

# 图强化学习在配电网优化中的应用综述

江昌旭<sup>1</sup>, 林骏驰<sup>1</sup>, 郭辰<sup>1</sup>, 袁羽娟<sup>2</sup>, 林俊杰<sup>1</sup>

(1. 福州大学电气工程与自动化学院, 福建省 福州市 350108;

2. 广东电网有限责任公司惠州供电局, 广东省 惠州市 516000)

## A Review on Application of Graph Reinforcement Learning in Power Distribution Network Optimization

JIANG Changxu<sup>1</sup>, LIN Junchi<sup>1</sup>, GUO Chen<sup>1</sup>, YUAN Yujuan<sup>2</sup>, LIN Junjie<sup>1</sup>

(1. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, Fujian Province, China;

2. Huizhou Power Supply Bureau of Guangdong Power Grid Co., Ltd., Huizhou 516000, Guangdong Province, China)

**ABSTRACT:** With the advancement of the carbon peak and neutrality targets, the integration of large amounts of highly uncertain renewable energy sources into the grid has introduced the uncertainty of the power system and severely altered the power flow and voltage distribution in the distribution network, which poses new challenges for the operation and management of the distribution network. Artificial intelligence technology has experienced very rapid development in recent years. Graph reinforcement learning algorithms (GRL) effectively combine the feature extraction capabilities of graph neural network (GNN) for non-Euclidean structured data with the sequential decision-making capabilities of deep reinforcement learning (DRL). This enables GRL to possess a strong nonlinear fitting ability, excellent adaptability, outstanding scalability, and significant advantages in dynamic optimization decision-making. As a result, GRL is highly suitable for addressing complex uncertainty optimization and decision-making problems in distribution network with graph-based models. The principles and characteristics of GRL are introduced first by combining the principles of reinforcement learning with commonly used types of GNNs. Then, the applications of GRL in distribution network are analyzed in four aspects: optimal dispatch, dynamic reconfiguration, fault recovery, and collaborative optimization of coupled power-transportation networks. Then a comparison is made between the application status and performance of GRL and DRL, analyzing the advantages of GRL in distribution network optimization as well as the existing issues. Finally, the issues that need improvement in the application of GRL for power distribution network optimization are summarized, and the future prospect of its

application is discussed.

**KEY WORDS:** graph reinforcement learning; graph neural network; optimal dispatch of distribution network; distribution network reconfiguration; fault recovery; coupled power-transportation networks optimization

**摘要:** 随着“双碳”目标的推进,大量具有较强不确定性的可再生能源接入配电网,进一步加剧了电力系统的不确定性并严重改变了配电网潮流和电压分布,使得配电网的运营和管理面临新的挑战。近年来,人工智能技术迅速发展,图强化学习算法(graph reinforcement learning, GRL)有效地结合了图神经网络(graph neural network, GNN)的非欧式结构数据特征提取能力和深度强化学习(deep reinforcement learning, DRL)的序贯决策能力,使其具备较强的非线性拟合能力、出色的适应性、优异的可扩展性,以及在动态优化决策中表现出的显著优势,在解决具有图模型结构的配电网复杂不确定性优化和决策问题上展现出了优越的效果。该文首先结合强化学习的原理及常用的GNN类型总结了GRL的原理及特点;然后从配电网的优化调度、动态重构、故障恢复、电力-交通耦合网协同优化4个方面对GRL在配电网优化中的应用现状进行综述,并将其与传统算法的应用现状和应用效果进行对比,分析了其在配电网优化中的应用优势及存在的问题;最后对GRL在配电网优化中应用需改进的问题进行总结,并对应用前景进行了展望。

**关键词:** 图强化学习; 图神经网络; 配电网优化调度; 配电网重构; 故障恢复; 电力-交通耦合网协同优化

**DOI:** 10.13335/j.1000-3673.pst.2025.1681

## 0 引言

随着“双碳”目标战略的实施,高比例分布式电源(distributed generation, DG)接入电网,大量电动汽车(electric vehicle, EV)及其充电站(charging station, CS)投入运行推动了电力系统向低碳环保和

基金项目: 国家自然科学基金项目(72401069); 福建省自然科学基金项目(2022J05125)。

Project Supported by National Natural Science Foundation of China (72401069); Natural Science Foundation of Fujian Province (2022J05125).

可持续发展方向转型。然而,这也导致配电网的潮流分布与负荷随机性显著增强,对其经济性与稳定运行构成严峻挑战,亟需采取更为有效的优化控制方法<sup>[1]</sup>。

配电网运行优化问题的本质是高维、非线性、复杂随机优化问题。对该类问题的求解常见的算法主要包括数学优化、启发式算法和机器学习算法<sup>[2-6]</sup>。其中,数学优化算法的优点是求解的精度高、可解释性强,但计算时间较长,并且依赖于完整且静态的数学模型,对复杂动态系统和随机因素的响应能力不足,因此适用于机理清晰、可获取信息较完备的系统及较长时间尺度的优化。启发式算法包括蚁群算法、遗传算法和粒子群算法等<sup>[7]</sup>。文献[8-11]采用各种类型的启发式算法求解配电网的无功优化、拓扑重构等问题,能够比较有效搜索出较优解,但存在计算速度慢、容易陷入局部最优等明显缺点<sup>[12-13]</sup>。部分文献尝试将这些算法结合或改进以达到更好的求解效果,如文献[14]将粒子群算法和遗传算法结合并运用于EV充电引导,改进后的算法多目标优化能力增强,同时提高了配电网的稳定性与充电的经济性,但计算速度较慢的缺点仍然难以避免。

近年来,机器学习逐渐兴起,其中,强化学习依靠智能体与环境进行动态交互并根据反馈来调整行为,但问题规模扩大会使得系统状态的维度急剧升高,求解效率会严重下降,甚至无法求解。针对这个问题,深度强化学习(deep reinforcement learning, DRL)利用深度神经网络提高算法解决高维空间复杂实际问题的能力,被广泛运用到配电网运行优化的研究当中。

相比于传统数学优化算法和启发式算法,DRL对随机环境具有更强的适应性,能够处理延迟奖励,决策速度快,求解高维复杂非线性优化问题的能力。但面对部分配电网优化问题时,也会遇到收敛性较差、求解精度不够高等问题,主要原因是传统神经网络处理非欧式图结构的效率和准确性较低。作为一种针对非欧式图结构设计新型神经网络,图神经网络(graph neural networks, GNN)能够依靠聚合邻域的信息传递方式高效地处理图结构数据。DRL拥有的良好泛化性能与前向推理能力能够较好地满足大型系统实时优化决策的需求,而配电网的拓扑结构具有天然的非欧式图结构特征,因此,与GNN相结合的DRL算法,即图强化学习(graph reinforcement learning, GRL),为配电网优化问题的求解提供了新的策略。

当前国内外已有许多文献对DRL在电力系统优化决策中的应用现状进行了总结<sup>[15-17]</sup>,但这些文献尚未针对电力系统包含的拓扑特点进行分析,缺少对GRL在电力系统中的应用进行结构化评述的系统综述及展望,特别是针对配电网的细化应用场景和实际挑战的总结。因此,本文从配电网的图结构特点与建模方法入手,系统分析了图建模与配电网优化问题的结合机理并对当前GRL在配电网中的应用进行综述。本文首先结合配电网的图结构建模方法、GNN和DRL的基本原理等内容系统地介绍了算法原理,接着对基于GRL的配电网优化研究成果进行总结,并与传统算法进行性能对比,分析了GRL算法的优缺点,最后针对存在的缺点对未来的研究方向与应用前景进行展望。附录A表A1整理了本文的主要术语及其英文缩写对照。

## 1 图神经网络概述

### 1.1 配电网图结构建模方法

图结构是一种非欧式数据,由节点及连接不同节点的边组成,其数学表达式如下:

$$G = (V, E) \quad (1)$$

式中:  $V$  为节点的集合;  $E$  为节点之间边的集合。

配电网具有天然的图结构特点,在基础的图结构中,节点与边分别对应于配电网的物理节点及节点之间的闭合支路,节点特征与边特征分别对应于配电网节点上和支路上的信息。随着DG和EV的接入,配电网节点与边的数据表现出明显的电气耦合关系、异质性和时间关联特性等,可以根据不同特性将配电网构建为不同类型的图模型。配电网图结构建模的主要方式如下:

#### 1.1.1 基本图结构建模

在最基本的图结构当中,配电网的节点与支路直接对应于图结构中的节点和边。基本图结构数据主要由节点信息、边信息和拓扑信息组成。其中,节点信息和边信息都可以用向量表示,如式(2)所示。节点信息主要包括配电网节点的功率、电压幅值、发电功率等,边信息主要包括配电网支路的功率、电流大小、阻抗等。

$$\begin{cases} \mathbf{x}_i = [P_i, Q_i, U_i, G_i, \dots] \\ \mathbf{x}_{ij} = [P_{ij}, Q_{ij}, I_{ij}, Z_{ij}, \dots] \end{cases} \quad (2)$$

式中:  $\mathbf{x}_i$  为节点  $i$  的节点信息向量;  $P_i$ 、 $Q_i$  和  $U_i$  分别为节点  $i$  的有功功率、无功功率和电压幅值;  $G_i$  为节点  $i$  所接入的发电机的出力;  $\mathbf{x}_{ij}$  为支路  $ij$  的边信息向量;  $P_{ij}$ 、 $Q_{ij}$  和  $I_{ij}$  分别为支路  $ij$  的有功功率、无功功率和电流;  $Z_{ij}$  为支路  $ij$  的阻抗。

拓扑信息主要通过邻接矩阵来表达，其各个位置的元素计算式如下：

$$A_{ij} = \begin{cases} 1, & i=j \text{ 或节点 } i, j \text{ 直接相连} \\ 0, & \text{其他} \end{cases} \quad (3)$$

根据节点所邻接的节点数对邻接矩阵进行归一化后，其中的元素可以表示为

$$\hat{A}_{ij} = \frac{\alpha_{ij} A_{ij}}{\sum_{k=1}^N A_{ik}} \quad (4)$$

$$\sum_{k=1}^N \alpha_{ik} = 1 \quad (5)$$

式中： $\alpha_{ij}$ 为节点  $i, j$  之间的边权重； $N$ 为图结构的节点总数。

### 1.1.2 动态邻接矩阵建模

配电网的节点之间存在一些动态的电气耦合关系，如无功-电压灵敏度、有功-电压灵敏度、电气距离等信息，因此也可以对配电网进行动态邻接矩阵建模，以使得图数据能够对更加复杂的空间关联特性进行灵活表征。

以无功-电压灵敏度为例，动态邻接矩阵中的元素的计算表达式如下：

$$A_{ij} = \begin{cases} f\left(\frac{\partial U_i}{\partial Q_j}\right), & \frac{\partial U_i}{\partial Q_j} \geq \tau \\ 0, & \text{其他} \end{cases} \quad (6)$$

式中： $U_i$ 为节点  $i$  的电压幅值； $Q_j$ 为注入节点  $j$  的无功功率； $\tau$ 表示判断电气耦合强度的阈值。

式(6)的物理意义为：当节点  $j$  的无功功率变化对节点  $i$  的电压变化影响程度高于设定的阈值  $\tau$ ，则认为节点  $i, j$  之间的无功-电压灵敏度较高，电气耦合性较强，邻接矩阵中的元素  $A_{ij}$  根据无功-电压灵敏度进行一定规则的计算，赋予节点  $i, j$  由电气耦合特性所决定的关联性数值；反之，则认为二者关联性较低，因此将  $A_{ij}$  设置为 0。

### 1.1.3 异质图结构建模

配电网节点的负荷类型、节点接入的 DG 类型、可控资源类型、支路的设备类型等具有明显的异质性，观测数据也具有多源异构的特点，因此可以根据这些特点将配电网进一步构建为异质图结构，使得不同节点的调控能力或对于场景变化的不同响应特性得以区分和协调。异构图结构可以表示为

$$G = (V, E, M_V, M_E) \quad (7)$$

式中： $M_V$ 为节点类型集合； $M_E$ 为边类型集合。

基于分类集合，可以进一步建立节点和边与对应类型之间的映射关系：

$$\begin{cases} \varphi: V \rightarrow M_V \\ \psi: E \rightarrow M_E \end{cases} \quad (8)$$

式中： $V_i$ 为第  $i$  个节点； $E_e$ 为第  $e$  条边； $\varphi_i$ 为节点的类型映射； $\psi_e$ 为边的类型映射。边的类型可以由其两端的节点来确定，若边  $e$  两端的节点分别为  $i, j$ ，则  $\psi_e$  的表达式可以写为

$$\psi_e = \Psi_{ij} = (\varphi_i, \varphi_j) \quad (9)$$

建立分类映射后，一般需要进一步根据节点和边的类型对图结构进行子图拆分，以便于同类型的节点能够进行分类聚合<sup>[18-19]</sup>。

### 1.1.4 时空图结构建模

在大规模 DG 和 EV 接入的情况下，配电网的节点功率呈现出较强的时空不确定性，对配电网进行时空图结构建模有利于 GNN 对时间特性进行捕捉，使得智能体能够观察环境状态随时间的演变特性，从而做出更有利于全局的决策。

在配电网的时空图结构当中，节点信息与边信息均为时间序列：

$$\begin{cases} \mathbf{X}_i = [\mathbf{x}_{i,0}, \mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,T}] \\ \mathbf{X}_{ij} = [\mathbf{x}_{ij,0}, \mathbf{x}_{ij,1}, \dots, \mathbf{x}_{ij,T}] \end{cases} \quad (10)$$

式中： $\mathbf{X}_i$ 为节点  $i$  的时序特征向量； $\mathbf{X}_{ij}$ 为节点  $i, j$  之间的支路的时序特征向量； $T$ 为时序总数。

在处理时空图结构数据时，每个时序都可以视为一个单独的图结构，随着时间的推进，每个图结构的拓扑都可能产生变化，因此其邻接矩阵也可以构建为时间序列：

$$A_{ij} = [A_{ij,0}, A_{ij,1}, \dots, A_{ij,T}] \quad (11)$$

上述各种类型的配电网图结构建模方法的整体框架如附录 A 图 A1 所示。

## 1.2 GNN 原理及类型

### 1.2.1 神经网络基本原理

神经网络是一种通过反向传播机制进行训练从而获得输入和输出之间复杂非线性映射的机器学习工具，目前常用于传统 DRL 的神经网络类型主要可以分为前馈网络、循环网络和自注意力网络。其中，前馈网络的常见类型有人工神经网络 (artificial neural networks, ANN) 和卷积神经网络 (convolution neural network, CNN)，循环网络的常见类型有门控循环单元和长短时记忆网络，自注意力网络最主要为 Transformer。上述各种神经网络的具体原理及特点详见附录 B.1。

### 1.2.2 GNN 基本原理

上述神经网络大多不具备空间特征识别能力，CNN 虽然具备，但仅限于规则的欧式数据，如列表、

图像数据。对于电网拓扑这样的非欧式数据而言，节点间没有明显的先后顺序与网格特征，平整排列与网格化的数据处理方式缺乏明确的物理意义。对此，GNN 在输入特征中考虑了节点之间的邻接关系，充分利用了节点之间的空间关联性，可以有效地对图结构进行学习。GNN 通过传播模块在节点间传递信息，并通过将邻域信息聚合到目标节点的方式更新节点状态，省略了对节点进行排列的过程，因此可以快速提取非欧图结构特征<sup>[20]</sup>。

GNN 通过聚合函数和输出函数对图结构中的节点状态进行更新。其中，聚合函数按一定的规则对节点及其相邻节点的信息进行聚合，如求和、均值、池化等，得到包含邻域信息的节点状态；输出函数再根据当前节点状态和节点自身的属性输出新的节点状态。该过程可以表示为

$$h_v^{(l+1)} = f(x_v, x_{ne(v)}, x_{co(v)}, h_v^{(l)}) \quad (12)$$

$$o_v = g(h_v^{(l+1)}, x_v) \quad (13)$$

式中： $h_v^{(l)}$  为图神经网络中节点  $v$  在第  $l$  层的状态； $f$  为聚合函数，该部分是 GNN 能够捕捉图结构拓扑信息的关键； $x_v$  为节点  $v$  的特征； $x_{ne(v)}$  为节点  $v$  的邻居节点特征， $x_{co(v)}$  为与节点  $v$  相连的边的特征； $g$  为输出函数，能够根据聚合后的图结构信息计算出所需要的输出； $o_v$  为节点  $v$  的状态输出结果。图 1 对 GNN 的基本原理进行了图解。

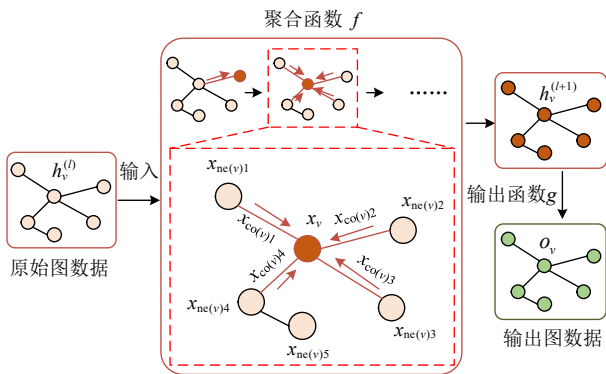


图 1 GNN 节点状态更新过程

Fig. 1 Node status update process of GNN

### 1.2.3 GNN 主要类型

GNN 是 GRL 的核心组成部分，在现有的基于 GRL 的配电网优化研究中，常用的 GNN 类型主要有图卷积网络(graph convolution network, GCN)、图注意力网络(graph attention network, GAT)、图胶囊卷积网络(graph capsule convolutional neural networks, GCAPCN)和时空图网络(spatio-temporal graph neural network, STGNN)等。上述各种 GNN 类型的原理及公式详见附录 B.2—B.5。

