

基于多标签分类与卷积神经网络的配电网拓扑辨识

龙 寰¹, 石子晴¹, 赵景涛², 郑 舒², 张晓燕², 谢文强³

(1. 东南大学电气工程学院, 南京 210096; 2. 南瑞集团(国网电力科学研究院)有限公司, 南京 211106;
3. 国网江苏省电力有限公司, 南京 210024)

摘要: 为适应新一代配电网运行特性, 配电网开关需频繁动作调整网络结构, 难以及时、准确获取配电网的实时拓扑结构, 给配电网的态势感知带来一定困难。鉴于传统以状态估计为框架的配电网拓扑识别方法计算复杂度、在线应用困难, 同时大规模配电网拓扑结构多样化, 该文提出了基于多标签分类与卷积神经网络的配电网拓扑辨识方法。通过配电网量测电压数据与开关状态间的多映射关系, 引入多标签分类机制, 对配电网拓扑结构进行多标签编码, 将配电网开关与拓扑辨识模型输出进行物理映射, 利用卷积神经网络搭建多标签分类器, 实现拓扑的准确辨识。基于改进的 IEEE 123 节点配电网算例对所提方法进行验证, 实验结果表明: 所提模型具有较高的拓扑识别准确率, 且对于在训练样本空间外的未知拓扑结构, 其具备更好的推理能力, 更适用于实际拓扑识别的场景, 证实了所提方法的优越性和鲁棒性。

关键词: 配电网; 拓扑辨识; 多标签分类; CNN; 知识外推

Topology Identification of Distribution Network Based on Multi-label Classification and CNN

LONG Huan¹, SHI Ziqing¹, ZHAO Jingtao², ZHENG Shu², ZHANG Xiaoyan², XIE Wenqiang³

(1. School of Electric Engineering, Southeast University, Nanjing 210096, China; 2. NARI Group Corporation (State Grid Electric Power Research Institute) Co., Ltd., Nanjing 211106, China; 3. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China)

Abstract: To adapt to the operation characteristics of the new distribution network, the distribution network switches require frequent adjustments to their structures. However, it is difficult to timely and accurately obtain the real-time topology of the distribution network, which poses challenges for situational awareness of the network. Traditional topology identification methods based on state estimation are difficult to apply online due to their high computational complexity and the large number of topology categories in large-scale distribution network. To address these challenges, this paper proposes a distribution network topology identification method based on multi-label classification and convolutional neural network (CNN). By exploring the multi-mapping relationship between measured voltage data and switch states, a multi-label classification mechanism is introduced to encode the distribution network topology. The switches are physically mapped to the topology identification model output and a CNN is used to build a multi-label classifier, achieving accurate topology identification. Verification of the proposed method is conducted using a revised IEEE 123-node distribution network, and experimental results show that it has a high topology recognition accuracy. Additionally, the model demonstrates better inference capability for unknown topologies outside the training sample space, making it more suitable for practical topology identification scenarios. The superiority and robustness of the proposed method can be verified.

Key words: distribution network; topology identification; multi-label classification; CNN; knowledge extrapolation

0 引言

配电网拓扑是进行潮流计算、建立伪量测、实现配电网态势感知的基础, 也是开展拓扑重构实现配电网经济优化运行的前提条件^[1-3]。随着可再生分布

式电源大规模并网发电, 配电系统的网络结构及运行特性趋于多样化、复杂化, 为配电网的安全可靠运行带来新的挑战。同时, 弹性负荷投切、系统故障、设备检修等事件使配网馈线上的联络开关或分段开关状态不定期发生变化^[4]。现场工作人员在维修后一般不能及时反馈拓扑变动, 局部拓扑的更改信息缺乏详实记录, 另外拓扑传感器数量有限, 遥信信号不满足实时性要求, 精确的拓扑结构通常是

未知的^[5]。因此, 配电网拓扑的实时精准辨识已成为一个重要的研究课题。

配电网拓扑识别方法可大致分为基于状态评估的方法和基于机器学习的方法。状态估计背景下基于物理模型的传统拓扑识别方法时效性差, 可运行拓扑的增加使其计算复杂度呈指数提升^[6]。为解决该问题, 部分学者将开关状态作为附加状态变量利用加权最小二乘法识别拓扑^[7]。总体来说, 基于状态估计的拓扑辨识方法在低量测冗余的配电网中不易实现。针对有限的量测数据, 基于混合整数规划的拓扑辨识方法^[8-9]应运而生, 但伪量测存在较大误差时, 这种方法性能欠佳。

近年来, 随着先进计量基础设施和微型同步相量测量单元(micro-phasor measurement unit, μ PMU)在配电网中的持续部署^[10], 基于机器学习的配电网拓扑辨识方法备受关注。如何在大量多源异构的量测数据中选取模型的输入特征, 是解决数据驱动下拓扑辨识问题的关键^[11-12]。文献[13-14]将节点注入功率和电压量测作为输入, 搭建多分类模型挖掘量测信息空间和配网拓扑间的关联关系。文献[15]利用最小绝对值收敛和选择算子(least absolute shrinkage and selection operator, LASSO)回归研究配电网拓扑估计, 但全部节点配置 μ PMU的假设限制了其实用性。文献[16]基于时间序列签名验证法, 以有限的 μ PMU量测检测拓扑变动, 但该算法不能用于多个断路器同时动作的场景。考虑不同量测数据对拓扑识别贡献的差异, 文献[17-18]筛选出部分电压幅值作为分类模型输入, 以尽量低维度的输入实现拓扑辨识。综上, 节点电压的分布更能反映出配电网拓扑的结构信息。

目前, 基于监督学习算法的拓扑辨识研究大多采用多分类器学习拓扑结构^[13-14, 17-19]。然而, 对于大型配电网, 候选拓扑数量过多, 多分类模型的识别准确度会大幅度降低。同时, 实际中历史样本并不能覆盖所有可能拓扑结构下的运行数据, 通过直接捕捉输入与输出间映射关系的多分类器并不具备足够的泛化能力来推理未知拓扑结构。因此, 本文提出基于多标签(multi-label, ML)分类的配电网拓扑辨识方法, 其特点在于:

1) 与多类别(multi-classification, MC)分类相比, 多标签分类能大幅降低拓扑辨识模型的输出空间维度, 从而降低拓扑类别过多时对分类器学习能力的要求。

