

基于安全引导图强化学习的故障交通电气化耦合网络应急调度策略

邢 强, 唐建艇, 葛 辉, 张国立

南京邮电大学自动化学院 江苏南京

【摘要】针对交通电气化耦合网络在突发故障下的恢复难题, 本文提出一种基于安全引导图强化学习的故障耦合网络应急调度策略。首先, 构建故障耦合网络的双层恢复架构, 上层实施配电网开关重构与交通潮汐车道调控, 确立交通流量与电网功率边界。下层依托上层优化的通行条件, 引导电动汽车驶向目标充电站, 利用移动储能支撑电网负荷缺口。其次, 通过图注意力网络刻画电力-交通-充电站的状态传播与交通流演变过程, 精准挖掘耦合网络的拓扑内嵌故障特征信息。然后, 采用改进 Dueling 双重 Q 网络算法, 求解出配电网开关重构序列、潮汐车道调控指令与充电站引导决策的最优调度策略。构建考虑拉格朗日安全约束的奖励机制, 引导智能体主动规避不安全动作。并建立基于混合整数线性规划专家知识的引导学习机制, 提升策略初始探索效率。最后, 仿真验证表明, 该策略提升了故障恢复效率, 降低了配电网网损与交通拥堵程度, 同时满足多维物理安全约束, 有效增强了耦合网络的故障应急韧性。

【关键词】交通电气化耦合网络; 应急调度策略; 图强化学习; 拉格朗日安全约束; 引导学习机制

【基金项目】江苏省基础研究计划自然科学基金 (BK20240652)

【收稿日期】2026 年 5 月 15 日

【出刊日期】2026 年 6 月 20 日

【DOI】10.12208/j.jeea.20260016

An emergency dispatch strategy for faulted transportation electrification coupled networks based on safety-guided graph reinforcement learning

Qiang Xing, Jianting Tang, Hui Ge, Guoli Zhang

Nanjing University of Posts and Telecommunications College of Automation, Nanjing, Jiangsu

【Abstract】To address the recovery challenges of transportation electrification coupled networks under sudden faults, this paper proposes an emergency dispatch strategy for faulted coupled networks based on safety-guided graph reinforcement learning. Firstly, a double layer recovery framework for faulted coupled networks is constructed: the upper layer implements distribution network switch reconfiguration and traffic tidal lane regulation to define the boundaries of traffic flow and power grid power; the lower layer leverages the optimized traffic conditions from the upper layer to guide electric vehicles to target charging stations and utilizes mobile energy storage to offset the power grid load gaps. Secondly, the Graph Attention Network(GAT) is adopted to characterize the state propagation and traffic flow evolution processes of the power-traffic-charging station, accurately mining the topology-embedded fault feature information of the coupled network. Then, a modified Dueling Double Q-Network(MD3QN) algorithm is utilized to solve the optimal dispatch strategy including distribution network switch reconfiguration sequences, tidal lane regulation commands, and charging station guidance decisions. A reward mechanism considering Lagrange safety constraints is established to guide the agent to actively avoid unsafe actions, and a guided learning mechanism based on mixed-integer linear programming(MILP) expert knowledge is developed to improve the initial exploration efficiency of the strategy. Finally, simulation results show that the proposed strategy effectively improves the fault recovery efficiency, reduces the power loss of distribution network and traffic congestion, meets the multi-dimensional physical safety constraints, and significantly improves the fault emergency resilience of the coupled network.

【Keywords】Transportation electrification coupled networks; Emergency dispatch strategy; Graph reinforcement learning; Lagrange safety constraints; Guided learning mechanism

引言

随着交通电气化进程加速, 城市配电网与交通网的时空耦合日益紧密, 形成了互联互通的复杂耦合系统。然而, 地震、台风等极端灾害给此类耦合系统的稳定运行带来了严峻挑战^[1]。而电动汽车(Electric Vehicle, EV)具备移动储能的灵活性, 其通过车载电池向电网受控放电, 可在故障期间为配电网提供关键的紧急功率支撑, 能够缓解节点电压越限并保障重要负荷的持续供应。因此, 如何协同多源资源并挖掘EV的动态调节潜力, 已成为提升城市基础设施韧性的关键^[2]。

文献[3-5]研究配电网故障恢复的多资源协调供电能力, 以保障重要负荷快速恢复。文献[3]提出基于改进哈里斯鹰算法的双层模型, 统筹抢修路径与配电网重构。文献[4]搭建多源资源动态适配双层框架, 实现故障恢复资源精准调度。文献[5]将功率流约束松弛为二阶锥形式, 构建混合整数二阶锥规划模型优化故障恢复效率。

文献[6-8]依托电动汽车、移动储能、微电网等灵活可调资源, 实现多场景故障恢复与负荷平滑。文献[6]构建分层动态协同恢复模型, 协同电动汽车与分布式资源提升地震灾害下配电网韧性。文献[7]提出两阶段故障恢复策略, 构建移动应急电源充电站预定位模型, 实现高权重负荷快速恢复。文献[8]建立微电网韧性评估指标体系, 量化极端条件下输配电网协同韧性提升效应。

近年来, 深度强化学习因其动态决策的优势, 被广泛用于故障调度领域。而图神经网络则成为捕捉网络拓扑演变、增强状态表征能力的核心技术。文献[9-12]重点围绕强化学习、图神经网络及其改进架构展开, 旨在提升模型对动态拓扑的适配性与场景适应性。文献[9]结合蒙特卡洛树搜索与深度Q学习, 构建异步信息场景下鲁棒故障恢复模型。文献[10]提出图强化学习框架, 融合配电网拓扑与图卷积网络, 适配大规模系统故障恢复。文献[11]通过改进图注意力网络(Graph Attention Networks, GAT)的权重分配机制, 增强了节点特征的深层提取能力。文献[12]则提出一种时空关联图神经网络方法, 建立低观测节点优选与时空注意力模块, 实现大规模配电网故障区段精准定位。