## 2 图强化学习概述

### 2.1 强化学习

强化学习是一种不断获取外部环境反馈信息并作为经验数据实现对序贯决策问题优化求解的机器学习算法。其过程涉及环境、状态、动作、奖励值等多个要素，可以使用马尔可夫决策过程(Markov decision process, MDP)进行描述。

MDP 可以表示为 5 元组  $(S, A, P, R, \gamma)$ 。其中， $S = \{s^{(1)}, s^{(2)}, \dots, s^{(m)}\}$  表示系统环境的状态集合； $A = \{a^{(1)}, a^{(2)}, \dots, a^{(m)}\}$  表示智能体的动作集合； $P = P(s^{(t+1)} | s^{(t)}, a^{(t)})$  表示状态转移概率分布； $R = R(s^{(t)}, a^{(t)})$  为奖励函数，表示环境对智能体做出的数值评价； $\gamma$  为小于 1 的折扣因子，表示奖励函数值的权重随时序推移而逐步衰减的快慢程度。

强化学习的具体原理及常见类型详见附录 C。

### 2.2 深度强化学习

传统强化学习需要对环境中的所有状态及其对应的所有动作值函数进行存储，在连续状态与动作空间中无法遍历所有状态及动作，对此，DRL 运用了神经网络的非线性特征提取能力。与传统强化学习类似，DRL 也由智能体、状态、动作、状态转移概率和奖励函数等要素构成，区别在于使用了神经网络对价值和策略函数进行拟合，能够有效解决传统算法无法处理连续状态空间与连续动作空间的问题。DRL 的基本架构如图 2 所示。

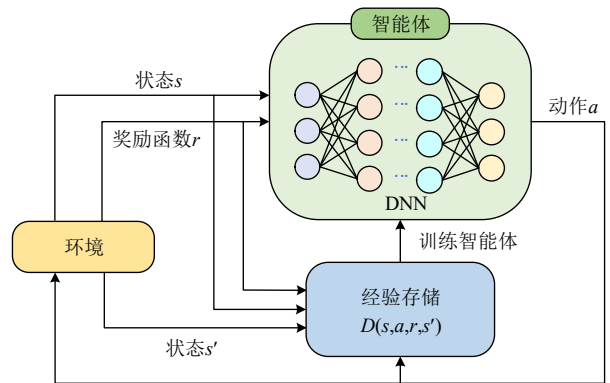


图 2 DRL 基本框架

Fig. 2 Basic framework of DRL

DRL 的类型主要可分为基于价值的算法、基于策略的算法和基于价值与策略结合的算法。其中，在基于价值的算法中，神经网络主要用于对所有动作的价值都进行评估，并从拟合出的动作价值空间中选取价值最高的动作。最常见的基于价值的算法主要有深度 Q 网络(deep Q network, DQN)和深度双 Q 网络(double DQN, DDQN)。基于策略的算法直接对策略进行建模和优化，通过神经网络参数化

策略函数，根据当前状态直接输出动作或动作的概率分布，没有价值函数拟合的过程。最常见的算法为近端策略优化(proximal policy optimization, PPO)。基于价值和策略结合的算法一般设计为动作-评价(actor-critic, AC)框架，由 Actor 网络和 Critic 网络相互协同。其中，Actor 根据输入的状态直接输出动作，为策略部分；Critic 则对所选动作的价值进行评估并指导策略更新，为价值部分。这种框架不需要对所有动作进行评价，并且可以直接输出连续值，因此适用于进行连续变量的决策。基于 AC 框架的典型算法主要有深度确定性策略梯度(deep deterministic policy gradient, DDPG)和双延迟 DDPG(twin delayed DDPG, TD3)等。

### 2.3 图强化学习

尽管 DRL 能够有效处理复杂优化决策问题，但在处理非欧式图结构时容易忽略部分特征，导致优化结果不够准确或者策略更新效率低。为了更有效地求解图结构上的优化问题，GRL 发挥了 GNN 处理图结构能力强的优势，可以对图结构数据进行高效学习<sup>[21]</sup>。其核心思想是在智能体的神经网络结构当中加入图处理层进行表征处理，再将处理后的特征输入到传统网络层中进行进一步的处理，即对特征与所需输出数据之间的映射关系进行学习。图 3 对比了 GRL 与 DRL 决策过程的主要区别。

对于本章介绍的各类算法，附录 C 表 C1 总结了一些具有代表性的算法类型及其特点。

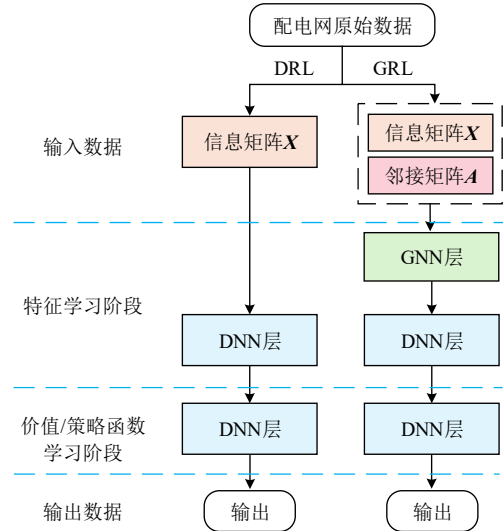


图 3 DRL 与 GRL 决策过程  
Fig. 3 Decision making process of DRL and GRL

### 3 图强化学习在配电网优化中的应用

GRL 在配电网优化领域已经得到较多应用并取得了良好成效，发挥出了拓扑动态适应性强、决策速度快、免模型等显著优势。目前该算法在配电网优化问题中应用的场景主要集中于配电网优化调度、配电网动态重构、配电网故障恢复和电力-交通耦合网系统优化等，因此本文选取这 4 种场景对基于 GRL 的配电网优化问题研究成果进行综述。图 4 结合算法在配电网优化中的基本应用过程与 4 种具体优化问题构建了 GRL 在配电网优化中的应用框架。

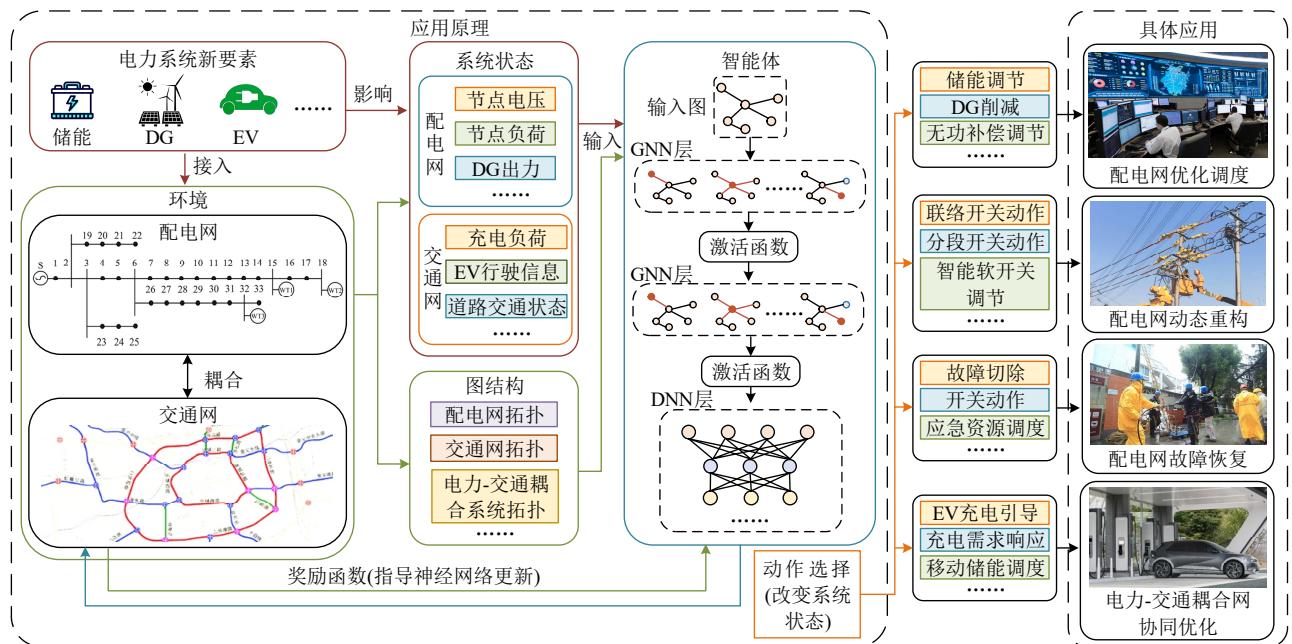


图 4 基于 GRL 的配电网优化应用框架  
Fig. 4 Application framework of distribution network optimization based on GRL

#### 3.1 基于 GRL 的配电网优化调度研究

优化调度是指对接入配电网的储能、发电机以

及无功补偿设备等可控资源的功率进行调整，达到降低运行成本和提高电压稳定性等目标。文献[22]

构建了电-氢-交通耦合的配电网调度模型,考虑了能源自洽率指标及氢能汽车与共享氢储能,平衡负荷且降低碳排。文献[23]考虑了配电网优化过程中多主体的意愿差异,构建了基于多主体利益划分的分层优化调度模型,实现多主体协同优化。从上述文献中可以看出数学优化具有求解精度高的优势,但建模复杂、求解速度慢。随着 DG 和 EV 的大规模接入,配电网运行中的随机性不断增加,调度问题日趋复杂,传统数学优化算法难以满足优化调度的实时性需求。此外,数学优化的建模需要依赖于配电网参数信息,然而在工程实际中,电力系统的信息往往难以完整且准确地获取,尤其是电压等级较低的配电网<sup>[24]</sup>。因此,数学优化算法在配电网优化调度中的局限性比较明显。

相比于数学优化方法,DRL 算法具有更快的决策速度,并且不依赖于准确的系统参数,因此被应用于含多重不确定性因素的配电网优化调度问题中<sup>[25]</sup>。文献[26]采用 DDPG 进行配电网的动态经济调度,方法简单有效,并且算法对随机因素的适应性较好,但是训练所耗费的时间较长。对于训练速度慢的问题,文献[27]采用了 TD3 求解配电网低碳经济调度问题,使用世界模型对配电网环境进行近似并利用历史数据训练,实现对含风电随机环境的有效模拟。文献[28]提出了一种三时间尺度的光储配电网优化调度方法,针对电容器、静止无功补偿器、储能等设备不同特性进行时间尺度划分,并根据不同设备动作的连续性与离散性差异采用不同算法进行优化,实现更高效有序的调度。DRL 的一个缺点是难以精确地服从约束条件,在执行阶段容易违反配电网安全约束。对此,文献[29]采用动作-双评价算法求解,在传统 AC 框架的基础上增加一个评价网络对安全约束成本进行评价。

上述配电网优化调度方法具有决策速度快、响应随机变化能力强的优点。然而,DRL 算法仅关注配电网的节点信息,并未捕捉到图结构特点。对此,部分学者利用了 GRL 算法能够处理拓扑信息的优势进行配电网优化调度研究。文献[30]采用基于 GCN 的柔性动作-评价算法(soft AC, SAC)求解配电网优化调度问题,并设计了集中式与分布式两种不同框架。其中,集中式框架直接对整个系统进行求解;分布式框架对配电网进行分区并为每个子区域设置智能体,智能体在训练阶段利用全局信息,在执行阶段仅需观测本地信息。仿真结果表明,分布式算法的训练速度与最优性低于集中式算法,但其优势在于各个智能体不依赖

全局观测数据,有利于优化调度的就地决策与协调。文献[31]采用基于多智能体(multi-agent, MA)的 DDPG 求解无功优化问题,其中神经网络分别利用门控循环单元和 GCN 以提取时间特征和拓扑特征。相比于传统算法,所提出的算法在 33 节点与 141 节点系统上优化后的电压越限率、网损和电压偏移程度均更低。然而,该文献并没有考虑到多智能体之间的配合以及不同智能体的动作对全局奖励的贡献程度。对此,文献[32]在基于 GCN 的 MATD3 中引入了元学习,使得智能体能够通过有限的观测数据对其他智能体的动作进行预测,从而加强智能体之间的配合。此外,该文献利用了自注意力网络评估各个智能体对全局优化的贡献程度,提高优化精度。

上述文献基于固定边权重进行特征聚合,未利用到决策变量与待优化指标之间的动态耦合特性。对此,基于 GAT 的算法所具有的自适应边权调节能力受到研究者的广泛关注。文献[33]将 GAT 与 DDPG 结合,求解主动配电网实时经济调度问题,发挥了图注意力机制的优势。在 33 节点配电网与基于 GCN 的算法相比,所提出的算法收敛所需的训练回合数减少约一半,验证了训练效率上的优势。该文献也在 69 节点和 123 节点系统上验证了算法的适应性。文献[34]将 GAT 与 MASAC 结合进行配电网的有功-无功协同优化并利用物理模型预先获收集经验以辅助算法的训练。该文献采用 33 节点配电网进行仿真,并在仿真过程中改变配电网的拓扑,从而验证了算法的动态拓扑适应能力。与传统的 MADRL 相比,所提算法优化后的电压偏移程度与网损均减少,并且在拓扑变化时仍然保持较好的优化结果。文献[35]依据配电网节点间的电压-无功灵敏度进行动态邻接矩阵建模,采用基于 GAT 的 MDDPG 进行无功电压优化,将动态边权机制与动态邻接矩阵相结合。所提的算法在 33 节点系统上优化后的结果与基于基本邻接矩阵建模的算法相比,电压偏移程度大幅降低,验证了动态邻接矩阵建模与图注意力机制在配电网优化调度中的独特适配性。文献[36]将 GAT 和门控循环单元组合,并将其加入 MASAC 中求解无功优化问题。相比于各种传统算法,所提算法在 33 节点与 123 节点系统上优化后的网损与电压偏移均为最低,收敛所需的回合数也更少。通过上述文献可以得知,GRL 能够充分利用系统拓扑信息,更好地求解配电网优化调度问题,优势主要体现在动态拓扑适应性强与调度结果更优,并且可拓展性良好。附录 D 表 D1 整

理了基于 GRL 的配电网优化调度研究的具体实验数据及与其他算法的对比情况。

### 3.2 基于 GRL 的配电网动态重构研究

配电网重构通过控制开关的状态来改变配电网的拓扑结构，核心目标为降低网损和电压偏移。从时间层面划分，配电网重构分为面向中长期规划的静态重构和面向实时优化的动态重构。其中，动态重构面临频繁的负荷变化以及 DG 出力不确定性，且开关的开合变化属于离散变量，数学优化求解此类问题的时间成本高，实时性差。因此，许多学者利用了 DRL 的快速决策能力进行求解。

文献[37]提出了一种基于 DQN 的配电网重构方法，采用了一种基于回路的编码方式进行动作筛选，为许多配电网重构研究提供了重要参考。文献[38]构建了配电网重构双层优化模型，上层采用 DQN 优化各个时段内的配电网拓扑结构，下层在确定的拓扑下对智能软开关进行调节。文献[39]通过分级决策的方式对动态重构的动作空间进行降维，再采用 DQN 进行优化。从上述文献中可以看出，DRL 在动态重构问题中的收敛速度整体上较慢，其原因主要是重构的动作数量众多，寻优难度大，即便这些文献采取了自调参和动作空间降维等措施。针对该问题，文献[40]提出了一种批约束 SAC 求解配电网重构，对经验回放进行了改进。离线算法利用经验数据进行训练，解决在线算法由于无经验初始状态所导致的训练不稳定问题，但容易陷入局部最优。针对该问题，文献[41]将训练过程分为两个阶段，离线阶段采用多智能体保守 Q 学习算法，利用部分经验数据先行试错，在线阶段则采用 DDPG，两种方式结合增强训练效果。文献[42]同样采用两阶段算法求解配电网重构，在探索安全动作的前提下提高性能，降低试错成本。

上述各种算法都能够较好地求解配电网重构问题，但存在收敛难度较大、不稳定等缺点。文献[41-42]的分段训练方式尽管具有较好的优化效果，但算法的复杂度大且对于先验数据具有较强的依赖。针对这些缺点，GRL 依靠充分提取图结构特征的优势实现对经验信息的高效利用，以此提高训练效率。文献[43]采用基于 GCN 的 DDPG 进行配电网重构的优化，在含高渗透率光伏的 33 节点系统上的仿真仿真结果显示所提算法能够较快地收敛到近似最优解。相比于 DDQN，所提算法优化后的平均网损与电压偏移更低，并且仅比基于数学优化的理论最优解高出 1.17%和 0.98%，表明了 GRL 能够更明显地提高配电网的经济性与可靠性。

文献[44]提出了一种结合了 PPO 和 GCAPCN 的配电网重构方法，对比基于传统神经网络的算法，所提出的算法收敛性明显更强，并且相比于离散二进制粒子群算法，所用的决策时间明显更短。该文献未与 DRL 优化前后的网损与电压偏移指标进行详细对比，但奖励函数收敛情况的显著差异体现出了图数据处理对拓扑重构决策的促进作用。文献[45]将 GCN、门控循环单元与基于内点策略优化的安全 DRL 结合。所提与遗传算法相比，优化后总运行成本降低的同时无电压电流越限现象发生，说明所提算法在得到较优结果的同时发挥出了安全特性。该文献未对比不同算法下的具体网损值和电压偏移，但对比了 GRL 与传统算法在参数丢失场景下的优化结果，验证了算法的鲁棒性。