2) 通过引入 ML 机制, 搭建起数学模型与物理模型之间的桥梁, 可以增强深度学习方法的可解释性。

3) 基于卷积神经网络(convolutional neural network, CNN)建立多标签分类拓扑辨识模型, 捕捉配电系统量测电压相量与网络开关状态间的非线性关系, 可实现对未知拓扑结构的推理。

本文基于改进的 IEEE 123 节点配电网算例进行仿真分析, 将所提方法与基于随机森林(random forest, RF)、多层感知机(multilayer perceptron, MLP)分类的拓扑辨识算法进行对比, 并对训练集外的未知样本进行测试, 验证了所提方法的优越性和泛化能力。

1 多类别分类与多标签分类

基于机器学习分类算法的拓扑辨识, 需要将历史断面数据对应的拓扑结构进行编码, 用于分类器的构建与训练, 因此编码方式会直接影响模型的学习效果, 特别是多分类问题。

针对多分类问题, 常采用 MC 分类方法。最简单的编码策略即对所有类别按照数字顺序编码, 但对于电网拓扑辨识问题, 由于电网拓扑的类别数量随着电网开关数量呈指数增长, 因此该编码策略不适用于大型配电网的拓扑识别。除单标签数字顺序编码外, 通常使用一对多或一对一的模型架构。

一对多策略将 i 个类别的多分类问题转化成多个二分类问题, 如图 1(a)所示, 图 1 中 T_i 表示第 i 类拓扑样本, L_m 表示第 m 个标签。这种方法依次选择某一类别为正类, 其余类别样本归为负类, 共训练 $i-1$ 个二分类器输出置信度最大的标签作为分类结果。一对多下的负类样本数目明显多于正类样本, 当 i 较大时, 正、负类样本数量相差悬殊, 样本不均衡使二分类器在训练过程中产生偏向性, 分类效果不稳定^[20]。

一对一策略是一种较为稳健的扩展方法, 令不同类别的样本两两组合进行训练, 产生 $i(i-1)/2$ 个二分类器进行投票判定输入样本所属类别, 如图 1(b)所示。一对多架构缓解了上述不平衡问题, 但仍会受到各类别样本本身不均衡问题的影响。此外, 每个二分类器的训练样本仅包含 2 个类别, 存在一定程度的信息丢失问题^[21]。

当分类类别很多时, 这 2 种多类别分类策略都需要训练大批量二分类器来执行分类任务, 内存和时间的消耗往往是难以承受的。此外, 考虑配电网

运行时, 对于负荷大幅度波动、配电网故障运维检修、极端天气等情况, 存在配电网的运行数据样本稀缺的问题。反映到数据模型中, 等价于数据样本缺失引起的多分类模型输出类别的缺失, 导致多类别分类器性能不佳。

分类器的主要任务是建立标签信息空间与输入特征空间的对应关系。MC 分类模型假设单个样本仅属于某一特定类别(单标签), 且不同标签互斥, 而 ML 分类中一个样本可以属于多个类别(多标签), 且不同标签之间存在关联。基于 MC 分类的拓扑辨识模型是学习量测特征和多个开关状态整体组合之

间的映射关系, 本文提出的基于 ML 分类的拓扑辨识模型是学习量测特征和单个联络开关状态之间的多映射关系。

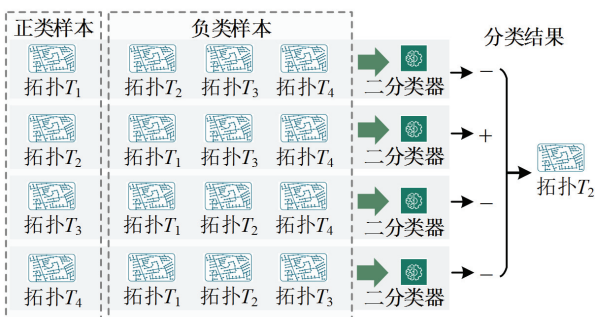
图 1(c)展示了 ML 分类的编码过程, 其中将 5 个标签对应的 0-1 编码作为输入标签进行训练, 样本通过 ML 分类器输出得到各标签对应的分类结果。因此, M 个标签至多可以包含 2^M 个类别, 当类别数目达到成百上千种时, ML 编码可以实现输出编码降维。

2 基于多标签分类与卷积神经网络的配电网拓扑辨识

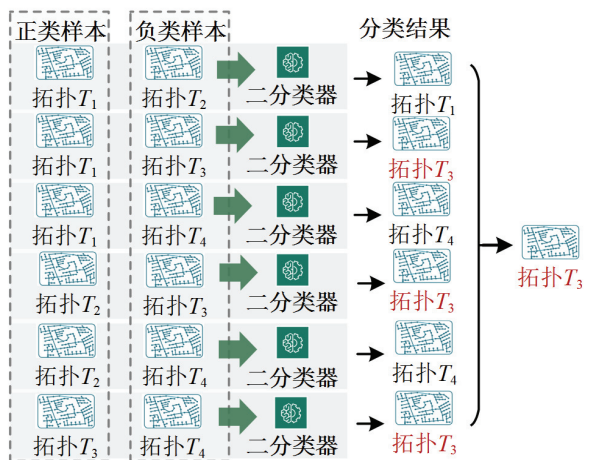
本节将拓扑辨识问题转化为 ML 分类问题, 使用深度学习方法来捕获由开关状态变化引起的共现依赖性。首先介绍基于 ML 分类的拓扑辨识流程, 然后阐释 CNN 的架构和功能, 最后详细说明该拓扑辨识模型的优势。

2.1 基于多标签分类的配电网拓扑辨识框架

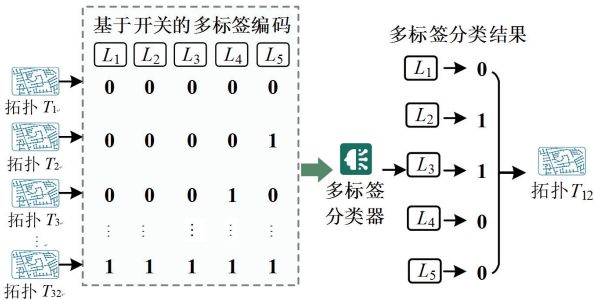
基于 ML 分类的配电网拓扑辨识框架如图 2 所示, 主要由数据集构建、离线训练、在线辨识 3 部分组成。



(a) 多类别分类-一对多策略



(b) 多类别分类-一对一策略



(c) 多标签分类

图 1 多类别分类与多标签分类的比较

Fig.1 Comparison between multi-class classification and multi-label classification