针对强化学习在大规模状态空间收敛慢的问题, 文献[13-14]依靠先验经验或机理知识优化训练, 加速智能体训练并提升决策安全性。文献[13]设计了一种改进的示教型强化学习方法, 通过收集先验经验与实时数据加速策略学习。文献[14]将专家机理知识嵌入算法

以屏蔽无效动作, 显著缩短了模型训练周期。

尽管上述方法取得了一定成效, 但仍存在以下局限: 其一, 现有研究多侧重于电网侧重构或单一的交通引导, 未能充分挖掘交通网潮汐车道与配电网开关动作的深度协同潜力。其二, 现有算法在处理高维决策空间时往往面临探索效率低下、收敛缓慢的问题, 且缺乏对系统物理安全边界的动态硬性约束机制, 导致生成的策略存在安全风险。

为解决上述问题, 本文提出了一种基于安全引导图强化学习的故障交通电气化耦合网络应急调度策略, 主要贡献如下:

1) 构建基于 GAT-Dueling 双重 Q 网络(Dueling Double Deep Q-Network, D3QN)的故障耦合网络的双层恢复架构。上层统筹配电网开关操作与交通潮汐车道调控, 确立全局流量与功率边界。下层实施电动汽车移动资源调度, 精准响应上层恢复需求。

2) 构建基于 GAT 的特征感知模型, 挖掘电力-交通-充电站的拓扑关联与动态特征。引入多头注意力机制构建状态感知模块, 增强对关键节点的高阶特征提取与辨识能力。

3) 构建故障约束马尔可夫决策过程(Constrained Markov Decision Process, CMDP)模型, 采用改进D3QN(Modified D3QN, MD3QN)算法求解。考虑拉格朗日乘子法转换电压越限等物理约束, 实现网损最小化的同时主动规避不安全动作。建立基于混合整数线性规划(Mixed Integer Linear Programming, MILP)专家知识的引导学习机制, 提升复杂故障场景下策略的初始求解效率。

1 故障交通电气化耦合网络应急调度策略架构设计

图1为基于 GAT-D3QN 的故障交通电气化耦合网络应急调度策略整体架构, 由交通电气化耦合网络模块、GAT 特征感知模块与 MD3QN 算法求解模块组成。交通电气化耦合网络模块负责上传配电网、交通网及充电站的拓扑结构与运行数据。GAT 特征感知模块通过图注意力网络挖掘耦合网络关键故障特征, 刻画耦合网络间的动态耦合关系。MD3QN 算法求解模块将应急调度问题转化为 CMDP 求解, 构建安全约束与引导学习机制, 输出最优应急调度策略。

A. 数学模型

(1) 上层目标函数

上层考虑网损最小化、交通拥堵缓解和充电站服务均衡构建目标函数 F_1 :

$$\min F_1 = L^{\text{PG}} + C^{\text{TN}} + U^{\text{CS}} \quad (1)$$

$$L^{\text{PG}} = \sum_{(i,j) \in \mathcal{E}^{\text{PG}}} R_{i,j} \cdot \frac{P_{i,j}^2 + Q_{i,j}^2}{V_i^2} \quad (2)$$

$$C^{\text{TN}} = \frac{1}{M} \sum_{(a,b) \in \mathcal{E}^{\text{TN}}} \frac{f_{a,b}}{C_{a,b}} \quad (3)$$

$$U^{\text{CS}} = \sqrt{\frac{1}{K} \sum_{k=1}^K (P_k - \bar{P})^2} \quad (4)$$

式中, L^{PG} 为配电网总网损, 其中 $R_{i,j}$ 为线路 (i,j) 的电阻; $P_{i,j}/Q_{i,j}$ 为线路 (i,j) 的有功/无功功率; V_i 为节点 i 的电压幅值; $\mathcal{E}^{\text{PG}}$ 为电网线路集合。 C^{TN} 为交通网平均拥堵指数, $f_{a,b}$ 为路段 (a,b) 交通流量; $C_{a,b}$ 为路段 ab 的通行能力; $\mathcal{E}^{\text{TN}}$ 为交通路段集合; M 为交通路段总数。 U^{CS} 为充电站服务不均衡度, 其中 K 为充电站数量; P_k 为充电站 k 负荷功率; $\bar{P}$ 为所有充电站平均负荷功率。