上述文献均采用规模较小的配电网进行仿真，验证了 GRL 在收敛性与寻优精度方面的优势，但并未测试其对较大规模配电网的适应性。对此，文献[46]将双向图神经网络与 SAC 结合以实现配电网重构与无功优化的协同，并在 33 节点、69 节点与 123 节点配电网均进行了仿真。双图网络分别将图结构的节点信息与边信息作为聚合中心进行特征聚合，并将提取的特征进行整合。相比普通的 SAC 算法，所提出的算法在 33 节点系统上收敛所需回合数减少约 35%，优化后的网损与电压偏移明显降低，在 69 节点与 123 节点系统上网损与电压偏移同样更低，验证了算法的可拓展性以及比传统算法更突出的性能。文献[47]对配电网进行了多时间尺度优化，将多类型设备的计算压力分摊到不同时间尺度的终端，其中长时间尺度阶段采用了基于 GCN 的 DDQN 对配电网的重构开关与离散无功补偿设备进行协同调节。与传统 DDQN 相比，所提的算法在 33 节点系统上收敛后的奖励值提高，电压偏移与网损显著降低。此外，虽然所提算法优化后的网损比模型驱动方法高，但电压偏移程度减少近一半，安全性更高。同时，所提出的算法能够通过灵活调节离散设备以减少 DG 削减量，有利于促进新能源消纳。文献[48]将 GCN 与 MASAC 结合，在 8500 节点特大型系统上优化后的网损相比普通 MASAC 降低了 11.1%，进一步验证了算法在大型系统中的性能。该文献未考虑电压偏移，但对电压的三相不平衡度进行了优化，同样提升了系统安全性。从上述文献可以看出，GRL 与 DRL 在配电网重构中应用的主要区别在于，前者除了将节点电压、功率等作为智能体的输入，还将配电网的图结构特征也作为观测数据的一部分。附录 D 表 D2 整

理了基于 GRL 的配电网重构研究的具体实验数据及与其他算法的对比情况。

### 3.3 基于 GRL 的配电网故障恢复研究

优化调度与动态重构都基于正常运行状态进行优化,而针对故障状态进行应急优化同样也是重要的优化问题。故障切除会使得部分负荷因与源端失去连接而停电,需要调动联络开关、储能、备用发电机和维修队等应急资源恢复供电。如今,在大量 DG 和 EV 导致的系统状态随机性影响下,故障恢复问题变得更为复杂多维<sup>[49]</sup>。

配电网故障恢复问题对实时性的要求极高,因此适合采用数据驱动方法进行求解,其中 DRL 方法被广泛研究。文献[50]提出了一种基于 MADDPG 的配电网恢复重构策略,其考虑了配电网与通信系统之间的联系,将无人机基站作为智能体对配电网故障后的恢复进行决策。该文献所提出的算法收敛性较弱,难以适应较大规模的系统。文献[51]提出了基于竞争 DQN 的主动配电网故障恢复策略,以降低故障后的经济损失和恢复正常运行后的经济成本为优化目标构建了高恢复力决策模型,但在面对大规模系统时,该方法的求解效率较低。文献[52]采用 MADQN 求解配电网负荷恢复问题,发挥了多智能体的优势,提高了算法对大型系统的求解能力。但从其仿真结果中可以看出其训练速度较慢,且稳定性较差,说明在拓扑结构变化的情况下,DRL 依然表现出了局限性。对此,许多学者开始采用 GRL 算法进行求解。

文献[53]将 GCN 与 TD3 算法结合,并融入了元学习,提高了训练的效率与迁移能力。该文献仅在规模较小的 37 节点系统仿真,而故障恢复能力对于较大型的配电网也非常关键,难度也更高,因此不少文献倾向于在更大的系统中对算法的故障恢复能力进行衡量。文献[54]采用基于 GCN 的 DQN 求解故障恢复问题,在 69 节点系统上进行仿真。该算法优化后的平均负荷恢复率与传统 DRL 相近,但当拓扑结构变化后,所提算法的负荷恢复率仍然达到较高的值,且显著高于传统算法,验证了 GRL 的动态拓扑适应能力。文献[55]将 GCAPCN 与 PPO 算法结合求解主动配电网故障恢复问题,发挥了高维的特征提取能力,使融合了图结构特征的主动配电网的状态信息得到更充分的利用。与普通 PPO 算法相比,训练相同步数后,所提算法在 13 节点、34 节点、123 节点系统上的奖励值均更高,且优化后在 13 节点与 34 节点系统上的负荷恢复量显著提高,展现出了更强的寻优能力。文献[56]考虑了交

通道损坏的问题,采用了数据-模型混合驱动方法以协调移动应急资源的调度决策,对配电网与路网进行抢修。其中,数据驱动部分采用基于 GAT 的多智能体竞争深度双 Q 网络算法(dueling DDQN, D3QN)对移动应急资源进行路由调度。相比于传统 DRL,所提算法在 33 节点配电网与 12 节点交通网所组成的仿真环境中收敛所需回合数增加,但优化后的经济损失、负荷削减量和道路损坏持续时间均更低,恢复效果更明显。在由 69 节点配电网与 24 节点交通网组成的更大的仿真环境中,所提算法各项优化结果同样更优。文献[57]将 GAT 与 SAC 结合,在 148 节点系统上优化后的负荷恢复率比传统 DRL 更高,开关动作次数也更少,且没有电压越限发生,验证了算法的优化精度及在较大系统上的适应性。文献[58]提出了基于记忆增强 GRL 的故障恢复方法,将长短时记忆网络的时序记忆能力与算法的拓扑感知能力进行了结合。在 123 节点系统上的仿真结果显示,所提算法与仅基于长短时记忆网络的算法和传统算法相比,收敛速度较慢,但收敛后的奖励值明显更高。同时,与传统算法相比,所提算法优化后的负荷恢复量也提高了约 11%。文献[59]将 GCN 与 MADQN 结合,采用基于开关分区的图简化方法缩小图结构的规模,并在 123 节点与 8500 节点系统中验证这种方式能够提高算法对大型系统的适应能力。在 8500 节点的系统中,所提出的算法仍然能够完整恢复所有负荷,而 MADQN 与 DQN 分别只能恢复 92.32%和 73.53%。通过上述文献的实验结果可以看出,GRL 算法在恢复强度与对大系统的适应能力上都要优于传统算法。附录 D 表 D3 整理了基于 GRL 的配电网故障恢复研究的具体实验数据及与其他算法的对比情况。

### 3.4 基于 GRL 的电力-交通耦合网协同优化研究

近年来,EV 数量大规模增长,减少碳排的同时也加深了交通系统对电力系统的影响,主要体现在大规模 EV 的无序充电导致的负荷不确定性增加、电压波动加剧,对配电网的影响尤为明显,因此针对电力-交通耦合网进行优化也十分有必要。由于交通路况、充电价格和用户充电习惯等因素的影响,充电行为表现出明显的随机性。大规模 EV 无序充电会影响用户的利益,同时也会加剧配电网的负荷波动,可能导致充电站所在节点的电压过低,对配电网的安全性造成严重威胁<sup>[60]</sup>。因此,电力-交通耦合网协同优化是一个热门的研究课题。

DRL 算法因其擅长处理复杂随机优化问题而

被许多学者用于解决电力-交通耦合网协同优化。文献[61]提出了一种基于 DDPG 的充放电智能管理方法，在满足用户充电需求的前提下降低充电成本、减少充电时间，构建了动态能量边界模型以调节 EV 充放电功率边界，有效优化充电站数量受限情况下的充电策略。文献[62]提出了一种基于 TD3 的有序充电优化方法，对各时段的充电优先级进行评定，以充电费用最少和负荷标准差最小为目标进行充电行为优化。文献[63]提出了一种基于 MAPPO 的充电站定价博弈策略，降低用户充电成本的同时平衡充电需求的分布，减少交通拥堵，且可以促进风电的消纳，提高配电网的成本效率。

上述文献并未对配电网与交通路网的拓扑特性进行利用，容易忽略数据之间的空间耦合性<sup>[64]</sup>。拓扑结构对于配电网与交通网的协同运行而言同样是重要信息，因此，采用 GRL 解决电力-交通耦合网协同优化问题是一大研究热点。文献[65]提出了一种双时间尺度的 EV 充电引导方法，其中，长时间尺度以降低配电网的发电成本与购电成本为目标求解最优潮流，并根据求解结果计算节点电价；短时间尺度采用基于 GAT 的多智能体 AC 进行实时充电引导，也有效降低了用户的充电等待时间与充电费用。与传统 DRL 相比，所提出的算法收敛性更好，收敛后的策略更稳定。文献[66]将 EV 的行驶和充电行为表示为电力-交通耦合网络中的节点信息，利用 GCN 提取数据特征，并采用彩虹算法进行训练，实现对多用户的实时充电引导，对包含数百个节点的大系统的求解效果突出。相比于多种传统 DRL，所提出的算法优化后的平均充电成本均降低了 10% 以上。

上述文献都能够有效降低充电成本，但仅从用户经济性角度出发，未充分考虑充电行为对配电网的影响。对此，文献[67]采用基于 GAT 的竞争 DQN 进行充电引导优化，以降低配电网电压偏移为目标。与最短路径算法相比，所提出的算法优化后充电站所在节点的最低电压升高了约 0.01，越下限的风险降低。该文献仅从配电网的角度出发，本质上仍然是单目标优化，而电力-交通耦合网协同优化问题涉及多个主体的不同利益，因此有必要对多个目标进行综合考虑。对此，文献[68]建立了包含 EV、充电站、道路和配电网的多目标优化模型，采用基于 GCN 的 DQN 进行快速充电引导，减少用户充电时间的同时也提高了充电站利用均衡度及配电网的电压稳定性。与传统的就近推荐算法相比，所提算法的优化效果更佳，用户充电时间、充电站服务

不均衡度均减少，交通拥挤程度、配电网电压偏移程度相近，并且在 EV 数量增加时仍然能够保持较好的优化效果。文献[69]基于图论将 EV 之间的联系转化为动态网络图模型，并采用基于 GAT 的 MADQN 进行充电引导以达到降低充电成本、提高站点利用率的目标。该文献在 33 节点配电网与 108 节点交通网组成的耦合系统中进行仿真，普通的 MADQN 算法难以收敛，而所提出的算法收敛性良好。相比于最短路径算法，所提算法优化后的用户平均充电排队时间减少了约 10min，负荷峰谷差也有所降低，节省用户充电成本的同时也提高了配电网的运行稳定性。上述文献并未关注到电网数据与交通网数据从属于不同管理部门，涉及到十分重要的数据隐私性问题，将用户信息直接与其他 EV 或电网信息进行交互会使得用户面临严重的隐私泄露风险<sup>[70-71]</sup>。针对这些问题，文献[72]建立了用户里程焦虑模型以模拟用户在产生充电需求时对前往充电站的里程所产生的焦虑值，并采用不同 GNN 分别处理站、网信息和用户信息以保护用户隐私。将所提算法与最小后悔值算法、多目标加权平均算法和最短路径算法进行比较。在 33 节点配电网与 25 节点交通网组成的耦合系统中，所提算法的充电成本、交通拥堵程度与配电网电压偏移率等指标均为最低，所产生的用户里程焦虑明显更小，有助于提高用户接受充电引导的意愿，产生正反馈。对于大规模的调度，EV 的决策过程之间往往存在一些共性，但上述文献未利用到这一点，对不同用户进行一些重复处理，影响算法的训练效率。对此，文献[73]将 GCN 与基于参数共享的 DDPG 结合求解 EV 充放电调度问题，将碳排放强度纳入观测空间并作为路由及充放电决策的依据。利用参数共享机制实现智能体之间的策略互通。相比于普通 DDPG 和未进行参数共享的算法，所提算法收敛后的策略明显更优，能够更有效地降低调度成本与系统整体碳排放。从上述文献可以得知，现有的图结构处理机制在 GRL 算法中已经发挥出较为明显的作用，使智能体能够更加充分地对电力-交通耦合网的信息进行利用，使训练行为更高效、协同多主体利益的能力更加突出。附录 D 表 D4 整理了基于 GRL 的电力-交通耦合网协同优化的具体实验数据及与其他算法的对比情况。

表 1 对本节所综述的几种基于 GRL 的配电网优化问题的核心目标、典型算法以及算法的主要优势进行了简单总结。各文献的创新点与不足之处详见附录 D 表 D5。

表 1 基于 GRL 的配电网优化主要特点总结  
Table 1 Summary of GRL based distribution network optimization methods

场景	文献	主要目标	典型算法		GRL 主要优势
			GNN 部分	DRL 部分	
配电网优化调度	[30-36]	运行成本、电压偏移	GCN、GAT、STGNN	DDPG、TD3、SAC 等	能自适应捕捉节点有功功率和无功功率与电压和成本之间的灵敏度关系，主要体现在优化后的运行成本和电压偏移更低
配电网动态重构	[43-48]	网损、电压偏移、开关动作频次	GCN、STGNN 等	DQN、DDQN、MASAC 等	对拓扑变化适应性强，主要体现在重构后网损和电压偏移更低且开关动作次数控制得更好
配电网故障恢复	[53-59]	负荷恢复率	GCN、GAT、STGNN 等	PPO、MADQN、MAD3QN 等	对拓扑变化的适应性强，能够捕捉待恢复负荷与线路开合、DG 出力等因素之间的相关性，主要体现在负荷恢复率更高
电力-交通耦合网协同优化	[65-69,72,73]	EV 充电成本、充电有序性	GCN、GAT	DQN、MADQN 等	对整体交通路况和充电需求具有更全面的感知，主要体现在优化后的 EV 充电成本更低以及充电更有序，使得充电负荷更均衡

## 4 总结与展望

### 4.1 GRL 在配电网优化中的应用总结

根据前文整理的文献，GRL 在配电网优化问题中的应用主要有以下几个特点：

1) 从网络类型的选取来看，大部分文献采用结构简单的 GCN 或 GAT，并且能够取得较好的优化效果，说明基本图处理机制能够显著提高算法对图结构数据的适应能力，且泛用性较强；而 GCAPCN、STGNN 等针对性较强的类型能够进一步增强优化效果，但训练的时间成本也更高。

2) 从算法类型的选取来看，在配电网的优化调度研究中，决策变量多为连续变量，因此大部分文献将 GNN 与基于策略或价值-策略的算法进行结合；而与基于价值的算法的结合则较多地应用于配电网重构、EV 充电路径规划等离散决策问题。

3) 采用了多智能体算法的文献大多会基于较大的系统进行仿真验证，如节点数量超过 100 的配电网或电力-交通耦合网，这说明对配电网进行分布式的图结构建模与优化同样可行，MAGRL 对大型系统具有显著的适应性。

4) 配电网动态重构、故障恢复等问题涉及到拓扑的变化，此时，算法除了对图信息进行直接利用，还需要发挥出响应拓扑结构时空变化特性的能力，因此部分学者运用了 GCAPCN、STGNN 等能提取复杂特征的 GNN。

本文从所综述的文献当中总结出了 GRL 算法的优缺点以及基于 GRL 的配电网优化方法的创新点与不足之处，详见附录 D.2。

### 4.2 GRL 在配电网优化中的主要挑战

针对前文所总结的 GRL 在配电网优化中存在的局限性，未来可以针对以下几方面的挑战进行深入研究：

1) 前文提到 GNN 的计算速度慢于部分传统神经网络，尤其是当图结构规模和参数空间扩大时，

节点的状态依赖于越来越多其他节点，需要处理的空间关联性越来越复杂，计算效率进一步降低。同时，GRL 在动态拓扑场景下的收敛时间也比较长。降低计算复杂度最常见的方法是进行分布式处理，即采用多智能体，但图结构规模扩大仍会导致智能体过多、并行计算压力大、收敛难度大等问题。针对上述问题，未来需要研究能够有效提高 GNN 处理大规模与动态图结构能力的方法。

2) 违反约束条件的风险也是制约 GRL 在实际中应用的一大因素。配电网的优化问题中存在各种约束条件，如电压约束、电流约束、拓扑结构约束等，这些约束本质上是对动作和状态进行限制，常采用的做法之一是筛选动作，使智能体可选择动作均满足约束条件，但这种方式仅适用于动作空间离散的情况。此外，在奖励函数中设置违反约束条件的惩罚项也是常见的做法。这种方式将数学模型中的目标函数与约束条件进行集成，使得智能体更加容易陷入局部最优解，并且训练结束后仍存在违反约束的风险，缺乏硬性约束能力。

3) 现实中读取到的数据存在噪声影响、完备性差和通信延迟等问题，理想化的算法设计存在鲁棒性不足的风险，未来需要在数据处理方面对 GRL 算法进行改进，其核心在于提高数据质量，可行的解决方式是将 GRL 与表征学习算法相结合，例如采用自编码模型或生成模型对含有噪声、缺损及延迟的数据进行处理。

4) 从现实层面出发，除了前文总结的几个理论局限性之外，GRL 在配电网优化中的应用还存在信息隐私方面的挑战。算法需要对用户负荷、DG 发电情况、EV 充电情况等多方数据进行读取，这些数据可能隐性包含了电力用户的隐私信息，如居民生活习惯、企业生产计划等<sup>[84]</sup>。采用 GRL 进行优化时，由于额外加入拓扑信息观测，隐私信息也随之增加。在电力-交通耦合系统当中，EV 用户的信息更是会直接与电网、路网及充电运营商进行交

互，隐私问题更加明显。

本文结合现有的前沿算法及相关研究成果对上述挑战的解决思路进行了分析与展望，并举例阐述了部分潜在可行的方案，详见附录 E。

### 4.3 GRL 在配电网优化中的应用前景展望

GRL 的独特性能使其在配电网的运行优化领域具有广阔的应用前景。除了本文综述的 4 个配电网优化问题，未来还可以在以下几个问题中进一步探索 GRL 的应用。

#### 4.3.1 电能质量优化问题

随着 DG 大规模接入与配电网电力电子化程度的不断提高，电能质量治理问题日趋关键，主要包括三相不平衡抑制、谐波畸变抑制、电压控制等，对于改善用户用电质量、提高电网运行稳定性与电能管控水平具有重要作用，对实时性的要求也比较高。在新型电力系统背景下，电能质量优化问题日益复杂，且受限于通信条件与设备异构化影响，智能体的信息观测范围难以定夺，存在数据处理不平衡的问题，影响电能质量优化的效果<sup>[85-86]</sup>。针对上述问题，未来可以尝试研究 GRL 在电能质量优化控制中的应用，结合图结构信息提高数据质量及智能体对观测信息的利用效率，提高电能质量优化的效果。例如，可以采用异构图的形式对配电网进行图结构建模，或者将三相分别建模为不同的节点和边，使得 GRL 读取基本节点信息的同时也能够捕捉到设备间以及相间的相关性，对 GRL 的训练产生更具体的指导作用，提高电能质量治理水平。