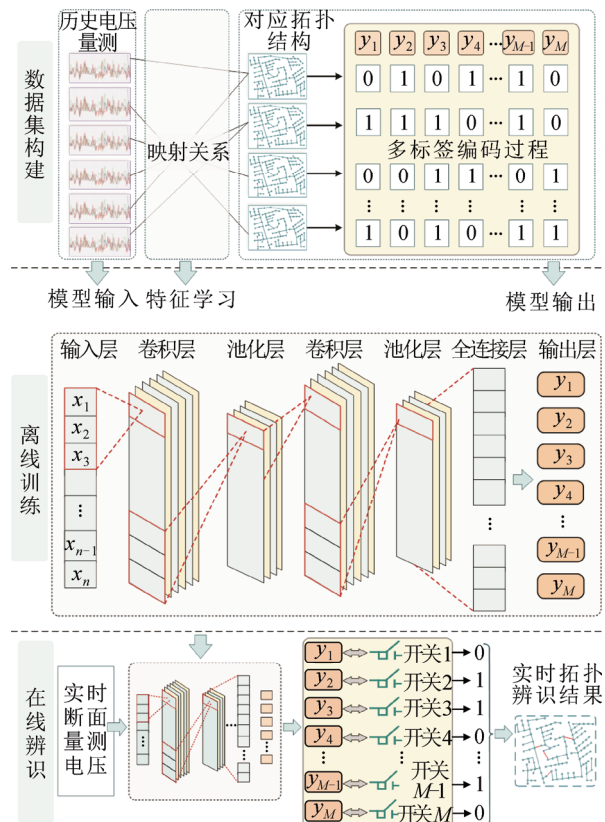


图 2 基于多标签分类的配电网拓扑辨识框架

Fig.2 Framework of distribution network topology identification based on multi-label classification

在数据集构建阶段, 基于配电网监测控制和数据采集(supervisory control and data acquisition, SCADA)与 μ PMU 混合量测系统获得历史电压量测数据库及对应的拓扑结构样本。对历史量测数据集进行数据清理、数据归一化等预处理工作。将样本中节点电压作为模型输入, 对样本拓扑结构进行 ML 编码, 作为模型输出, 将数据集划分为模型训练数据集和测试集。

在离线训练阶段, 基于模型训练集与测试集进行 ML-CNN 的离线训练, 通过模型调参, 训练建立拓扑辨识模型, 完成对历史量测电压数据与全网联络开关工作状态间多重复杂映射关系的学习。

在线辨识阶段, 将训练完成的模型用于在线辨识, 将实时量测数据作为数据输入, 即可得到当前配电网各联络开关的状态, 进行拓扑辨识。

2.2 基于卷积神经网络的多标签分类模型

CNN 作为深度学习的代表算法, 能够从输入信息中提取高阶特征, 具有相当可观的表征学习能力。CNN 架构如图 3 所示, 主要包括特征提取器和分类器 2 大部分, 前者由输入层、卷积层、池化层构成。后者由全连接层构成, 对提取出的特征进行非线性组合以完成学习目标。

卷积层中存在卷积核以水平滑动窗格的方式扫描前一层输入的特征图, 通过卷积运算、批处理归一化以及非线性变换来学习深层特征。卷积层产生卷积计算量的同时, 利用权值共享来限制网络参数量, 使得卷积层间神经元的连接具有稀疏性, 有利于分类模型的训练且不易陷入局部最优^[22]。卷积层提取高阶特征的过程为:

$$C_{l,j} = f_{con}(C_{l-1} * W_{l,j} + b_{l,j}) \quad (1)$$

式中: $C_{l,j}$ 表示第 l 层卷积层、第 j 个卷积核的特征映射; C_{l-1} 表示第 l 层卷积层的输出, 当 $l=1$ 时, C_0 为输入特征空间 X ; $W_{l,j}$ 表示第 l 层卷积层第 j 个卷积核的权值向量; “*” 表示卷积运算符, 进行卷积操作后与第 l 卷积层第 j 个卷积核的偏移向量 $b_{l,j}$ 叠加, 最后将线性结果嵌套非线性卷积层激活函数 $f_{con}(\cdot)$ 。

池化层通常处于 2 卷积层之间, 旨在通过降低特征映射的分辨率来实现移位不变性^[23]。池化操作又称为下采样, 本文采用最大池化方式, 池化层的每个特征映射都与卷积层的相应特征映射相连接, 第 $l+1$ 层池化层对应的输出特征图的计算式为:

$$P_{l+1,j} = f_{pool}(k_{l+1}p(C_{l,j}) + b_{l+1,j}) \quad (2)$$

式中: k_{l+1} 和 $b_{l+1,j}$ 分别表示池化输出中的权重和偏置; $p(\cdot)$ 表示最大池化操作即下采样函数; $f_{pool}(\cdot)$ 表示池化层激活函数。

分类器的构建, 是将最后一个卷积层所提取的特征扁平化为 1 维向量并输入到具有激活函数的全连接隐藏层中。全连接层中神经元与相邻 2 层的所有神经元直接连接, 从而推理全局信息。采用 softmax 函数以实现 MC 分类输出, 嵌套 sigmoid 函数以实现 ML 分类。第 n 个样本对应输出如下:

$$Y_n = f_{output}(w_d C_F + b_d) \quad (3)$$

式中: C_F 表示扁平层特征映射向量; w_d 和 b_d 分别表示前一层全连接层对应的权重矩阵和偏置向量; $f_{output}(\cdot)$ 为输出层激活函数。其中, 设置分类器判别阈值为 0.5, 大于该阈值则输出 1, 不满足则输出 0, 将 ML 分类模型概率输出二值化。

以 $Y_n=[y_{n,1}, y_{n,2}, \dots, y_{n,M}]$ 表示样本 n 的 ML 模型输出, $Z_n=[z_{n,1}, z_{n,2}, \dots, z_{n,M}]$ 表示样本 n 的真实标签向量。ML 分类的加权交叉熵损失函数 L_{ml} 定义为:

$$\begin{cases} L_{ml} = \frac{1}{N} \sum_{n=1}^N L_n \\ L_n = \frac{1}{M} (\lambda \sum_{n \in S_{0,n}} l_{n,m} + (1-\lambda) \sum_{n \in S_{1,n}} l_{n,m}) \\ l_{n,m} = z_{n,m} \ln y_{n,m} + (1-z_{n,m}) \ln(1-y_{n,m}) \end{cases} \quad (4)$$