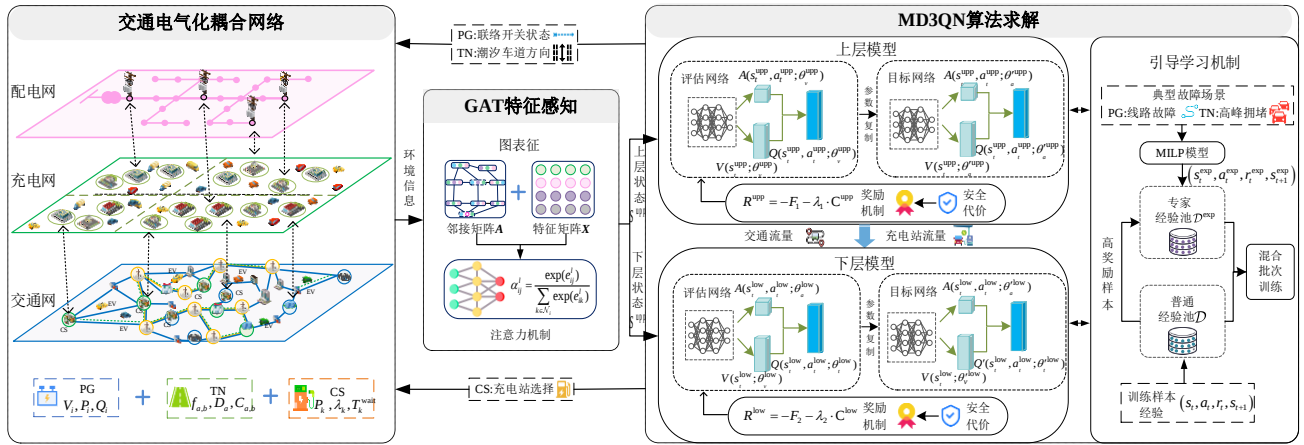


图1 所提故障交通电气耦合网络应急调度策略整体架构。

上层网络需满足以下几个核心约束, 确保决策可行与配电系统安全:

1) 配电网功率平衡约束

$$P_i^{\text{gen}} - P_i^{\text{load}} - P_i^{\text{CS}} = \sum_{j \in \mathcal{V}_i^{\text{nei}}} (G_{i,j} V_i V_j \cos \theta_{i,j} + B_{i,j} V_i V_j \sin \theta_{i,j}) \quad (5)$$

$$P_i^{\text{gen}} - P_i^{\text{load}} - P_i^{\text{CS}} = \sum_{j \in \mathcal{V}_i^{\text{nei}}} (G_{i,j} V_i V_j \cos \theta_{i,j} + B_{i,j} V_i V_j \sin \theta_{i,j}) \quad (6)$$

式中, $P_i^{\text{gen}}/Q_i^{\text{gen}}$ 为节点 i 的有功/无功发电; $P_i^{\text{load}}/Q_i^{\text{load}}$ 为节点 i 的有功/无功负荷; $P_i^{\text{CS}}/Q_i^{\text{CS}}$ 为节点 i 与关联充电站的有功/无功交互功率; $G_{i,j}/B_{i,j}$ 为线路 (i,j) 的电导/电纳; $\theta_{i,j}$ 为节点 i,j 间电压相角差; $\mathcal{V}_i^{\text{nei}}$ 为节点 i 的邻接节点集合。

2) 节点电压约束

$$V_{\min} \leq V_i \leq V_{\max}, \quad \forall i \in \mathcal{V}^{\text{PG}} \quad (7)$$

式中, $V_{\min}$, $V_{\max}$ 分别为节点电压允许的上下限, $\mathcal{V}^{\text{PG}}$ 为电网节点集合。

3) 交通流量守恒约束

$$\sum_{(a,b) \in \mathcal{E}^{\text{TN}}} f_{a,b} - \sum_{(c,a) \in \mathcal{E}^{\text{TN}}} f_{c,a} = D_a, \quad \forall a \in \mathcal{V}^{\text{TN}} \quad (8)$$

式中, D_a 表示交通节点 a 的交通需求, $\mathcal{V}^{\text{TN}}$ 为交通节点集合。

4) 决策变量约束

$$x_{i,j} \in \{0,1\}, \quad \forall (i,j) \in \mathcal{E}^{\text{PG}} \quad (9)$$

$$y_{a,b} \in \{-1,0,1\}, \quad \forall (a,b) \in \mathcal{E}^{\text{tid}} \quad (10)$$

式中, $x_{i,j}$ 为电网线路 (i,j) 的状态变量, 1 表示连通, 0 断开; $y_{a,b}$ 为交通网潮汐车道 (a,b) 的状态变量, 1 表示正向开放, -1 反向开放, 0 关闭, $\mathcal{E}^{\text{tid}}$ 为潮汐车道集合。

(2) 下层目标函数

下层目标基于上层交通流量与充电站负荷信息, 优化电动汽车充电站调度, 目标函数 F_2 如下:

$$\min F_2 = C^{\text{dis}} + T^{\text{wait}} \quad (11)$$

$$C^{\text{dis}} = \sum_{k=1}^K \sum_{t=1}^T c(t) \cdot P_k(t) \cdot \Delta t \quad (12)$$

$$T^{\text{wait}} = \frac{1}{K} \sum_{k=1}^K \frac{\lambda_k}{\mu_k (\mu_k - \lambda_k)} \quad (13)$$

式中, C^{dis} 为电动汽车总放电成本, $c(t)$ 为时刻 t 电价; $P_k(t)$ 为充电站 k 时刻 t 的放电功率; Δt 为时间间隔。 T^{wait} 为充电站平均用户等待时间; λ_k 是充电站 k 的电动汽车到达率; μ_k 为服务率; $T = 96$ 。

下层网络包含以下几个约束条件:

1) 充电站功率约束

$$P_{k,\min} \leq P_k(t) \leq P_{k,\max}, \quad \forall t \in [1, T] \quad (14)$$

式中, $P_{k,\min} / P_{k,\max}$ 为充电站 k 的最小/最大允许放电功率; $\mathcal{K}$ 为充电站集合。

2) 电网交互功率约束

$$\left| \sum_{k=1}^K P_k(t) \right| \leq P_{\max}^{\text{PG}}, \quad \forall t \in [1, T] \quad (15)$$

