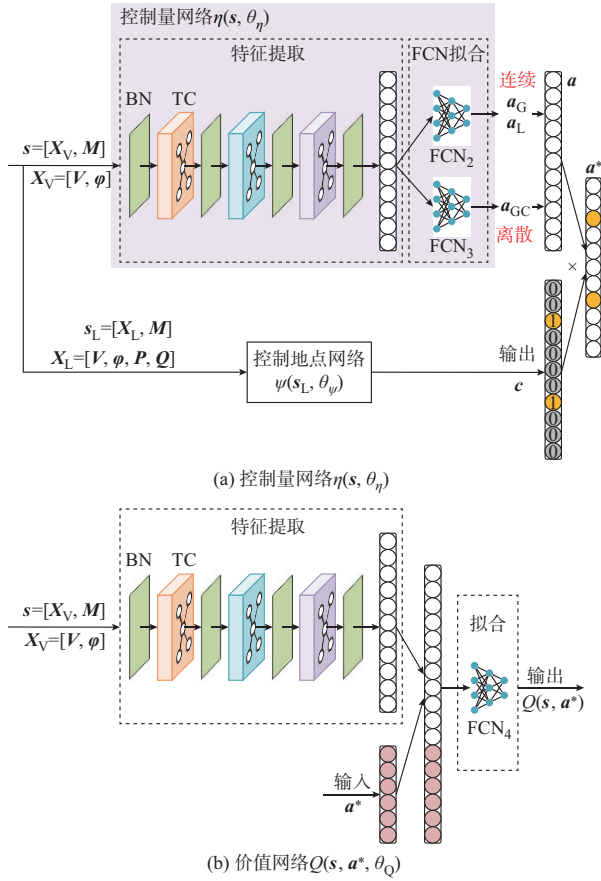


图3 控制量网络和价值网络结构  
Fig. 3 Structure of control value network and value network

式中:  $V_{k, \max}$  和  $V_{k, \min}$  分别为节点  $k$  的电压上、下限。

$R_3$  用于评估原过载线路是否控制到限值的80%以下,防止控制后受到轻微扰动再次引发过载,如式(18)所示。

$$R_3 = \frac{\sum_{l=1}^{N_l} \Delta r_l}{N_{80\%}} \quad (18)$$

$$\Delta r_l = \begin{cases} 0 & I_l \leq I_{l, \text{limit}} \\ -5 \left( \frac{I_l}{I_{l, \text{limit}}} - 0.8 \right) & I_l > I_{l, \text{limit}} \end{cases} \quad (19)$$

式中:  $N_{80\%}$  为过载线路电流超过限值的80%的数量;  $l$  为线路编号索引;  $N_l$  为线路数量;  $\Delta r_l$  为第  $l$  条线路的线路电阻修正量,取值在  $[-1, 0]$  间;  $I_l$  和  $I_{l, \text{limit}}$  分别为第  $l$  条线路的实际运行电流和其电流安全限值。

控制成本回报  $R_4$  反映对控制措施调节量的惩罚,防止过度调节或切除,如式(20)所示。

$$R_4 = - \left( \sum_m |\Delta P_m| + \sum_d |\Delta P_d| \right) \quad (20)$$

式中:  $\Delta P_m$  为发电机节点功率的调节量;  $\Delta P_d$  为负荷节点功率的调节量。切除第  $m$  台发电机视为把发电机功率调为0。

5) 价值网络。使用行动价值函数  $Q(s, a^*)$  衡量在状态  $s$  下选择动作  $a^*$  的期望收益,并选用 T-GNN 来表示,记为  $Q(s, a^*, \theta_Q)$ ,其中,  $\theta_Q$  为  $Q$  网络的参数。 $Q$  网络结构如图3(b)所示,其由4层BN层、3层TC层、一次特征拼接和一个FCN组成,激活函数采用ReLU。网络输入为状态  $s = [X_V, M]$  和动作  $a^*$ 。 $s$  先经过TC层和BN层进行特征提取;然后,将特征提取后的  $s$  展开成一维向量,再与  $a^*$  拼接;最后,经过  $FCN_4$  拟合行动价值  $Q$ 。

### 3.2 控制量网络 DRL 算法

DRL 算法包括环境决策部分和控制量网络强化学习部分,如图4所示。

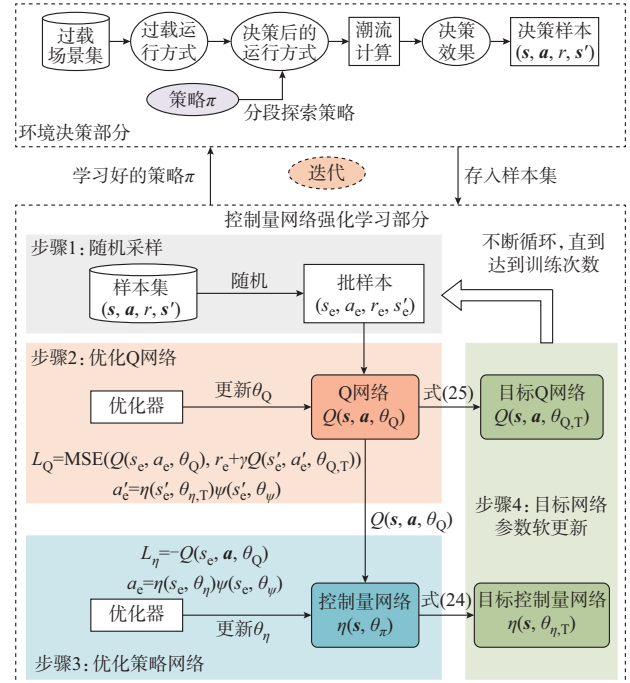


图4 控制量网络  $\eta$  的 DRL 算法原理  
Fig. 4 Principle of DRL algorithm for control value network  $\eta$

环境决策部分的作用是产生用于强化学习的样本,其流程可概括为针对所有过载运行方式,采用策略  $\pi$  决策产生  $a$ ,作用于运行方式,并通过潮流计算获得决策效果  $(s, a, r, s')$ 。然而,  $\pi$  未收敛时,其产生的  $a$  一般效果较差,甚至长时间决策出负回报。如果一直按照  $\pi$  决策,会停留在局部较差空间而难以收敛。因此,决策时通过探索跳出原有经验,尝试新动作,有利于找到较大回报的决策,指引策略网络向高回报调整。

设计包含随机与启发式的分段探索策略,如式

(21)所示。

$$a = \begin{cases} f_r(a) & N_{De} \leq N_{E,1}, N_{PR} < N_{P,1} \\ f_g(a) & N_{De} \leq N_{E,1}, N_{PR} \geq N_{P,1} \\ \pi(s, \theta_\pi) & N_{De} > N_{E,1} \end{cases} \quad (21)$$