式中: N 为总样本数; $S_{1,n}$ 和 $S_{0,n}$ 分别表示样本 n 的正类标签集合和负类标签集合; $l_{n,m}$ 表示样本 n 的第 m 个标签的值; 引入权重 λ , 旨在不影响计算复杂度的前提下调整正、负类标签数量不均的情况。

2.3 基于多标签分类与卷积神经网络拓扑辨识模型分析

基于 ML-CNN 的配电网拓扑辨识方法, 其核心思想是将配电网中影响拓扑结构的联络开关的闭合/关断状态作为多个标签项构建 ML 分类模型。这种模型架构具有多方面优势:

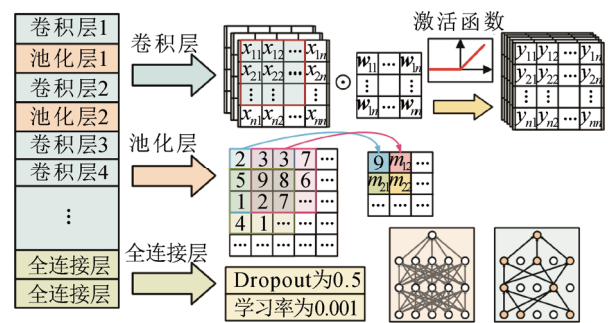


图 3 CNN 架构图

Fig.3 Convolution neural network architecture

1) ML 分类机制的引入可以克服 MC 分类中输出特征空间过大对模型的不良影响。

2) 针对机器学习解释性差的不足, 所提出模型基于联络开关进行 ML 编码, 具备物理依托, 分类模型结果与开关状态一一对应, 增强了模型的可解释性。

3) 考虑符合配电网运行规律的拓扑结构中各开关状态间潜在的关联性, 在部分配电网拓扑样本不足或缺失的情况下, 可以根据训练数据集学习到的局部信息判断开关状态, 更好地利用了可采集样本在局部信息上的丰富度和差异性。

因此, ML-CNN 拓扑辨识模型可以通过特征提取学习到各个开关附近节点电压的局部特征, 同时又能涵盖各开关状态之间的关系。

2.4 分类模型评价指标

MC 分类下的分类结果无需考虑基于标签度量和基于样本度量的差异, 而 ML 分类下的标签分类度量与样本分类度量有所不同。因此, 针对 ML 分类模型, 本节分别建立了基于标签和基于样本的准确度评价指标。

分类模型评价指标计算以混淆矩阵为基础, 常采用准确度、查准率、召回率、F1-score 作为评价指标。式(5)中, A_{Lal} 表示基于标签的准确度, 该指标反映了开关状态的平均判别情况。式(6)定义 A_{Samp} 为基于样本的准确度, 要求模型输出标签向量与真实标签完全一致。因此, A_{Samp} 能更直观地反映出模型分类的精确程度, 面向具有先验信息的网络拓扑整体来说, 正确判断所有开关的状态是样本分类正确的必要条件。

$$A_{Lal} = \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M d(z_{n,m}, y_{n,m}) \quad (5)$$

$$A_{Samp} = \frac{1}{N} \sum_{n=1}^N d(\mathbf{Z}_n, \mathbf{Y}_n) \quad (6)$$

式中: $z_{n,m}$ 与 $y_{n,m}$ 表示标量, 当 $z_{n,m}=y_{n,m}$ 时, $d(z_{n,m}, y_{n,m})=1$, 否则为 0; \mathbf{Z}_n 与 \mathbf{Y}_n 表示向量, 当 $\mathbf{Z}_n=\mathbf{Y}_n$ 时, $d(\mathbf{Z}_n, \mathbf{Y}_n)=1$, 否则为 0。

同理, 计算每个样本的查准率、召回率、F1-score, 累加求和后与样本总数相比, 得到基于样本的各项平均分类指标, 如式(7)—式(9)所示, 这 3 个评价指标均与分类效果正相关。因此, 单个样本的查准率表示正确判断开关闭合的标签占所有被判断为闭合的开关标签的比例; 单个样本的召回率表示正确判断开关闭合的标签占所有真实情况下闭合

开关标签的比例。基于标签的以上 3 个指标表征含义类似, 不在此处详细展开。

$$P = \frac{1}{N} \sum_{n=1}^N \frac{l_n^{TP}}{l_n^{TP} + l_n^{FP}} \quad (7)$$

$$R = \frac{1}{N} \sum_{n=1}^N \frac{l_n^{TP}}{l_n^{TP} + l_n^{FN}} \quad (8)$$

$$F_1 = \frac{2PR}{P+R} \quad (9)$$

式中: P 、 R 与 F_1 分别表示查准率、召回率、F1-score; l_n^{TP} 和 l_n^{FP} 分别表示第 n 个样本中分类为正的标签数量和负标签数量; l_n^{FN} 表示第 n 个样本中分类为负的正标签数量。

3 算例分析

本文采用增设联络开关后的 IEEE 123 节点配电系统进行实验验证。基于 ML 和 MC 2 种编码方式分别搭建了 MC-RF、MC-MLP、MC-CNN、ML-RF、ML-MLP、ML-CNN 共 6 个分类模型。实验主要从 4 个方面展开: (1)比较不同分类算法对应模型的拓扑辨识效果; (2)对比 ML 和 MC 2 种分类编码方式下拓扑辨识效果; (3)验证存在分布式电源出力时所提出模型的鲁棒性; (4)验证 ML-CNN 拓扑辨识模型针对未知拓扑数据样本的泛化能力。所有实验在配备处理器为 AMD Ryzen 74 800U with Radeon Graphics 和内存配置为 16.0 GB RAM 的笔记本电脑上实现。

3.1 实验设置

3.1.1 实验仿真设置

改进的 IEEE 123 节点配电网如图 4 所示, 该系

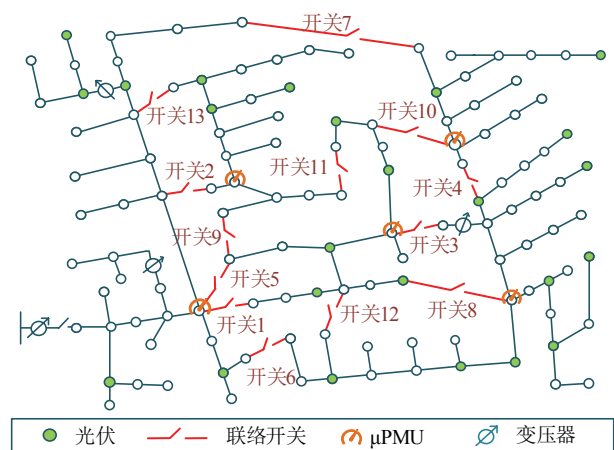


图 4 IEEE 123 配电网结构

Fig.4 IEEE 123 distribution network

统包含 118 条支路和 13 个联络开关。假设全网电压幅值数据由 SCADA 系统提供, 鉴于 μ PMU 成本较高, 仅在少数联络开关附近装设该装置, 本文选取图 4 所示的 5 个台区母线处布置 μ PMU 量测节点电压相量。