#### 4.3.2 需求侧响应问题

需求侧响应是鼓励用户参与配电网调节的措施，通过电价影响用户的用电行为及电力资源的分配，间接改变用户负荷，进而起到削峰填谷、平衡电压、提高 DG 消纳等作用，具有较强的灵活性。在大规模 DG 和 EV 接入的背景下，配电网的负荷波动性增加，针对随机性强的负荷特性对电价进行调整实现需求侧响应是改善配电网运行状况的重要手段。现有的需求侧响应多数基于模型驱动或传统 DRL 算法求解，未来可以尝试发挥 GRL 在需求侧响应中的作用。例如，在图结构建模过程中将电价信息加入到节点信息当中，利用 GAT 自适应地捕捉不同节点的电价与负荷之间的相关性，进而对电价调节产生指导作用。

#### 4.3.3 故障场景的路径-资源联合规划问题

在配电网故障恢复问题当中，应急资源的种类多样，其中包含以车辆为载体的可移动应急资源，如移动储能、维修车队等，使得故障恢复策略更加

复杂灵活。前文提到的文献[56]虽然采用 GRL 对移动应急资源进行了调度，但其将路径规划和其他资源调度的决策过程进行分层求解，全局最优性有待提高。路径决策与资源分配之间需要较强的协同性，尤其是在车队规模、储能容量和通信等资源受限的条件下，因此有必要对其进行联合规划。路径-资源的联合决策是一个混合整数规划问题，传统算法的求解效率较低，不利于灾后快速恢复。GRL 特有的空间特征学习能力不仅能用于诸如电网、路网等实际拓扑所形成的物理连接图结构，还能够用于处理由各要素之间隐含的相关性所形成的耦合关联图结构。具体而言，可以将故障线路、资源供应点和可选路径等作为节点、路况和资源配额约束等作为节点信息来构建图结构，并设计路径-资源联合动作空间以及用于对耦合关联图进行编码的 GNN，利用全局奖励进行训练，使得算法能够收敛到整体恢复能力更强的路径-资源联合策略。

### 4.4 GRL 工程化落地展望

前文从理论层面分析了 GRL 在配电网优化中应用的可行性及挑战，其应用对于提高电网智能化水平具有重要意义。然而，GRL 的工程化落地还需要克服一些现实场景中的挑战，其中核心问题包括如下几点：

1) 可操作化。GRL 需要经过训练与测试后才能投入实际工程中运行，而训练阶段需要试错，不允许在实际系统中进行，因此调度中心需要开发出测试系统或提供经验数据供 GRL 进行训练与测试。该挑战的关键在于测试系统的拟真程度与经验数据的充足和完备，因此可以结合工业大数据、数字孪生等前沿技术进行研究部署。此外，测试阶段需要有真实系统中的应用结果作为支撑，因此也需要做好试点与示范工程的部署工作。

2) 规范化。为了保证配电网运行管理的规范性，GRL 算法需要与现有的 SCADA 和 EMS 等成熟的管理系统进行对接，这要求对算法的输入数据和输出数据进行标准化。GRL 的输入数据中包含拓扑数据，在算法中以邻接矩阵形式参与计算，因此调度中心也需要针对这一特点开发统一的拓扑识别模块以及相关计算程序，以增强算法应用过程中的数据规范性。

3) 安全化。前文提到了隐私保护问题的算法改进方向，但是仅在算法层面改进依然存在信息安全挑战。前文提到的联邦 GRL 框架采用去中心化的训练方式确保隐私安全，但各个数据处理单元依旧需要收集大量本地数据并存储以供经验回放，意

意味着这些数据虽然不会被调度中心与运营商直接读取,但需要上传至本地数据库,同样具有受攻击的风险,因此需要设置多道“防线”,避免数据交互的同时也要增强数据库的信息安全防御水平。

附录见本刊网络版(<http://www.dwjs.com.cn/CN/1000-3673/current.shtml>)。

## 参考文献

- [1] 朱介北,徐思昉,李炳森,等.基于电网专家策略模仿学习的新型电力系统实时调度[J].电网技术,2023,47(2):517-528.  
ZHU Jiebei, XU Siyang, LI Bingsen, et al. Real-time security dispatch of modern power system based on grid expert strategy imitation learning[J]. Power System Technology, 2023, 47(2): 517-528(in Chinese).
- [2] BEHBAHANI M R, JALILIAN A, BAHMANYAR A, et al. Comprehensive review on static and dynamic distribution network reconfiguration methodologies[J]. IEEE Access, 2024, 12: 9510-9525.
- [3] LALLJITH S, FLEMING I, PILLAY U, et al. Applications of flower pollination algorithm in electrical power systems: a review[J]. IEEE Access, 2022, 10: 8924-8947.
- [4] MISHRA A, TRIPATHY M, RAY P. A survey on different techniques for distribution network reconfiguration[J]. Journal of Engineering Research, 2024, 12(1): 173-181.
- [5] AL-OGAILI A S, HASHIM T J T, RAHMAT N A, et al. Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: challenges and recommendations[J]. IEEE Access, 2019, 7: 128353-128371.
- [6] 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 Guanzhong, YANG Shengchun, GUO Xiaorui, 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(in Chinese).
- [7] 张玉敏,孙鹏凯,叶平峰,等.基于CNN的配电网快速重构方法[J].智慧电力,2022,50(11):112-118.  
ZHANG Yumin, SUN Pengkai, YE Pingfeng, et al. Fast reconfiguration method of distribution network of based on convolutional neural network[J]. Smart Power, 2022, 50(11): 112-118(in Chinese).
- [8] ZHAO B, GUO C X, CAO Y J. A multiagent-based particle swarm optimization approach for optimal reactive power dispatch[J]. IEEE Transactions on Power Systems, 2005, 20(2): 1070-1078.
- [9] KHAN N H, WANG Yong, TIAN De, et al. Design of fractional particle swarm optimization gravitational search algorithm for optimal reactive power dispatch problems[J]. IEEE Access, 2020, 8: 146785-146806.
- [10] 徐岩,张荟,马天祥,等.含分布式电源的配电网故障紧急恢复与抢修协调优化策略[J].电力系统自动化,2021,45(22):38-46.  
XU Yan, ZHANG Hui, MA Tianxiang, et al. Coordinated optimization strategy of fault emergency recovery and repair for distribution network with distributed generators[J]. Automation of Electric Power Systems, 2021, 45(22): 38-46(in Chinese).
- [11] WANG Ning, LI Bo, DUAN Yan, et al. A multi-energy scheduling strategy for orderly charging and discharging of electric vehicles based on multi-objective particle swarm optimization[J]. Sustainable Energy Technologies and Assessments, 2021, 44: 101037.
- [12] WANG Chaoxue, ZHAO Anjun, DONG Hui, et al. An improved immune genetic algorithm for distribution network reconfiguration [C]//Proceedings of 2009 International Conference on Information Management, Innovation Management and Industrial Engineering. Xi'an, China: IEEE, 2009: 218-223.
- [13] QIAN Cheng, WANG Aiyuan. Distribution network reconfiguration based on improved differential evolution ant colony algorithm[C]//Proceedings of 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). Xi'an, China: IEEE, 2022: 234-240.
- [14] KANG Qi, WANG Jiabao, ZHOU Mengchu, et al. Centralized charging strategy and scheduling algorithm for electric vehicles under a battery swapping scenario[J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(3): 659-669.
- [15] LI Yuanzheng, YU Chaofan, SHAHIDEHPOUR M, et al. Deep reinforcement learning for smart grid operations: algorithms, applications, and prospects[J]. Proceedings of the IEEE, 2023, 111(9): 1055-1096.
- [16] MASSAOUDI M S, ABU-RUB H, GHRAYEB A. Navigating the landscape of deep reinforcement learning for power system stability control: a review[J]. IEEE Access, 2023, 11: 134298-134317.
- [17] CAO Di, HU Weihao, ZHAO Junbo, et al. Reinforcement learning and its applications in modern power and energy systems: a review[J]. Journal of Modern Power Systems and Clean Energy, 2020, 8(6): 1029-1042.
- [18] 朱思婷,管霖,郭梦轩,等.基于异质边图注意力网络的电力系统振荡评估模型[J].电网技术,2022,46(7):2581-2592.  
ZHU Siting, GUAN Lin, GUO Mengxuan, et al. Power system oscillation evaluation model based on heterogeneous edge graph attention network[J]. Power System Technology, 2022, 46(7): 2581-2592(in Chinese).
- [19] LV Zhe, WANG Bin, GUO Qinglai, et al. Short-term voltage stability assessment based on heterogeneous edge-integrated graph attention network[J]. IEEE Transactions on Power Systems, 2025, 40(1): 421-434.
- [20] 吴博,梁循,张树森,等.图神经网络前沿进展与应用[J].计算机学报,2022,45(1):35-68.  
WU Bo, LIANG Xun, ZHANG Shusen, et al. Advances and applications in graph neural network[J]. Chinese Journal of Computers, 2022, 45(1): 35-68(in Chinese).
- [21] 侯磊,刘金环,于旭,等.图神经网络研究综述[J].计算机科学,2024,51(6):282-298.  
HOU Lei, LIU Jinhuan, YU Xu, et al. Review of graph neural networks[J]. Computer Science, 2024, 51(6): 282-298(in Chinese).
- [22] 苏小玲,陈来军,赵超凡,等.计及能源自洽率和共享氢储能的电-氢-交通耦合配电网低碳经济运行[J].高电压技术,2024,50(6):2424-2432.  
SU Xiaoling, CHEN Laijun, ZHAO Chaofan, et al. Low carbon economic dispatch for electricity hydrogen transport coupling network considering self-sufficiency and shared hydrogen storage[J]. High Voltage Engineering, 2024, 50(6): 2424-2432(in Chinese).
- [23] 肖传亮,彭克,陈佳佳,等.基于多利益主体划分的配电网互动博弈电压优化控制[J/OL].电网技术,2024:1-14[2025-7-10].  
<https://doi.org/10.13335/j.1000-3673.pst.2024.1783>.  
XIAO Chuanliang, PENG Ke, CHEN Jijia, et al. Multiple stakeholder partition based interactive-game voltage control for distribution networks[J/OL]. Power System Technology, 2024: 1-14 [2025-7-10]. <https://doi.org/10.13335/j.1000-3673.pst.2024.1783>(in Chinese).

- [24] 杨龙月, 任焯辰, 蔡智鹏, 等. 高光伏渗透率配电网电压控制策略研究综述[J]. 电网技术, 2024, 48(12): 5056-5070.  
YANG Longyue, REN Xuanchen, CAI Zhipeng, et al. Review on voltage control strategies for high PV penetration distribution networks[J]. Power System Technology, 2024, 48(12): 5056-5070(in Chinese).
- [25] 冯斌, 胡轶婕, 黄刚, 等. 基于深度强化学习的新型电力系统调度优化方法综述[J]. 电力系统自动化, 2023, 47(17): 187-199.  
FENG Bin, HU Yijie, HUANG Gang, et al. Review on optimization methods for new power system dispatch based on deep reinforcement learning[J]. Automation of Electric Power Systems, 2023, 47(17): 187-199(in Chinese).
- [26] LIU Zhicheng, LIU Yipeng, XU Hao, et al. Dynamic economic dispatch of power system based on DDPG algorithm[J]. Energy Reports, 2022, 8: 1122-1129.
- [27] 陈实, 朱亚斌, 刘艺洪, 等. 基于世界模型深度强化学习的含风电电力系统低碳经济调度[J]. 电网技术, 2024, 48(8): 3143-3154.  
CHEN Shi, ZHU Yabin, LIU Yihong, et al. Low-carbon economic dispatch of wind-containing power systems based on world model deep reinforcement learning[J]. Power System Technology, 2024, 48(8): 3143-3154(in Chinese).
- [28] 黄冬梅, 余京朋, 崔承刚, 等. 多时间尺度深度强化学习光储配电网电压优化[J]. 中国电机工程学报, 2025, 45(17): 6709-6722.  
HUANG Dongmei, YU Jingpeng, CUI Chenggang, et al. Voltage optimization for PV-ES distribution network with multi-timescale deep reinforcement learning method[J]. Proceedings of the CSEE, 2025, 45(17): 6709-6722(in Chinese).
- [29] 徐业琰, 姚良忠, 廖思阳, 等. 基于多智能体 Actor-Double-Critic 深度强化学习的源-网-荷-储实时优化调度方法[J]. 中国电机工程学报, 2025, 45(2): 513-527.  
XU Yeyan, YAO Liangzhong, LIAO Siyang, et al. Real-time optimal dispatch method of source-grid-load-storage based on multi-agent actor-double-critic deep reinforcement learning[J]. Proceedings of the CSEE, 2025, 45(2): 513-527(in Chinese).
- [30] 陈俊斌, 余涛, 潘振宁. 面向主动配电网实时优化调度的图强化学习方法[J]. 控制理论与应用, 2024, 41(6): 999-1008.  
CHEN Junbin, YU Tao, PAN Zhenning. Graph reinforcement learning for real-time optimal dispatch of active distribution network[J]. Control Theory & Applications, 2024, 41(6): 999-1008(in Chinese).
- [31] MU Chaoxu, LIU Zhaoyang, YAN Jun, et al. Graph multi-agent reinforcement learning for inverter-based active voltage control [J]. IEEE Transactions on Smart Grid, 2024, 15(2): 1399-1409.
- [32] GE Leijiao, LI Jingjing, HOU Luyang, et al. Autonomous voltage regulation for smart distribution network with high-proportion PVs: a graph meta-reinforcement learning approach[J]. IEEE Transactions on Sustainable Energy, 2025, 16(4): 2768-2781.
- [33] XING Qiang, CHEN Zhong, ZHANG Tian, et al. Real-time optimal scheduling for active distribution networks: a graph reinforcement learning method[J]. International Journal of Electrical Power & Energy Systems, 2023, 145: 108637.
- [34] CHEN Yongdong, LIU Youbo, ZHAO Junbo, et al. Physical-assisted multi-agent graph reinforcement learning enabled fast voltage regulation for PV-rich active distribution network[J]. Applied Energy, 2023, 351: 121743.
- [35] LUO Fengzhang, WANG Shengyuan, LV Yunqiang, et al. Domain knowledge-enhanced graph reinforcement learning method for Volt/Var control in distribution networks[J]. Applied Energy, 2025, 398: 126409.
- [36] HU Daner, LI Zichen, YE Zhenhui, et al. Multi-agent graph reinforcement learning for decentralized Volt-VAR control in power distribution systems[J]. International Journal of Electrical Power & Energy Systems, 2024, 155: 109531.
- [37] WANG Beibei, ZHU Hong, XU Honghua, et al. Distribution network reconfiguration based on NoisyNet deep Q-learning network[J]. IEEE Access, 2021, 9: 90358-90365.
- [38] 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 Lei, WU Yi, ZHANG Tao, 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(in Chinese).
- [39] 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 Zihan, GAO Hongjun, GAO Yiwen, 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(in Chinese).
- [40] GAO Yuanqi, WANG Wei, SHI Jie, et al. Batch-constrained reinforcement learning for dynamic distribution network reconfiguration[J]. IEEE Transactions on Smart Grid, 2020, 11(6): 5357-5369.
- [41] JIANG Siyuan, GAO Hongjun, WANG Xiaohui, et al. Deep reinforcement learning based multi-level dynamic reconfiguration for urban distribution network: a cloud-edge collaboration architecture [J]. Global Energy Interconnection, 2023, 6(1): 1-14.
- [42] GAO Hongjun, WANG Renjun, HE Shuaijia, et al. A cloud-edge collaboration solution for distribution network reconfiguration using multi-agent deep reinforcement learning[J]. IEEE Transactions on Power Systems, 2024, 39(2): 3867-3879.
- [43] ZHAN Hua, JIANG Changxu, LIN Zheng. A novel graph reinforcement learning-based approach for dynamic reconfiguration of active distribution networks with integrated renewable energy[J]. Energies, 2024, 17(24): 6311.
- [44] JACOB R A, PAUL S, LI Wenyuan, et al. Reconfiguring unbalanced distribution networks using reinforcement learning over graphs[C]// Proceedings of 2022 IEEE Texas Power and Energy Conference (TPEC). College Station, TX, USA: IEEE, 2022: 1-6.
- [45] WU Minghe, HONG Lucheng, ZHU Jin, et al. Dynamic network reconfiguration in power distribution network based on spatial-temporal aware safe reinforcement learning[C]//Proceedings of 2024 3rd International Conference on Power Systems and Electrical Technology (PSET). Tokyo, Japan: IEEE, 2024: 887-894.
- [46] WANG Ruoheng, BI Xiaowen, BU Siqi. 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.
- [47] JIANG Changxu, GUO Chen, LIN Junchi, et al. Enhancing economic efficiency and operational stability of high-penetration renewable distribution networks: a multi-timescale coordinated optimization method leveraging multi-agent graph reinforcement learning[J]. Renewable Energy, 2026, 256: 124345.
- [48] LIU Hong, LI Qizhe, ZHANG Qiang, et al. Soft open points scheduling in unbalanced active distribution networks based on multi-agent graph reinforcement learning[J]. Sustainable Energy, Grids and Networks, 2025, 42: 101689.
- [49] 张家美, 孙凯, 李洪涛, 等. 考虑光-车-氢接入的新型城市配电网

- 规划研究综述与展望[J]. 高电压技术, 2024, 50(3): 1067-1079.  
ZHANG Jiamei, SUN Kai, LI Hongtao, et al. Review and prospect of future city distribution networks considering the integration of PV-EV-hydrogen[J]. High Voltage Engineering, 2024, 50(3): 1067-1079(in Chinese).
- [50] QI Xianglong, CHEN Jian, ZHAO Haoran, et al. Post-disaster distribution system restoration considering UAV-based communication recovery based on multi-agent reinforcement learning [C]//Proceedings of IECON 2023-49th Annual Conference of the IEEE Industrial Electronics Society. Singapore, Singapore: IEEE, 2023: 1-6.
- [51] 罗欣儿, 杜进桥, 田杰, 等. 基于深度强化学习的主动配电网高恢复力决策方法[J]. 南方电网技术, 2022, 16(1): 67-74.  
LUO Xiner, DU Jinqiao, TIAN Jie, et al. High resilience decision-making method of active distribution network based on deep reinforcement learning[J]. Southern Power System Technology, 2022, 16(1): 67-74(in Chinese).
- [52] VU L, VU T, VU T L, et al. Multi-agent deep reinforcement learning for distributed load restoration[J]. IEEE Transactions on Smart Grid, 2024, 15(2): 1749-1760.
- [53] FAN Bangji, LIU Xinghua, XIAO Gaoxi, et al. Enhancing adaptability of restoration strategy for distribution network: a meta-based graph reinforcement learning approach[J]. IEEE Internet of Things Journal, 2024, 11(14): 25440-25453.
- [54] 张沛, 陈玉鑫, 王光华, 等. 基于图强化学习的配电网故障恢复决策[J]. 电力系统自动化, 2024, 48(2): 151-158.  
ZHANG Pei, CHEN Yuxin, WANG Guanghua, et al. Fault recovery decision of distribution network based on graph reinforcement learning[J]. Automation of Electric Power Systems, 2024, 48(2): 151-158(in Chinese).
- [55] JACOB R A, PAUL S, CHOWDHURY S, et al. Real-time outage management in active distribution networks using reinforcement learning over graphs[J]. Nature Communications, 2024, 15(1): 4766.
- [56] 江昌旭, 周龙灿, 庄鹏威, 等. 基于数据-模型混合驱动方法的多类型移动应急资源优化调度策略[J/OL]. 电网技术, 2024: 1-21 [2025-07-16]. <https://doi.org/10.13335/j.1000-3673.pst.2024.1516>.  
JIANG Changxu, ZHOU Longcan, ZHUANG Pengwei, et al. Scheduling optimization strategy for multi-type mobile emergency resources based on a data-model hybrid driven approach[J/OL]. Power System Technology, 2024: 1-21[2025-07-16]. <https://doi.org/10.13335/j.1000-3673.pst.2024.1516>(in Chinese).
- [57] 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.  
HE Xiaolong, GAO Hongjun, WANG Renjun, et al. Fault recovery method of active distribution network based on graph deep reinforcement learning[J]. Power System Technology, 2025, 49(10): 4342-4352(in Chinese).
- [58] FAN Bangji, LIU Xinghua, XIAO Gaoxi, et al. A memory-based graph reinforcement learning method for critical load restoration with uncertainties of distributed energy resource[J]. IEEE Transactions on Smart Grid, 2025, 16(2): 1706-1718.
- [59] ZHAO Tianqiao, WANG Jianhui. Learning sequential distribution system restoration via graph-reinforcement learning[J]. IEEE Transactions on Power Systems, 2022, 37(2): 1601-1611.
- [60] 刘志坚, 戴景, 杨灵睿. 考虑电力-交通耦合网动态协调的EV集群灵活性挖掘与优化调度[J]. 电力系统自动化, 2024, 48(7): 127-137.  
LIU Zhijian, DAI Jing, YANG Lingrui. Flexibility mining and optimal scheduling for electric vehicle clusters considering dynamic coordination of power-transportation coupling network[J]. Automation of Electric Power Systems, 2024, 48(7): 127-137(in Chinese).
- [61] LI Hang, LI Guojie, LI Shidan, et al. Optimal EV charging scheduling considering the lack of charging facilities based on deep reinforcement learning[C]//Proceedings of 2023 8th Asia Conference on Power and Electrical Engineering (ACPEE). Tianjin, China: IEEE, 2023: 1825-1829.
- [62] 喻磊, 谈竹奎, 王扬, 等. 基于深度强化学习的电动汽车有序充电优化方法[J]. 南方电网技术, 2024, 18(12): 148-155.  
YU Lei, TAN Zhukui, WANG Yang, et al. Deep reinforcement learning based optimization method for ordered charging of electric vehicle[J]. Southern Power System Technology, 2024, 18(12): 148-155(in Chinese).
- [63] YANG Xiaoying, CUI Tianxiang, WANG Hongru, et al. Multiagent deep reinforcement learning for electric vehicle fast charging station pricing game in electricity-transportation nexus[J]. IEEE Transactions on Industrial Informatics, 2024, 20(4): 6345-6355.
- [64] 叶宇剑, 吴奕之, 胡健雄, 等. 城市电力-交通耦合系统的联合推演与协同优化: 研究综述、挑战与展望[J]. 中国电机工程学报, 2025, 45(11): 4144-4162.  
YE Yujian, WU Yizhi, HU Jianxiong, et al. Joint prediction and coordinated optimization of integrated urban power distribution and transportation systems: literature review, challenges and prospects[J]. Proceedings of the CSEE, 2025, 45(11): 4144-4162(in Chinese).
- [65] LI Yujing, SU Su, ZHANG Minghao, et al. Multi-agent graph reinforcement learning method for electric vehicle on-route charging guidance in coupled transportation electrification[J]. IEEE Transactions on Sustainable Energy, 2024, 15(2): 1180-1193.
- [66] XING Qiang, XU Yan, CHEN Zhong, et al. A graph reinforcement learning-based decision-making platform for real-time charging navigation of urban electric vehicles[J]. IEEE Transactions on Industrial Informatics, 2023, 19(3): 3284-3295.
- [67] CHEN Wenlei, LIU Junqi, CAO Junwei. Graph reinforcement learning method for electric vehicle charging recommendation considering voltage deviation[C]//Proceedings of 2024 IEEE International Conference on Energy Internet (ICEI). Zhuhai, China: IEEE, 2024: 608-613.
- [68] 袁红霞, 张俊, 许沛东, 等. 基于图强化学习的电力交通耦合网络快速充电需求引导研究[J]. 电网技术, 2021, 45(3): 979-986.  
YUAN Hongxia, ZHANG Jun, XU Peidong, et al. Fast charging demand guidance in coupled power-transportation networks based on graph reinforcement learning[J]. Power System Technology, 2021, 45(3): 979-986(in Chinese).
- [69] 江昌旭, 卢玥君, 邵振国, 等. 基于图神经网络多智能体强化学习的电力-交通融合网协同优化运行[J]. 高电压技术, 2023, 49(11): 4622-4631.  
JIANG Changxu, LU Yuejun, SHAO Zhenguo, et al. Collaborative optimization operation of integrated electric power and traffic network based on graph neural network multi-agent reinforcement learning [J]. High Voltage Engineering, 2023, 49(11): 4622-4631(in Chinese).
- [70] WU Tong, SCAGLIONE A, RAVI N, et al. Optimal control of differentially private EV charging: a scalable learning approach under uncertainty[J]. IEEE Internet of Things Journal, 2025, 12(24): 55061-55075.
- [71] 周政, 杨祺铭, 卞艺衡, 等. “车-商-网”模式下面向配网弹性提升的分布式车网协同应急供电策略[J/OL]. 高电压技术, 1-23 [2025-12-06]. <https://doi.org/10.13336/j.1003-6520.hve.20250642>.  
ZHOU Zheng, YANG Qiming, BIAN Yiheng, et al. Distributed vehicle-grid collaborative emergency power supply strategy for

- distribution network resilience enhancement under vehicle-operator-grid framework[J/OL]. High Voltage Engineering, 1-23[2025-12-06]. <https://doi.org/10.13336/j.1003-6520.hve.20250642>(in Chinese).
- [72] 江昌旭, 袁羽娟, 刘晨曦, 等. 用户综合满意度驱动的多智能体图强化学习电动汽车充电引导策略[J/OL]. 电网技术, 2024: 1-15 [2025-07-18]. <https://doi.org/10.13335/j.1000-3673.pst.2024.1915>. JIANG Changxu, YUAN Yujuan, LIU Chenxi, et al. Multi-agent graph reinforcement learning-based electric vehicle charging guidance strategy driven by user comprehensive satisfaction[J/OL]. Power System Technology, 2024: 1-15[2025-07-18]. <https://doi.org/10.13335/j.1000-3673.pst.2024.1915>(in Chinese).
- [73] QIU Dawei, WANG Yi, DING Zhaohao, et al. Graph reinforcement learning for carbon-aware electric vehicles in power-transport networks[J]. IEEE Transactions on Smart Grid, 2024, 15(4): 3919-3935.
- [74] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc., 2017: 1025-1035.
- [75] 顾雪平, 刘彤, 李少岩, 等. 基于改进双延迟深度确定性策略梯度算法的电网有功安全校正控制[J]. 电工技术学报, 2023, 38(8): 2162-2177. GU Xueping, LIU Tong, LI Shaoyan, et al. Active power correction control of power grid based on improved twin delayed deep deterministic policy gradient algorithm[J]. Transactions of China Electrotechnical Society, 2023, 38(8): 2162-2177(in Chinese).
- [76] 冯昌森, 汤飞霞, 王国烽, 等. 基于融合经验安全强化学习的配电网电压控制[J]. 电力系统自动化, 2025, 49(8): 169-177. FENG Changsen, TANG Feixia, WANG Guofeng, et al. Volt-Var control for distribution network based on safe reinforcement learning with mixed experiences[J]. Automation of Electric Power Systems, 2025, 49(8): 169-177(in Chinese).
- [77] WU Pengfei, CHEN Chen, LAI Dexiang, et al. Real-time optimal power flow method via safe deep reinforcement learning based on primal-dual and prior knowledge guidance[J]. IEEE Transactions on Power Systems, 2025, 40(1): 597-611.
- [78] 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): 8764-8778. HONG Lucheng, WU Minghe, ZHU Jin, et al. Constraint-enhanced safe reinforcement learning-based decision-making method for re/active power optimization in highly penetrated pv-storage-charging distribution network[J]. Proceedings of the CSEE, 2025, 45(22): 8764-8778(in Chinese).
- [79] SU Tong, WU Tong, ZHAO Junbo, et al. A review of safe reinforcement learning methods for modern power systems[J]. Proceedings of the IEEE, 2025, 113(3): 213-255.
- [80] KIPF T N, WELING M. Variational graph auto-encoders[J]. arXiv preprint arXiv: 1611.07308, 2016.
- [81] YANG Ting, XU Zheming, JI Shijie, et al. Cooperative optimal dispatch of multi-microgrids for low carbon economy based on personalized federated reinforcement learning[J]. Applied Energy, 2025, 378: 124641.
- [82] MUKHERJEE S, HOSSAIN R R, MOHIUDDIN S M, et al. Resilient control of networked microgrids using vertical federated reinforcement learning: designs and real-time test-bed validations [J]. IEEE Transactions on Smart Grid, 2025, 16(2): 1897-1910.
- [83] 安芸峽, 崔明建, 韩一宁, 等. 基于数据隐私保护自适应联邦学习的分布式光伏有功可调节能力评估方法[J]. 中国电机工程学报, 2025, 45(21): 8281-8294. AN Yunzhi, CUI Mingjian, HAN Yining, et al. A distributed PV active power adjustability evaluation method based on self-adaptive privacy-preserving federated learning[J]. Proceedings of the CSEE, 2025, 45(21): 8281-8294(in Chinese).
- [84] 汤凌峰, 谢海鹏, 别朝红. 联邦分割强化学习驱动的配电网-建筑群隐私安全协同运行方法[J/OL]. 中国电机工程学报, 1-14[2025-10-30]. <https://link.cnki.net/urlid/11.2107.TM.20250820.1129.004>. TANG Lingfeng, XIE Haipeng, BIE Zhaohong. Federated split reinforcement learning-driven privacy-preserving coordinated operation method for distribution network and buildings[J/OL]. Proceedings of the CSEE, 1-14[2025-10-30]. <https://link.cnki.net/urlid/11.2107.TM.20250820.1129.004>(in Chinese).
- [85] 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(in Chinese).
- [86] 欧杰宇, 张逸, 辛荣, 等. 基于多智能体深度强化学习的低压配电网电能质量分散式协同调控策略[J]. 中国电机工程学报, 2025, 45(21): 8308-8322. OU Jieyu, ZHANG Yi, XIN Rong, et al. Decentralized coordinated control strategy for power quality in low voltage distribution networks based on multi-agent deep reinforcement learning[J]. Proceedings of the CSEE, 2025, 45(21): 8308-8322(in Chinese).
- [87] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations. Toulon, France: ICLR, 2017.
- [88] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//Proceedings of the 6th International Conference on Learning Representations. Vancouver, BC, Canada: ICLR, 2018.
- [89] VERMA S, ZHANG Zhili. Graph capsule convolutional neural networks[J]. arXiv preprint arXiv: 1805.08090, 2018.
- [90] YAN Sijie, XIONG Yuanjun, LIN Dahua. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA: AAAI, 2018: 7444-7452.



江昌旭

在线出版日期: 2026-01-08。

收稿日期: 2025-11-06。

作者简介:

江昌旭(1990), 男, 博士, 副教授, 硕士生导师, 研究方向为人工智能技术在电力系统中的应用、电力-交通融合网, E-mail: cxjiang@fzu.edu.cn;

林骏驰(2002), 男, 硕士研究生, 研究方向为基于强化学习的主动配电网运行优化, E-mail: linjc2182@126.com;

林俊杰(1992), 男, 通信作者, 博士, 副教授, 硕士生导师, 研究方向为电力系统分析与控制技术、状态估计, E-mail: linjunjie@fzu.edu.cn。

(责任编辑 赵梓含)

表 A1 主要术语及英文缩写对照表

Table A1 Main terms and English abbreviations reference

术语	英文缩写	术语	英文缩写	术语	英文缩写
图强化学习	GRL	图卷积网络	GCN	动作-评价框架	AC
深度强化学习	DRL	图注意力网络	GAT	深度 Q 网络	DQN
图神经网络	GNN	图胶囊卷积网络	GCAPCN	深度双 Q 网络	DDQN
分布式电源	DG	时空图神经网络	STGNN	近端策略优化	PPO
电动汽车	EV	图采样聚合网络	GraphSAGE	深度确定性策略梯度	DDPG
人工神经网络	ANN	图自编码器	GAE	双延迟深度确定性策略梯度	TD3
卷积神经网络	CNN	多智能体	MA	柔性动作-评价	SAC

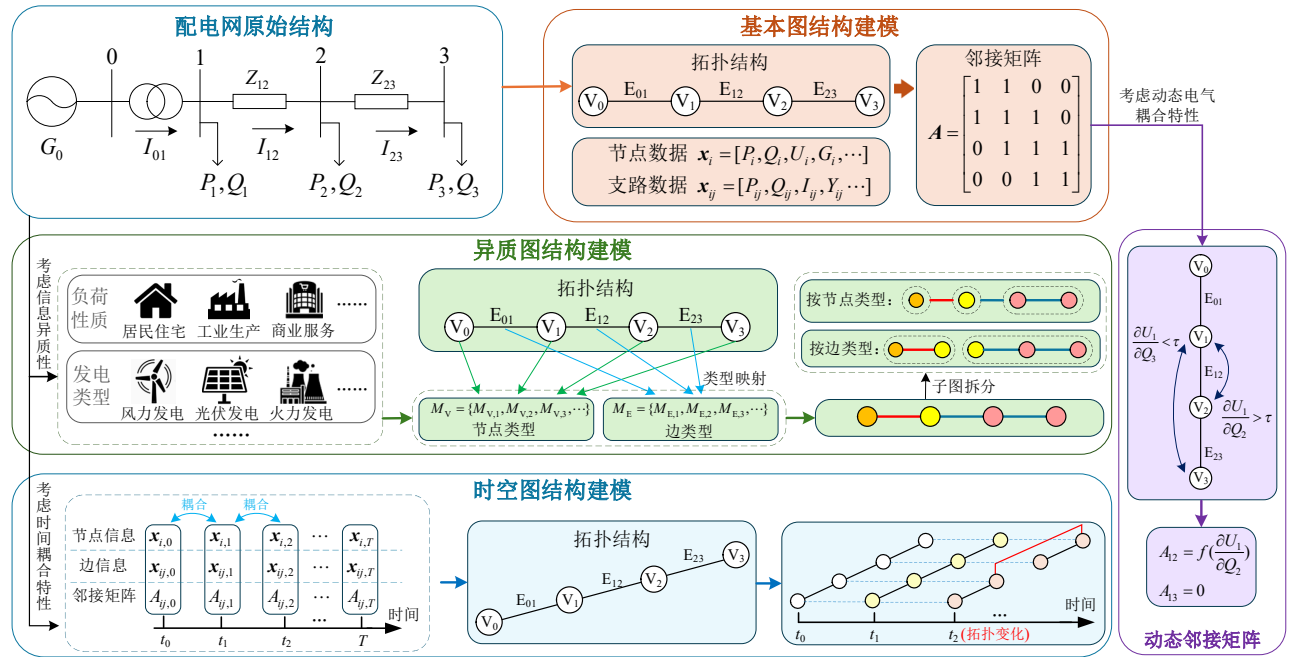


图 A1 配电网图结构建模原理

Fig. A1 Graph modeling principles of distribution network

附录 B

附录 B.1 常见神经网络总结

前馈网络的基础形式是 ANN，依靠多个全连接层的堆叠来拟合非线性映射关系。采用其处理电网数据时，将所有节点进行平整排列，能够对包含任意节点信息的数据进行全连接计算，但是不具备局部特征提取能力。前馈网络的另一种类型是 CNN，能够利用卷积、池化等操作对输入的一维数据或多维网格数据进行处理。采用其处理电网数据时，往往也将其转化为一维平整排列数据，通过一维卷积核的滑动能够实现局部特征提取。前馈网络具有结构简单、泛化能力强的优势，因此被广泛嵌入 DRL 并运用于配电网的优化决策。然而，前馈网络忽略了电网节点间的物理连接关系，不具备提取空间特征的能力，虽然 CNN 能够通过卷积核扫描局部特征，但这种局部特征在配电网中并不具备明确物理意义，仅能感知到平整化排列后的相邻数据，局限性较明显。

循环网络针对于具有明显顺序特征的序列数据，比较常用的类型有门控循环单元和长短时记忆网络，均通过循环机制与门控机制对序列数据的先后关系和时序特征进行学习。循环网络适用于提取配电网节点信息中的时序耦合关系，例如 DG 出力、负荷等数据的时序特性，但仅局限在节点内的不同时刻之间，依然难以捕捉节点间的空间相关性。

运用了自注意力机制的网络最主要为 Transformer，是目前最前沿的神经网络类型之一，通过位置编码机制代替复杂循环操作，同时融合了自注意力机制与多头注意力机制，提高了并行计算能力以及对大规模数据的处理能力。Transformer 同样适用于提取配电网的时序特征，同时也能够对不同节点之间的相关性进行捕捉，但这种相关性并不基于配电网节点之间的物理连接关系，且计算复杂度高，容易拟合出较多冗余信息，影响计算效率与结果。上述各种神经网络的主要原理及特点参见表 B1。



表 B1 DRL 与 GRL 中常见神经网络类型

Table B1 Common types of neural networks in DRL and GRL

类型	名称	基本原理	优点	缺点	适用数据结构类型
前馈网络	ANN	堆叠全连接层	结构简单, 可以拟合任意的非线性函数	难以提取局部特征; 参数量随数据规模的增加而显著增加	平整排列的一维数据
	CNN	通过卷积核滑动进行局部数据处理	参数量少, 具有局部特征提取能力, 泛用性强	难以提取非欧数据的空间特征	平整排列的一维数据或网格矩阵数据
循环网络	GRU	通过循环机制与门控机制实现时序特征处理	能够捕捉序列数据的顺序相关特性	对于顺序特性不够明显的数据会产生冗余, 泛用性较弱	较短的序列数据
	LSTM	通过循环机制与记忆单元实现时序特征处理	处理长序列的能力强	计算复杂度高, 面对短序列容易过拟合	较长的序列数据
自注意力网络	Transformer	位置编码、自注意力与多头注意力机制结合	并行计算能力强, 对大规模数据的适应性较好, 泛用性强	内存消耗量大, 计算复杂度高	超大规模任意类型数据
图神经网络	GCN	将节点自身及其邻域的信息进行聚合	最基础的 GNN 类型, 计算复杂度低, 泛用性强	节点信息聚合方式单一, 缺乏灵活性	基本图结构, 即仅考虑物理连接的配电网数据
	GAT	在 GCN 的基础上引入了注意力机制	能自适应捕捉节点间的动态相关性, 灵活性强	注意力权重计算复杂度较高	基本图结构和基于动态邻接矩阵的图结构, 例如考虑节点间动态电气耦合的配电网数据
	GCAPCN	引入向量神经元, 将标量特征拓展为向量特征	能够提取节点的高阶特征及更复杂的空间关联	多维向量特征计算复杂度高, 泛用性较差	异质图结构, 例如需要考虑负荷类型、异质性或调压设备异质性的数据
	STGNN	将其他时刻的特征输入当前时刻	能够捕捉图结构数据的时序耦合特性	计算复杂度高, 需进行时序耦合建模, 泛用性较弱	时空图结构, 例如包含各时刻 DG 出力、负荷大小、拓扑变化的配电网数据

各种常见传统神经网络和 GNN 的主要原理及特点如表 B1 所示。常用的 GNN 类型及其具体原理和特点如下:

### 附录 B.2 GCN 基本原理

GCN 是最基础的 GNN 类型, 其聚合函数是对节点自身及其邻居节点的特征进行归一化后再相加, 这一过程又被称作图卷积操作, 其表达式如下<sup>[87]</sup>:

$$\mathbf{H}^{(l+1)} = f(\mathbf{H}^{(l)}, \hat{\mathbf{A}}) = \sigma(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}) \quad (\text{B1})$$

$$\hat{\mathbf{A}} = \hat{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} \quad (\text{B2})$$

式中:  $\hat{\mathbf{A}}$  为归一化后的邻接矩阵;  $\tilde{\mathbf{A}}$  为包含节点自身在内的邻接矩阵;  $\hat{\mathbf{D}}$  为节点的度矩阵, 仅对角元素非零, 第  $i$  个对角元素为  $d_i$ , 表示包括节点  $i$  自身在内的与节点  $i$  相连接的节点数量;  $\mathbf{H}^{(l)}$  为第  $l$  个隐含层的节点特征矩阵, 其大小为  $n \times d$ ,  $n$  为节点数量;  $d$  为每个节点的特征数量;  $\mathbf{H}^{(0)}$  为初始特征矩阵;  $\sigma$  为激活函数;  $\mathbf{W}^{(l)}$  为第  $l$  个隐含层的权重矩阵。

在使用 GCN 处理配电网数据时, 一般仅需要将配电网构建为基本图结构, 简单高效, 泛用性强, 因此在配电网优化领域得到了广泛应用。但由于其聚合函数比较单一, 在处理较为复杂的图结构或未知的图结构时缺乏足够的灵活性与适应性。

### 附录 B.3 GAT 基本原理

GAT 在 GCN 的基础上引入了注意力机制, 在聚合邻域节点信息时对邻居节点施加了可学习的注意力系数, 使得 GAT 能够动态地学习节点之间的关联程度, 并自适应地对正式(4)中边的权重  $\alpha_{ij}$  进行调整, 从而关注到更加重要的邻居节点<sup>[88]</sup>。

GAT 状态更新的表达式如下:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{j \in N(i)} \alpha_{ij}^{(l)} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}\right) \quad (\text{B3})$$

式中:  $N(i)$  为节点  $i$  的邻居节点集合;  $\mathbf{h}_j^{(l)}$  为第  $l$  个隐含层中节点  $j$  的特征向量;  $\alpha_{ij}^{(l)}$  为第  $l$  个隐含层中节点  $i, j$  之间的注意力系数, 是 GAT 中图注意力机制的关键参数, 其计算公式如下:

$$\alpha_{ij}^{(l)} = \frac{\exp(\sigma_{\text{LR}}(F[\mathbf{W}^{(l)} \mathbf{h}_i^{(l)} \parallel \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}]))}{\sum_{k \in N(i)} \exp(\sigma_{\text{LR}}(F[\mathbf{W}^{(l)} \mathbf{h}_i^{(l)} \parallel \mathbf{W}^{(l)} \mathbf{h}_k^{(l)}]))} \quad (\text{B4})$$

式中:  $\sigma_{\text{LR}}$  为 LeakyReLU 激活函数;  $F$  是一个单层前馈神经网络;  $\parallel$  符号表示向量的拼接操作。

式(B4)的物理意义是将节点  $i, j$  经隐含层更新后的特征拼接为一个能够同时表征两个节点的联合特征向量, 再经过单层前馈神经网络将高维联合特征映射为低维特征, 计算节点  $i$  与其所有邻居节点的联合特征并归一化, 即可计算出节点  $i$  的所有邻边权重。可以看出, 边权重的自适应更新通过权重矩阵  $\mathbf{W}$  与前馈网络层  $F$  协同实现。

图 B1 对比了 GCN 与 GAT 的邻域聚合过程。

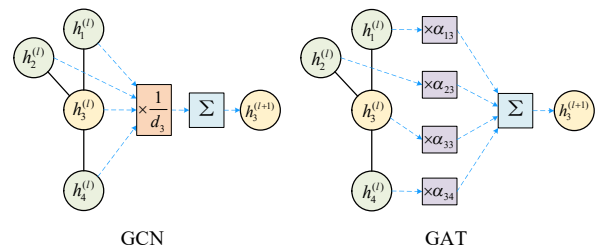


图 B1 GCN 与 GAT 节点状态更新过程

Fig. B1 State update process of GCN and GAT

相比于 GCN, GAT 引入的注意力机制虽然增加了计算复杂度,但在处理配电网数据时会对支路的权重进行动态自适应调整,这种方式有利于捕捉到配电网节点之间或节点与支路之间的电气耦合关系,如无功-电压灵敏度、故障电流灵敏度等,能够提高 GAT 对配电网状态识别的丰富性与准确性,并且对物理关系进行了利用,使得特征提取更加灵活。

#### 附录 B.4 GCAPCN 基本原理

将 GCN 与胶囊网络结合,其核心是胶囊神经元,这种神经元将传统的基于标量的节点特征更新方式转化为基于向量的更新方式,其优点是能够有效捕捉到更高阶的图结构信息,如节点特征中包含的方向信息、非相邻节点之间的多跳邻接关系等,在面对大规模复杂图结构时往往能够提取到更丰富的信息,得到更准确的特征输出<sup>[89]</sup>。

GCAPCN 的节点特征更新过程表达式如下:

$$\mathbf{H}^{(l+1)} = \sigma\left(\sum_{k=0}^K \mathbf{L}^k \mathbf{H}^{(l)} \mathbf{W}_k^{(l)}\right) \quad (\text{B5})$$

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (\text{B6})$$

式中:  $K$  表示胶囊神经元中由标量特征拓展得到的向量特征的维数,是一个可以灵活设置的参数;  $\mathbf{L}$  为图结构的拉普拉斯矩阵;  $\mathbf{D}$  为节点的度矩阵;  $\mathbf{A}$  为节点的邻接矩阵。

从式(18)可以得到,由标量特征拓展而来的节点特征可表示为:

$$\mathbf{g}(\mathbf{H}^{(l)}, \mathbf{L}) = [\mathbf{H}^{(l)}, \mathbf{L}\mathbf{H}^{(l)}, \mathbf{L}^2\mathbf{H}^{(l)}, \dots, \mathbf{L}^K\mathbf{H}^{(l)}] \quad (\text{B7})$$

式中:  $\mathbf{g}$  表示特征拓展函数。从上式可以看出,拓展后每个节点的每个特征都由一个标量变为一个长度为  $K$  的向量,结合隐含层的  $K$  个权重矩阵可以实现高维的节点表征。

GCAPCN 在处理复杂且高维的图结构特征时具有显著优势,使得特征提取更具层次化。其缺点主要是计算复杂度高且针对性强,训练速度较慢,在面对结构比较简单的图结构时容易产生冗余的计算,训练效果可能劣于 GCN、GAT 等更基础的 GNN 类型。

#### 附录 B.5 STGNN 基本原理

上述 GNN 类型仅提取图结构在空间上的特征,无法提取其时序动态演化规律,如配电网中的 DG 出力时序变化特征和负荷时序变化特征等。STGNN 将循环网络的时序特征提取能力运用到图域,融合了空间图卷积与时间卷积,能够针对性地处理时空图结构数据。其核心思想是将时序神经元

融入到 GNN 当中,通过分层的方式捕获由相邻节点关联产生的空间拓扑依赖和同一节点特征的时序变化特性<sup>[90]</sup>。

STGNN 的主要架构分为空间特征聚合与时间特征聚合。其中,空间特征聚合将图结构按时间顺序解耦成多个图结构分别独立进行邻域聚合,时序特征聚合则对空间特征聚合后的图结构进行线性特征变换。时序特征聚合过程中,每个节点在当前时刻的特征都会与上一个时刻的特征进行融合,从而实现时序特征提取。ST-GNN 节点特征更新过程的基本表达式如下:

$$\mathbf{H}_{t+1}^{(l)} = \sigma(\hat{\mathbf{A}}_{t+1} \mathbf{X}_{t+1} \mathbf{W}_{e,t+1}^{(l)} + \mathbf{H}_t^{(l)} \mathbf{W}_{m,t}^{(l)} + \mathbf{b}) \quad (\text{B8})$$

式中:  $\mathbf{H}_{t+1}^{(l)}$  表示第  $l$  个隐含层在第  $t+1$  个时序的节点特征矩阵;  $\mathbf{X}_{t+1}$  表示第  $t+1$  个时序的节点原始信息矩阵;  $\mathbf{W}_{e,t+1}^{(l)}$  和  $\mathbf{W}_{m,t}^{(l)}$  分别表示第  $l$  个隐含层用于处理当前时序节点状态和上一个时序节点状态的不同权重矩阵;  $\mathbf{b}$  为偏置矩阵。ST-GNN 的基本原理如图 B2 所示。

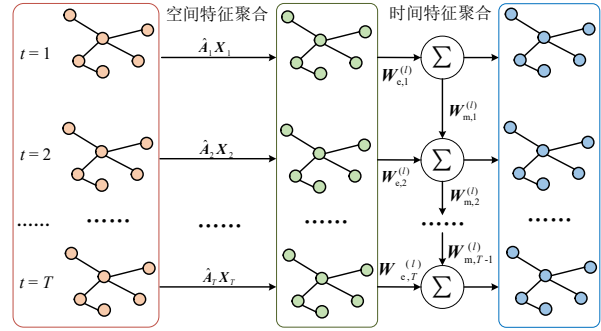


图 B2 STGNN 节点信息更新过程

Fig. B2 Node information update process of STGNN

从式(B8)和图 B2 中可以看出,ST-GNN 在处理节点状态时,会将前一时刻的节点特征聚合到当前时刻,从而提取节点状态的时间关联特性。STGNN 的时间特征提取模块具有拓展性与通用性,可以与前文介绍的 GCN、GAT 等各种类型 GNN 进行拼接,组合为 STGCN、STGAT 等 GNN。

#### 附录 C

强化学习的目标是寻找出一组最优的动作使得长期累计回报  $G^{(l)}$  的期望值最大。若整个过程所持续的时间为  $T$ ,则在智能体采用策略  $\pi$  确定动作的情况下,累计回报期望  $E_{\pi}(G^{(l)})$  的计算如下:

$$E_{\pi}(G^{(l)}) = E_{\pi}(\gamma^0 r^{(t+1)} + \gamma^1 r^{(t+2)} + \dots + \gamma^{T-t} r^{(T)}) \quad (\text{C1})$$

式中:  $r^{(t+1)}$  表示智能体执行动作  $a^{(t)}$  后从状态  $s^{(t)}$  变化到状态  $s^{(t+1)}$  时的奖励函数值,可由奖励函数  $R$  确定:

$$r^{(t+1)} = R(s^{(t)}, a^{(t)}) \quad (\text{C2})$$

式(C1)表示在状态  $s^{(t)}$  下, 往后  $T-t$  个时间的累计期望回报值, 可将其看作是一个以状态为自变量的状态值函数  $V_{\pi}(s^{(t)})$ , 则

$$V_{\pi}(s^{(t)}) = E(\gamma^0 r^{(t)} + \gamma V_{\pi}(s^{(t+1)})) \quad (C3)$$

若在当前状态  $s^{(t)}$  和动作  $a^{(t)}$  已知的情况下, 累计回报期望也称为动作值函数, 用  $Q_{\pi}(s^{(t)}, a^{(t)})$  表示。在策略  $\pi$  确定的情况下, 状态值函数与动作值函数取值相等。在策略  $\pi$  为随机策略的情况下, 状态值函数则为所有可能采取的动作所对应的动作值函数加权之和:

$$V_{\pi}(s^{(t)}) = \sum_{a \in A} \pi(a|s) Q_{\pi}(s, a) \quad (C4)$$

式中: 权重  $\pi(a|s)$  为随机策略下状态为  $s$  时采取动作  $a$  的概率, 即

$$\pi(a^{(t)}|s) = P(a^{(t)}), a^{(t)} \in A \quad (C5)$$

动作值函数也可以依据状态值函数与策略  $\pi$  进行计算, 其计算式表示如下:

$$Q_{\pi}(s^{(t)}, a^{(t)}) = r^{(t+1)} + \gamma \sum_{s^{(t+1)} \in S} P(s^{(t+1)}|s^{(t)}, a^{(t)}) V_{\pi}(s^{(t+1)}) \quad (C6)$$

