

基于关系图卷积神经网络的新能源配台区 拓扑识别方法

谢宏¹, 张华赢¹, 梁晓锐¹, 陈煜², 杨林立², 周斌²

(1. 南方电网公司新型智慧城市高品质供电联合实验室(深圳供电局有限公司), 广东 深圳 518020;
2. 湖南大学 电气与信息工程学院, 长沙 410082)

摘要: 针对传统拓扑识别方法难以适应高比例分布式光伏接入下低压配台区电气耦合特性复杂的问题, 提出了一种基于关系图卷积神经网络的新能源配台区拓扑识别方法。文章分析了分布式光伏接入对低压台区线户关系识别的影响机理, 提出了高渗透率分布式光伏接入下配台区的自适应线户关系识别方法, 通过电压皮尔逊相关系数矩阵建模和全局自适应聚类方法实现线户关系识别。基于新能源台区拓扑关联特性将配网节点关联分类匹配为分隔、上下、并列及光伏节点接纳关系, 建立了适应分布式光伏接入的台区拓扑邻接矩阵模型。提出了基于关系图卷积神经网络的配台区拓扑生成算法, 通过提取电压量测数据形成台区节点特征矩阵, 基于关系图链接预测挖掘潜在节点关联关系逐步生成配台区拓扑。算例仿真对比验证了所提拓扑识别方法的有效性, 与传统算法相比可提升识别准确率4.3%以上。

关键词: 配台区; 分布式光伏; 线户关系识别; 拓扑生成; 图神经网络

DOI: 10.19753/j.issn1001-4390.2024.07.014

中图分类号: TM73

文献标识码: B

文章编号: 1001-4390(2024)07-0094-09

A topology identification method based on relational-graph convolutional network for distribution substation area with high renewables

XIE Hong¹, ZHANG Huaying¹, LIANG Xiaorui¹, CHEN Yu², YANG Linli², ZHOU Bin²

(1. New Smart City High-Quality Power Supply Joint Laboratory of China Southern Power Grid (Shenzhen Power Supply Co., Ltd.), Shenzhen 518020, Guangdong, China.

2. School of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: A proposed topology identification method, based on a relational-graph convolutional network, addresses the challenges faced by traditional methods in adapting to the complex electrical coupling characteristics of low-voltage distribution substation areas with a high proportion of distributed photovoltaic (PV) integration. This paper analyzes the influence mechanism of distributed PV integration on the identification of line-user relationships in low-voltage substation areas, and presents an adaptive method for identifying line-user relationships in distribution substation areas with high-penetration distributed PV integration. This method achieves line-user relationship identification through voltage Pearson correlation coefficient matrix modeling and global adaptive clustering. Considering the topology association characteristics of distribution substation areas with high renewables, the paper classifies and matches distribution network node associations into separated, hierarchical, parallel, and PV node acceptance relationships. An adjacent matrix model of substation area topology is established which is specifically adapted to distributed PV integration. This paper proposes a distribution substation area topology generation algorithm based on a relational-graph convolutional network. By extracting voltage measurement data to create node feature portraits for the substation area and employing graph link prediction to uncover potential node association relationships, the distribution substation area topology is progressively generated. A case simulation comparison validates the effectiveness of the proposed topology identification method, which improves identification accuracy by more than 4.3% when compared to traditional algorithms.

基金项目: 中国南方电网有限责任公司科技项目(090000KK52222133/SZKJXM20222115)

Keywords: distribution substation area, distributed photovoltaic, line-user relationship identification, topology generation, graph neural network

0 引言

在“双碳”国家战略目标和“整县光伏”政策背景下,大量分布式户用光伏接入低压配电网,给新能源配电台区拓扑识别带来了全新挑战^[1]。高比例分布式光伏在配电台区的广泛接入使得台区智能电表量测数据量和维度增加,并且引起台区拓扑节点间存在双向潮流,加剧了台区拓扑网络中电压相关性与电气耦合分析难度,成为影响低压配电台区态势感知能力提升的关键瓶颈^[2-3]。研究适应高比例分布式光伏接入下低压配电台区的精准拓扑识别算法,对保障分布式户用光伏友好接入低压配电网和促进台区精益化、智能化运维管控具有重要现实意义。

台区拓扑识别问题可分为两阶段子问题进行求解,包括线户关系识别与小范围节点拓扑生成^[4]。基于数据驱动的线户关系识别方法可分为状态估计、关联分析和数据学习。状态估计类方法需要完整准确的线路参数,且耗时较长^[5-7];关联分析类方法最为常用,通过挖掘智能电表实时电气量测数据识别不同节点之间相关性,来判定节点间的连接关系^[8-11];数据学习类方法将上述基础识别方法与机器学习算法相结合,如 t -SNE + BIRCH 聚类^[12]、GMM 聚类^[13]、DBSCAN 聚类^[14]等。

目前配电台区拓扑生成算法大多通过各类传统搜索或排序方法提取目标数据中的特征分析节点连接关系,如贪婪算法^[15]、马尔可夫随机场^[16]、生成树算法^[17]等。此外人工智能技术在拓扑生成领域也有了更广泛的应用,如贝叶斯网络^[18]、知识图谱技术^[19]、深度神经网络^[20] (Deep Neural Networks, DNN)、支持向量机^[21]等。现有拓扑识别技术大多采用关联分析法提取配电台区电气量间的相关特性,未考虑到大量分布式光伏接入后配电台区双向潮流和电压抬升导致的复杂节点电气量关联特性,因此适应高比例分布式光伏接入的配电台区拓扑识别是一个亟需解决的科学问题。

针对分布式光伏接入低压配电网为台区线户关联分析与拓扑识别带来的问题与挑战,文中提出了一种基于关系图卷积神经网络的新能源配电台区拓扑识别方法,主要创新点如下:1)提出了高渗透率分布式光伏接入下配电台区的自适应线户关系识别方法,分析了分布式光伏接入对低压台区线户关系识别的影响机理,通过电压皮尔逊相关系数矩阵建模和全局自适应聚类方法实现线户关系识别;2)提出了基于关系图卷积神经网络的配电台区