式中, $P_{\max}^{\text{PG}}$ 为充电站与上级电网交互最大允许功率。

3) 用户等待时间约束

$$T_k^{\text{wait}} \leq T_{k,\max}^{\text{wait}} \quad (16)$$

式中, $T_{k,\max}^{\text{wait}}$ 为充电站 k 允许的最大用户等待时间。

B. 基于 GAT 的耦合系统特征感知

为捕捉耦合系统的拓扑结构与动态特征, 采用 GAT 构建异构图模型, 实现环境信息精准感知。图结构定义:

$$G = (\mathcal{V}, \mathcal{E}) \quad (17)$$

$$\mathcal{V} = \mathcal{V}^{\text{PG}} \cup \mathcal{V}^{\text{TN}} \cup \mathcal{V}^{\text{coup}} \quad (18)$$

$$\mathcal{E} = \mathcal{E}^{\text{PG}} \cup \mathcal{E}^{\text{TN}} \cup \mathcal{E}^{\text{coup}} \quad (19)$$

式中, $\mathcal{V}, \mathcal{E}$ 分别为节点集和边集; $\mathcal{V}^{\text{PG}}, \mathcal{V}^{\text{TN}}, \mathcal{V}^{\text{coup}}$ 分别为配电网节点、交通网节点和充电站耦合节点, $\mathcal{E}^{\text{PG}}, \mathcal{E}^{\text{TN}}, \mathcal{E}^{\text{coup}}$ 分别为配电网线路、交通网路段和充电站与电网/交通网的耦合边。

(1) 图矩阵构建

1) 邻接矩阵

$$A^{\text{PG}}(i, j) = \begin{cases} 1, & \text{电网节点 } i, j \text{ 相连} \\ 0, & \text{否则} \end{cases} \quad (20)$$

$$A^{\text{TN}}(a, b) = \begin{cases} 1, & \text{路段 } (a, b) \text{ 通行能力} \geq \text{最小阈值} \\ 0, & \text{否则} \end{cases} \quad (21)$$

耦合邻接矩阵: $A^{\text{coup}}(i, k) = 1$ (充电站 k 接入电网节点 i), $A^{\text{coup}}(a, k) = 1$ (充电站 k 位于交通节点 a 附近)。

2) 特征矩阵

$$X^{\text{PG}} = [V_i, P_i, Q_i] \quad (22)$$

$$X^{\text{TN}} = [f_{a,b}, D_a, C_{a,b}] \quad (23)$$

$$X^{\text{coup}} = [P_k, \lambda_k, T_k^{\text{wait}}] \quad (24)$$

将三类矩阵整合为统一的图表示:

$$A = \begin{bmatrix} A^{\text{PG}} & 0 & A^{\text{PG-coup}} \\ 0 & A^{\text{TN}} & A^{\text{TN-coup}} \\ A^{\text{coup-PG}} & A^{\text{coup-TN}} & 0 \end{bmatrix}, \quad X = \begin{bmatrix} X^{\text{PG}} \\ X^{\text{TN}} \\ X^{\text{coup}} \end{bmatrix} \quad (25)$$

(2) GAT 特征提取

采用多层多头注意力机制, 对输入特征矩阵 X 进行线性变换与注意力加权, 提取节点高阶特征:

1) 线性变换:

$$h_i^l = W^l x_i \quad (26)$$

式中, h_i^l 为第 l 层变换后的特征向量; W^l 为第 l 层权重矩阵; x_i 为节点 i 的初始特征向量。

2) 注意力系数计算:

$$e_{ij}^l = \text{ReLU}(\bar{a}^T [h_i^l \parallel h_j^l]) \quad (27)$$

式中, $\bar{a}$ 为注意力参数向量; $\parallel$ 表示向量拼接。

3) 归一化:

$$\alpha_{ij}^l = \frac{\exp(e_{ij}^l)}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik}^l)} \quad (28)$$

式中, $\mathcal{N}_i$ 为节点 i 邻域集。

4) 多头注意力融合:

$$h_i^{l+1} = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{l,m} W^{l,m} h_j^l \right) \quad (29)$$

式中, M 为注意力头数量; $\alpha_{ij}^{l,m}, W^{l,m}$ 分别为第 l 层 m 个注意力头的注意力权重和权重矩阵; σ 为激活函数。

最终输出高阶特征矩阵 $H = [H^{\text{upp}}, H^{\text{low}}]^T$, 作为上、下层 CMDP 的状态输入。

C. 基于双层 CMDP 的故障交通电气化耦合网络应急调度策略制定

传统 FMDP (Factored Markov Decision Process) 模型仅聚焦奖励最大化, 易导致为追求网损降低、拥堵缓解等性能目标而违反电压越限、功率过载等安全约束, 引发二次故障。因此, 本节将跨系统动态交互过程建模为双层 CMDP 模型, 整合优化目标与安全约束:

(1) 状态空间

由 GAT 提取出的 H 表示, 上层含配电网电压、线

路功率、交通流量、充电站负荷等特征信息; 下层含充电站功率、电动汽车到达率、电网交互功率等特征信息:

$$S^{\text{upp}} = H^{\text{upp}} \quad (30)$$

$$S^{\text{low}} = H^{\text{low}} \quad (31)$$

(2) 动作空间

上层网络的动作 A^{upp} 定义为配电网联络开关操作和潮汐车道调控:

$$A^{\text{upp}} = \{x_{i,j}, x_{a,b}\} \quad (32)$$

下层动作 A^{low} 由选择的充电站组成:

$$A^{\text{low}} = \{k_1, k_2, \dots, k_n\} \quad (33)$$

式中, k_n 为用户 n 选择的充电站。

(3) 代价函数

考虑电网电压、功率越限与交通网过度拥堵等安全约束, 设计代价函数如下:

$$C^{\text{PG}} = \max(0, V_{\min} - V_i) + \max(0, V_i - V_{\max}) \quad (34)$$

$$C^{\text{TN}} = \max(0, \frac{f_{a,b}}{C_{a,b}} - 1.2) \quad (35)$$

$$C^{\text{CS}} = \max(0, P_k - P_{k,\max}) + \max(0, P_{k,\min} - P_k) \quad (36)$$

$$C = C^{\text{PG}} + C^{\text{TN}} + C^{\text{CS}} \quad (37)$$

式中, C^{PG} 表示电网安全代价, 仅当电压越限时产生正代价, 越限幅度越大代价越高。 C^{TN} 为交通网安全代价。 C^{CS} 为充电站安全代价, 考虑功率越限。 C 为总安全代价。