选择合适的策略能够使得所有状态下对应的值函数取值均为最大, 最大的值函数即为最优值函数。将最优动作值函数作为 MDP 的优化目标并以此确定最优策略  $\pi^*$ , 则强化学习的优化问题可以表示为:

$$\pi^*(a|s) = \arg \max_{\pi} Q_{\pi}^*(s, a) \quad (C7)$$

典型的传统强化学习算法如表 C1 所示。

表 C1 RL、DRL 与 GRL 代表性算法  
Table C1 Representative algorithms of RL, DRL, and GRL

类型	代表性算法	主要原理	主要目标	优点	缺点
强化学习	Q 学习	利用 Q 值表存储所有状态-动作组合的价值并更新	准确计算所有状态-动作组合的价值并取最优	无需准确的系统模型	无法处理高维动作或状态空间
	动态规划	利用贝尔曼方程和值函数迭代来更新策略	不断更新值函数并求解最优策略	计算准确, 能够适应高维动作和状态空间	依赖准确的系统模型
DRL	DQN	神经网络代替 Q 值表, 拟合状态-动作组合的价值函数	不断提高神经网络拟合出的 Q 值准确性	解决了 Q 学习无法处理高维状态空间的问题	容易过高估计 Q 值导致局部最优, 无法处理连续动作空间
	DDQN	在 DQN 的基础上将动作选取与动作评价分离	进一步提高神经网络拟合出的 Q 值准确性	缓解了 DQN 容易过高估计 Q 值的问题	无法处理连续动作空间
	PPO	通过策略梯度和重要性采样裁剪直接优化策略	使策略梯度不断朝更优的区域更新	解决了动态规划依赖完整模型的缺点	稳定性较差, 需要的样本数量较多
	DDPG	设计两组 Actor-Critic 分别求解当前和目标动作-价值	提高 Critic 网络评价的准确性并据此指导 Actor 网络收敛至最优策略	能够处理连续动作空间	Critic 容易过高估计状态-动作的价值
	TD3	在 DDPG 基础上额外设置一组 Critic, 两组 Critic 计算出的价值取较小值	进一步提高 Critic 网络评价的准确性	能够处理连续动作空间且缓解了 DDPG 容易过高估计价值的问题	网络数量多, 更新时的计算复杂度增加
GRL	GCN+DRL	将图卷积操作集成到 DRL 智能体的价值函数拟合或策略优化的过程中	赋予 DRL 智能体空间特征感知能力	最简单的 GRL 形式, 计算复杂度低, 泛用性强	图计算方式单一, 缺乏处理动态相关性的能力, 灵活性较差
	GAT+DRL	在 GCN+DRL 的基础上加入图注意力机制	赋予 DRL 智能体自适应学习图结构节点间关联性的能力	学习数据间动态相关性的能力强, 泛用性强	计算复杂度较高
	GCAPCN+DRL	在 GCN+DRL 的基础上加入向量神经元, 将观测数据映射至高维	捕捉数据中隐含的高维特征, 对策略优化起到更强的指导作用	能够将观测到的数据映射到高维空间, 学习高维隐含特征的能力强	计算复杂度较高, 容易产生冗余计算
	STGNN+DRL	在其他 GRL 的基础上加入顺序特征聚合的过程	捕捉图结构数据的顺序相关性	能够同时学习数据间的时空相关性	仅针对具有顺序耦合性的数据, 泛用性较弱

附录 D

附录 D.1 各文献仿真环境及结果总结

表 D1 基于 GRL 的配电网优化调度方法  
Table D1 GRL based optimal dispatching methods of distribution network

文献	优化目标	节点数量	DG 渗透率/%	随机因素设置	超参数调优方式	算法	算法性能		
							收敛	收敛速度	优化结果
[30]	发电成本、储能成本、弃风弃光成本、购电成本	33	/	DG 波动服从 $N(0,0.03^2)$ ; 负荷波动服从 $N(0,0.2^2)$	无具体调参和对比过程	数学优化	/	/	运行总成本为 4237.75 \$, 决策时间 16.12s
						集中式 SAC	×	未收敛	运行总成本为 202505.19 \$, 决策时间 0.004s
						集中式 GCN+SAC*	√	约 200 回合	运行总成本为 4325.9 \$, 决策时间 0.006s
						分布式 SAC	×	未收敛	运行总成本为 190264.6 \$, 决策时间 0.0003s
						分布式 GCN+SAC*	√	约 300 回合	运行总成本为 4968.91 \$, 决策时间 0.005s
[31]	网损、电压偏移	33	250	采用真实 DG 和负荷历史数据, 未提供随机数设置方式	无具体调参和对比过程	数学优化	/	/	无电压越限, 网损指标 0.056, 电压偏移 0.02 pu
						MADDPG	√	约 220 回合	电压越限比例 0.072, 网损指标 0.056, 电压偏移 0.014 pu
						GCN+GRU+MADDPG*	√	约 150 回合	电压越限比例 0.023, 网损指标 0.049, 电压偏移 0.013 pu
		141	400			数学优化	/	/	无电压越限, 网损指标 0.819, 电压偏移 0.021 pu
						MADDPG	√	约 300 回合	电压越限比例 0.041, 网损指标 1.597, 电压偏移 0.017 pu
						GCN+GRU+MADDPG*	√	约 200 回合	电压越限比例 0.008, 网损指标 1.153, 电压偏移 0.013 pu
[32]	网损、电压偏移	33	250	采用真实 DG 和负荷历史数据, 未提供随机数设置方式	无具体调参和对比过程	MADDPG	√	约 1200 回合	收敛奖励值约-0.012, 电压越限率约 1.5%, 决策时间约 36ms
						MATD3	√	约 1200 回合	收敛奖励值约-0.01, 电压越限率约 3.14%, 决策时间约 31ms
						GCN+MATD3*	√	约 1100 回合	收敛奖励值约-0.009, 电压越限率约 0.768%, 决策时间约 35ms
						MADDPG	√	约 400 回合	收敛奖励值约-0.046, 电压越限率约 5.26%, 决策时间约 57ms
		141	400			MATD3	√	约 550 回合	收敛奖励值约-0.052, 电压越限率约 3.76%, 决策时间约 59ms
						GCN+MATD3*	√	约 600 回合	收敛奖励值约-0.042, 电压越限率约 2.92%, 决策时间约 53ms
						MADDPG	√	约 850 回合	收敛奖励值约-0.0062, 电压越限率约 2.79%, 决策时间约 81ms
						MATD3	√	约 700 回合	收敛奖励值约-0.0039, 电压越限率约 1.53%, 决策时间约 88ms
322	250	GCN+MATD3*	√	约 600 回合	收敛奖励值约-0.0033, 电压越限率约 1.42%, 决策时间约 78ms				
		DDPG	√	约 2000 回合	网损 6.82MWh, 电压偏移 0.026 pu, 运行成本 143.3k¥				
		GCN+DDPG	√	约 1400 回合	网损 6.6MWh, 电压偏移 0.023 pu, 运行成本 139.63k¥				
		GAT+DDPG*	√	约 650 回合	网损 6.54MWh, 电压偏移 0.019 pu, 运行成本 138.5k¥				
[33]	网损、电压偏移、购电成本、光伏成本、储能成本、柔性负载成本	33	161.5	动作噪声服从 $N(0,0.6)$	设置不同 GAT 层数并对比损失函数变化以确定最佳层数	PSO	/	/	网损 2.57MWh, 电压偏移 0.0253 pu, 运行成本 24.04k¥
						DDPG	/	/	网损 3.02MWh, 电压偏移 0.0288 pu, 运行成本 25.10k¥
		69	32.9			DDPG	/	/	网损 3.02MWh, 电压偏移 0.0288 pu, 运行成本 25.10k¥
						GAT+DDPG*	√	约 250 回合	网损 2.27MWh, 电压偏移 0.0243 pu,

						运行成本 23.39k¥			
	123	64.8			PSO	/	/	网损 2.03MWh, 电压偏移 0.0296 pu, 运行成本 21.77k¥	
					DDPG	/	/	网损 2.25MWh, 电压偏移 0.0312 pu, 运行成本 23.3k¥	
					GAT+DDPG*	√	约 300 回合	网损 1.92MWh, 电压偏移 0.0279 pu, 运行成本 20.58k¥	
[34]	网损、电压偏移	33	64.6	负荷率在 [0.2,0.8]内均匀随机变化	无具体调参和对比过程	MASAC	√	约 1000 回合	网损 0.466MWh, 电压最值为 0.966、1.031pu
						GAT+MASAC*	√	约 400 回合	网损 0.41MWh, 电压最值为 0.974、1.022pu
						优化前	/	/	网损 1.71MWh, 电压最值为 0.962、1.052pu
		136	45.2			GAT+MASAC*	/	/	网损 0.957MWh, 电压最值为 0.989、1.025pu
[36]	网损、电压偏移	33	36.3	负荷波动服从 $N(0,0.05^2)$	无具体调参和对比过程	DQN	×	未收敛	网损 14.23MW, 总电压偏移 5.97 pu
						MADDPG	√	约 1600 回合	网损 6.33MW, 总电压偏移 4.81 pu
						MAAC	√	约 900 回合	网损 4.59MW, 总电压偏移 4.06 pu
		STGAT+MASAC*	√			约 900 回合	网损 4.13MW, 总电压偏移 3.51 pu		
		DQN	×			未收敛	网损 10.89MW, 总电压偏移 7.21 pu		
		MADDPG	√			约 900 回合	网损 4.92MW, 总电压偏移 5.87 pu		
		123	133.2			MAAC	√	约 1200 回合	网损 2.79MW, 总电压偏移 5.14 pu
						STGAT+MASAC*	√	约 800 回合	网损 2.56MW, 总电压偏移 4.06 pu

注：1、表格中的“\*”表示文献所提出的 GRL 算法，“/”表示该算法没有提到该内容或指标。

2、“DG 渗透率”计算方式：(DG 总额定功率/配电网总额定负荷)×100%

表 D2 基于 GRL 的配电网动态重构方法

Table D2 GRL based dynamic reconfiguration methods of distribution network

文献	优化目标	节点数量	DG 渗透率/%	随机因素设置	超参数调优方式	算法	算法性能		
							收敛	收敛速度	优化结果
[43]	网损、电压偏移	33	96.9	负荷波动服从 $N(0,0.05^2)$	无具体调参和对比过程	数学优化	/	/	网损 86.86kW, 总电压偏移 0.75 pu, 决策时间 34.94s
						GA	/	/	网损 83.73kW, 总电压偏移 0.84 pu, 决策时间 28.34s
						DDQN	×	未收敛	网损 98.91kW, 总电压偏移 0.88 pu, 决策时间 0.12s
						GCN+DDPG*	√	约 400 回合	网损 91.7kW, 总电压偏移 0.78 pu, 决策时间 0.11s
[44]	网损	13	/	/	无具体调参和对比过程	BPSO	/	/	网损 105.73kW, 决策时间 25.571s
						PPO	×	未收敛	未收敛, 训练后期奖励值约 0.1~0.3
						GCAPCN+PPO*	√	约 40000 步	收敛后奖励值约 0.7, 网损 157.9kW, 决策时间 0.0019s
		BPSO	/			/	网损 56.66kW, 决策时间 245.398s		
		PPO	×			未收敛	未收敛, 训练后期奖励值约 0~0.1		
		GCAPCN+PPO*	√			约 70000 步	收敛后奖励值约 0.7, 网损 58.51kW, 决策时间 0.002s		
[45]	网损成本、开关动作成本	14	172.4	采用真实 DG 和负荷历史数据, 未提供随机数	无具体调参和对比过程	GA	/	/	相比优化前, 运行成本降低 23.3%
						数学优化	/	/	三种程度数据丢失运行成本 141.5 \$、305.9 \$、505.1 \$
		STGCN+安全 DRL*	√			约 500 回合	相比优化前运行成本降低 33.2%; 三种程度数据丢失后运行成本 15.3 \$、105.6 \$、199.7 \$		
		GA	/			/	相比优化前, 运行成本降低 31.9%		
		数学优化	/			/	三种程度数据丢失后运行成本 31.5、43.7、66.8		
		33	193.4			STGCN+安全 DRL*	√	约 750 回合	相比优化前, 运行成本降低 41.8%; 三种程度数据丢失后运行成本 6.8、14.9、24.5

[46]	网损成本、开关动作成本、电压偏移	33	500	负荷功率因数在±5%范围内波动	无具体调参和对比过程	SAC	/	/	网损 980kWh, 平均电压偏移 0.023 pu, 奖励值-30.4
						GCN+SAC	/	/	奖励值-29.4
		双向 GNN+SAC*	√			约 1.5 万回合	网损 957kWh, 平均电压偏移 0.02 pu, 奖励值-28.7		
		SAC	/			/	网损 1203kWh, 平均电压偏移 0.019 pu, 奖励值-33.9		
		GCN+SAC	/			/	奖励值-33.5		
		双向 GNN+SAC*	√			约 1.5 万回合	网损 1197kWh, 平均电压偏移 0.016 pu, 奖励值-32.8		
[47]	网损成本、开关动作成本、电压偏移	33	107.7	负荷与 DG 出力在±5%范围内波动	手动设置不同的学习率和折扣因子并对比损失函数变化以确定最佳参数组合	ANN+DDQN	×	未收敛	平均电压偏移约 0.013pu, 网损约 55kW
						CNN+DDQN	×	未收敛	平均电压偏移约 0.02pu, 网损约 70kW
		GCN+DDQN*	√			约 800 回合	平均电压偏移约 0.012pu, 网损约 58kW, DG 削减量 45.18kW		
		GCN+DQN	×			未收敛	平均电压偏移约 0.01pu, 网损约 170kW, DG 削减量 4.642kW		
		GCN+DDPG	√			约 3000 回合	平均电压偏移约 0.014pu, 网损约 220kW, DG 削减量 3.197kW		
		GCN+DDQN*	√			约 4800 回合	平均电压偏移约 0.008pu, 网损约 150kW, DG 削减量 3.053kW		
[48]	网损、电压偏移	123	60.5	采用真实 DG 和负荷历史数据, 未提供随机数	无具体调参和对比过程	PSO	/	/	网损 467.6kW, 电压偏移 0.0247 pu, 决策时间 1064.4s
						MATD3	√	约 3000 回合	收敛后奖励值与 GCN+MASAC 相近
		MASAC	√			约 2100 回合	收敛后奖励值与 GCN+MASAC 相近		
		GCN+MASAC*	√			约 1100 回合	网损 465.4kW, 电压偏移 0.0247 pu, 决策时间 0.015s		
		SAC	√			约 6000 回合	网损 1589.442kW, 电压偏移 0.026 pu		
		MASAC	√			约 6000 回合	网损 1593.716kW, 电压偏移 0.026 pu		
		8500	16.7			GCN+MASAC*	√	约 4500 回合	网损 1416.658kW, 电压偏移 0.024 pu

表 D3 基于 GRL 的配电网故障恢复方法

Table D3 GRL based fault recovery methods of distribution network

文献	优化目标	节点数量	DG 渗透率/%	随机因素设置	超参数调优方式	算法	算法性能		
							收敛	收敛速度	优化结果
[53]	负荷恢复量	37	66.3	设置训练集中未出现过的故障场景并随机抽取测试	无具体调参和对比过程	DDPG	√	约 2500 回合	负荷恢复量 748.8kW
						TD3	√	约 800 回合	负荷恢复量 769.601kW
						GCN+TD3*	√	约 600 回合	负荷恢复量 1045.115kW
[54]	负荷恢复量	69	52.7	负荷和 DG 出力分别在 ±10%和 ±50%范围内波动; 拓扑变化后随机选取故障	无具体调参和对比过程	DRL	/	/	拓扑变化前后负荷恢复率分别为 94.3%、72.4%
						GCN+DQN*	√	约 5000 回合	拓扑变化前后负荷恢复率分别为 94.6%、90.3%
[55]	负荷恢复量	13	75.4	负荷和 DG 随机波动但未提供具体随机数范围或分布	无具体调参和对比过程	数学优化	/	/	负荷恢复量约 3800kW, 决策时间 0.304s
						PPO	√	约 80 万步	负荷恢复量约 2200kW, 决策时间 0.0039s
		GCAPCN+PPO*	√			约 200 万步	负荷恢复量约 3000kW, 决策时间 0.0049s		
		数学优化	/			/	负荷恢复量约 1700kW, 决策时间 0.5676s		
		PPO	√			约 150 万步	负荷恢复量约 800kW, 决策时间 0.002s		
		34	28.7			GCAPCN+PPO*	√	约 150 万步	负荷恢复量约 1400kW, 决策时间

									0.0025s
								数学优化	/ / 负荷恢复量约 3500kW, 决策时间 0.92s
	123	64.3						PPO	× 未收敛 负荷恢复量<300kW, 决策时间 0.005s
								GCAPCN+PPO*	√ 约 40 万步 负荷恢复量约 3200kW, 决策时间 0.0078s
[56]	故障损失、应急响应资源调度成本	33	129.2	负荷和 DG 出力在±5% 范围内波动	无具体调参和对比过程			MAD3QN	√ 约 2000 回合 总损失成本 4053.64, 负荷削减量 1070.49kW
								GAT+MAD3QN*	√ 约 3000 回合 总损失成本 1560.81, 负荷削减量 374.23kW
	69	231.5					MAD3QN	/ / 总损失成本 18604.97, 负荷削减量 5053.419kW	
							GAT+MAD3QN*	/ / 总损失成本 2995.89, 负荷削减量 637.11kW	
[57]	负荷恢复量、网损、电压越限	33	53.84	GRL 及对比算法均在 5 组随机种子下训练, 结果取平均	无具体调参和对比过程			SAC	√ 约 50 回合 负荷恢复率 95.67%, 决策时间 0.13s
								GAT+SAC*	√ 约 30 回合 负荷恢复率 97.92%, 决策时间 0.08s
	148	/					D3QN	√ 约 120 回合 负荷恢复率 98.38%, 网损率 3.62%, 电压越下限	
							SAC	√ 约 110 回合 负荷恢复率 98.38%, 网损率 3.47%, 电压无越限	
							GAT+SAC*	√ 约 80 回合 负荷恢复率 100%, 网损率 3.56%, 电压无越限	
[58]	负荷恢复量	123	/	将 DG 出力数据划分为训练集与测试集, 模拟测试阶段的 DG 出力随机性	无具体调参和对比过程			SAC	× 未收敛 负荷恢复量 2610.86kW
								GCN+SAC	√ 约 900 回合 收敛奖励值约 1600
								STGCN+SAC*	√ 约 500 回合 收敛奖励值约 1900, 负荷恢复量 2886.77kW
[59]	负荷恢复量	123	/	负荷率在[0.1,2.0]的范围内均匀随机变化	无具体调参和对比过程			DQN	√ 约 1300 回合 轻负荷恢复量 100%, 重负荷恢复量 68%
								MADQN	√ 约 750 回合 轻负荷恢复量 100%, 重负荷恢复量 92%
								GCN+MADQN*	√ 约 300 回合 轻负荷恢复量 100%, 重负荷恢复量 100%
	8500	/			DQN	/ / 负荷恢复量 73.53%			
					MADQN	/ / 负荷恢复量 92.32%			
					GCN+MADQN*	√ 约 1000 回合 负荷恢复量 100%			

表 D4 基于 GRL 的电力-交通耦合网协同优化方法

Table D4 GRL based collaborative optimization methods of power-transportation system

文献	优化目标	节点数量		EV 渗透率/%	随机因素设置	超参数调优方式	算法	算法性能		
		电网	路网					收敛	收敛速度	优化结果
[65]	EV 充电时间与费用	69	373	82.7	GRL 及对比算法均在 5 组随机种子下训练, 结果取平均	无具体调参和对比过程	SAC	×	未收敛	未收敛
							MADDPG	√	约 300 回合	奖励值收敛后波动较大
							GAT+MAAC*	√	约 1200 回合	奖励值收敛后波动较小
[66]	EV 充电时间与费用	/	588	/	将充电数据划分为训练集与测试集, 模拟测试阶段 EV 充电行为的随机性	设置不同 GCN 层数并对比损失函数变化以确定最佳层数	DDQN	√	约 500 回合	奖励值约-45, 平均充电总成本约 48.25¥
							Dueling DQN	√	约 600 回合	奖励值约-45, 平均充电总成本约 46.23¥
							GCN+Rainbow*	√	约 500 回合	奖励值约-35, 平均充电总成本约 41.13¥
[67]	电压偏移	33	28	/	随机设置 EV 起点与目的地	无具体调参和对比过程	最短路径算法	/	/	优化后充电节点最低电压约 0.895
							Deuling DQN	√	约 30 回合	约 50 回合后奖励值开始下降
[68]	EV 充电时间、电	33	25	54.6	仿真平台随机生成 EV 起点与目的地;	无具体调参和对比过程	GAT+Deuling DQN*	√	约 30 回合	奖励值稳定, 优化后充电节点最低电压约 0.905
							最短路径算法	/	/	用户成本指标 1.46, 充电不均衡度指标 8.51, 交通拥堵指标 1.04,

		EV 初始 SOC 服从正态分布				电压偏移指标 1.43		
	压偏移、交通拥挤程度、充电服务均衡度							用户成本指标 0.96, 充电不均衡度指标 0.16, 交通拥堵指标 1.04, 电压偏移指标 1.44
[69]	EV 充电时间与费用、配电网购电成本	33	108	64.5	/	对比了不同学习率和折扣因子下的损失函数	最短路经算法 / / MADQN × 未收敛 GAT+MADQN* ✓ 约 12 万步	平均充电排队时间 11.46min, 负荷峰谷差最大值比例 70.86% 平均奖励值约为 10 平均奖励值约为 55, 平均充电排队时间 1.04min, 负荷峰谷差最大值比例 69.87%
[72]	EV 充电时间与费用、用户里程焦虑	33	25	30	EV 有充电需求的 SOC 阈值服从 $N(0.25, 0.05^2)$	无具体调参和对比过程	最小后悔值算法 / / 加权平均算法 / / 最短路经算法 / / GCN+GAT+MADQN* ✓ 约 4 万步	总时间 43.92, 充电成本 18.23, 里程焦虑 1.79, 交通拥堵程度 0.153, 电压偏移率 0.046 总时间 43.01, 充电成本 18.21, 里程焦虑 1.82, 交通拥堵程度 0.148, 电压偏移率 0.048 总时间 52.51, 充电成本 17.86, 里程焦虑 0.6, 交通拥堵程度 0.024, 电压偏移率 0.049 总时间 40.15, 充电成本 17.36, 里程焦虑 0.62, 交通拥堵程度 0.013, 电压偏移率 0.046
[73]	度电碳排放、EV 调度的成本	33	12	/	采集真实交通、负荷和 DG 出力并划分训练集与测试集, 模拟测试阶段的随机性	无具体调参和对比过程	DDPG ✓ 约 8000 回合 GCN+DDPG ✓ 约 6000 回合 GCN+DDPG (参数共享)* ✓ 约 6000 回合 优化前 / / GCN+DDPG (参数共享)* / /	收敛后奖励值约-140 收敛后奖励值约-104 收敛后奖励值约-78 度电碳排放量 147.14 g CO <sub>2</sub> /kW 度电碳排放量 83.69 g CO <sub>2</sub> /kW

注：“EV 渗透率” 计算方式：(充电站总额定充电功率/配电网总额定负荷)×100%

## 附录 D.2 GRL 优缺点总结

根据前文整理的相关文献，可以总结出 GRL 主要有以下几个显著优势：

1) 从算法训练的角度出发，GRL 最显著的特点是能够有效利用非欧图结构的拓扑信息。训练过程中，图结构信息的加入能够间接提高经验数据的质量，帮助智能体更好地拟合节点之间的相互关系，提高智能体对环境状态的感知深度，进而提高算法训练的稳定性及调度方案的优质性。

2) 从算法实际使用的角度出发，GRL 的首要特点与 DRL 一致，即决策速度远远高于数学优化和启发式算法等传统算法，具备更强的实时性，有利于算法适应 DG 出力和 EV 负荷的动态变化。

3) 作为一种数据驱动算法，GRL 的另一个重要优点是不依赖于准确的配电网参数信息，能够适应实际电力系统中参数信息不完整或有误的情况。

4) 在算法设计层面，GRL 与 DRL 一脉相承。得益于 GNN 的强适应性，GRL 可以在已有的大量成熟 DRL 框架的基础上进行设计与部署，能够充分利用过往的研究成果。从前文综述的文献中可以看出，GNN 与 DRL 具有多样化的组合方式，并且

都能够取得良好的效果，印证了 GRL 算法较强的可拓展性。

即使 GRL 在配电网优化问题中已经展现出多种优势，但从部分文献的仿真结果中也可以看出算法仍然存在一些局限性，制约了 GRL 在实际工程中的应用。局限性主要有以下几点：

1) 从[31,32]等文献中可以看出，GRL 算法优化后仍然会出现电压越限等违反约束条件的现象，说明 GRL 仍然难以完全保证满足约束，存在违反安全运行需求的隐患。现实中的配电网状态变化迅速，安全约束类型多样，未能实现硬性且全面约束的 GRL 算法存在明显的安全风险，难以在实际工程中应用。

2) 从[30,36,44,55]等文献中可以看出，GRL 在执行阶段的决策速度慢于 DRL，这主要是由于 GRL 嵌入了图结构编码的环节，导致计算复杂度增加，前向传播时间延长。虽然在这些文献的仿真结果中仅有微秒级的差距，但当系统规模扩大时，会增大 GNN 层的前向传播计算量，使得所需决策时间显著增加，影响了 GRL 在实际工程中决策的实时性。因此，计算复杂度问题不可忽略。

3) 从[46,55,56]等文献中可以看出, 部分 GRL 算法训练所耗费的回合数明显较高, 并且部分 GRL 收敛的回合数多于 DRL, 虽然最终能够收敛到更高的奖励值, 但收敛时间的增加会在训练阶段给配电网管理部门增加额外的成本。同时, 这也说明了 GRL 在收敛机制上仍然存在一定的不足。这些文献的共同点在于所处场景为故障场景或拓扑优化重构场景, 训练过程中拓扑结构变化的频率较高。在变化频率低的拓扑结构下, 节点间的相关性往往更加稳定, 而在频繁改变的拓扑结构下, 节点间的相关性将会持续变化, 不稳定性增加, GNN 需要耗费更高的时间成本对这种动态相关性进行拟合, 因此收敛速度较慢。

4) GRL 在鲁棒性方面存在理论局限性。从理论

层面分析, 虽然数据驱动算法不依赖固有参数信息的完整性与准确性, 但对于观测数据的准确性要求较高。GRL 的观测数据中加入了拓扑信息, 这一问题更加凸显。在现实中, 受限于配电网的基础设施完备性, 配电网可获取数据的准确性难以完全保证, 数据噪声及数据缺失会影响算法的决策效果。此外, 通信设备的延迟问题会使得终端接收到的数据实时性降低, 同样影响了算法决策的准确性。前文整理的大部分文献中的 GRL 算法在假设了观测数据准确的前提下进行设计, 比较理想化, 未将现实世界中的数据质量问题考虑进去, 也未进行相关的验证, 因此, 现有的 GRL 框架存在鲁棒性不足的潜在风险。

表 D5 总结了文献中基于 GRL 的配电网优化方法的创新点与不足之处。

表 D5 基于 GRL 的配电网优化方法创新点及缺点总结

Table D5 Contributions and disadvantages of GRL based distribution network optimization methods

场景	文献	算法	优点及创新	不足
配电网优化调度	[30]	GCN+SAC(集中式)	训练速度快, 集中式执行优化精度高	集中式控制需观测全局状态
		GCN+SAC(分布式)	分布式执行, 降低部署难度	训练速度与优化精度均劣于集中式算法
	[31]	STGCN+MADDPG	引入了时序特征提取, 提高优化精度	训练时间较长
	[32]	GCN+MATD3	引入元学习, 提高多智能体协同准确性	对历史数据具有一定的依赖性
	[33]	GAT+DDPG	首次将图注意力机制与 DRL 结合	集中式控制需观测全局状态
	[34]	GAT+MASAC	采用物理模型辅助经验积累	需要准确的物理模型支持
	[35]	GAT+MADDPG	根据无功-电压灵敏度信息进行动态邻接矩阵建模, 并与图注意力机制结合	计算复杂度高, 决策速度较慢
	[36]	STGAT+MASAC	时序特征提取与注意力机制相结合	分布式训练使得平稳性较差
配电网动态重构	[43]	GCN+DDPG	减小了动作空间规模	需预先筛选可行动作
	[44]	GCAPCN+PPO	采用 GCAPCN, 收敛性好, 适应性强	未与基于其他 GNN 的算法比较, GCAPCN 的优势未充分体现
	[45]	STGCN+SDRL	引入时序特征提取与安全强化学习, 且考虑了参数不完整的影响	未与传统 DRL 或未引入时序特征提取、安全机制的 GRL 进行比较
	[46]	双向 GNN+SAC	采用双向 GNN 提高含拓扑的特征提取准确性, 且适应性强	算法复杂度较高, 训练时间较长
	[47]	GCN+DDQN	日前 GRL 与日内模型驱动滚动优化相结合, 提高多时间尺度协同效率	系统规模扩大后训练稳定性明显下降
	[48]	GCN+MASAC	采用分布式重构, 对大型配电网适应性强	未考虑开关动作成本
	[53]	GCN+TD3	引入元学习且考虑了三相不平衡性	未测试算法对较大系统的适应性
	[54]	GCN+DQN	在变化拓扑下测试了动态拓扑适应性	对源荷不确定性模拟的准确性不够高
配电网故障恢复	[55]	GCAPCN+PPO	对大规模系统的适应性强	训练时间较长
	[56]	GAT+MAD3QN	同时考虑了电力网与交通网损坏的影响	收敛速度慢于 DRL
	[57]	GAT+SAC	在变化的拓扑下测试了动态拓扑适应性; 引入不可行动作掩码机制	未考虑孤岛供电的场景
	[58]	STGCN+SAC	引入了时序特征提取, 优化精度高	计算复杂度高, 决策速度较慢
	[59]	GCN+MADQN	采用相邻参数共享, 提高智能体之间的协同性	在大型系统中需依靠预训练维持稳定性
	[65]	GAT+MAAC	引入图注意力, 收敛的策略比 DRL 稳定	未考虑电网侧利益, 收敛速度慢于 DRL
	[66]	GCN+Rainbow	采用 Rainbow 算法, 优化精度高	未考虑电网侧利益
	[67]	GAT+Deuling DQN	考虑了电压越限影响	未考虑用户侧利益
电力-交通耦合网协同优化	[68]	GCN+DQN	优化多主体的多个目标, 综合性强	对电压偏移的优化精度不够高
	[69]	GAT+MADQN	基于 EV 之间的相互影响关系构建图结构	训练时间较长
	[72]	GCN+GAT+MADQN	考虑了用户隐私保护与里程焦虑	训练时间较长
	[73]	GCN+DDPG	引入了参数共享, 优化精度高	未考虑电网侧安全可靠目标



## 附录 E

1) 针对 GNN 计算速度慢以及处理大规模动态图难度大的问题, 未来需要研究能够有效提高 GNN 处理大规模图结构与动态图结构能力的方法, 例如图采样聚合框架 (graph sample aggregate, GraphSAGE)<sup>[74]</sup>。该框架在训练过程中抽取部分节点形成子图, 聚合操作仅基于子图进行, 这种方式可能损失部分信息, 但能够显著降低计算复杂度, 提高训练效率。例如, 在节点数量较多的配电网图结构当中利用 GraphSAGE 框架进行子图采样, 使得参与图特征编码的图规模减小, 降低图卷积操作或边权计算的计算量。更进一步考虑, 可以针对子图采样进行采样优先级的划分, 例如依据其他节点与目标节点之间的电气距离或灵敏度等相关性指标来确定各个节点的采样优先级, 在提高图计算效率的同时也能够确保采样到的子图节点信息对目标节点有较强的重要性, 减少采样随机性的影响。

2) 针对安全约束问题, 当前已有许多关于安全 DRL 及其在电力系统中应用的研究<sup>[75-78]</sup>。安全约束手段主要可以归纳为动作修正与安全策略改进两种类型<sup>[79]</sup>。动作修正指对智能体输出的动作进行安全性判别, 通过安全层和投影等方式将不安全动作调整为安全动作; 安全策略改进方式则利用数学模型对策略更新进行指导, 确保在安全的约束域中更新。这些改进方式同样适合运用于 GRL, 未来可以结合 GRL 所输入的图信息特点设计安全约束模块。例如, 可以依据支路容量和支路潮流设计动态邻接矩阵, 对支路的过载现象进行动态建模, 并利用 GAT 的图注意力机制对设备动作与支路过载程度之间的相关性进行自适应学习, 对高过载风险的动作进行修正。

3) 针对数据处理问题, 核心在于提高数据质量, 可行的解决方式是将 GRL 算法与表征学习算法相结合, 例如采用自编码器、生成对抗网络等深

度学习方法对含有噪声、缺损及延迟的数据进行处理。自编码器是一种经典的生成模型, 能够以无监督学习的方式对输入数据进行表征学习, 生成对抗网络则将生成模型与判别器相结合, 采用对抗的方式实现数据生成和数据增强。对于配电网所特有的动态拓扑数据而言, 也可以采用更加前沿的图自编码器 (graph auto encoder, GAE) 进行针对性增强<sup>[80]</sup>。例如, 考虑到重构开关的开合需要一定时间, 导致拓扑信息的更新产生延迟或误差, 因此可以利用 GAE 对节点信息与邻接矩阵进行自编码处理, 解决拓扑数据延迟问题的同时也能够对拓扑延迟和节点噪声之间的关联性进行捕捉。上述方法的缺点在于需要预先收集历史数据进行训练, 对数据量的要求较高, 但能够显著降低 GRL 算法对数据准确性和完整性的依赖, 提高 GRL 在工程实际中应用的可能性, 因此值得进行相关研究与实践。

4) 目前已有许多学者聚焦于 DRL 的隐私保护问题进行了针对性研究, 其中最常见措施是引入联邦学习<sup>[81-84]</sup>。联邦学习是一种去中心化的分布式学习算法, 各个分布式模型独立训练, 观测数据不会被中央模型直接读取, 也不会与其他模型交互, 能够有效保护隐私。中央模型只获取各模型的训练参数, 形成一个集中模型, 再将梯度返回给各个体。对于 GRL 而言, 拓扑信息也可以与其他状态信息一同由各模型分散观测并加密, 因此未来可以针对拓扑信息隐私的特点对联邦 GRL 进行研究。例如, 在算法当中将各 EV 用户、各充电站和各个配电网分区分别作为不同联邦模型分别进行本地训练, 中央模型仅获取各联邦模型的训练参数以及充电站、EV 与配电网物理连接关系构成的异质图, 并利用 GNN 进行训练, 不会获取具体数据信息, 由联邦模型与中央模型组成针对于电力-交通耦合系统的联邦 GRL 框架。该框架如附录 E 图 E1 所示。

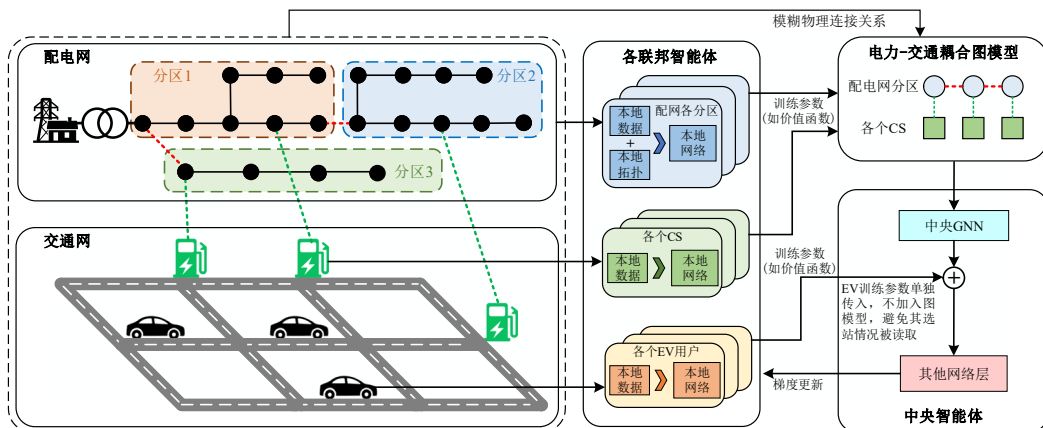


图 E1 面向电力-交通耦合系统的联邦 GRL 框架

Fig. E1 Federal GRL framework designed for coupled power-transportation networks optimization