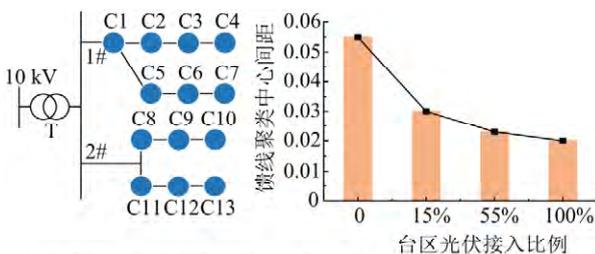
拓扑生成算法,将台区节点关联分类匹配为分隔、上下、并列及光伏节点接纳关系,建立了适应分布式光伏接入的台区拓扑邻接矩阵模型,通过提取台区电压特征矩阵和关系图链接预测挖掘潜在节点关联关系逐步生成配电台区拓扑。

1 新能源配电台区的自适应线户关系识别方法

1.1 分布式光伏接入对台区线户关系识别影响分析

传统无源配电网沿线节点电压幅值在同一时间断面呈现逐渐下降的规律,相邻节点间的电压变化关联度总是趋近1,并高于非相邻节点,故现有配电台区拓扑识别方法多利用节点电压关联特性进行分析^[20-21]。高比例分布式光伏接入后配电台区由单源变为多源,同时由于户用光伏的装机容量、内部结构及控制方式等不同,光伏接入点处节点电压幅值出现不同程度的抬升,台区节点电压关联特性发生改变,传统拓扑识别方法准确性受到影响,不再适用。

以图1(a)所示典型配电台区为例,分析不同分布式光伏接入比例对台区线户关系识别的影响规律,计算节点电压的皮尔逊相关系数矩阵,并从相关系数矩阵中选取初始聚类中心。采用 K-means 聚类算法分析台区不同馈线的聚类中心间距如图1(b)所示。



a) 典型配电台区拓扑结构图 b) 不同光伏接入比例下聚类效果图

图1 分布式光伏接入对配电台区的影响分析

Fig. 1 Analysis of the impact of distributed photovoltaic integration on distribution substation areas

由结果可知,随着分布式光伏接入比例升高,馈线电压聚类中心逐渐接近,不同馈线簇边界模糊化,其中最接近的一组聚类中心相距仅0.02,使得关联系数阈值设定的微量偏移可能导致识别结果出现较大误差。此外,除了同一馈线上的相邻节点电压相关性高,不同馈线首端节点电压也存在较高的相关性,使得馈线簇内距离不平衡,不同馈线首端节点也可能被分到同一簇中,导致线户关系识别准确率下降。因此需提出一种基于自适应相关系数法的线户关系识别算法,以实现高比例分布式光

伏接入下台区节点电压关联特性的准确识别。

1.2 考虑分布式光伏接入的自适应线户关系识别

针对传统关联分析方法依赖单组或多组局部相关性数据,台区线户关系识别准确性受分布式光伏接入影响的问题,本节提出了基于皮尔逊相关系数和AP(affinity propagation)聚类的自适应算法。充分考虑节点电压数据集中整体特征,对所有节点而非某个特定节点间的相关系数进行全局自适应,以完成新能源配电台区的线户关系识别业务。通过计算皮尔逊相关系数矩阵度量节点电压数据间的相似度距离。首先利用一阶差分处理电压时序数据,消除序列中的随机性趋势,如噪声干扰、量测误差等,并对数据进行 min-max 标准化,以便于特征提取及后续算法运行,然后进行皮尔逊相关性矩阵^[10]计算,其矩阵元素作为后续节点电压聚类算法的数据基础。

采用 AP 聚类算法将所有台区节点都当作潜在聚类中心,从全局角度综合考虑所有节点间的相关系数,捕捉数据的内在特征。通过对节点电压相关系数矩阵进行聚类运算,迭代遍历台区内所有节点找寻合适的聚类中心及分类结果,自适应地完成分类任务^[22]。具体过程为根据节点电压数据间的相似度,迭代计算吸引信息和归属信息,最终得到聚类结果,主要参数设置如下:阻尼系数 $l=0.8$,各节点参考度 $P(i) = \min(S)$,表征节点 i 作为聚类中心的可能性,其中 S 为相似度矩阵中的主对角元素 $S(i, i)$ 。文中考虑量测数据集中不包含各馈线首端量测信息,将皮尔逊相关系数与 AP 聚类算法相结合,提出一种新能源台区线户关系自适应关联性识别方法,根据相关系数矩阵及数据内在关联特性动态地推理节点与馈线之间的归属关系,具有全局自适应、无需预定义聚类中心及考虑相似度距离的优势,适用于处理新能源台区线户相关系数高度接近的线户关系识别问题。算法流程图如图 2 所示。

2 基于关系图卷积神经网络的台区拓扑生成算法

2.1 考虑分布式光伏接入的台区拓扑邻接矩阵建模

分布式光伏接入下配电台区拓扑信息较传统台区更为丰富,不同节点间关联特性不能简单划分为连接或不连接,应将台区拓扑建模为多关系异质图来全方位构建配电台区节点特征矩阵,挖掘节点关联特性,从而提升台区拓扑识别的准确性。在进行拓扑生成时,为避免图内信息冗杂,对配电台区拓扑建模做出如下简化处理:1) 删除多余的测量点和分支汇合点,保留待识别节点,如用户节点、光伏节点;2) 开关仅保留常闭开关,并简化为普通的电力线路;3) 合并各种多余的电力线路,相邻的不同节点间仅保留一条连接线。

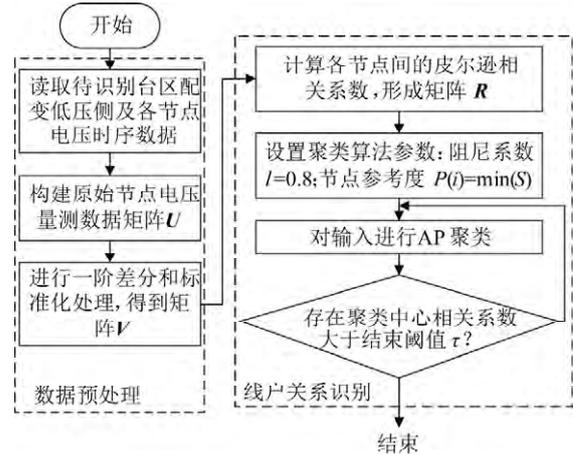


图 2 自适应线户关系识别方法流程图

Fig. 2 Flow chart of adaptive line-user relationship identification method

本节通过设置时间窗长度,对节点电压时序数据进行等长的时间序列截取,构建配电台区节点特征矩阵,在此基础上通过知识匹配对新能源台区节点实体与关联关系进行分类,并按照(头部节点,连接关系,尾部节点)三元组形式建模。为了清晰细致地刻画新能源配电台区节点间的连接关系,基于拓扑关联知识将三元组实体关联关系划分为分隔关系、上下关系、并列关系、接纳关系。节点连接关系类型如表 1 所示,示意图如图 3 所示。