式中： $N_{De}$ 为环境决策次数； $N_{E,1}$ 为探索次数限制； $N_{PR}$ 为探索中产生正回报的次数； $N_{P,1}$ 产生正回报的次数阈值； $\theta_\pi$ 为策略 $\pi$ 的参数； $f_r(a)$ 和 $f_g(a)$ 均为探索策略，前者为随机探索策略，随机在动作空间选择动作，后者为基于梯度的启发式探索策略，按照回报梯度方向选择动作。当 $N_{De}$ 不超过 $N_{E,1}$ 时进行探索；否则，按原策略 $\pi$ 决策。探索类型将由 $N_{P,1}$ 来界定，在决策初期通过 $f_r(a)$ 策略探索正回报决策，当探索出 $N_{P,1}$ 个正回报决策后，采用 $f_g(a)$ 策略，利用正回报决策进一步基于梯度指导找到更高回报动作，实现决策前期快速积累正回报样本。

启发式探索策略 $f_g(a)$ 的流程如下：

步骤1：假设通过随机探索策略 $f_r(a)$ 已找到正回报的动作集合为 $A_+$ 。

步骤2：从 $A_+$ 中选出两个回报最高动作，假设回报最高和次高动作分别为 $a_1$ 和 $a_2$ ，回报分别为 $r_1$ 和 $r_2$ ， $r_1 > r_2$ 。由于希望往高回报的方向选择动作，故采用式(22)来确定本次决策的动作 $a$ 。

$$a = a_1 + \alpha(a_1 - a_2) \quad (22)$$

式中： $\alpha$ 为梯度更新的步长。

步骤3： $a$ 作用于环境后，产生回报 $r$ 。如果 $r$ 为正回报，则将该次决策添加到集合 $A_+$ 中。

步骤4：判断 $N_{De}$ 是否小于 $N_{E,1}$ 。如果是，则跳转到步骤2，否则结束。

通过上述环境决策形成样本后，对这些样本进行强化学习，概述如下：

步骤1：从样本集随机选取一批样本 $(s_e, a_e, r_e, s'_e)$ ，其中， $s_e$ 为样本状态， $a_e$ 为本次决策的动作， $r_e$ 为本次决策的回报， $s'_e$ 为动作后的状态。

步骤2：基于时序差分思想对Q网络 $Q(s, a, \theta_Q)$ 进行迭代更新，如式(23)所示，将均方误差(mean squared error, MSE)作为损失函数，更新参数 $\theta_Q$ 。

$$Q(s_e, a_e, \theta_Q) = r_e + \gamma Q(s'_e, a'_e, \theta_Q) \quad (23)$$

式中： $a'_e$ 为在状态 $s'_e$ 通过策略所选出的动作； $\gamma$ 为折扣因子。

为防止迭代时式(23)等号两端因两个 $\theta_Q$ 同时更新引起训练不稳定，进一步设计软更新机制，在迭代中引入目标Q网络 $Q(s, a, \theta_{Q,T})$ 和目标策略网络 $\eta(s, \theta_{\eta,T})$ ，结构分别与 $Q(s, a, \theta_Q)$ 和 $\eta(s, \theta_\eta)$ 相同，参数分别为 $\theta_{Q,T}$ 和 $\theta_{\eta,T}$ 。把式(23)中 $Q(s'_e, a'_e, \theta_Q)$ 替换

为 $Q(s'_e, a'_e, \theta_{Q,T})$ ，此时损失函数如图4的 $L_Q$ 所示。 $a'_e$ 由 $\pi(s_e, \theta_{\pi,T})$ 产生，根据图4所示框架和流程， $\psi$ 网络为给定。在迭代更新 $\theta_Q$ 和 $\theta_\eta$ 时，目标网络参数 $\theta_{Q,T}$ 和 $\theta_{\eta,T}$ 均被固定。在迭代完 $\theta_Q$ 和 $\theta_\eta$ 后，再对目标网络参数 $\theta_Q$ 和 $\theta_\eta$ 进行软更新。

步骤3：在给定 $\psi$ 下，采用 $-Q(s_e, a, \theta_Q)$ 作为迭代 $\eta(s, \theta_\eta)$ 的损失函数，更新参数 $\theta_\eta$ ，相关表达式如图4所示。

步骤4：基于式(24)和式(25)，软更新目标网络 $\theta_{Q,T}$ 和 $\theta_{\eta,T}$ 。

$$\theta_{\eta,T} \leftarrow \tau \theta_\eta + (1 - \tau) \theta_{\eta,T} \quad (24)$$

$$\theta_{Q,T} \leftarrow \tau \theta_Q + (1 - \tau) \theta_{Q,T} \quad (25)$$

式中： $\tau$ 为软更新参数。

步骤5：不断循环上述步骤1至4，直到达到训练次数。

上述环境决策部分和强化学习部分进行快速迭代，通过采用学习后的策略在环境中决策获得决策效果，以此来不断学习到控制量网络。

#### 4 DS与PM动态迭代实施流程

3.2节的控制量网络 $\eta$ 基于给定的控制地点网络 $\psi$ 学习，然而当DS被过度裁剪时，可能学不出解决过载场景的 $\eta$ ，此时需要扩大DS，增加控制地点。

本节提出DS-PM-HL方法，对 $\eta$ 和 $\psi$ 进行联合迭代，如图5所示。

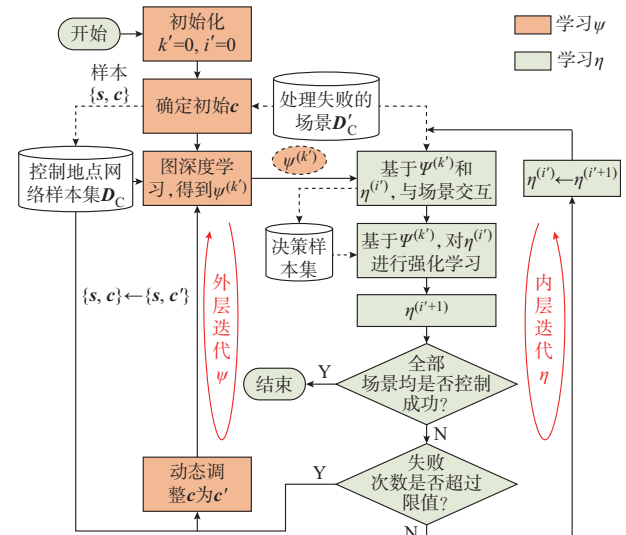


图5 DS-PM-HL方法中 $\eta$ 和 $\psi$ 的联合实施流程  
Fig. 5 Joint implementation process of  $\eta$  and  $\psi$  in DS-PM-HL method