(4) 奖励函数

基于目标函数与安全代价设计上下层奖励, 表达式如下:

$$R^{\text{upp}} = -F_1 - \lambda^{\text{upp}} \cdot C^{\text{upp}} \quad (38)$$

$$R^{\text{low}} = -F_2 - \lambda^{\text{low}} \cdot C^{\text{low}} \quad (39)$$

式中, $\lambda^{\text{upp}}, \lambda^{\text{low}}$ 为拉格朗日乘子。

(5) 动作-价值函数

调度动作的质量可以用 K 个时间步长内未来奖励的期望总和来评估, 表示为:

$$Q^{\pi}(s_t, a_t) = \mathbb{E}^{\pi} \left[\sum_{k=0}^K \gamma^k r(s_{t+k}, a_t) \mid s_t = s, a_t = a \right] \quad (40)$$

式中, $Q^{\pi}(s_t, a_t)$ 代表动作-价值函数; π 为当前电力-交通耦合调度策略; $\gamma \in [0, 1]$ 为折扣因子, 表示未来奖励相对于当前奖励的重要性。

该调度问题的目的是寻找到最优策略 π^* 以得到

最优动作-值函数 $Q^*(s_t, a_t)$, 上下层网络同理:

$$Q^*(s_t, a_t) = \max_{\pi} Q^{\pi}(s_t, a_t) \quad (41)$$

2 所提 MD3QN 算法的解决方案

(1) D3QN 算法

采用 D3QN 替代传统 DQN, 通过分离状态价值与动作优势函数, 降低 Q 值估计偏差:

$$Q(s_t, a_t; \theta) = V(s_t; \theta_v) + A(s_t, a_t; \theta_a) - \frac{1}{|\mathcal{A}|} \sum_{a' \in \mathcal{A}} A(s_t, a'; \theta_a) \quad (42)$$

式中, $\theta = (\theta_v, \theta_a)$ 为评估网络参数 (θ_v 为状态价值网络参数, θ_a 为优势函数网络参数); $|\mathcal{A}|$ 为动作总数, 平均优势项用于消除尺度模糊性。

构建目标网络 $Q'(s, a; \theta')$, 定期同步评估网络参数, 得到目标值 y_t :

$$y_t = r_t + \gamma Q(s_{t+1}, \arg \max_{a'} Q(s_{t+1}, a'; \theta'); \theta') \quad (43)$$

采用 TD-error 计算策略损失 $L(\theta)$:

$$L(\theta) = (y_t - Q(s_t, a_t; \theta))^2 \quad (44)$$

最后, 利用梯度下降法最小化损失, 并更新网络参数。

(2) 拉格朗日安全转换

CMDP 的核心是在“最大化累积奖励”的同时, 满足安全代价的期望不超过阈值, 即:

$$\begin{cases} \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] \\ \text{s.t. } \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t C(s_t, a_t) \right] \leq \Gamma \end{cases} \quad (45)$$

本文使用拉格朗日方法, 将 CMDP 的安全约束转化为无约束优化求解, 引导智能体在优化性能的同时主动规避不安全动作。

1) 拉格朗日函数构建

建立拉格朗日对偶变量 $\lambda \geq 0$, 将安全约束嵌入目标函数, 构建拉格朗日函数:

$$\mathcal{L}(\pi, \lambda) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] - \lambda \cdot \left(\mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t C(s_t, a_t) \right] - \Gamma \right) \quad (46)$$

当满足约束时, 惩罚项非正, 聚焦性能优化; 反之, $\mathcal{L}(\pi, \lambda)$ 随违反程度增大而减小, 抑制不安全动作。

2) 原-对偶问题转化

通过强对偶性假设, 原 CMDP 的带约束优化等价

于对偶问题的无约束优化:

$$\min_{\lambda \geq 0} \max_{\pi} \mathcal{L}(\pi, \lambda) \quad (47)$$

先固定 λ , 找到使 $\mathcal{L}(\pi, \lambda)$ 最大的策略 π ; 再调整 λ , 最小化该最大值, 最终得到同时满足性能与安全的最优策略。

3) 对偶变量 λ 的自适应更新

根据训练过程中安全代价的满足情况, 动态调整 λ 以平衡性能与安全:

$$\lambda_{t+1} = \max(0, \lambda_t + \eta \cdot (\bar{C}_t - \Gamma)) \quad (48)$$

式中, $\bar{C}_t = \mathbb{E}_{\mathcal{B}}[C(s, a)]$ 为当前训练批次 $\mathcal{B}$ 的平均安全代价; $\eta > 0$, 为更新步长; $\lambda_{t+1} \geq 0$, 确保对偶变量非负, 避免奖励不安全动作。

(3) 引导学习机制

为解决强化学习初始探索低效、收敛缓慢的问题, 本文设计基于专家经验的引导学习机制, 在训练初期利用专家策略约束智能体探索边界, 加速算法收敛:

1) 专家经验生成

针对典型故障场景(如线路故障、交通高峰拥堵), 采用基于线性化潮流模型的 MILP 求解上层优化与下层调度问题, 生成近优策略序列 $\{(s_t^{\text{exp}}, a_t^{\text{exp}}, r_t^{\text{exp}}, s_{t+1}^{\text{exp}})\}$, 存储于专家经验池 $\mathcal{D}^{\text{exp}}$, 作为引导学习的先验知识库。

2) 混合批次训练

为在利用专家知识与自主探索之间取得动态平衡,

设计了基于时间步长的混合采样策略。训练初期(前 20% 回合), 训练批次由 70% 专家经验与 30% 普通经验组成。随着训练推进, 专家经验比例线性衰减, 至 50% 训练回合时降至 10%, 后续阶段维持该比例直至训练结束。

3) 专家经验更新

考虑到 MILP 求解的静态性可能无法完全覆盖动态环境的随机演变, 因此建立专家经验池的动态更新机制。在训练过程中, 实时监控智能体生成轨迹的累积奖励值。当普通经验池 $\mathcal{D}$ 中的样本轨迹奖励值超过历史最优轨迹的 90% 分位数时, 判定该策略优于现有专家策略。系统将该高优样本存储到专家经验池 $\mathcal{D}^{\text{exp}}$ 中, 替换原有的低效样本。

3 仿真结果与分析

(1) 仿真设置

为验证本文方法在复杂场景下的有效性, 本研究采用 IEEE 33 节点配电网系统。该系统包含 33 个节点, 并设置了 5 组联络开关以支持拓扑重构, 配电网拓扑结构如图 2 所示, 基准电压为 12.66kV。交通网络拓扑如图 3 所示, 共有 24 个节点, 5 条道路设有潮汐车道(分别为路段 3-4、路段 8-9、路段 10-11、路段 19-20、路段 22-23)。5 个充电站分别部署在交通节点 4、7、10、19、23 处, 配电网节点 6、8、12、19、22 处。所有仿真在硬件配置 CPU I7-12700KF, GPU RTX4060 和 RAM 32GB 以及 Python 环境中进行验证。

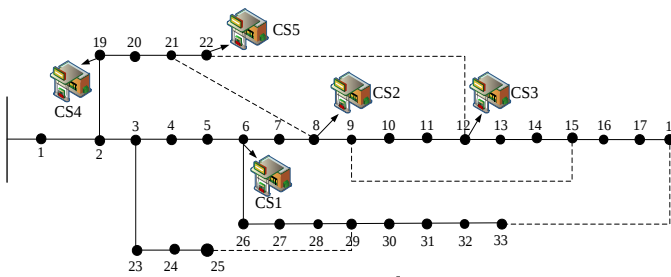


图 2 IEEE-33 节点拓扑图

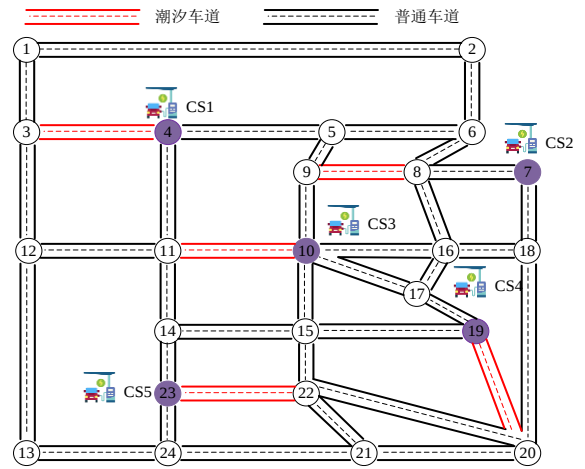


图 3 交通网络拓扑图

(2) 训练过程

图 4 和图 5 分别展示了采用所提方法决策获得的上下层每回合训练的奖励值。由图可知, 二者的奖励曲线从剧烈波动逐渐过渡到平稳收敛, 表明算法初期通

过探索积累有效策略, 后期进入局部最优区间稳定优化。上层网络在 448 回合进入收敛区间, 奖励值从-114 升至-58。下层网络在 286 回合进入收敛区间, 奖励值从初始-155 提升至-63。由于上层动作空间涉及复杂的

拓扑重构组合,其探索难度高于下层充电资源调度,因此收敛所需回合数较多。

(3) 实际测试结果

针对不同区域线路故障的场景,本文选取了配电

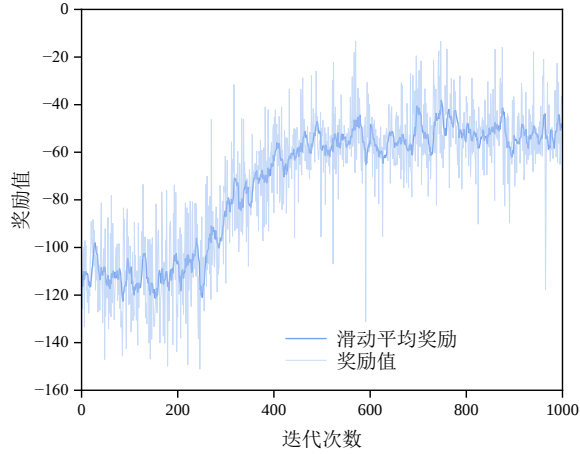


图4 所提方法上层训练过程获得的奖励值

网中 4 个代表性位置的线路故障,选取了涵盖早高峰(8:00)、晚高峰(17:00)及平峰时段(12:00, 15:00)的典型时刻进行测试,实验结果如表 1 所示。