表 1 节点连接关系类型

Tab. 1 Node connection relationship type

关系类型	分隔关系	上下关系	并列关系	接纳关系
关系表示	$x_{ij} = x_{ji} = 0$	$x_{ij} = -x_{ji} = 1$	$x_{ij} = x_{ji} = -2$	$x_{ij} = -x_{ji} = 0.5$
电气距离	较远	较近	非常接近	非常接近

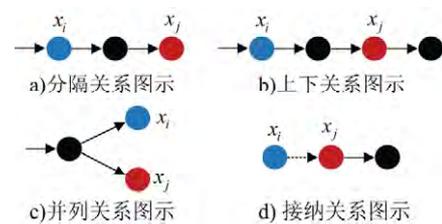


图 3 节点连接关系类型示意图

Fig. 3 Schematic diagram of node connection relationship types

节点连接关系具体表示如下:1) 分隔关系:若两节点不存在直接的电气连接,则认为二者互为“分隔关系”。此处“直接电气连接”指以两节点为端点的连线上不存在任何其他节点;2) 上下关系:若两节点存在直接的电气连接,且存在稳定的功率输送方向,则认为二者互为“上下关系”,其中电气距离上较接近配变的节点为上级节点;3) 并列关系:若两节点不存在直接的电气连接,但具有共同的功率来

源方向 则认为二者互为“并列关系”,用于描述台区拓分支点附近线路连接情况;4) 接纳关系:若两节点存在直接的电气连接,且存在不稳定的功率输送,则认为二者互为“接纳关系”。接纳关系专门用来描述光伏节点与其接入的用户节点间的关系。

基于上述节点连接关系类型及表示方式,可采用有向邻接矩阵描述配电台区多关系异质图,并将其作为后续拓扑信息的存储形式,如图 4 所示。

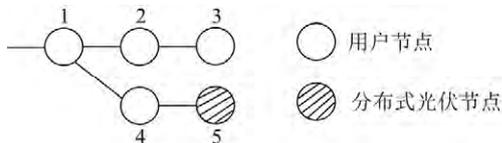


图 4 新能源配电台区局部拓扑

Fig. 4 Local topology of the renewable energy distribution substation area

以图 4 所示的新能源配电台区局部拓扑为例,其可表示为矩阵 Y :

$$Y = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ -1 & 0 & 1 & -2 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ -1 & -2 & 0 & 0 & -0.5 \\ 0 & 0 & 0 & 0.5 & 0 \end{bmatrix} \quad (1)$$

需要指出的是,文中考虑先验拓扑知识中不包含节点类型信息,故对配电台区的拓扑生成也包含了节点类型的预测任务。

2.2 基于关系图卷积网络的小范围拓扑生成算法

文中采用关系图卷积神经网络进行分布式光伏接入下配电台区的小范围拓扑生成。关系图卷积神经网络 (relational-graph convolutional network, R-GCN) 是图卷积神经网络 (graph convolutional network, GCN) 的推广,针对的是异质图建模,以 GCN 节点更新方式为基础,考虑到了节点间不同关系类型,在处理邻居节点时对于每一种关系的邻居引入不同的权重参数,分别对属于同一关系类型的邻居聚合之后进行总聚合,能够承担分布式光伏接入下配电台区的拓扑生成问题^[23-24]。

为充分挖掘新能源台区节点间复杂电气量关联特性,通过 R-GCN 网络对新能源配电台区节点连接关系进行链接预测^[25]。以 R-GCN 网络为编码器,通过 R-GCN 网络提取节点电压时序数据特征,将节点时序电压的高维特征映射为图中实体节点的低维向量,并获取节点连接关系矩阵。R-GCN 编码器根据输入的节点特征矩阵更新节点信息,计算公式如下:

$$H^0 = V = [V_1, V_2, \dots, V_N] \quad (2)$$

$$H_i^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in N_r^i} \frac{1}{c_{i,r}} W_r^{(l)} h_j^l + W_0^{(l)} h_i^l \right) \quad (3)$$

式中 H^0 为输入网络的初始信息,即节点时序电压数据; $H_i^{(l+1)}$ 表示台区节点 v_i 隐藏特征信息; σ 为 ReLU 激活函数; $W_r^{(l)}$ 为正则化权值矩阵; $W_0^{(l)}$ 为自身权值矩阵; $R = \{ \text{分隔关系, 上下关系, 并列关系, 接纳关系} \}$; r 表示在集合 R 下的第 r 种关系; N_r^i 表示在关系为 r 时,与节点 i 相连的邻居节点集合; $c_{i,r}$ 为预先设定的标准化常数,通常设定 $c_{i,r} = |N_r^i|$ 。

台区内每个节点对的关联关系唯一,通过对节点关联关系进行推理可生成台区局部拓扑图。首先通过聚合每种关系下邻居节点信息来完善每个节点实体特定的关联关系,然后将所有关系下的邻居节点信息进行向量归一化叠加,以获取每个节点实体的总邻居节点信息,最后整合本节点实体历史信息及总邻居节点信息获取当前节点 v_i 的隐藏特征信息。