具体流程如下：

步骤1：初始化 $\eta$ 的迭代次数 $i' = 0$ 和 $\psi$ 的迭代次数 $k' = 0$ 。

步骤2:进入 $\psi$ 的外层迭代,针对每一过载场景,首先基于2.2节步骤1裁剪对过载线路功率控制效果相反或无效的控制地点;然后,基于步骤2和步骤3确定初始 $c$ ,得到样本 $\{s, c\}$ ,样本集记为 $D_c$ 。

步骤3:根据 $D_c$ ,通过图深度学习得到 $\psi^{(k)}$ 。

步骤4:进入 $\eta$ 的内层迭代,在给定 $\psi^{(k)}$ 下不断学习 $\eta^{(i)}$ ,并令 $i' \leftarrow i' + 1$ 。如果过载场景控制失败次数超过限值,则进入 $\psi$ 的外层迭代,跳转至步骤5;如果均控制成功,则跳转至步骤6。

步骤5:在外层迭代中向 $\psi$ 动态添加控制地点,扩大DS。新增控制地点需考虑功率平衡,所选发电机和负荷节点的数量应相对平衡。当前控制地点向量 $c$ 中已选发电机节点数量大于负荷节点数量时,优先选择 $P_{ij}S_{ij,k} < 0, |S_{ij,k}|$ 最大且 $c$ 中元素为0的负荷节点;否则优先选择 $|S_{ij,k}|$ 最大、 $c$ 中元素为0且满足表1和表2控制调节需求的发电机节点。新增控制地点后, $c$ 动态调整为 $c'$ ,并用新样本 $\{s, c'\}$ 替换原样本 $\{s, c\}$ ,得到新样本集 $D'_c$ ,基于 $D'_c$ 训练 $\psi^{(k)}$ 得到 $\psi^{(k+1)}$ ,令 $k' \leftarrow k' + 1$ ,跳转至步骤4。

步骤6:记录 $\eta^{(i')}$ 和 $\psi^{(k)}$ ,学习结束。

## 5 算例分析

本文在改进的IEEE 39节点、IEEE 300节点系统与中国某省级电网中分析和验证所提DS-PM-HL方法,并与其他算法进行比较,包括基于决策空间裁剪图深度强化学习的过载主导型连锁故障紧急控制方法(记为DSP-GDRL算法)<sup>[21]</sup>、DDPG算法、DDQN算法,以验证本文方法有效性。实验用计算机采用CPU为Intel Core i7-10 875H, 2.30 GHz, GPU为RTX 2060,内存为6 GB。

### 5.1 IEEE 39节点系统算例分析

对IEEE 39节点系统进行改进,系统拓扑如图6所示。改进后的系统含5台火电机组、2台水电机组、2台配有储能的风机和1台无储能风机,并在节点4中新增独立储能。由于风电和光伏的功率调节方式与效果一致,算例中仅设置风电。

#### 5.1.1 算例设计

改进的IEEE 39节点系统算例设计如下:

1)过载场景生成。在某一基础运行方式 $O_0$ 下,模拟线路 $N-1$ 、发电机 $N-1$ 以及随机多个新能源和负荷在 $[-50\%, 50\%]$ 范围内随机波动的情况,生成大量不同线路状态、机组投停运、源荷分布的运行方式,场景集合记为 $O_D$ 。然后,进行线路过载扫描得到过载场景集合 $O_O$ 。

2)单一场景学习与验证分析。对上述 $O_O$ 中线

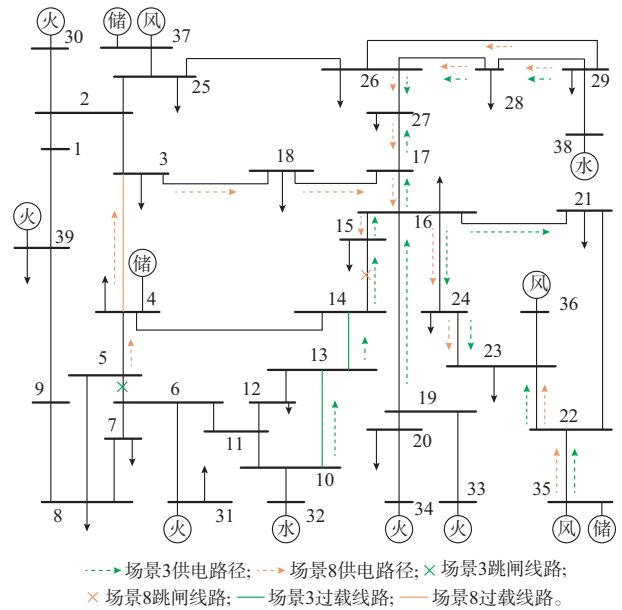


图6 改进的IEEE 39节点系统拓扑结构  
Fig. 6 Topology of modified IEEE 39-bus system

路 $N-1$ 后的过载场景分别进行线路过载紧急控制强化学习。把各场景下学到的策略网络用于该场景的紧急控制,与其他算法比较来分析控制效果。

3)多场景混合学习与验证分析。对上述 $O_O$ 中所有过载场景进行多场景混合学习,其中, $O_O$ 中场景的线路通断状态不同,验证了本文所提方法可在不同运行方式间进行混合学习。将学习到的策略网络用于 $O_O$ 中所有过载场景的紧急控制验证,与其他算法比较来分析控制效果。

4)泛化性能验证分析。基于 $O_O$ 中的过载场景,模拟多个负荷或新能源在 $[-20\%, 20\%]$ 范围内随机波动生成大量过载场景,集合记为 $O_F$ 。采用3)中训练出的策略网络对 $O_F$ 中的场景进行紧急控制,与其他算法比较来分析控制效果,验证了本文方法训练出的策略网络在源荷波动下也适用,具有泛化性能。

在对比算法方面,设置考虑发电机功率调节和切机切负荷措施的DSP-GDRL算法和DDPG算法(DDPG-G),其中,DSP-GDRL算法设置电流控制裕度分别为0.7和1,分别记为DSP-GDRL-1和DSP-GDRL-2;设置仅考虑传统切机切负荷措施的DDPG算法和DDQN算法,其中,DDQN算法中每个负荷节点的切除量离散成10个动作,按10%递增。

#### 5.1.2 单一场景学习与验证分析

集合 $O_O$ 中线路 $N-1$ 的过载场景及控制地点选择见附录A表A1,其中,控制电流裕度 $\beta$ 设为1,所

选出的控制地点可使线路电流  $I_{ij}$  最大降低到限值  $\beta I_{ij,limit}$  以下,说明所选择的控制地点的有效性。

对  $O_0$  中线路  $N-1$  后过载的 9 个场景进行紧急控制强化学习,并分析场景 3 和 8,如图 7 所示。

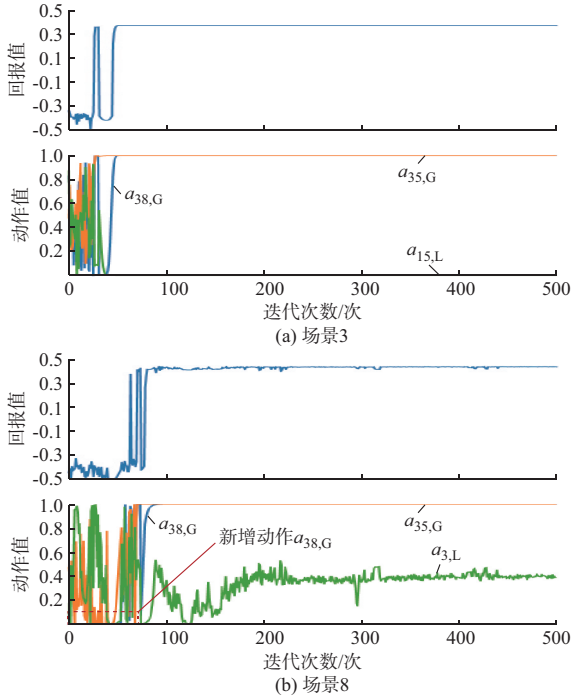


图 7 部分场景回报和动作曲线  
Fig. 7 Reward and action curves of partial scenarios

图 7(a)为场景 3 的回报和动作曲线,两者均很快达到稳定值。针对过载线路  $L_{10,13}$  和  $L_{13,14}$ ,如图 6 场景 3 下供电路径可知,增加节点 35 风储出力,可使节点 23 的部分供电需求由  $L_{24,23}$  转移到  $L_{22,23}$ ;切除节点 15 的全部负荷以减少来源于  $L_{14,15}$  的供电功率;增加节点 38 的水电出力,使节点 27 的部分供电需求由  $L_{17,27}$  转移到  $L_{26,27}$ 。上述措施均可减少  $L_{10,13}$  和  $L_{13,14}$  的过载功率。

图 7(b)为场景 8 的回报和动作曲线,可以看出,回报和动作均收敛。针对过载线路  $L_{3,4}$ ,如图 6 场景 8 下供电路径可知,通过切除节点 3 约 40% 的负荷能降低  $L_{3,4}$  的功率传输需求,同时增加节点 35 的出力,使来自  $L_{18,17}$  的供电路径转移到  $L_{22,23}$ 。然而,学习中发现仅通过两个控制地点难以控制成果。因此,通过第 4 章步骤 5 动态添加控制地点  $a_{38,G}$ 。最终,经过强化学习后,学到通过增加节点 38 的出力,使原本来自  $L_{18,17}$  的供电路径转移到来自  $L_{27,17}$  的供电路径。上述措施均可减少  $L_{3,4}$  的过载功率。

由于篇幅原因,表 3 给出各场景下各种算法的学习效果,不具体展开分析。本文通过给定总调节

量  $\Delta P$ 、控制地点数量  $N_a$ 、综合效果  $E$ 、电压偏差合格占比和控制后电压恢复的节点占比五种指标来分析控制效果。由于所有场景的电压指标均正常,故表 3 仅给出 3 种指标,不展示其余两项电压指标。可以看出,本文所提 DS-PM-HL 方法优先调节发电机功率解决线路过载,在总调节量、控制地点数量和综合效果上,均显著优于对比算法。

总调节量和控制地点数量是紧急控制中为避免大范围控制所设定的需求,总调节量越小、控制地点数量越少越好;综合效果  $E$  综合考虑 3.1 节中电网风险回报  $R_1$ 、节点电压约束回报  $R_2$ 、控制成本回报  $R_4$ ,以反映本次控制质量,如式(26)所示。其中,  $R_4$  与  $N_a$  相乘体现控制地点数量越少越好的原则;电压偏差合格占比和控制后电压恢复的节点占比均用于衡量控制后稳态条件下的电压质量,参考国标 GB/T 12325<sup>[23]</sup>、现有研究<sup>[24]</sup>及地方调度规程,设置事故后电压偏差范围的标幺值为 0.9~1.1。

$$E = w_1 R_1 + w_2 R_2 + w_4 R_4 N_a \quad (26)$$

虽然 DSP-GDRL 算法也有一定控制效果,但其效果与电流裕度  $\beta$  强相关。DSP-GDRL-1 算法的  $\beta$  较小,虽能解决过载问题,但控制地点数量较多,综合效果偏低;DSP-GDRL-2 算法的  $\beta$  较大,部分场景因控制地点数量不足而难以找到解决方案,如场景 8。显然,  $\beta$  的设定难以界定,设定后控制地点将固定,导致 DSP-GDRL 方法遭遇瓶颈。而本文所提 DS-PM-HL 方法设置控制地点动态增加机制,DS 可变,未能找到合适解时可适当增加控制地点,扩大 DS,效果更好。

DDPG-G 算法和 DDPG 算法虽能解决过载问题,但未进行 DS 裁剪,难以找到合适的解,使  $\Delta P$  和  $N_a$  均较大;DDQN 算法只能处理离散动作,在连续场景中需要把动作离散化,使 DS 变大,导致其在部分场景中失效。

综上,采用本文所提算法进行线路过载紧急控制的效果较好,可优先调节发电机出力,总调节量较少,参与控制的节点数量较少,综合效果较大,与对比算法相比均体现出优越性。

### 5.1.3 多场景混合学习与验证分析

对附录 A 表 A1 中 20 个过载场景进行多场景混合学习,把学习后的策略网络用于这些场景的紧急控制,结果如表 4 所示。由于表 3 中 DDQN 算法和 DDPG 算法效果较差,后续不进行对比。

从表 4 可以看出,本文所提 DS-PM-HL 算法均能解决  $O_0$  中不同运行方式下的场景过载问题,效果较好。然而,对比算法的效果较差,DSP-GDRL-1

表3 IEEE 39节点系统在单一场景学习下算法控制效果  
Table 3 Control effect of algorithms under single-scenario learning in IEEE 39-bus system

本文算法( $\beta=1$ )			DSP-GDRL-1( $\beta=0.7$ )			DSP-GDRL-2( $\beta=1$ )			DDPG-G			DDQN			DDPG		
$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$
32.5 (1.00)	1	0.65	202.5 (1.00)	2	0.50	32.5 (1.00)	1	0.65	2 574.0 (0.28)	17	-45.03	274.5	6	-4.70	2 840.7	8	-26.15
348.3 (0.57)	2	-0.34	353.5 (0.66)	3	-0.89	344.4 (0.58)	2	-0.33	2 048.2 (0.31)	11	-31.01	290.7	10	-11.3	2 142.2	5	-14.78
362.5 (0.56)	3	-0.50	362.5 (0.56)	3	-0.50	362.5 (0.56)	3	-0.50	1 304.3 (0.62)	18	-27.07	247.3	9	-6.32	1 248.5	3	-10.10
154.1 (1.00)	1	0.65	152.5 (1.00)	1	0.65	155.3 (1.00)	1	0.65	2 552.0 (0.35)	11	-22.17	/	/	/	2 138.0	6	-19.60
186.3 (1.00)	1	0.64	198.0 (1.00)	1	0.63	184.3 (1.00)	1	0.64	2 522.5 (0.48)	9	-23.88	1 161.2	9	-9.08	1 583.1	6	-15.15
377.0 (0.86)	2	-0.01	406.5 (0.86)	3	-0.60	364.5 (1.00)	2	-0.01	2 688.7 (0.33)	11	-27.12	/	/	/	1 187.7	5	-7.82
697.4 (0.72)	3	-1.39	669.2 (0.75)	4	-2.05	665.3 (0.71)	3	-1.33	1 901.5 (0.47)	11	-27.12	1 027.4	9	-11.51	2 134.5	6	-17.07
298.1 (0.68)	3	-0.14	298.9 (0.68)	3	-0.14	/	/	/	2 098.3 (0.24)	10	-19.50	277.4	3	-0.74	2 850.1	7	-29.25
32.5 (1.00)	1	0.65	135.8 (0.24)	2	0.21	32.5 (1.00)	1	0.65	2 032.0 (0.47)	12	-38.30	/	/	/	3 022.8	5	-12.10

注: $\Delta P$ 数值后括号里为发电机调节量占比;“/”表示调节后未能解决过载问题,下同。

表4 IEEE 39节点系统在多场景混合学习下算法控制效果  
Table 4 Control effect of algorithms under multi-scenario mixed learning in IEEE 39-bus system

本文算法			DSP-GDRL-1			DSP-GDRL-2			DDPG-G		
$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$	$\Delta P/MW$	$N_a$	$E$
32.5	1	0.65	207.5	3	0.37	/	/	/	2 323.4	10	-20.16
340.6	2	-0.32	319.0	4	-1.24	341.0	2	-0.32	2 310.5	10	-19.95
362.5	3	-0.50	390.5	4	-0.60	/	/	/	2 405.7	10	-21.29
156.7	1	0.65	170.8	1	0.64	174.3	1	0.64	2 304.1	10	-19.90
158.5	1	0.65	176.1	1	0.64	175.2	1	0.64	2 303.5	10	-19.86
404.6	2	-0.04	514.6	3	-0.91	355.9	2	-0.01	2 307.9	10	-19.87
640.2	3	-1.18	679.5	5	-2.79	574.6	3	-1.13	2 323.7	10	-20.09
193.5	2	0.01	296.3	3	-0.13	/	/	/	2 310.5	10	-20.04
32.5	1	0.65	132.0	2	0.23	/	/	/	2 314.9	10	-20.04
750.5	6	-5.61	756.6	8	-6.68	/	/	/	2 552.1	11	-20.81
589.5	5	-3.05	827.6	8	-6.84	/	/	/	2 316.7	10	-20.29
1 013.2	7	-10.43	914.5	10	-12.44	/	/	/	2 376.6	11	-23.75
605.6	2	-0.26	837.0	6	-1.74	/	/	/	2 466.6	11	-22.89
605.6	2	-0.26	820.6	6	-1.56	/	/	/	2 391.3	11	-22.89
605.6	2	-0.26	798.5	6	-1.53	/	/	/	2 820.1	11	-22.88
605.6	2	-0.26	803.7	6	-1.45	/	/	/	2 825.1	11	-22.92
558.2	5	-4.09	849.1	9	-8.66	/	/	/	3 114.4	10	-20.08
698.8	3	-1.23	880.2	6	-2.50	/	/	/	2 895.9	11	-24.47
605.6	2	-0.26	800.0	6	-1.45	/	/	/	2 825.7	11	-22.92
605.6	2	-0.26	804.3	6	-1.47	/	/	/	2 820.1	11	-22.88

算法和DDPG-G算法在总调节量 $\Delta P$ 、控制地点数量 $N_a$ 、综合效果 $E$ 方面均不如本文所提方法。而DSP-GDRL-2算法中大部分场景失效,原因是部分场景DS过度裁剪,控制地点数量不足,混合学习时甚至影响其他场景的学习效果。

表5为学习的统计效果。表中: $\Delta P_{ave}$ 为平均总

调节量; $N_{ave}$ 为参与紧急控制的平均节点数; $E_{ave}$ 为平均综合效果。可以看出,本文算法学习耗时仅为3 min,即使是先进的DSP-GDRL方法,也需要至少29 min,且学习时间受 $\beta$ 影响。同时,在平均调节量、平均控制节点数、平均综合效果上,本文算法均对比算法有优势。

表5 IEEE 39节点系统在多场景混合学习下算法性能统计结果

Table 5 Statistical results of performance for algorithms under multi-scenario mixed learning in IEEE 39-bus system

算法	学习耗时/min	处理成功场景比例/%	平均调节量 $\Delta P_{ave}$ /MW	平均控制节点数 $N_{ave}$ /个	平均综合效果 $E_{ave}$
本文算法	3	100	478.3	2.70	-1.27
DSP-GDRL-1	29	100	598.9	5.15	-2.51
DSP-GDRL-2	39	25(失效)	/	/	/
DDPG-G	146	100	2 515.5	10.45	-21.40

综上,在多场景混合学习下,本文所提方法可针对不同运行方式下的多场景进行混合学习,效果较好,所训练出来的策略网络能同时解决多场景的线路过载问题,且与对比算法相比展现出优越性。

#### 5.1.4 泛化性能验证与分析

基于 $O_o$ 生成的 $O_F$ 共有1 600个场景。将5.1.3节中各算法所学紧急控制策略应用于 $O_F$ 中场景,统计紧急控制效果,如表6所示。表中: $N_T$ 为 $O_F$ 中场景数; $N_S$ 为可完全控制过载的场景数; $N_R$ 为过载未能