根据联络开关的状态变化, 可筛选出 240 种辐射状拓扑结构。配电网在常规运行场景下的历史数据量所占比例更高, 而极端情况下可能只有少量历史断面数据甚至不存在数据记录, 这种情况下对应的网络结构即为未知拓扑。为仿真配电网真实运行情况, 随机选取上述拓扑中的 1/12, 生成少量数据集作为未知拓扑推理实验样本。其余 220 种拓扑结构随机平均分为 22 组, 并按组划分生成数据量层次, 从而反映各拓扑结构的运行概率差异。

负荷数据源于真实居民负荷数据。由该数据计算得到平均归一化的负荷曲线, 以 IEEE 123 节点配电网标准负荷为基准值模拟系统潮流。负荷波动服从 $N(0, \sigma^2)$ 的正态分布, σ^2 的取值有 0.03、0.06、0.09、0.012, 选取不同方差仿真、不同时间断面的负荷波动程度, 从而更好地模拟配电网负荷变化。在节点电压幅值数据中加入 1% 的高斯噪声来模拟 SCADA 系统量测误差, 以标准差为 0.001° 的正态分布模拟 μ PMU 相角量测误差。生成共 236 160 个样本用于常规拓扑辨识, 额外生成 1 920 个样本用于推理未知拓扑。

3.1.2 实验超参数设置

ML-RF 与 MC-RF 模型参数设置相同, 树的最大生长深度设置为 13。ML-MLP 与 MC-MLP 模型参数设置也保持一致, 其中隐藏层神经元参数为 (120, 200)。ML-CNN 与 MC-CNN 的模型参数如表 1 所示, 其中 MC-CNN 模型的损失函数选取多类别交叉熵函数。为提高收敛速度, 避免过拟合, 采用小批量梯度下降法训练, 每批次 70% 数据为训练集, 30% 数据为验证集, 使用 Adam 优化器, 初始学习率为 0.001。

3.2 拓扑识别实验结果及分析

为避免数据集差异造成的偶然性, 各分类模型的训练和测试均采用相同数据集。表 2 汇总了 MC 和 ML 分类编码方式下 RF、MLP 和 CNN 3 种算法的测试结果。

首先在分类算法对比上, 无论采用 MC 编码还是 ML 编码, CNN 相较 RF 和 MLP 的分类效果都更佳。其中, RF 和 MLP 分类器基于样本的拓扑分

表 1 基于 ML-CNN 与 MC-CNN 模型架构的参数设置

Table 1 Parameters settings of ML-CNN and MC-CNN

分类算法	层类型	尺寸	激活函数
ML-CNN	卷积层	121×90	ReLU
	卷积层	119×180	ReLU
	扁平层	21 420×1	—
	全连接层	13×1	Sigmoid
MC-CNN	卷积层	121×60	ReLU
	卷积层	119×120	ReLU
	扁平层	14 280×1	—
	全连接层	240×1	Softmax

表 2 多类别与多标签分类器分类结果

Table 2 Classification results of multi-class classifier and multi-label classifier

分类算法	$A_{\text{Samp}}/\%$	$P/\%$	$R/\%$	$F_1/\%$	
MC 编码	MC-RF	84.42	84.42	84.42	84.42
	MC-MLP	86.14	86.15	86.15	86.15
	MC-CNN	98.91	99.22	99.22	99.22
ML 编码	ML-RF	85.02	94.64	94.64	94.64
	ML-MLP	86.89	97.90	97.05	97.47
	ML-CNN	99.49	99.96	99.91	99.94

类准确度在 85% 左右, 而 CNN 分类模型分类效果准确度达到 99% 左右。相比于 CNN, RF 与 MLP 算法均缺乏特征提取模块, 分类准确度相对较低。实验结果充分展现出 CNN 卷积层特征提取在学习复杂映射上的重要作用。所提出 ML-CNN 方法的 P 、 R 和 F_1 均达到了 99.9% 以上, 其严格指标 A_{Samp} 也较高, 表明该模型可应用在高精度要求下的配电网拓扑辨识场景。

相比于 MC 分类, 采用 ML 分类的 3 种算法模型在 P 、 R 和 F_1 上的表现具有不同程度的提升。考虑综合性能指标 F_1 , RF 模型提高了 10.22%, MLP 模型提高了 11.32%, 但对于 CNN 模型的提升并不明显, 表明在数据样本足够丰富时, 2 种模型性能差异不大。

为验证 ML-CNN 模型在分布式电源出力波动场景下的鲁棒性, 参考文献[24]选择图 4 所示的 25 个节点处接入分布式光伏系统, 装机容量在 8~32 kW 之间, 功率因数设为 0.95。随机选取 60 种可行拓扑模拟场景, 其中光伏出力波动以 $N(0, 0.01)$ 的正态偏差模拟。设置模型的超参数不变, 通过设置不同装机容量改变光伏渗透比例, 测试在 10%、20%、30%、40% 渗透率下模型的辨识效果。

如图 5 所示, 随着光伏渗透率的增加, 模型分类评价指标中 A_{Lal} 和 F_1 基本保持在同一水平, 表明分布式电源出力的增加对开关状态的辨识和模型分类性能基本没有影响。同时, 严格性评价指标 A_{Samp} 在 99.6% 上下波动, 波动程度小且识别精度不随渗透率的增加而降低, 说明 ML-CNN 模型在不同光伏渗透率下具备良好的鲁棒性。

3.3 未知拓扑推理实验结果及分析

在 3.2 节中, ML-CNN 模型与 MC-CNN 模型的性能相近, 未能充分体现出所提出模型的优势。因此, 本节重点分析针对未知拓扑样本的实验结果, 以验证基于 ML-CNN 的拓扑辨识模型在训练样本输出特征空间外的泛化能力。

对于 MC 分类, 由于其训练集中没有这些样本标签, 无法直接计算出有效的样本评价指标。因此通过比较分类器输出的拓扑与真实标签间开关状态的偏差来计算各指标。在表 3 所示的未知拓扑推理实验结果中, MC-CNN 模型的 A_{Lal} 为 69.50%, 对未知拓扑中开关状态的判断仍有高达 98.73% 的平均准确度, 且其 A_{Samp} 为 85.47%, 具备了 RF 或 MLP 在常规情况下的分类效果。 A_{Samp} 是在 A_{Lal} 基础上得到的, 对于拓扑整体的辨识准确度来说, 每个开关状态的识别准确度都至关重要, 更提高了对开关状态分类准确度的要求。

尽管针对标签的平均分类结果并不能直观地体现出拓扑辨识效果, 但可以找出最大程度上影响整体拓扑识别效果的短板, 即分类效果最差的开关。图 6 展示了 MC-CNN 和 ML-CNN 模型在外推实验中对各开关状态的评价指标得分。MC-CNN 在各开关状态上的分类效果不稳定, 对联络开关 1、4、5、6 尚有不错的分类效果, 基本达到了 80% 或以上; 半数以上的开关分类指标得分在 60% 以下; 开关 11 的评价指标得分最低, 不超过 10%, 很大程度上拉低了整体拓扑辨识精确度。MC-CNN 方法在未知拓扑推理上存在较大限制。

ML-CNN 模型对大多数联络开关状态的分类效果稳定, 能保证完全分类正确。除开关 8、10 外, 其他开关对应的 P 都是 100%, 意味着模型辨识开关为闭合时实际开关状态也是闭合状态; 开关 3 的 R 较低, 说明实际情况下闭合的开关 3 易被分类为关断状态。选取 2 类模型辨识错误较多的拓扑样本进行分析, 图 7 展示了 ML-CNN 在 58 号和 118 号拓扑上的辨识情况, 可直观地看出各开关状态的

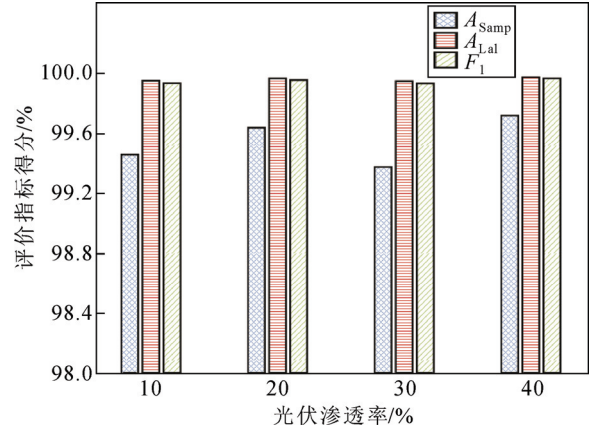


图 5 不同光伏渗透率下拓扑辨识效果

Fig.5 Topology identification effects under different photovoltaic permeabilities

表 3 MC-CNN 和 ML-CNN 在未知拓扑样本的分类结果

Table 3 Results of MC-CNN and ML-CNN in unknown topology samples

分类模型	$A_{Samp}/\%$	$A_{Lal}/\%$	$P/\%$	$R/\%$	$F_1/\%$
MC-CNN	—	69.50	50.43	50.43	50.43
ML-CNN	85.47	98.73	99.23	96.69	97.94

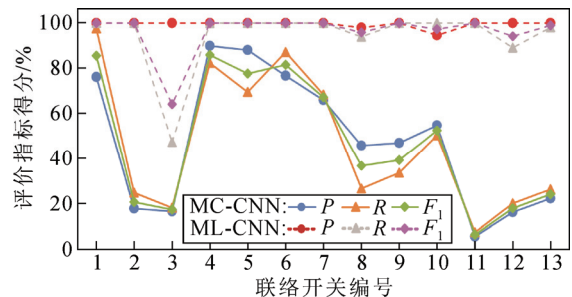


图 6 MC-CNN 和 ML-CNN 在未知拓扑推理实验中标签分类结果

Fig.6 Label-based classification results in unknown sample experiments of MC-CNN and ML-CNN classifier

类结果。针对拓扑 58, 分类错误表现为开关 3 误判为关断, 以及包含上述情况时开关 10 误判为闭合。拓扑 118 对应的错误分类样本分为 3 类: 开关 12 误判为断开; 开关 12 误判为断开且开关 8 误判为闭合; 开关 13 误判为断开。通过对网络拓扑中开关位置及潮流流向进行分析, 推断所提出模型在开关 3 和开关 10、开关 8 和开关 12 的状态划分上有所混淆。

结合表 3、4 可知, ML-CNN 分类模型在未知样本推理实验中, 当判别阈值取 0.5 时 P 高达 99.23%, R 相对于 P 偏低, 为 96.69%。由于在开关

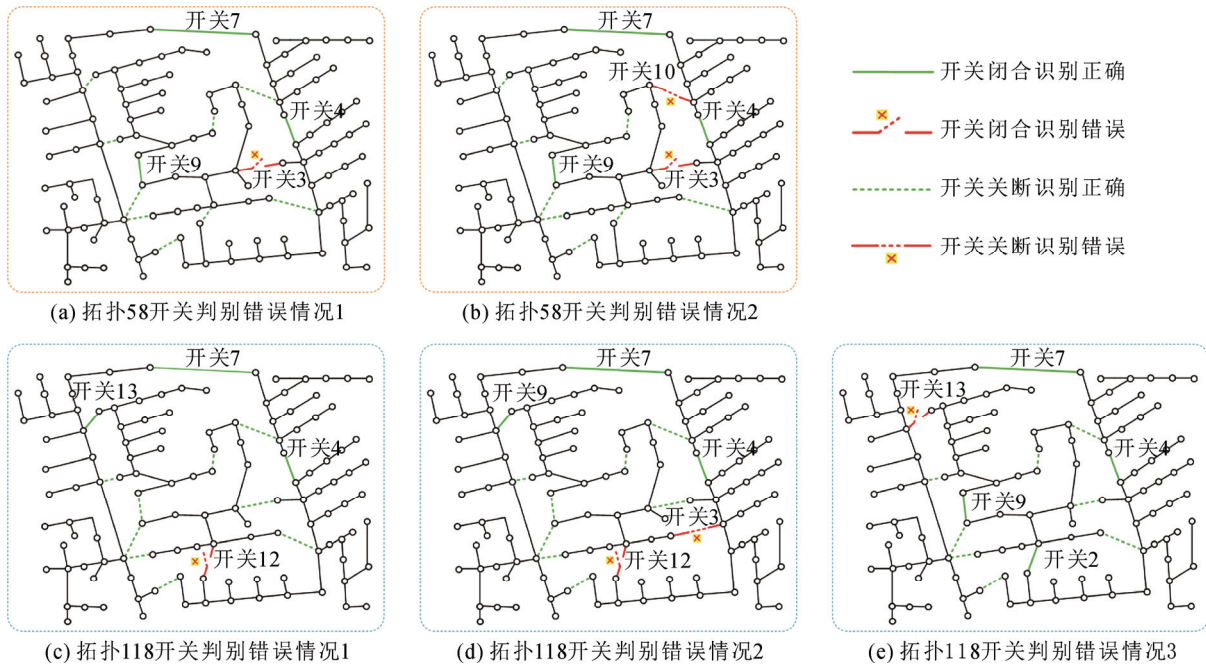


图 7 ML-CNN 在拓扑 58 和拓扑 118 上的开关状态判别情况

Fig.7 Switch state classification results of ML-CNN on topology 58 and 118

状态的分类上有着接近 100% 的 P , 因此开关断开时模型分类更敏感, 可以较好地判断开关的状态。假设某一个真实状态为闭合的开关被判断为断开, 则根据分类结果得到的拓扑结构很可能存在独立区域, 而无法保证整个系统的连通性, 显然这种分类结果是不合理的。因此, 在将模型分类结果进行二值化与真实标签比较时, 可以通过适当地降低分类器输出判别阈值来实现 R 的提升, 同时综合各项指标选择更合适的判别阈值以提升模型的拓扑推理效果。

ML-CNN 分类器输出离散化过程采用不同判别阈值时, 未知拓扑样本的分类效果如表 4 所示, 分别设置判别阈值为 0.15、0.25、0.35、0.45、0.50。可以看出随着判别阈值的减小, A_{Lal} 和 F_1 略有下降, P 随之下降, R 则相反呈上升趋势。当判别阈值取 0.35 时, 虽然标签平均准确度小幅降低, 但 A_{Samp} 提高至 87.19%, 且 P 和 R 也更加均衡。

表 5 给出了 ML-CNN 分类器输出二值化采用不同判别阈值时, 各联络开关对应的 F_1 , 表中 h 为判别阈值。显然随着 h 的递减, 开关 3 的 F_1 逐级上升, 与此同时发现开关 10 的得分逐次下降, 说明分类器在外推数据样本上容易将开关 3 和开关 10 的状态相互混淆, 结合拓扑网络中开关的位置和被错误分类的样本可以发现, 当开关 4 和 7 闭合时, 有功潮流由右侧子网络流向中心区域, 开关 3 或开关 10 闭合对网络末端附近节点造成的电压降落可能十分相

表 4 不同输出判别阈值下 ML-CNN 未知拓扑推理实验中的分类结果

Table 4 Results of ML-CNN with different output discriminant thresholds in the unknown sample experiment

判别阈值	$A_{Samp}/\%$	$A_{Lal}/\%$	$P/\%$	$R/\%$	$F_1/\%$
0.15	82.92	98.24	96.76	97.96	97.36
0.25	85.73	98.47	97.68	97.59	97.63
0.35	87.19	98.65	98.46	97.25	97.85
0.45	85.99	98.68	98.91	96.86	97.87
0.50	85.47	98.73	99.23	96.69	97.94

表 5 不同输出判别阈值下 ML-CNN 标签分类的 F1-score

Table 5 F1-score of labels under different output discriminant thresholds of ML-CNN classifier

开关 编号	$F_1/\%$				
	$h=0.15$	$h=0.25$	$h=0.35$	$h=0.45$	$h=0.50$
1	100	100	100	100	100
2	100	100	100	100	100
3	74.62	72.85	70.27	66.36	64.15
4	100	100	100	100	100
5	100	100	100	100	100
6	100	100	100	100	100
7	100	100	100	100	100
8	96.39	96.44	95.82	94.74	95.88
9	100	100	100	100	100
10	87.76	91.28	94.74	96.32	97.24
11	100	100	100	100	100
12	97.46	96.08	95.65	94.36	94.21
13	99.63	99.33	99.10	99.02	99.02

近, 可能导致模型将开关 3 闭合的状态判断成了开关 10 闭合。

综上所述, 未知拓扑样本推理实验结果充分体现了所提出模型在学习局部关键特征信息上的优越性, 验证了基于 ML-CNN 的拓扑辨识模型的推理能力, 一定程度上缓解了实际中样本丰富度不足引起的基于数据驱动方法应用困难的问题。

4 结论

本文采用数据驱动方法将配电网拓扑辨识问题转换为多标签分类问题, 在此基础上, 设计了基于 ML-CNN 的配电网辨识模型。主要结论如下:

1) 通过引入 ML 机制建立了数学模型到物理层面的联系, 增强了机器学习的可解释性, 同时避免了大规模配电网下多类别分类的输出样本空间过大问题。

2) 在分布式电源出力场景下, ML-CNN 模型在不同光伏渗透率下仍能保证有较高精度的拓扑辨识效果, 模型鲁棒性良好。

3) 通过结合 ML 编码的机制与 CNN 局部信息学习能力, 在训练样本丰富度不足时, 对训练样本空间外的未知拓扑数据样本仍能保持较高的拓扑识别能力, 相比传统的 MC 分类, 其具有更强的分析推理能力, 更适用于配电网运行的实际场景。

参考文献 References

- [1] 葛磊蛟, 李元良, 陈艳波, 等. 智能配电网态势感知关键技术及实施效果评价[J]. 高电压技术, 2021, 47(7): 2269-2280.
GE Leijiao, LI Yuanliang, CHEN Yanbo, et al. Key technologies of situation awareness and implementation effectiveness evaluation in smart distribution network[J]. High Voltage Engineering, 2021, 47(7): 2269-2280.
- [2] 刘永梅, 王金丽, 杨红磊, 等. 计及柔性负荷调节能力的有源配电网动态优化方法[J]. 高电压技术, 2021, 47(1): 73-80.
LIU Yongmei, WANG Jinli, YANG Honglei, et al. Dynamic optimal method of distribution network in consideration of flexible load adjustment capability[J]. High Voltage Engineering, 2021, 47(1): 73-80.
- [3] 盛万兴, 刘科研, 李昭, 等. 新型配电系统形态演化与安全高效运行方法综述[J]. 高电压技术, 2024, 50(1): 1-18.
SHENG Wanxing, LIU Keyan, LI Zhao, et al. Review of basic theory and methods of morphological evolution and safe & efficient operation of new distribution system[J]. High Voltage Engineering, 2024, 50(1): 1-18.
- [4] 刘迪, 张强, 吕干云, 等. 基于支路有功功率的配电网拓扑辨识方法[J]. 电力工程技术, 2021, 40(3): 92-98.
LIU Di, ZHANG Qiang, LYU Ganyun, et al. Distribution network topology identification method based on branch active power[J]. Electric Power Engineering Technology, 2021, 40(3): 92-98.
- [5] 黄毕尧, 张明, 李建岐, 等. 联合高低频电力线通信的中压配电网拓扑自动识别方法[J]. 高电压技术, 2021, 47(7): 2350-2358.
HUANG Biyao, ZHANG Ming, LI Jianqi, et al. Automatic identification of medium-voltage power distribution network topology based on high and low frequency power line communication[J]. High Voltage Engineering, 2021, 47(7): 2350-2358.
- [6] WU F F, LIU W H E. Detection of topology errors by state estimation (power systems)[J]. IEEE Transactions on Power Systems, 1989, 4(1): 176-183.
- [7] KORRES G N, KATSIKAS P J. Identification of circuit breaker statuses in WLS state estimator[J]. IEEE Transactions on Power Systems, 2002, 17(3): 818-825.
- [8] TIAN Z, WU W C, ZHANG B M. A mixed integer quadratic programming model for topology identification in distribution network[J]. IEEE Transactions on Power Systems, 2016, 31(1): 823-824.
- [9] FARAJOLLAHI M, SHAHSAVARI A, MOHSENIAN-RAD H. Topology identification in distribution systems using line current sensors: an MILP approach[J]. IEEE Transactions on Smart Grid, 2020, 11(2): 1159-1170.
- [10] 尚博阳, 罗国敏, 茹嘉昕, 等. 基于有限量测信息的多分支配电网故障定位方法[J]. 高电压技术, 2023, 49(6): 2308-2317.
SHANG Boyang, LUO Guomin, RU Jiaxin, et al. Fault location method of multi-branch distribution lines based on limited measurement information[J]. High Voltage Engineering, 2023, 49(6): 2308-2317.
- [11] 刘羽霄, 张宁, 康重庆. 数据驱动的电力网络分析与优化研究综述[J]. 电力系统自动化, 2018, 42(6): 157-167.
LIU Yuxiao, ZHANG Ning, KANG Chongqing. A review on data-driven analysis and optimization of power grid[J]. Automation of Electric Power Systems, 2018, 42(6): 157-167.
- [12] 杨挺, 耿毅男, 郭经红, 等. 人工智能在新型电力系统智能传感、通信与数据处理领域应用[J]. 高电压技术, 2024, 50(1): 19-29.
YANG Ting, GENG Yinan, GUO Jinghong, et al. Applications of artificial intelligence in sensing, communication, and data processing in the new power system[J]. High Voltage Engineering, 2024, 50(1): 19-29.
- [13] 杨冬峰, 付强, 刘晓军, 等. 配电网动态拓扑与线路参数联合在线辨识方法[J]. 电力系统自动化, 2022, 46(2): 101-108.
YANG Dongfeng, FU Qiang, LIU Xiaojun, et al. Joint online identification method for dynamic topology and line parameters of distribution network[J]. Automation of Electric Power Systems, 2022, 46(2): 101-108.
- [14] 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.
YANG Xiu, JIANG Jiafu, LIU Fang, et al. Distribution network topology identification based on attention mechanism and convolutional neural network[J]. Power System Technology, 2022, 46(5): 1672-1682.
- [15] ARDAKANIAN O, WONG V W S, DOBBE R, et al. On identification of distribution grids[J]. IEEE Transactions on Control of Network Systems, 2019, 6(3): 950-960.
- [16] CAVRARO G, ARGHANDEH R. Power distribution network topology detection with time-series signature verification method[J]. IEEE Transactions on Power Systems, 2018, 33(4): 3500-3509.
- [17] 叶剑华, 罗凤章, 杨理. 基于改进蝠鲼觅食优化 SVM 的配电网拓扑辨识[J]. 电力系统及其自动化学报, 2021, 33(10): 43-50.
YE Jianhua, LUO Fengzhang, YANG Li. Distribution network topology identification based on SVM optimized by improved manta ray foraging optimization algorithm[J]. Proceedings of the CSU-EPSSA, 2021, 33(10): 43-50.
- [18] 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.

- [19] 陈沛东, 曹华珍, 何 璇, 等. 中压配电网近邻交互式分布式拓扑辨识算法[J]. 电力工程技术, 2023, 42(2) : 139-146.

CHEN Peidong, CAO Huazhen, HE Xuan, et al. Distributed topology identification algorithm of medium-voltage distribution network based on neighboring interaction[J]. Electric Power Engineering Technology, 2023, 42(2): 139-146.

- [20] 丁世飞, 张 健, 张谢锴, 等. 多分类孪生支持向量机研究进展[J]. 软件学报, 2018, 29(1): 89-108.

DING Shifei, ZHANG Jian, ZHANG Xiekai, et al. Survey on multi class twin support vector machines[J]. Journal of Software, 2018, 29(1): 89-108.

- [21] KRAWCZYK B, KOZIARSKI M, WOŹNIAK M. Radial-based over-sampling for multiclass imbalanced data classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(8): 2818-2831.

- [22] ZANG H X, LIU L, SUN L, et al. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations[J]. Renewable Energy, 2020, 160: 26-41.

- [23] ZUBAIR S, YAN F, WANG W W. Dictionary learning based sparse coefficients for audio classification with max and average pooling[J]. Digital Signal Processing, 2013, 23(3): 960-970.

- [24] 任鹏哲, 刘友波, 刘挺坚, 等. 基于互信息贝叶斯网络的配电网拓

扑鲁棒辨识算法[J]. 电力系统自动化, 2021, 45(9): 55-62.

REN Pengzhe, LIU Youbo, LIU Tingjian, et al. Robust identification algorithm for distribution network topology based on mutual-information Bayesian network[J]. Automation of Electric Power Systems, 2021, 45(9): 55-62.



LONG Huan

Ph.D.

Associate professor
Corresponding author

龙 寰(通信作者)

1992—, 女, 博士, 副教授

主要从事人工智能及其在电力系统应用方面的研究

E-mail: hlong@seu.edu.cn



SHI Ziqing

石子晴

1998—, 女, 硕士生

主要从事基于数据驱动的配电网优化运行方面的研究

E-mail: 1061079293@qq.com

收稿日期 2023-06-05 修回日期 2024-02-23 编辑 程子丰