将式 (2) 和式 (3) 应用于多关系台区节点数据计算时, R-GCN 网络参数迅速增多导致出现过拟合问题,为此引入基函数分解和块对角分解来定义模型的权重更新^[23]。基函数使得不同关系类型的权重可以共享,分解公式如下所示:

$$W_r^l = \sum_{b=1}^B a_{rb}^l V_b^l \quad (4)$$

式中 W_r^l 为基 V_b^l 与系数 a_{rb}^l 的线性组合,系数 a_{rb}^l 只依赖于关系 r 。块对角分解为对每个关系类型的权重矩阵进行稀疏约束,可通过对低维矩阵直接求和得到,公式如下所示:

$$W_r^l = \bigoplus_{b=1}^B Q_{br}^l \quad (5)$$

式中 W_r^l 为由低维稀疏矩阵 Q_{br}^l 直接求和得到的对角矩阵,可看作不同关系类型的稀疏约束。

由于所提方法已事先定义节点间的所有可能连接关系,故样本中只有正例没有负例,需通过均匀采样等概率地替换三元组中任一节点构造负例以提升模型泛化能力^[24]。对某一确定关系,判断两个节点实体间是否存在该关系,若存在则将该三元组作为正样本,通过改变正样本中任一节点生成错误的三元组作为负样本。假设 T 为所有节点连接关系三元组集合, F 为生成的负例集合, E 为所有节点实体集合,则生成方法如下所示:

$$F_r = \left\{ (s, r, \rho) \mid r_F \in E \wedge r_F \neq r \wedge (s, r, \rho) \in E \right\} \quad (6)$$

式中 (s, r, ρ) 为节点连接关系真实三元组, (s, r_F, ρ) 为节点连接关系错误三元组。

根据 R-GCN 网络输出的节点关联关系邻接矩阵,结合 DistMult 解码器对台区节点间的连接关系进行链接预测。通过对正负样本三元组的相似性进行评分,获取节点实体与关系间的相关程度,以评估训练过程中出现的三元组 (s, r, ρ) 、 (s, r_F, ρ) 成

立的可能性,分值越高代表事实成立的可能性越大。以评估正样本三元组为例,计算公式为:

$$f(s, r, \rho) = e_s^T R_r e_\rho \quad (7)$$

式中 e_s, e_ρ 分别表示节点三元组头尾实体向量, R_r 为关系 r 的矩阵向量表示。

将 DistMult 解码器评分结果使用交叉熵损失进行优化,如式(8)所示,使模型可观测的正三元组样本评分高于负三元组样本。根据评分结果进行排序,得分最高的三元组作为预测拓扑结果输出,即:

$$L = - \frac{\sum_{(s, r, \rho) \in T} y \log(f(s, r, \rho)) + (1 - y) \log(1 - l(f(s, r, \rho)))}{(1 + \omega) |\mathcal{E}|} \quad (8)$$

式中 T 表示真实三元组的集合; l 为 sigmoid 函数; y 为负样本指示器,当 $y = 1$ 时表示该样本为正样本,当 $y = -1$ 时表示该样本为负样本。

综上,基于 R-GCN 网络的台区拓扑生成算法步骤流程图如图 5 所示。

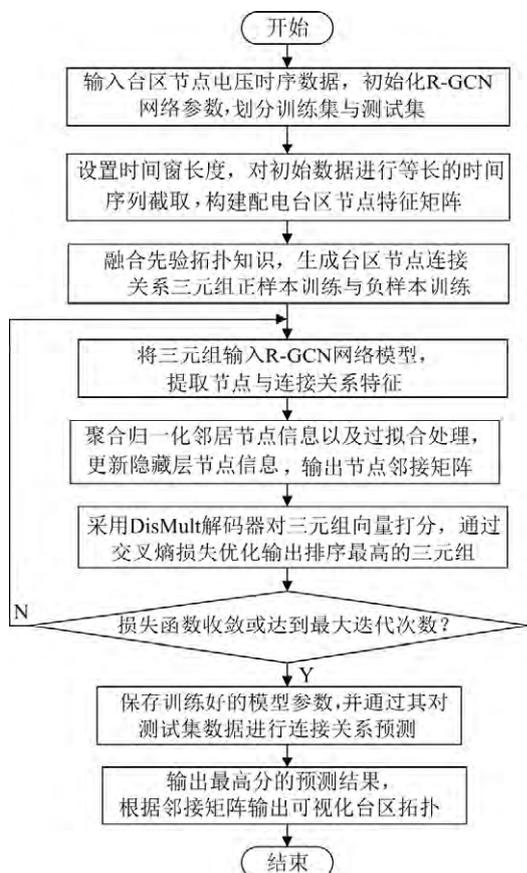


图 5 基于 R-GCN 网络的台区拓扑生成流程图

Fig. 5 Flow chart for generating substation area topology based on R-GCN network

3 算例分析

3.1 系统参数设置

选取不同节点规模、不同采样时段、不同光伏

接入比例的 6 个典型农村和城市低压配电台区作为拓扑识别算例来验证所提算法性能。主要算例参数说明如下: 1) 采样时段即为量测周期内的时刻点数量; 2) 分布式光伏接入比例定义为配电台区中分布式户用光伏的并网容量与总装机容量的比值。算例参数设置见表 2^[4, 9, 26]。台区主干线采用 LGJ-70 的架空导线, 分支线采用 LGJ-50 的架空导线, 下户线采用 LGJ-16 的架空导线。用户节点之间的线路长路设置 10 m ~ 200 m 不等。

表 2 算例典型配电台区参数设置

Tab. 2 Typical parameters setting for case study of distribution substation areas

算例编号	台区配变编号	用户节点数	光伏节点数	节点总数	光伏接入比例/%	出线数
1	A	13	2	15	13.3	2
2	B	30	28	58	48.3	2
3	C	59	10	69	14.5	3
4	D	81	35	116	30.2	4
5	E	92	28	120	23.3	6
6	F	110	20	130	15.4	4

算例采用 MATLAB/Simulink 平台建模, 计算环境为 Intel Core i5-10300 CPU, 2.50 GHz, 内存 16 GB, 操作系统 Windows 10, 在 TensorFlow 环境中实现 R-GCN 网络运行, 训练参数设置见表 3。其中以所搭建算例台区总节点和边的数量为 R-GCN 样本数量, 并按照 4:1 的比例划分训练集与测试集。