完全控制时,能降低过载风险的场景数; $N_V$ 为电压合格场景数; $N_S/N_T$ 为可完全控制的场景比例,指能把线路过载完全控制到限值以下的场景比例; $N_R/(N_T - N_S)$ 为可降低风险的场景比例,指未能控制过载的场景中能降低线路过载程度的场景比例; $P_r$ 为 $N_R$ 中至少降低风险率,指可降低风险的场景中至少能降低的过载功率百分比; $N_V/N_T$ 为电压合格的场景比例,指场景中电压均在合格范围内的场景比例。

表6 IEEE 39节点系统中算法泛化性能

Table 6 Generalization performance of algorithms in IEEE 39-bus system

算法	$N_T$ /个	$N_S/N_T$ /%	$N_R/(N_T - N_S)$ /%	$P_r$ /%	$N_{ave}$ /个	$\Delta P_{ave}$ /MW	$E_{ave}$	$N_V/N_T$ /%	$N_{RV}/(N_T - N_S)$ /%
本文算法	1 600	96.3	100.0	39.6	2.70	477.8	-1.31	100.0	
DSP-GDRL-1	1 600	95.1	91.0	21.3	5.17	602.3	-2.60	100.0	
DSP-GDRL-2	1 600	18.5	23.0	0	4.16	1 027.6	-5.99	94.1	0
DDPG-G	1 600	99.8	0	0	10.46	2 529.7	-21.59	100.0	

表6中数据表明,本文所提DS-PM-HL算法可完全控制过载的比例为96.3%,即使在过载不可完全控制的场景中,过载程度也能全部得到有效改善,至少降低风险率为39.6%,总调节量、控制节点数量均较少,综合效果较好,所有场景的电压指标均合格。上述指标与两种DSP-GDRL算法相比均体现出优越的泛化性能。虽然DDPG-G算法可处理较多场景,但其总调节量和控制节点数量均较大,综合效果很低,不满足紧急控制需求。

综上,本文所提算法具有有效性以及优秀的泛化性能。

## 5.2 IEEE 300节点系统算例分析

对IEEE 300节点系统进行改进,形成含30台火电机组、15台水电机组、10台配有储能的风机、5台无储能风机和8台独立储能的系统。算例

设计与5.1.1节39节点系统算例的3)和4)相同,旨在更大系统中验证所提方法的有效性和泛化性。

### 5.2.1 多场景混合学习与验证分析

在此算例中,集合 $O_o$ 共有33个过载场景。对33个过载场景进行多场景混合学习,把学习后的策略网络用于紧急控制,控制效果的统计数据如表7所示。可以看出,在300节点系统中,本文算法仅需15 min即可学到针对所有过载场景的控制策略,而其他算法用时均较长。其中,DSP-GDRL-1算法所选出的控制地点较多,DS较大,耗时为39 min,且 $N_{ave}$ 和 $E_{ave}$ 均差于本文所提算法。DSP-GDRL-1算法中由于过度裁剪DS,部分场景失效,在92 min内学不到有效策略。DDPG-G的算法针对所有场景均失效。

综上,验证了本文所提算法相比现有常见强化

表7 IEEE 300节点系统多场景混合学习算法性能统计结果

Table 7 Statistical results of performance for algorithms under multi-scenario mixed learning in IEEE 300-bus system

算法	学习耗时/min	处理成功场景比例/%	$\Delta P_{ave}/MW$	$N_{ave}/个$	$E_{ave}$
本文算法	15	100	526.8	1.70	-0.34
DSP-GDRL-1	39	100	471.6	2.45	-0.53
DSP-GDRL-2	92	90.9(失效)	/	/	/
DDPG-G	167	0(失效)	/	/	/

学习算法的优越性,可应用于含大DS的大系统场景中。

### 5.2.2 泛化性能验证与分析

基于 $O_0$ 生成的 $O_F$ 共有2310个场景,其中,只有73.6%的场景电压指标合格。

将5.2.1节中各算法所学紧急控制策略应用于 $O_F$ 中的场景,效果如表8所示。表中: $N_{RV}$ 为电压质量整体提升的场景数; $N_{RV}/(N_T - N_S)$ 为电压质量整体提升的场景比例,指场景中部分节点虽不在合格范围内,但经过控制有改善的场景比例。本文

DS-PM-HL算法可完全控制过载的比例达到99.7%,可降低过载风险比例达100%,至少能降低69.6%的过载功率,且平均参与控制的节点数量仅为1.71个,说明所提算法在大系统场景中训练出来的控制策略具有优秀泛化性能,且各种指标效果好。从电压指标可以看出,80.2%的场景电压指标合格,而剩余场景中有60.0%的场景的电压质量能得到整体改善,40%的场景中也仅有少数个别节点电压超出限值。

表8 IEEE 300节点系统中算法的泛化性能

Table 8 Generalization performance of algorithms in IEEE 300-bus system

算法	$N_T/个$	初始电压合格比例/%	$N_S/N_T/ %$	$N_R/(N_T - N_S)/ %$	$P_r/ %$	$N_{ave}/个$	$\Delta P_{ave}/ MW$	$E_{ave}$	$N_V/N_T/ %$	$N_{RV}/(N_T - N_S)/ %$
本文算法	2310	73.6	99.7	100.0	69.6	1.71	528.0	0.10	80.2	60.0
DSP-GDRL-1	2310	73.6	98.2	66.7	0	2.51	477.3	-0.58	80.5	64.5
DSP-GDRL-2	2310	73.6	87.7	45.8	0	1.39	361.4	-0.16	80.9	51.8
DDPG-G	2310	73.6	0	0	0	/	/	/	/	/

相比之下,DSP-GDRL-1和DSP-GDRL-2算法的泛化性能比本文所提方法要差,控制节点数量 $N_{ave}$ 和综合效果 $E_{ave}$ 均不及本文所提算法。DDPG-G算法甚至全部失效。

综上,本文所提算法在大系统中具有有效性以及优秀的泛化性能。

### 5.3 某省级电网系统算例分析

本节在中国某省级电网系统中验证本文方法的有效性及其所学策略网络的优秀泛化性能。该省级电网包含1451个节点和770条线路,其中有236个负荷节点、93个火电节点、10个光伏节点、3个风电节点,其拓扑简化示意图见附录A图A1。算例设计与5.1.1节IEEE 39节点系统算例的3)和4)基本相同。从IEEE 39节点系统和300节点系统案例中已经看出,本文所提方法较对比算法具有优越性,故本节只对所提DS-PM-HL方法进行验证,即在实际城市电网系统中验证所提方法的有效性和泛化性。

#### 5.3.1 多场景混合学习与验证分析

选择该省级电网系统的某一运行场景作为基础运行方式 $O_0$ ,基于 $O_0$ 进行线路 $N-1$ 、发电机 $N-1$ 、

以及随机新能源和负荷在 $[-50\%, 50\%]$ 内随机波动生成24个过载场景,记为集合 $O_0$ 。对24个过载场景进行多场景混合学习,把学习后的策略网络用于紧急控制,控制效果的统计数据如表9所示。可以看出,在该省级系统中,本文算法仅需11min即可学得针对所有过载场景的控制策略,且处理成功场景比例为100%,即全部处理成功。平均综合效果 $E_{ave}$ 为-20.67,这是因为该省级电网规模较大,需要多个控制节点共同参与才能解决过载问题,平均控制节点数 $N_{ave}$ 为6.9个,由式(26)可知,当参与控制节点数量较大时,综合效果会降低。

表9 某省级电网系统中本文算法的有效性

Table 9 Efficient performance of proposed algorithm in a provincial power grid system

参数	数值
学习耗时	11 min
处理成功场景比例	100%
平均调节量 $\Delta P_{ave}$	511.8
平均控制节点数 $N_{ave}$	6.9个
平均综合效果 $E_{ave}$	-20.67

综上,本文所提算法在规模为1 451个节点和770条线路的省级电网系统中依然具有有效性,可应用于含大DS的大系统场景中。

### 5.3.2 泛化性能验证与分析

基于 $O_o$ 生成的 $O_F$ 共有1 440个场景,初始电压合格场景比例为0,平均每个场景中节点电压合格比例为96.2%。

将5.3.1节中本文算法所学紧急控制策略应用于 $O_F$ 中的场景处理,效果如表10所示。本文所提DS-PM-HL算法可完全控制过载的比例达到99.8%,可降低过载风险比例达100%,至少能降低50.0%的过载功率,各种指标效果好。从电压指标可以看出,节点电压合格平均比例为96.2%,与初始节点电压合格平均比例相近,不合格的节点电压中有85.9%的节点的电压质量得到提高。

**表 10 某省级电网系统中本文算法的泛化性能**  
**Table 10 Generalization performance of proposed algorithm in a provincial power grid system**

参数	数值
场景数 $N_T$	1 440(=24×60)个
初始电压合格比例	0
$N_T$ 中平均每个场景节点电压合格比例	96.2%
可完全控制场景比例 $N_S/N_T$	99.8%
可降低风险场景比例 $N_R/(N_T - N_S)$	100%
至少降低风险率 $P_r$	50%
$N_{ave}$	6.92个
$\Delta P_{ave}$	511.8 MW
$E_{ave}$	-20.59
节点电压合格平均比例	96.2%
节点电压质量平均提升比例	85.9%

综上,采用所提算法在大规模省级电网系统中训练出来的控制策略具有优秀泛化性能。

## 6 结语

本文提出一种DS与PM动态迭代的线路过载紧急控制混合学习方法,可实现控制地点裁剪选择、控制量网络高效学习以及两者的高效迭代,从而有效支撑线路过载紧急控制策略在线生成。在IEEE 39节点和IEEE 300节点系统中的算例表明,相比DSP-GDRL、DDPG-G、DDPG和DDQN等算法,采用本文所提算法学习的线路过载紧急控制策略可优先调节发电机出力,参与控制的节点数量较少,综合效果较高,且当控制地点数量不足时可以动态增加,扩大DS,实现较好的控制效果。在某省级电网系统

中的算例进一步验证了本文所提算法可应用于大规模城市电网。

本文方法后续将进一步推广至电压稳定、频率稳定等其他电力系统安全稳定问题的紧急控制应用中。

附录见本刊网络版,点击<http://www.aeps-info.com/aeps/article/abstract/20250625004>,或扫描英文摘要后二维码,可阅读全文。

## 参考文献

- [1] 姜盛波,杨军,王建雄,等.基于预防-紧急协调控制的大电网连锁故障防御策略[J].电力自动化设备,2019,39(12):148-154.  
JIANG Shengbo, YANG Jun, WANG Jianxiong, et al. Defense strategy against large power grid cascading failure based on coordinated preventive-emergency control [J]. Electric Power Automation Equipment, 2019, 39(12): 148-154.
- [2] XIE J, SUN W. Distributional deep reinforcement learning-based emergency frequency control [J]. IEEE Transactions on Power Systems, 2022, 37(4): 2720-2730.
- [3] GORDON S, MCGARRY C, TAIT J, et al. Impact of low inertia and high distributed generation on the effectiveness of under frequency load shedding schemes [J]. IEEE Transactions on Power Delivery, 2022, 37(5): 3752-3761.
- [4] HUANG Q H, HUANG R K, HAO W T, et al. Adaptive power system emergency control using deep reinforcement learning [J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1171-1182.
- [5] 张哲,秦博宇,高鑫,等.基于CNN-LSTM网络的电网电压稳定紧急控制策略[J].电力系统自动化,2023,47(11):60-68.  
ZHANG Zhe, QIN Boyu, GAO Xin, et al. Emergency control strategy of power grid voltage stability based on convolutional neural network and long short-term memory network [J]. Automation of Electric Power Systems, 2023, 47(11): 60-68.
- [6] CHEN C Y, CUI M J, LI F X, et al. Model-free emergency frequency control based on reinforcement learning [J]. IEEE Transactions on Industrial Informatics, 2021, 17(4): 2336-2346.
- [7] 卢恒光,林碧琳,温步瀛.基于深度强化学习的切机控制策略研究[J].电器与能效管理技术,2023(3):11-15.  
LU Hengguang, LIN Bilin, WEN Buying. Research on generator tripping control strategy based on deep reinforcement learning [J]. Electrical & Energy Management Technology, 2023 (3): 11-15.
- [8] LI J, CHEN S, WANG X Y, et al. Load shedding control strategy in power grid emergency state based on deep reinforcement learning [J]. CSEE Journal of Power and Energy Systems, 2022, 8(4): 1175-1182.
- [9] 刘杰,石访,宋雪萌,等.基于多变量样本卷积交互网络的电力

- 系统频率安全性评估[J]. 电力系统自动化, 2024, 48(22): 160-170.
- LIU Jie, SHI Fang, SONG Xuemeng, et al. Frequency safety assessment of power systems based on multivariable-sample convolution and interaction network[J]. Automation of Electric Power Systems, 2024, 48(22): 160-170.
- [10] 王建, 吴昊, 张博, 等. 不平衡样本下基于迁移学习-AlexNet的输电线路故障辨识方法[J]. 电力系统自动化, 2022, 46(22): 182-191.
- WANG Jian, WU Hao, ZHANG Bo, et al. Fault identification method for transmission line based on transfer learning-AlexNet with imbalanced samples[J]. Automation of Electric Power Systems, 2022, 46(22): 182-191.
- [11] MORADZADEH A, MOHAMMADI-IVATLOO B, POURHOSSEIN K, et al. Data mining applications to fault diagnosis in power electronic systems: a systematic review[J]. IEEE Transactions on Power Electronics, 2022, 37(5): 6026-6050.
- [12] ZHANG H T, SUN X F, LEE M H, et al. Deep reinforcement learning-based active network management and emergency load-shedding control for power systems[J]. IEEE Transactions on Smart Grid, 2024, 15(2): 1423-1437.
- [13] 李宏浩, 张沛, 刘翌. 基于深度强化学习的暂态稳定紧急控制决策方法[J]. 电力系统自动化, 2023, 47(5): 144-152.
- LI Honghao, ZHANG Pei, LIU Zhao. Decision-making method for transient stability emergency control based on deep reinforcement learning [J]. Automation of Electric Power Systems, 2023, 47(5): 144-152.
- [14] LIU Z, LIU Y Q, HE J H, et al. Double DQN-based power system transient stability emergency control with protection coordinations [C]// IEEE 6th International Electrical and Energy Conference, May 12-14, 2023, Hefei, China.
- [15] MA Q F, ZHANG H, HE X Q, et al. Emergency frequency control strategy using demand response based on deep reinforcement learning [C]// 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference, September 20-23, 2020, Nanjing, China.
- [16] ZHANG J Y, LUO Y H, WANG B Y, et al. Deep reinforcement learning for load shedding against short-term voltage instability in large power systems [J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(8): 4249-4260.
- [17] HOSSAIN R R, HUANG Q H, HUANG R K. Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control [J]. IEEE Transactions on Power Systems, 2021, 36(5): 4848-4851.
- [18] 万庆祝, 袁润娇, 李伊梦, 等. 风光储参与电网预防-紧急控制策略研究[J]. 太阳能学报, 2024, 45(4): 404-415.
- WAN Qingzhu, YUAN Runjiao, LI Yimeng, et al. Research on power grid prevention-emergency control strategy with wind and solar storage participation[J]. Acta Energetica Sinica, 2024, 45(4): 404-415.
- [19] QU C H, ZHANG X W, HU S W, et al. Research and engineering field-test of a renewable energy stations rapid power control technology for improving power grid stability [C]// International Conference on Power System Technology, December 8-9, 2021, Haikou, China.
- [20] THAPA K B, JAYASAWAL K. Pitch control scheme for rapid active power control of a PMSG-based wind power plant [J]. IEEE Transactions on Industry Applications, 2020, 56(6): 6756-6766.
- [21] 陈戈, 张俊勃, 彭颖, 等. 基于决策空间裁剪强化学习的连锁故障调切结合紧急控制[J]. 电力系统自动化, 2025, 49(6): 144-156.
- CHEN Ge, ZHANG Junbo, PENG Ying, et al. Emergency control combined with adjustment and tripping against cascading failures based on decision space pruning reinforcement learning [J]. Automation of Electric Power Systems, 2025, 49(6): 144-156.
- [22] 孙仲卿, 刘福锁, 李威, 等. 基于特征匹配的暂态稳定紧急控制策略快速生成[J]. 电力系统自动化, 2024, 48(2): 167-175.
- SUN Zhongqing, LIU Fusuo, LI Wei, et al. Rapid generation of transient stability emergency control strategy based on feature matching[J]. Automation of Electric Power Systems, 2024, 48(2): 167-175.
- [23] 电能质量 供电电压偏差: GB/T 12325—2008[S]. 北京: 中国标准出版社, 2008.
- Power quality—deviation of supply voltage: GB/T 12325—2008[S]. Beijing: Standards Press of China, 2008.
- [24] 王锐, 顾伟, 万秋兰, 等. 考虑二重预想事故的在线电压稳定预防控制[J]. 电力自动化设备, 2010, 30(10): 62-66.
- WANG Rui, GU Wei, WAN Qiulan, et al. Online preventive control for voltage stability considering supervenient contingency [J]. Electric Power Automation Equipment, 2010, 30(10): 62-66.

张寿志(2000—), 男, 博士研究生, 主要研究方向: 电力系统紧急控制与决策优化、电力系统小扰动稳定性分析与决策调节、强化学习在电力系统的应用。E-mail: 202411083518@mail.scut.edu.cn

陈戈(1995—), 男, 博士, 主要研究方向: 电力系统紧急控制与调度决策、连锁故障紧急控制、强化学习在电力系统的应用、云计算与任务调度。E-mail: epchenge@163.com

张俊勃(1986—), 男, 通信作者, 博士, 教授, 博士生导师, 主要研究方向: 新型电力系统稳定性、数字电网智能化应用、大型电力软件系统、人工智能在大型软件工程的应用。E-mail: epjbzhang@scut.edu.cn

(编辑 王梦岩)

## Hybrid Learning for Line Overload Emergency Control with Dynamic Iteration of Decision Space and Strategy Model

ZHANG Shouzhi<sup>1</sup>, CHEN Ge<sup>2</sup>, ZHANG Junbo<sup>1</sup>, PENG Ying<sup>1</sup>

(1. School of Electric Power Engineering, South China University of Technology, Guangzhou 510641, China;

2. Energy Development Research Institute, China Southern Power Grid, Guangzhou 510663, China)

**Abstract:** Renewable energy can rapidly regulate the power in the new power system, demonstrating the potential of participating in overload emergency control for lines. However, when it is adopted, generation methods for the emergency control strategy based on deep reinforcement learning face the challenges of excessively large decision space and high solution complexity. To address this issue, a hybrid learning method for emergency control with dynamic iteration of decision space and strategy model is proposed. First, a dual-network model comprising a control location network and a control value network is constructed, and an iterative learning framework for both networks is designed. Second, the control location network and its learning objectives are introduced, and a sensitivity-based sample generation method is designed to learn the control location network. Then, a deep reinforcement learning method for the control value network is proposed, and a segmented exploration strategy is designed for efficient learning of the control value network. Next, a dynamic iteration implementation process between the control value network and control location network is designed. Finally, the effectiveness of the proposed method is validated in the IEEE 39-bus system, IEEE 300-bus system, and a provincial power grid of China.

This work is supported by National Natural Science Foundation of China (No. U22B6007, No. 52277101) and Fundamental Research Funds for the Central Universities (No. 2024ZYGXZR109).

**Key words:** emergency control; renewable energy; deep reinforcement learning; line overload; hybrid learning; dynamic iteration; decision space; sample generation