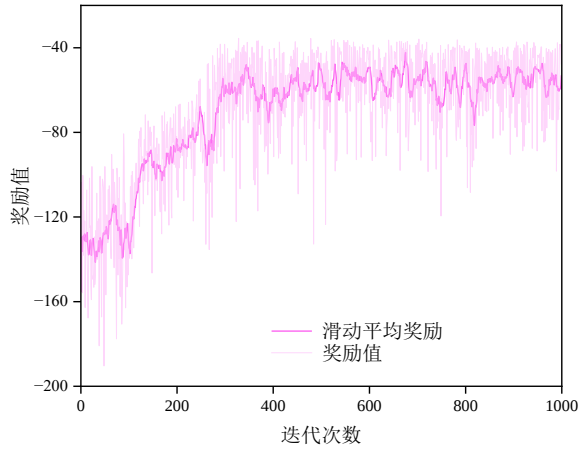


图5 所提方法下层训练过程获得的奖励值

表 1 典型故障场景下的应急调度策略测试结果

故障线路	故障发生	联络开关动作	潮汐车道方向	CS 选择	负荷恢复率
6-26	08:00	闭合(8-21) 断开(7-8)	正向开放路段 3-4	CS 1	95.2%
16-17	17:00	闭合(18-33) 断开(16-17)	反向开放路段 19-20	CS 4	98.5%
28-29	12:00	闭合(25-29) 断开(28-29)	关闭调控路段 8-9	CS 2	96.8%
2-3	15:00	闭合(8-21)	反向开放路段 22-23	CS 3	94.7%

表 1 的仿真结果表明,本文所提策略具有显著的有效性与场景适应性。系统负荷恢复率稳定在 94.7%至 98.5%之间,其中晚高峰时段故障恢复率最高为 98.5%,表明策略在交通与电力负荷双重压力下仍能有效支撑系统恢复。早高峰通过开放正向潮汐车道(路段 3-4)并调用邻近充电站(CS1)疏导车流、支撑电网。晚高峰则切换路段 19-20 反向并适配 CS4,以匹配交通潮汐特征。平峰期则侧重电网重构与充电站经济调度,减少对交通系统的主动干预。

为分析本文所提方法对配电网韧性恢复的效果,图 6 展示了策略实施前后配电网电压差值。由图可知,策略实施前配电网电压偏差较高,电压偏差平均值为 5.15×10^{-2} p.u.。最大电压偏差达到 8.69×10^{-2} p.u.,电压稳定性不足。经本文方法调控后,电压偏差范围显著收窄。电压偏差平均值降低至 3.26×10^{-2} p.u.,降低了 36.69%。最大电压偏差降至 5.88×10^{-2} p.u.,且整体分布更趋均衡。因此,本文方法有效降低配电网电压偏差幅

度,提升了电压分布稳定性。

(4) 不同方法测试

为验证所提策略的优越性,本节选取 MILP 作为静态全局优化的理论基准;选取 DQN 作为强化学习的常规对照组;并选取未引入引导机制的 D3QN 进行消融对比。图 7 展示了上述三种基准方法与本文所提 MD3QN 在典型故障场景下的电压偏差分布情况。由图可知,MILP 作为传统优化方法,电压偏差平均值为 3.92×10^{-2} p.u.,表现稳定但优化速度受限。DQN 与 D3QN 因缺乏安全引导与专家经验,电压偏差分别为 4.68×10^{-2} p.u.与 4.15×10^{-2} p.u.,波动较大且收敛性不足。相比之下,本文所提 MD3QN 方法电压偏差平均值最低,仅为 3.26×10^{-2} p.u.,且分布最为集中,最大偏差亦控制在 5.88×10^{-2} p.u.以内。综上,MD3QN 通过拉格朗日安全约束与专家经验引导,有效抑制了无效动作探索,在优化电网性能的同时动态保障电压安全,显著提升了电压稳定性和策略可靠性。

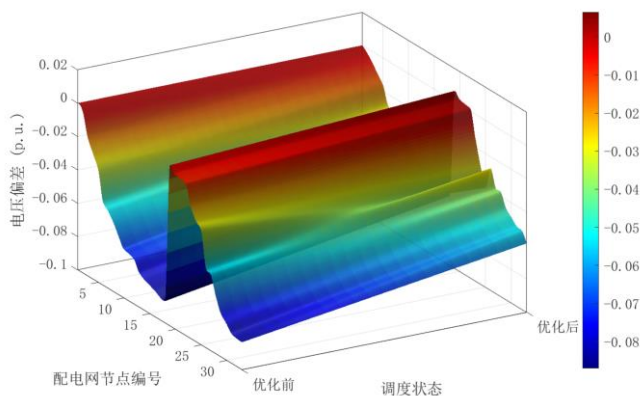


图6 策略实施前后配电网电压偏差

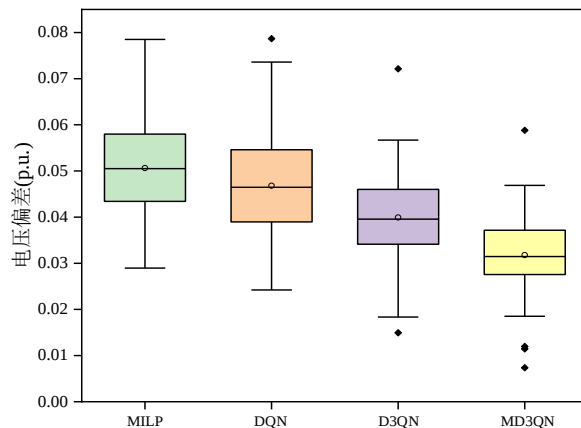


图7 不同方法下的电压偏差对比

(5) 不同场景测试

图8展示了在不同故障规模与交通压力组合场景下，采用本文方法后的负荷恢复率变化。由图可知，当并发故障从1处增加至5处时，因为专家经验引导下的快速重构，负荷恢复率仅从97.5%缓慢下降至91.2%。这证明了本文所提双层协同架构与图注意力机制能够动态感知耦合系统状态，在故障规模扩大或交通压力增强时，仍能通过潮汐车道调度、充电站响应与电网重构的联动，维持较高的负荷恢复水平，体现出良好的场景适应性与系统韧性。