表 3 R-GCN 训练参数表

Tab. 3 Training parameters table of R-GCN

系统参数	变量符号	参数值
初始学习率	a_0	0.01
隐藏层层数	n	1
输入特征维度	in_feat	72
输出特征维度	out_feat	12
隐藏层单元个数	h_dim	20
迭代训练预设值	$Epoch$	300
优化器设置	$Optimizer$	Adam
自环丢失率	$Self_dropout\ rate$	0.2
其他边丢失率	$Other_dropout\ rate$	0.4
训练集样本数量	$Train_num$	70 758
测试集样本数量	$Test_num$	17 689

为对比各算法优劣, 文中引入混淆矩阵来评估拓扑识别准确率^[3]。对于线户关系识别问题, 可将各节点相对出线的识别结果分为“属于该出线”与“不属于该出线”两类; 对于拓扑生成问题, 亦可将各节点相对其他节点的识别结果简化为“与该节点有关联”与“与该节点无关联”两类。由此定义算法识别准确率 ACC 如下:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

式中 TP 为实际有关联且识别为有关联的节点数; FP 为实际无关联但识别为有关联的节点数; FN

为实际有关联但识别为无关联的节点数; TN 为实际无关联且识别为无关联的节点数。准确率反映了算法在所有节点中正确分类的比例。

拓扑识别准确性与 FP 、 FN 均有关联,故引入 F_1 分数,作为 ACC 的辅助评价指标,其计算流程为:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (10)$$

式中 P 和 R 分别代表识别结果的精确率和召回率,其中精确率反映算法对正例预测的准确性,计算式为 $P = TP / (TP + FP)$,召回率反映算法对正例的覆盖程度,计算式为 $R = TP / (TP + FN)$ 。由于文中所涉及分类问题均为多分类问题,故在求 ACC 、 P 、 R 及 F_1 分数时采用宏平均值来进行综合评价。

3.2 算例结果分析

从 6 个算例台区获取节点电压量测数据用于线户关系识别,并将聚类结果与实际拓扑进行比较,在不同采样时段下的识别效果如图 6 所示。结果表明,所有台区在获取 3 天(72 个数据点)的条件下识别准确率均已达到 100%,说明算法对采样时段要求不高。其中台区 A 在各采样时段下均能维持 100% 的识别准确率;台区 B、E、F 在采样时段为 48 时也能实现 100% 的准确识别,证明了所提考虑分布式光伏接入的自适应线户关系识别算法的有效性。

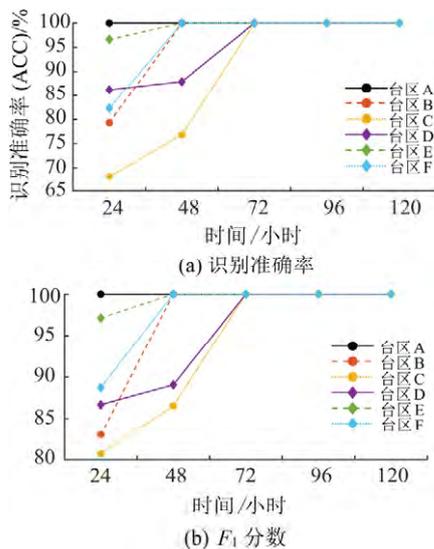


图 6 各台区在不同采样时段下的线户关系识别效果
Fig. 6 Line-user relationship recognition effectiveness of each substation area under different sampling periods

然后对基于关系图卷积网络的小范围拓扑生成算法有效性进行验证。将节点特征(截取电压时序数据获取)和邻接矩阵(同一馈线节点连接关系矩阵为元素均为 1 的方阵)输入模型,利用 R-GCN 网络对连接关系进行预测,将输出的新能源台区有向邻接矩阵可视化生成配电台区拓扑,其中典型

130 节点台区 F 拓扑生成结果如图 7 所示。



图 7 典型 130 节点台区拓扑生成结果图
Fig. 7 Typical topology generation result diagram of a 130-node substation area

通过各算例台区拓扑生成结果与真实拓扑的对比可知,经训练的 R-GCN 网络模型拥有良好的泛化能力,可在不同规模、不同光伏接入比例的配电台区中精准刻画台区拓扑。

在不同节点特征维度下进行各算例台区的拓扑生成,对结果取平均值进行对比分析,识别效果如表 4 所示。本算法中输入模型的节点特征维度即节点电压时序采样时段。

表 4 不同特征维度下拓扑生成准确率

Tab. 4 Topology generation accuracy in different characteristic dimensions 单位: %

特征维度	12	24	36	48	60	72
台区 A	86.15	93.71	99.48	99.60	99.70	99.78
台区 B	80.59	93.15	99.04	99.30	99.45	99.51
台区 C	85.34	92.88	97.95	99.19	99.30	99.48
台区 D	73.94	91.39	97.13	98.97	99.16	99.37
台区 E	80.69	89.99	95.78	98.92	99.04	99.21
台区 F	78.05	89.58	94.20	98.80	99.03	99.17

随着节点特征维度的增加,拓扑生成准确率总体上保持增长趋势。当节点维度为 12 时,数据集未能涵盖全天的数据特征,光伏节点接纳关系判断不稳定,因此具有较高光伏接入比例的台区 B 和台区 D 在节点特征为 12 时识别准确率较低。当节点维度为 24 时,拓扑生成准确率快速上升,所有台区的识别准确率均达到了 89%,但仍不能满足工程实践要求。当节点维度达到 36 时,所有台区的识别精度趋于稳定,但在复杂台区中还存在进一步提升的空间。而当节点维度大于等于 48 时,全算例台区拓扑生成准确率均达到 98% 以上,可满足各种拓扑生成场景的需求,且再继续增大节点特征维度对识别准确率的提高效果不再明显。此外台区 F 各馈线下节点均为并列关系,而训练集中并联关系标签最少,导致台区 F 几乎在所有数据维度下的识别准确率均为最低。因此,在进行模型训练时,一方面为了保证计算速度和识别效果,选定节点维度为

48 最优; 另一方面对于各种节点连接关系, 在条件允许的范围内, 应尽可能增加数据集中样本容量, 以提升拓扑识别的效果。

图 8 为训练损失函数值随迭代次数变化的曲线。在迭代次数小于 150 时, 模型训练损失函数值整体水平较高, 且存在较大波动; 当迭代次数大于 150 时趋向于平稳, 训练损失函数值约为 0.46。结合表 4 可知, 当训练迭代次数为 300 时, 算法可以有效准确地生成台区拓扑, 在节点维度高的情况下各台区的生成准确率均能达到 99%, 证明了迭代计算收敛, 迭代损失较低, 能满足工程应用需求。

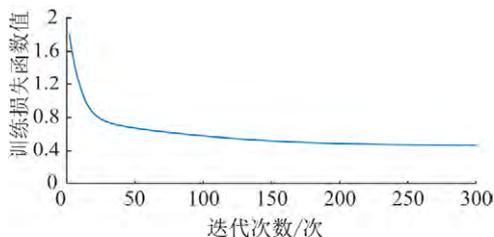


图 8 R-GCN 算法训练损失函数曲线

Fig. 8 Training loss function curve of R-GCN algorithm

3.3 算法对比分析

本节固定采样时段为 72, 基于表 2 的数据场景, 对比分析以下 4 种线户关系识别算法: 方法 1: 欧式距离 + AP 聚类算法^[27]; 方法 2: Pearson 相关系数 + K-means 聚类算法^[28-29]; 方法 3: Pearson 相关系数 + KNN 聚类算法^[30]; 方法 4: 文中所提出的自适应聚类方法。不同算法线户识别效果如图 9 所示。

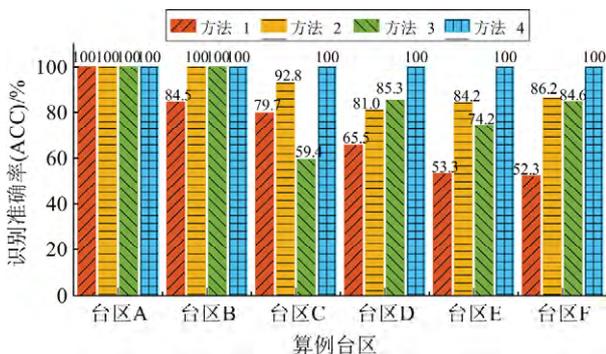


图 9 不同线户关系识别算法结果对比

Fig. 9 Comparison of different line-user relationship recognition algorithms results

可得如下结论:

1) 比较方法 1 与方法 4。随着节点数量的增加, 方法 1 识别准确率逐渐下降, 而方法 4 能适应不同节点规模的识别场景。这是由于分布式光伏接入下配台区不同节点电压曲线形态差异较小, 电压关联特性混乱, 以欧式距离作为聚类分析判断依据不再

适用, 采用皮尔逊相关系数的方法 4 更具优越性。

2) 比较方法 2 与方法 4。受分布式光伏接入影响, 方法 2 可疑节点的划分可能出现误差, 后续 K-means 聚类将进一步放大误差, 导致识别准确率下降。其中台区 D 各馈线的光伏接入比例高且接近, 相关系数的分布范围高度重合, 因此方法 2 在台区 D 效果最差。而方法 4 综合考虑了所有节点间的相关性, 识别准确率更加稳定。

3) 比较方法 3 与方法 4。随着馈线数量的增加, 方法 3 的应用效果逐渐下降, 在台区 E 中识别准确率低至 74.2%, 分析可知该台区馈线多, 不同馈线间的电气距离相较馈线上节点间的电气距离更接近, 大部分节点相关性集中, 导致初始聚类中心选择困难, 且位于馈线上游的节点易被划分到其他线路, 识别准确率降低。而方法 4 能自动搜寻合适的聚类中心, 克服因聚类中心及聚类参考点选取不当带来的偏差, 在各台区均能达到 100% 的识别准确率。

综上可知, 在固定时间长度的条件下, 方法 4 相较其他三种方法取得了更精准的识别效果, 在复杂台区中线户关系识别准确率可至少提升 7.2%, 证明了其在新能源配电台区线户关系识别中的优越性。

同样将文中所提配电台区拓扑生成算法与其他常用的非机器学习算法以及机器学习算法进行比较。所比较的算法如下: 贪婪算法^[15] (greedy algorithm, GA)、深度神经网络^[20] (deep neural networks, DNN)、支持向量机^[21] (support vector machines, SVM) 以及图卷积神经网络 (graph convolutional networks, GCN)。结果如表 5 ~ 表 7 所示。

表 5 不同算法的拓扑生成准确率

Tab. 5 Accuracy of topology generation for different algorithms 单位: %

台区	节点个数	GA	DNN	SVM	GCN	R-GCN
台区 A	15	95.29	94.63	97.47	96.93	99.60
台区 B	58	93.58	93.68	93.69	96.38	99.30
台区 C	69	87.31	93.83	92.53	95.50	99.19
台区 D	116	83.05	93.45	90.86	95.12	98.97
台区 E	120	74.73	93.17	88.57	95.12	98.92
台区 F	130	10.28	92.84	86.96	94.50	98.80

表 6 不同算法的拓扑生成精确率

Tab. 6 Precision of topology generation for different algorithms 单位: %

台区	节点个数	GA	DNN	SVM	GCN	R-GCN
台区 A	15	95.47	94.70	96.32	96.93	99.26
台区 B	58	93.32	93.05	93.28	96.17	98.75
台区 C	69	87.20	93.43	92.13	95.64	98.61
台区 D	116	83.98	92.82	91.69	95.43	98.44
台区 E	120	74.05	92.78	89.42	95.78	98.46
台区 F	130	9.71	92.82	88.16	95.03	98.37

表7 不同算法的拓扑生成召回率

Tab. 7 Recall of topology generation for different algorithms 单位: %

台区	节点个数	GA	DNN	SVM	GCN	R-GCN
台区 A	15	95.12	94.56	92.54	96.93	99.95
台区 B	58	93.85	94.31	95.04	96.59	99.85
台区 C	69	86.42	94.23	89.10	95.36	99.77
台区 D	116	84.12	94.08	87.93	94.82	99.51
台区 E	120	76.42	93.57	84.73	94.46	99.38
台区 F	130	10.50	92.86	80.79	93.97	99.23

从模型结构方面对各算法进行分析,GA对相关性进行排序后按照序次进行连接,无法识别并列关系;DNN、SVM存在因计算量过大导致的模型训练过拟合、训练时间过长等缺陷;GCN网络层间连接稀疏及权值共享机制可大幅度减少训练参数的个数,降低模型计算量,且其池化层会对无关变量进行削减从而降低计算量。R-GCN充分聚合各个节点及其邻居节点的信息,且在GCN的基础上考虑了不同关系的不同权值,有利于节点连接关系的精准刻画,更适用于复杂连接关系的台区拓扑生成。

由生成结果可知,随着台区节点个数的增加,各算法拓扑生成准确率均有所下降,其中R-GCN算法在复杂台区中效果最佳。实际场景中台区用户规模为100左右,少数规模较大的台区能达到200以上,限制了机器学习算法的样本规模。R-GCN可通过负采样环节扩充样本容量,弥补了这方面的不足,相较其他方法更具优势。以负荷高度密集的典型城市台区F为例,由于无法识别并列关系,GA近乎失效,R-GCN较其他机器学习算法生成准确率、精确率、召回率分别可至少提升4.30%、3.34%、5.26%。

4 结束语

高比例新能源广泛接入低压配电台区,导致台区拓扑网络中电气耦合特性发生本质变化,传统拓扑识别方法准确率受影响。针对此问题,文中提出了基于关系图卷积神经网络的拓扑识别方法,能够适应分布式光伏接入下的配电台区拓扑识别场景。文中的主要贡献及结论如下:

1) 分析了分布式光伏接入对台区线户关系识别的影响规律;提出了自适应线户关系识别方法,综合皮尔逊相关系数与AP聚类算法的优势,克服了现有方法难以确定关联系数阈值的缺陷,对比其他算法,在复杂台区中线户关系识别准确率可至少提升7.2%。

2) 提出了基于关系图卷积神经网络的小范围节点拓扑生成方法。融合新能源配电台区特征及邻接矩阵定义,采用4种连接关系刻画台区拓扑连接图,以先验拓扑和电压数据构建配电台区的节点关系画像,通过关系图卷积神经网络进行链接预测,生成配电台区完整拓扑。在各类场景中拓扑生

成准确率、精确率和召回率较机器学习类算法可至少提升4.30%、3.34%、5.26%。

参考文献

- [1] 刘晟源,章天晗,林振智,等.数据赋能低压配用电系统精益化运行的关键技术与算法[J].电力系统自动化,2023,47(3):187-99.
LIU Shengyuan, ZHANG Tianhan, LIN Zhenzhi, et al. Key technologies and algorithms of data empowerment for lean operation of low-voltage power distribution and consumption system [J]. Automation of Electric Power Systems, 2023, 47(3): 187-99.
- [2] BHATTARAI BISHNU P, PAUDYAL SUMIT, LUO YUSHENG, et al. Big data analytics in smart grids: state of the art, challenges, opportunities, and future directions [J]. IET Smart Grid, 2019, 2(2): 141-54.
- [3] 徐明昕,赵健,王小宇,等.基于电压聚类 and 关联卷积的配电网户变关系识别方法[J].电力系统保护与控制,2022,50(4):103-111.
XU Mingxin, ZHAO Jian, WANG Xiaoyu, et al. Transformer-customer identification method for a low-voltage distribution network based on voltage clustering and incidence convolution [J]. Power System Protection and Control, 2022, 50(4): 103-111.
- [4] 胡选正,丛伟,古千硕,等.基于电表量测数据的低压供电网络拓扑识别方法[J/OL].电力自动化设备:1-17 [2024-05-13].
HU Xuanzheng, CONG Wei, GU Qianshuo, et al. Topology identification method for low-voltage power supply network based on meter measurement data [J]. Electric Power Automation Equipment, 1-17 [2024-05-13].
- [5] 张爱军,刘紫玉,邢华栋,等.基于GM估计的电-气-热综合能源系统分解协调鲁棒状态估计方法[J].智慧电力,2023,51(10):9-14,53.
ZHANG Aijun, LIU Ziyu, XING Huadong, et al. Decomposition and coordination robust state estimation method for electric-gas-heat integrated energy system based on GM estimation [J]. Smart Power, 2023, 51(10): 9-14, 53.
- [6] 萧展辉,邹文景,唐良运.基于多周期性MILP模型的新型配电系统拓扑辨识方法[J].电测与仪表,2023,60(2):117-125.
XIAO Zhanhui, ZOU Wenjing, TANG Liangyun. Topology identification method of novel distribution system based on multi-period MILP model [J]. Electrical Measurement & Instrumentation, 2023, 60(2): 117-125.
- [7] 刘凯,李镇海,吕利,等.基于聚类分析的配电台区拓扑识别方法[J].电力系统保护与控制,2022,50(6):165-171.
LIU Kai, LI Zhenhai, LYU Li, et al. Topology identification method for distribution areas based on clustering analysis [J]. Power System Protection and Control, 2022, 50(6): 165-171.
- [8] 陈培育,金尧,郑晓麟,等.基于可观性和可计算性的配电网关键量测识别方法[J].智慧电力,2024,52(3):110-116.
CHEN Xiuxiu, JIN Yao, ZHENG Xiaolin, et al. Identification of critical measurement for distribution network based on observability and computability [J]. Smart Power, 2024, 52(3): 110-116.
- [9] 唐捷,蔡永智,周来,等.基于数据驱动的低压配电网线户关系识别方法[J].电力系统自动化,2020,44(11):127-134.
TANG Jie, CAI Yongzhi, ZHOU Lai, et al. Data-driven based identification method of feeder-consumer connectivity in low-voltage distribution network [J]. Automation of Electric Power Systems, 2020, 44(11): 127-134.

- [10]李琮琮,范学忠,王清,等.基于用电信息采集系统的配电网台区识别[J].电测与仪表,2019,56(24):109-114.
- [11]王献志,李铁成,郭少飞,等.基于 μ PMU电压相位数据的配电网拓扑辨识[J].电网与清洁能源,2022,38(6):54-63.
WANG Xianzhi, LI Tiecheng, GUO Shaofei, et al. Identification of distribution network topology based on μ pmu voltage phase data [J]. Power System and Clean Energy, 2022, 38(6): 54-63.
- [12]连子宽,姚力,刘晟源,等.基于 t -SNE降维和BIRCH聚类的单相用户相位及表箱辨识[J].电力系统自动化,2020,44(8):176-184.
LIAN Zikuan, YAO Li, LIU Shengyuan, et al. Phase and meter box identification for single-phase users based on t -SNE dimension reduction and BIRCH clustering [J]. Automation of Electric Power Systems, 2020, 44(8): 176-184.
- [13]徐明昕,赵健,王小宇,等.基于电压聚类 and 关联卷积的配电网户变关系识别方法[J].电力系统保护与控制,2022,50(4):92-102.
- [14]崔雪原,刘晟源,金伟超,等.基于APAA和改进DBSCAN算法的户变关系及相位识别方法[J].电网技术,2021,45(8):3034-3043.
CUI Xueyuan, LIU Shengyuan, JIN Weichao, et al. Consumer-transformer relationship and phase identification based on APAA and improved DBSCAN algorithm [J]. Power System Technology, 2021, 45(8): 3034-3043.
- [15]BARIYA M, DEKA D, VON MEIER A. Guaranteed phase & topology identification in three phase distribution grids [J]. IEEE Transactions on Smart Grid, 2021, 12(4): 3605-3612.
- [16]ZHAO J, LI L, XU Z, et al. Full-scale distribution system topology identification using Markov random field [J]. IEEE Transactions on Smart Grid, 2020, 11(6): 4714-4726.
- [17]PAPPU S J, BHATT N, PASUMARTHY R, et al. Identifying topology of low voltage distribution networks based on smart meter data [J]. IEEE Transactions on Smart Grid, 2017, 9(5): 5113-5122.
- [18]ZHANG H, ZHAO J, WANG X, et al. Low-voltage distribution grid topology identification with latent tree model [J]. IEEE Transactions on Smart Grid, 2022, 13(3): 2158-2169.
- [19]ZEPU G, YONGJIAN L, ZIWEI X, et al. Knowledge graph based method for identifying topological structure of low-voltage distribution network [J]. The Journal of Engineering, 2020(12): 1177-1184.
- [20]裴宇婷,秦超,余贻鑫.基于LightGBM和DNN的智能配电网在线拓扑辨识[J].天津大学学报(自然科学与工程技术版),2020,53(9):939-950.
PEI Yuting, QIN Chao, YU Yixin. Online topology identification for smart distribution grids based on LightGBM and deep neural networks [J]. Journal of Tianjin University(Science and Technology), 2020, 53(9): 939-950.
- [21]杨秀,蒋家富,刘方,等.基于注意力机制和卷积神经网络的配电网拓扑辨识[J].电网技术,2022,46(5):1672-1682.
- [22]FREY B. J., DUECK D. Clustering by passing messages between data points [J]. Science, 2007, 315: 972-976.
- [23]徐冰冰,岑科廷,黄俊杰,等.图卷积神经网络综述[J].计算机学报,2020,43(5):755-780.
XU Bingbing, CEN Keting, HUANG Junjie, et al. A survey on graph convolutional neural network [J]. Chinese Journal of Computers, 2020, 43(5): 755-780.
- [24]侯中妮,靳小龙,陈剑赞,等.知识图谱可解释推理研究综述[J].软件学报,2022,33(12):4644-4467.
HOU Zhongni, JIN Xiaolong, CHEN Jianyun, et al. Survey of interpretable reasoning on knowledge graphs [J]. Journal of Software, 2022, 33(12): 4644-4467.
- [25]CAI L, LI J, WANG J, et al. Line graph neural networks for link prediction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 5103-5113.
- [26]闫卫国,王高猛,林济铿,等.基于AMI量测信息的低压配电网拓扑校验方法[J].中国电力,2019,52(2):125-133.
YAN Weiguo, WANG Gaomeng, LIN Jikeng, et al. Method for LV distribution network topology verification based on ami metering data [J]. Electric Power, 2019, 52(2): 125-133.
- [27]尹鹏,梁海深,顾志成,等.基于欧氏形态距离与AP聚类分析的配电网拓扑结构辨识方法[J].电力系统及其自动化学报,2023,35(9):95-102.
YIN Peng, LIANG Haishen, GU Zhicheng, et al. Identification method of topology structure in distribution area based on euclidean morphological distance and AP cluster analysis [J]. Proceedings of the CSU-EPSCA, 2023, 35(9): 95-102.
- [28]李彦兆,赵庆生,梁定康,等.基于量测信息的低压配电网阻抗拓扑模型构建[J].电网与清洁能源,2021,37(4):15-22,31.
LI Yanzhao, ZHAO Qingsheng, LIANG Dingkang, et al. Construction of impedance topology model for low voltage distribution network based on AMI measurement [J]. Information Power System and Clean Energy, 2021, 37(4): 15-22, 31.
- [29]郭亮,王晓卫,康乾坤,等.基于粒子群优化与K-means聚类的配网5G改造经济性评价方法[J].电网与清洁能源,2022,38(6):31-36,43.
GUO Liang, WANG Xiaowei, KANG Qiankun, et al. An economic evaluation method for 5G transformation of distribution network based on particle swarm optimization and K-means clustering [J]. Power System and Clean Energy, 2022, 38(6): 31-36, 43.
- [30]陈招安,黄纯,张志丹,等.基于T型灰色关联度和KNN算法的低压配电网台区拓扑识别方法[J].电力系统保护与控制,2021,49(1):163-169.

作者简介:



谢宏(1967—)男,硕士,高级工程师,研究方向为智能配电网运行与管理。
Email: xiehong@sz. csg. cn

张华赢(1981—)男,硕士,教授级高级工程师,研究方向为配电网电能质量监测与治理技术。Email: zhytyx@163. com

梁晓锐(1994—)男,博士,高级工程师,研究方向为新能源配电网运行优化、新型储能及碳中和研究。
Email: xrliang@mail. tsinghua. edu. cn

陈煜(2000—)男,博士研究生,研究方向为配电网运行规划与态势感知。Email: yuchen201128@163. com

杨林立(1997—)男,硕士研究生,研究方向为配电网运行规划与态势感知。Email: yllfendou@hnu. edu. cn

周斌(1984—)男,通信作者,教授,博士生导师,研究方向为智能配电网运行优化与管理。Email: binzhou@hnu. edu. cn

收稿日期:2023-08-24; 修回日期:2024-05-08

(任美辉 编发)