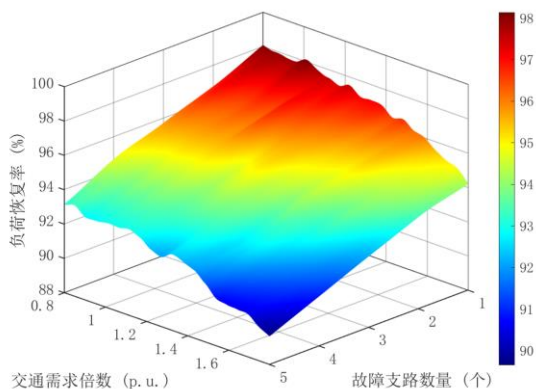


图8 不同故障规模与交通压力下的负荷恢复率

4 结语

本文提出了一种基于安全引导图强化学习的应急调度策略。通过图注意力机制特征感知、嵌入安全约束与引导学习的MD3QN算法，实现了多系统资源的动态协同调度，得到如下结论：

1) 本文构建了基于GAT-D3QN的故障耦合网络的双层恢复架构，实现了电网重构、潮汐车道调控与电

动汽车调度的动态协同，显著提升了故障恢复效率与系统韧性。

2) 本文建立了故障CMDP模型，利用GAT精准捕捉电力-交通异构拓扑的动态特征。仿真验证表明，在早晚高峰及多故障并发场景下，系统负荷恢复率稳定维持在91.2%~98.5%，体现了模型在应对复杂故障规模与交通压力时优异的场景适应性。

3) 本文提出了基于拉格朗日安全约束与专家经验引导的MD3QN算法，有效解决了强化学习初期探索低效的问题，并保障了物理安全边界。结果显示，该策略将配电网电压偏差平均值降低了36.69%，验证了算法在保障电压安全硬约束前提下的高效寻优能力。

未来研究将依托本文提出的安全引导机制，进一步探索策略在部分可观测环境及非理想通信条件下的鲁棒性研究。

参考文献

- [1] 孙科,陈文钢,陈佳佳,等. 基于电动汽车的极端场景多微电网韧性提升策略研究[J]. 电力系统保护与控制, 2023, 51(24): 53-65.
- [2] 朱晓荣,司羽. 考虑物理—信息—交通网耦合的配电网多时段动态供电恢复策略[J]. 电工技术学报, 2023, 38(12): 3306-3320.
- [3] 徐岩,郭佳睿,马天祥. 考虑韧性提升的配电网故障恢复与抢修协调优化[J]. 高电压技术, 2024, 50(12): 5516-5528.
- [4] 刘文泽,陈珂瑶,成润婷,等. 面向韧性提升的主动配电网灵活调度与故障修复协同策略[J]. 电力建设, 2025, 46(11): 10-23.

- [5] Li Y, Xiao J, Chen C, et al. Service Restoration Model With Mixed-Integer Second-Order Cone Programming for Distribution Network With Distributed Generations [J]. IEEE Transactions on Smart Grid, 2019, 10 (4): 4138-4150.
- [6] 田书欣,姚尚坤,符杨,等. 地震灾害下考虑交通路况的主动配电网动态协同恢复策略 [J]. 电力建设, 2024, 45 (01): 68-82.
- [7] 孔惠文,马静,程鹏,等. 基于灾害场景预估的配电系统韧性两阶段故障恢复策略 [J]. 电网技术, 2024, 48 (09): 3812-3821.
- [8] Liu X, Shahidehpour M, Li Z, et al. Microgrids for Enhancing the Power Grid Resilience in Extreme Conditions [J]. IEEE Transactions on Smart Grid, 2017, 8 (2): 589-597.
- [9] Bedoya J C, Wang Y, Liu C C. Distribution System Resilience Under Asynchronous Information Using Deep Reinforcement Learning [J]. IEEE Transactions on Power Systems, 2021, 36 (5): 4235-4245.
- [10] Zhao T, Wang J. Learning Sequential Distribution System Restoration via Graph-Reinforcement Learning [J]. IEEE Transactions on Power Systems, 2022, 37 (2): 1601-1611.
- [11] 何小龙,高红均,王仁浚,等. 基于图深度强化学习的有源配电网故障恢复方法 [J]. 电网技术, 2025, 49 (10): 4342-4352.
- [12] 马继龙,胡旭光,王杰,等. 面向大规模配电网的时空关联图神经网络故障定位方法研究 [J/OL]. 中国科学:技术科学, 2026,1-18.
- [13] 张浩然,许沛东,乔骥,等. 基于改进示教型强化学习的有源配电网故障恢复方法 [J/OL]. 电网技术, 2025,1-12.
- [14] 萧文聪,陈俊斌,余涛,等.基于专家知识嵌入强化学习的配电系统灾后恢复决策方法[J].电力系统自动化,2025, 49(12):91-100.

版权声明: ©2026 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS