基于LightGBM算法的配电网单相接地故障区段定位方法

郑一斌1,王慧芳1,张 磊2,姜 宽1,杨文斌2,周才全2

(1. 浙江大学 电气工程学院,浙江 杭州 310027;2. 中国电建集团华东勘测设计研究院有限公司,浙江 杭州 311122)

摘要:我国广泛分布的中性点不接地配电网受量测设备配置所限,长期存在故障区段定位困难问题。为此, 提出了一种量测量受限条件下的基于轻量级梯度提升机(LightGBM)算法的故障区段定位方法。该方法可在 不加装相量测量单元(PMU)、微型PMU、电压互感器等量测装置的情况下,利用常规的相电流有效值准确识 别中性点不接地配电网故障区段。通过对目标配电网的量测信息和网架结构分析,采用LightGBM算法离线 建立故障区段定位模型,实现在线快速定位故障区段。其中,故障前后各线路稳态电流有效值变化量被选定 为故障识别核心特征。此外,针对模型预测结果可能存在的偏差,提出了易错线路标注法和类别概率法辅助 识别误判情况,进一步提高了预测结果的可靠性。以修改后的IEEE 123节点系统作为算例,仿真验证了所 提方法的可行性和有效性。

关键词:故障区段定位;配电网;叠加原理;LightGBM 算法;人工智能 中图分类号:TM 713 文献标志码:A

DOI:10.16081/j.epae.202108033

0 引言

配电网为了提高供电可靠性,当系统的电容电 流不超过一定值时,多采用中性点不接地方式。当中 性点不接地配电网发生单相接地故障时,故障电流 较小,电流保护不会动作,为故障排除提供了时间,但 也造成了故障定位的困难^[1]。量大面广的中性点不 接地配电网为降低建设成本,多在线路上仅装设电 流互感器以及电流保护,而不装设电压互感器以及 相量测量单元 PMU(Phasor Measurement Unit)、微 型相量测量单元 μPMU(micro-Phasor Measurement Unit)等精确量测设备,不充足的量测信息条件使 得基于精确电气量的故障定位方法难以应用。尤其 当故障点接触到树枝、沙土、沥青、水泥等非理想介 质时,会形成高阻接地故障 HIF(High Impedance Fault),中性点不接地配电网的故障定位更加困难^[2]。

中性点不接地配电网单相接地故障区段定位问题是存在已久的难题,众多专家学者提出了多种定位方法^[3]。这些方法总体上可分为基于就地信息、基于广域信息的故障区段定位方法和其他故障区段定位方法。基于就地信息的故障区段定位方法借鉴故障选线的诸多方法,包括基于稳态特征的基波群体比幅比相法、零序电流相位法、零序无功功率方向法^[4],以及基于零序暂态特征在故障点上下游的差异^[5]、基于灰色关联度的模分量差异^[6]的识别方法等,这类方法虽然物理机理明确,但由于依赖的信息范围有限且特征微弱,故障区段定位的可靠性受较多因素影响,如阈值的选择。阈值需要适应各种可能的运行方式和故障情况,过高和过低均可能产生误判或漏判,所以阈值选择非常困难。基于广域信

息的故障区段定位方法包括矩阵法与人工智能法。 矩阵法^[7]根据配电网的拓扑结构及电流过流告警信 息构建故障信息矩阵,由此得出故障判断矩阵,定位 故障区间,但容错性欠佳。人工智能法包括粗糙 集^[8]、人工神经网络^[9]、Petri网^[10]等方法,其容错性 较好,但计算量大、定位速度慢。其他故障区段定位 方法,如基于故障投诉电话的区段定位方法等,多作 为前2种方法的补充。

随着算法、算力的持续改进和提高,人工智能 技术取得了长足的进步。本文针对中性点不接地系 统单相接地配电网的故障区段定位问题,从广域角 度出发,延续人工智能法的思路,提出一种基于轻 量级梯度提升机LightGBM(Light Gradient Boosting Machine)算法的中性点不接地配电网单相接地故障 区段定位方法。该方法基于实际容易获得的相电流 量测信息,通过故障仿真获取大量样本,每个样本包 含故障信息以及各线路在故障前、后的稳态电流有 效值变化量,采用LightGBM算法训练多分类故障区 段定位模型,实现在不同运行方式、不同过渡电阻、 不同干扰因素条件下的可靠故障区段定位。

1 中性点不接地配电网单相接地故障区段 定位原理

中性点不接地配电网正常和发生单相接地故障 后的运行状态如附录A图A1所示。发生单相接地 故障时,配电网从故障前的稳态经过一个暂态过程, 进入故障后新的稳态。假定整个过程中,保护装置 均未动作,则可以通过采集故障前、后稳态下各线路 的电流有效值,采用叠加原理表示故障带来的影响, 如式(1)所示。 式中:*I*、*I*′分别为故障前、后稳态下各线路的电流有 效值矩阵;Δ*I*为各线路的电流有效值变化量矩阵, 反映故障支路对各线路的影响,其数值大小与故障 位置、非故障线路对地电容、故障线路对地电容、故 障过渡电阻有关。由文献[11]可知:对于非故障线 路及故障线路位于故障点后的部分,其三相电流变 化量为对地电容电流,幅值相等、波形一致;对于故 障线路位于故障点前的部分,其非故障相的电流变 化量类似,而故障相的电流变化量则包含对地电容 电流与故障点对地电流2个部分,比非故障相的电 流变化量更大。因此,当中性点不接地配电网的不 同区段发生单相接地故障时,故障线路和非故障线 路的各相电流变化量是存在差别的,Δ*I*可以作为中 性点不接地配电网单相接地故障的特征。

人工智能技术中的机器学习方法可以通过大量 样本习得配电网的电流变化信息与故障位置间的关 系,进而通过新的电流变化量信息实现故障区段定 位。然而,采用机器学习方法解决实际问题时需要 满足以下3个条件:①每个样本中的特征对应的标 签具有唯一性;②样本在数量和质量上满足训练要 求;③有与待解决的问题相匹配的学习算法。中性 点不接地配电网的不同区段发生故障时,总会导致 全网线路在故障前、后的电流分布不同,即各线路电 流有效值变化量矩阵ΔI构成的信息矩阵隐含某个 确定的区段发生的故障,因此满足标签唯一的条件 ①。在训练阶段获得充足而不偏斜的训练数据通常 是人工智能技术应用的一大难题,本文使用高精度 的模拟仿真软件,并结合噪声叠加,可以实现对实际 电网各种情况的模拟,因此具备机器学习模型训练 和测试的条件:同时现有配电网都配置有电流互感 器,可以通过配电自动化等系统采集全网的线路电 流有效值信息,具备模型线上预测应用的条件,因此 满足条件②。条件③则是下文的研究和论证重点。

2 基于LightGBM 算法的故障区段定位模型

2.1 特征选择与分类模型选择

在样本数据充足的条件下,人工智能技术对于 大多数有规律、有单一解的问题,都能通过特征学习 与模型训练,实现对新样本的良好预测效果。本文 选择反映故障情况的故障前、后稳态下的电流有效 值变化量,以及反映网架拓扑状态的联络开关开断 情况作为样本特征,以故障线路编号为标签,建立样 本特征与标签之间的映射关系。

假设某一目标网络中,共有*l*条线路、*p*组联络 开关,则其特征*x*可表示为:

$$\mathbf{x} = \begin{bmatrix} \Delta I_{1A}, \Delta I_{1B}, \Delta I_{1C}, \Delta I_{2A}, \Delta I_{2B}, \Delta I_{2C}, \cdots, \\ \Delta I_{IA}, \Delta I_{IB}, \Delta I_{IC}, S_{BK,1}, S_{BK,2}, \cdots, S_{BK,p} \end{bmatrix}$$
(2)
式中: $\Delta I_{IA}, \Delta I_{iB}, \Delta I_{iC} (i=1,2,\cdots,l)$ 分别为第*i*条线路

的A、B、C相电流有效值变化量; $S_{BK,i}(i=1,2,\dots,p)$ 为第i组联络开关的开断状态,取值为0、1分别表示断开、闭合。下文用q表示特征x包含的特征数量,则 $q=3l+p_{\circ}$

当前机器学习领域算法种类繁多,不同算法在 回归、二分类、多分类等不同种类的具体问题上的表 现会有差别,选择合适的机器学习模型对配电网故 障区段定位方法的性能至关重要。针对故障区段定 位问题,特征x与标签y之间映射关系的建立有二分 类、多分类2种方法。二分类法可以对每条线路建 立1个二分类模型,以0、1分别代表线路正常、发生 故障。由于不同线路的二分类模型具有独立性,每 多1条线路就多1个模型,各模型之间的性能评价较 难达到统一,且网络支路越多,多个模型之间的故障 区段定位冲突造成的误判概率越大。多分类法以故 障线路编号为分类标签,再加一个编号0表示无故 障,即可由特征信息直接建立1个故障区段定位的 多分类模型,简单明了且结果具有一定的参考依据 和对比空间。本文经过对多种人工智能算法的研 究,选择LightGBM多分类算法处理故障区段定位 问题。

2.2 LightGBM 算法

LightGBM 算法是一种梯度提升决策树 GBDT (Gradient Boosting Decision Tree)的改进优化算 法^[12],其模型训练过程以足量的样本数据为基础,通 过建立多棵决策树(弱学习器),综合决策树群的输 出结果决定模型的最终输出。其实际训练过程可以 表述为:以迭代的方式生长增加决策树,当树增加带 来的准确度提升小于某个阈值时,停止迭代,获得由

 N_{tree} 棵决策树组成的LightGBM模型 $\varphi(\mathbf{x}_i) = \sum_{k=1}^{m} f_k(\mathbf{x}_i),$

其中, \mathbf{x}_i 为输入的特征向量; $f_k(\mathbf{x}_i)$ 为第k棵决策树。与适用性较强的随机森林RF(Random Forest)算法^[13] 以及适合二分类问题求解的支持向量机SVM (Support Vector Machine)^[14]相比,LightGBM 算法 具备良好的可修改性,在处理多分类问题时能够与问题实际情况相配合,且输出的概率矩阵可以为后续的误定位分析提供帮助;另外与同类算法(如XGBoost^[15-16])相比,LightGBM 算法采用直方图算法,并通过带深度限制的叶子生长策略(Leaf-wise)进行加速和防止过拟合。因此,理论上,利用LightGBM算法进行故障区段定位在准确性和快速性上更具优势。

处理多分类问题时,LightGBM 算法采用一对多 策略,对每个类别训练 N_1 个分类器,当标签类别总 数为Q时,训练的模型中共有 QN_1 棵树,对应输出Q个结果 $Z_1 - Z_q$,经Softmax函数实现结果的概率化输 出,即 $S_1 - S_q$ 为各区段发生故障的概率,概率最大的 为故障区段,将其编号y为定位结果进行输出。多 分类下基于LightGBM算法的故障区段定位原理如图1所示,图中Q=1+l。



图 1 多分类下基于LightGBM算法的 故障区段定位原理图

Fig.1 Principle diagram of fault section location based on LightGBM algorithm under multi-classification condition

LightGBM算法作为有监督学习算法,需要定义 损失函数和目标函数,损失函数用于衡量模型预测 值与真实值之间的偏差,目标函数是损失函数与正 则化项的叠加,用于最优化经验风险和结构风险, 而模型的训练目标是使目标函数*f*_{obj}的值尽可能小。 模型的目标函数通过泰勒二次展开推导得到,由文 献[17]可得,对于第*k* 棵决策树,其目标函数*f*_{obj,k}的 简化公式为:

$$f_{\text{obj},k} = \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (3)$$

式中:T为第k棵树的叶节点个数; λ 和 γ 用于调节正则项,一般而言, λ 取1,仅对 γ 进行必要调整; w_j 为叶 节点权重,是叶节点j对应的连续值; I_j 为所有映射 到叶节点j的样本编号集合。假设 $y_i^{*(k)}$ 表示第k次 迭代时前k-1棵树对第i个样本的预测值,则 g_i 是 损失函数 f_{loss} 对 $y_i^{*(k)}$ 的一阶导, h_i 为 f_{loss} 对 $y_i^{*(k)}$ 的二阶 导,大小可求。

故障区段定位问题中,可利用的特征信息种类 较少,信息密度不高,而对故障区段定位结果的准 确率有较高的要求。Softmax函数又称归一化指数 函数,作用在于将多分类结果以概率形式展现出来, 本文在利用LightGBM算法解决故障区段定位这一 多分类问题时,采用带Softmax函数的交叉熵来定义 损失函数 f_{loss},以增强结果可靠性,其形式如式(4) 所示。

$$f_{\rm loss} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{Q} L\{y_i = j\} \ln S_j^i$$
(4)

$$L\{y_{i}=j\} = \begin{cases} 1 & y_{i}=j \\ 0 & y_{i}\neq j \end{cases}$$
(5)

式中:N为训练样本数; S_i 为Softmax函数对第i个样本第j个类别的输出结果; y_i 为第i个样本的真实标签值。

3 应用问题分析与解决方案

3.1 类别不平衡问题权重调整

故障区段定位样本自动构建过程中,训练样本 集存在各类别数量不平衡问题。而机器学习处理多 分类任务时,不同类别的训练样本数量相差过大会 使算法的学习结果产生偏向,出现类别不平衡问题。

本文针对故障区段定位中的多分类任务,采用 代价敏感学习方法选择不同类别数据之间的比例分 布确定各类别权重 w。,作为模型训练参数加入训练 过程,调整不同类别标签 y 的权重,其表达式如式 (6)所示。

$$w_{\mathrm{c},j} = \frac{1}{\frac{1}{N} f_{\mathrm{num},j}} \tag{6}$$

式中:f_{num,i}为第j个类别在训练集中的样本数。

3.2 数据缺失问题下的数据补齐

实际信号在量测和传输过程中会小概率地出现 数据丢失或失真现象。由于样本特征包含全网各线 路的电流,各线路电流间包含了一定的冗余关系,所 以个别线路数据的失真对故障选线结果的影响非常 小。数据丢失会造成数据集中存在缺失数据,不同 的数据缺失处理方法会对模型预测结果产生不同 的影响。常用的数据补齐方法包括特殊值填充法、 条件平均值填充法、K最近距离邻法等。特殊值填 充法采用某个数值(如0)或某个特殊符号(如 "NaN")填充缺失数据:条件平均值填充法从缺失数 据所在的数据样本中,选择相关性较强的数据,取平 均值作为填充数据:K最近距离邻法选出欧氏距离 最小的K个样本中缺失部分对应的数据,对这K个 值进行加权平均后填充至缺失位置。针对故障区段 定位问题存在特征信息分散、标签分类并不完全依 赖于某个特征数据的特点,选择将缺失数据特殊化, 可以在LightGBM算法的分类过程中,有效避免缺失 数据的影响。

3.3 噪声模拟

针对仿真数据与实际数据之间存在误差的问题,采用在仿真数据中添加白噪声的方法进行模拟。 信噪比SNR(Signal-to-Noise Ratio)的计算方法如式 (7)所示,其数值越大,表示信号中的噪声含量越低。 $V_{\text{SNR}}=20 \log(I_{\text{S}}/I_{\text{N}})$ (7)

式中: *V*_{SNR} 为信噪比的值; *I*_s、*I*_N分别为电流实际信号、噪声信号的有效值。实际运行过程中,电网数据的信噪比一般控制在40 dB 以上,为了尽可能地模拟实际电网运行数据,并保留一定的裕度,在仿真数据中添加 30 dB 白噪声,参与模型的离线训练与在线预测过程。

3.4 误判分析与解决方案

经过训练和测试,发现LightGBM算法在故障区

段定位中发生的误判可分为3类。第1类是配电网整个状态的误判,将存在故障判断为无故障、将无故 障判断为某条线路故障。分析发现,当训练集样本 的过渡电阻范围设置过大,模型敏感度较高时,容易 出现该类误判。第2类是相邻位置线路的误判,当 相邻线路较短、线路阻抗不大时,由于电流差值相差 不大,容易造成对故障线路位置判断的偏移,误判的 线路之间具有区域集中的特点。这类误判的发生, 一般与网络结构较为特殊有关,而故障过渡电阻设 置过大也会增加该类误判发生的概率。第3类误判 属于随机性误判,模型的固有缺陷、电流的突然波 动、传输中出现的大偏差都有可能成为原因,且往往 难以避免,但该类误判发生的概率极低。

为减少故障区段定位过程中误判带来的影响, 本文提出了置信度阈值判别和易错线路标注相结合 的方法。

1)类别置信度阈值判断。Softmax 函数输出 LightGBM算法的多分类结果,且经过概率校准后, 模型的输出为一个*M×Q(M为*测试样本数)维的概率 输出矩阵,该矩阵包含了每个样本对应每个类别的 分类概率信息,可选择每个样本中概率最大的类别 作为分类结果。因此,最终类别对应的概率输出值, 可以辅助判断预测结果的类别置信度。对于置信度 低于一定阈值的分类结果,通过对其进行标注提示, 可有效识别故障区段定位结果的可疑性。

2)易错线路标注。第2类误判由于机制清晰, 大多发生在固定的几条线路之间,因此通过在样本 充足的情况下进行测试,可以提前对易错线路予以 标注,为确定真实故障位置提供参考。除此之外,在 易错区域安装 μPMU等高精度量测设备对该类易错 线路予以特殊监测,也可以大幅减少甚至避免第2 类误判的发生。

4 基于LightGBM 算法的故障区段定位流程

基于 LightGBM 算法的故障区段定位流程主要 分为离线训练、在线应用2个部分,其流程图见附录 A图A2。离线训练阶段需要大量训练样本,而实际 电网因故障次数有限无法满足要求,为此通过仿 真获取训练样本。首先根据已有的网络拓扑信息建 立仿真模型模拟常用的运行方式,通过修改系统运 行方式、故障发生位置、过渡电阻、联络开关状态并 添加白噪声,重复仿真获得充足样本,形成样本数据 集。然后,根据样本数据集中各类别(故障线路编 号)的比例求得各类别权重系数,代入 LightGBM 算 法,通过交叉验证的方式选取最优超参数组合,训练 得到目标模型。最后,对测试集中测试结果、误判样 本展开分析,通过线路误判频率确定易错线路区域, 完成离线训练工作。在线应用阶段,实时采集系统 当前各线路的三相电流有效值 I',通过与上一次采 集到的电流有效值 I 做差,由式(1)获得网络电流有 效值变化量 ΔI,与当前联络开关状态共同组成特征 数据。对缺失数据进行补齐后,将特征数据输入训 练好的模型,即可获得故障区段定位结果以及对应 的类别置信度。若输出线路编号为0,说明未发生 故障,继续监测下一时刻信息;若判定有线路发生故 障,则发出故障提醒,并将类别置信度低于阈值的故 障区段定位结果与离线获得的易错线路编号相比 较,若不属于易错线路,则定位结束,输出定位结果 及类别置信度,否则进一步提供可能的偏移范围。 经算例分析,阈值可设置为0.98。值得一提的是, LightGBM 算法受到网络规模大小的影响,在离线训 练阶段耗时较长,而在线应用时非常快速。

5 算例分析

5.1 算例情况

为模拟国内的中性点不接地配电网,同时兼顾 配电网架的通用性,本文在MATLAB/Simulink中搭 建 IEEE 123 节点系统模型,对该系统的接地方式、 线路长度进行了适合国情的修改:采用10.5 kV作为 系统侧电源电压,将频率调整为50 Hz,电源中性点 采用不接地方式,将线路长度修改为原有长度的2.5 倍,线路单位阻抗和负荷水平保持不变,系统拓扑图 及线路编号见附录A图A3。修改后的配电网对地 电容电流约为1.14 A,满足中性点不接地系统的要 求。通过大量仿真获得样本集,每个故障样本对应 的单相接地故障是随机发生在任意一相线路的任意 位置,最终样本涵盖所有线路。为分析不同因素对 本文方法的影响,建立样本集 Ω_0 , Ω_0 , Ω_0 , $\Omega_1 - \Omega_2$ 如表1所示。表中, Ω_0 与 Ω_0 由 Ω_0 按0.85:0.15的比 例划分得到,是 Ω_0 的互斥子集。每轮均仿真一次完 整的故障发生过程,包含故障前稳态、故障发生、故 障后稳态3个阶段,产生一个故障特征样本和一个 非故障特征样本,分别对应故障位置线路编号和0 作为标签。其中,故障特征样本由I'-I获得,非故 障特征样本由I分别叠加2个不同的随机噪声后做 差获得。每个样本对应的过渡电阻在设定的范围 内随机取值,系统阻抗在3+j4~7+j8Ω之间随机选 取[17]。随机断开5个联络开关中的1个,各负荷在 80%~120%之间随机波动,在最终采集的电流中添 加 30 dB 的白噪声干扰。除样本集 $\Omega_{18} - \Omega_{21}$ 外,其 他样本集中的负荷在单轮仿真中的故障前、后保持 不变,获取每次故障发生前、后的三相线路电流有效 值变化量以及联络开关情况作为样本特征x,以及 故障发生的线路编号作为标签值 y(未发生故障时 标签为0)。

| | Table I Data dis | stribution of sample | set |
|-----------------------------|---------------------------------|---------------------------------|-----------------------|
| 样本集 | 非故障 样本数量 | 故障 样本数量 | $R_{ m f}$ / Ω |
| $arOmega_0$ | 7 000 | 7 000 | [0,1400] |
| $arOmega_{01}$ | 5950 | 5950 | [0,1400] |
| $arOmega_{02}$ | 1 0 5 0 | 1 0 5 0 | [0,1400] |
| \varOmega_1 | 0 | 1 000 | [0,100) |
| $\Omega_{2} - \Omega_{13}$ | 0 | 1 000 | 上、下限每次 增大100Ω |
| $arOmega_{14}$ | 0 | 1 000 | [1300,1400) |
| $\Omega_{15} - \Omega_{15}$ | 1000(1%、2%、 5%缺失) | 1000(1%、2%、 5%缺失) | [0,1400] |
| $\Omega_{18} - \Omega_2$ | 1000(1%、2%、 3%、5%随机 负荷波动) | 1000(1%、2%、 3%、5%随机 负荷波动) | [0,1400] |
| $arOmega_{22}$ | 500 | 500 | [0,1400] |

表 1 样本集数据分布

5.2 交叉验证选取最优超参数

本文采用K折交叉验证法选取机器学习算法的 最优超参数,将训练集 Ω_{01} 均匀分为10份,轮流选择 其中的9份训练,剩余1份用于验证,计算平均预测 准确率,最终选择使得平均准确率最高的超参数组 合作为最优超参数,具体过程如图2所示。图中, Num boost round和Learning rate分别表示迭代次数 和学习率,两者均为LightGBM算法的超参数; P_i 为 第i轮迭代的预测准确率,每个超参数组合对应1个 平均预测准确率 P_o



图2 交叉验证过程示意图

Fig.2 Schematic diagram of cross-validation process

5.3 算法选择与特征选择对故障区段定位效果的 影响

1)算法选择的影响。

为验证所提基于 LightGBM 算法的故障区段定 位方法的可行性与准确性,将其与常用的 RF、SVM、 XGBoost 算法进行对比。4种算法对配电网故障区 段定位问题进行建模时,均采用 Python 编程通过交 叉验证得到最优超参数,分别将 Ω_{01} 与 Ω_{02} 作为训练 集与测试集,则4种算法的故障区段定位准确率如 表2所示。

表 2 不同算法的故障区段定位准确率

 Table 2
 Fault section location accuracy of different algorithms

| 算法 | 故障区段定位准确率 / % |
|----------|---------------|
| RF | 95.6 |
| SVM | 94.0 |
| XGBoost | 97.1 |
| LightGBM | 98.2 |

由表2可见,4种算法均能学习到特征与标签之 间的映射关系,其中LightGBM算法在解决该问题上 表现最佳,与理论分析结果一致。

2)特征选择的影响。

对比 $I' + S_{BK}$ 和 $\Delta I + S_{BK}$ 这2种特征选择对模型故障区段定位效果的影响,结果如表3所示。由表3可知,采用 $\Delta I + S_{BK}$ 的组合更有优势。

表3 特征选择对故障区段定位准确率的影响

Table 3 Influence of feature selection on

fault section location accuracy

| 特征选择 | 故障区段定位准确率 / % |
|-------------------------|---------------|
| $I' + S_{\rm BK}$ | 94.8 |
| $\Delta I + S_{\rm BK}$ | 98.2 |

5.4 关键因素对故障区段定位效果的影响

1)训练样本数量的影响。

机器学习方法的训练集样本数量会对建模效 果产生显著影响。在样本集 Ω_0 中分别挑选420、700、 1400、2800、5600、8400、11200、14000个样本形成8 组数据集,每组数据集中均包含同等数量的故障样 本和非故障样本,在测试集 Ω_{22} 中对比本文模型在不 同的训练样本数量下的故障区段定位准确率,结果 如图3所示。



图3 故障区段定位准确率与训练样本数量的关系

Fig.3 Relationship between fault section location accuracy and number of training samples

由图3可见,对于IEEE 123节点系统,当训练样本数量小于2000时,故障区段定位准确率随训练样本数量的增加而显著提升,训练样本数量达到2000时故障区段定位准确率接近95%;训练样本数量超过2000后,样本的数量增加对故障区段定位准确率的提升逐渐放缓,当训练集样本量达到8000个以上时,模型已取得较好的训练效果,训练样本数量即使进一步增加,对故障区段定位准确的效果提升也不

再明显。

2)过渡电阻的影响。

当线路发生单相接地故障时,线路电流会发生 变化,随着过渡电阻的增大,电流变化量不断减小, 此时负荷波动、系统阻抗以及噪声对电流的影响会 增大,模型的故障区段定位准确率必然会下降。美 国德克萨斯农机大学的研究^[18]表明,配电网发生高 阻接地故障时,过渡电阻一般在100 Ω以下,进行算 法验证时通常设定为500~1000 Ω^[19]。基于样本集 Ω_{01} 训练本文模型,对不同过渡电阻区段的样本集 $\Omega_{1} - \Omega_{14}$ 进行测试,获得的故障区段定位误判率如 表4所示。由表4可知,本文模型对于过渡电阻小于 800 Ω 的样本可进行准确的故障区段定位,对于过 渡电阻在800~1400 Ω范围内的样本,其误判率可控 制在5%以下。常见的配电网单相接地故障的过渡 电阻一般小于100 Ω,所以本文方法可用于常规配 电网的故障区段定位,且有较大的过渡电阻裕度。

表4 过渡电阻与模型误判率的关系

Table 4 Relationship between transition resistance and model misjudgment rate

| $R_{ m f}$ / Ω | 误判率 / % | $R_{ m f}$ / Ω | 误判率 / % |
|-----------------------|---------|-----------------------|---------|
| [0,100) | 0 | [700,800) | 0 |
| [100,200) | 0 | [800,900) | 0.6 |
| [200,300) | 0 | [900,1000) | 0.8 |
| [300,400) | 0 | [1000,1100) | 1.3 |
| [400,500) | 0 | [1100,1200) | 1.5 |
| [500,600) | 0 | [1200,1300) | 2.3 |
| [600,700) | 0 | [1300,1400) | 3.5 |

3)其他中性点接地方式下本文方法的可迁移性 分析。

除中性点不接地方式外,我国配电网常见的接 地方式还有中性点经消弧线圈接地、小电阻接地方 式。为测试本文方法的可迁移性,在5.1节设置的基 础上保持其他条件不变,将中性点接地方式修改为中 性点经小电阻接地方式和中性点经消弧线圈接地方 式,前者的中性线电阻为20Ω,后者的过补偿度为 5%。通过仿真得到过渡电阻在0~1400Ω范围内的 8000个样本,按0.85:0.15定位比例划分训练集和测 试集,应用本文方法得到的测试结果如表5所示。

表 5 中性点接地方式对故障区段定位准确率的影响

Table 5Influence of neutral point groundingmode on fault section location accuracy

| 中性点接地方式 | 故障区段定位准确率 / % |
|---------|---------------|
| 小电阻 | 99.5 |
| 消弧线圈 | 99.3 |

由表5可知,在中性点经小电阻接地方式和经 消弧线圈接地方式下,本文方法的故障区段定位准 确率更高,这与上述中性点接地方式下,配电网故障 前后电流稳态值变化量更大、特征更明显相一致,所 以本文方法可迁移应用于上述中性点接地方式。

5.5 其他因素对故障区段定位效果的影响

1)缺失数据处理方法的影响。

以样本集 $\Omega_{15} - \Omega_{17}$ 作为测试集,分别采用特殊 值(采用特殊符号"NaN")填充法、条件平均值填充 法、K最近距离邻法对测试集中的缺失数据进行补 齐,然后采用基于 Ω_{01} 样本集训练好的LightGBM模 型进行测试,故障区段定位准确率结果如表6所示。

表6 缺失数据处理方法对故障区段定位准确率的影响

 Table 6
 Influence of missing data processing method

 on fault section location accuracy

| 011 | 10011 000110 | | | |
|---------------|--------------|----------|-------|--|
| | 故 | 障区段定位准确。 | 率 / % | |
| 剱掂碳大 亥 / ∞ | 特殊值 | 条件平均值 | K最近距离 | |
| 平 / % | 填充法 | 填充法 | 邻法 | |
| 1 | 98.8 | 98.9 | 97.4 | |
| 2 | 98.4 | 98.4 | 94.5 | |
| 5 | 98.7 | 95.8 | 87.4 | |

由表6可见:针对故障区段定位问题,特殊值填 充法与条件平均值填充法明显优于K最近距离邻 法,且少量的数据缺失不会对本文模型的故障区段 定位效果产生明显影响;当数据缺失率大于2%后, 条件平均值填充法的故障区段定位准确率明显开始 下降,特殊值填充法则能保持稳定的缺失数据处理 效果,具有明显的优势。

2)故障前、后负荷波动的影响。

本文采用 $\Delta I + S_{BK}$ 组合作为特征,故障前、后的稳态负荷有可能发生变化,但由于中性点不接地配电网发生单相接地故障后,线电压仍然是对称的,所以可认为故障前、后的负荷变化是随机的正常波动。采用基于样本集 Ω_{01} 训练好的模型,将样本集 $\Omega_{18} - \Omega_{21}$ 作为负荷波动测试集,测试结果如表7所示。表中,±5%表示前后2次测量值对应的负荷在95%~105%之间随机取值,最大负荷变化可达10%,其他同。

表7 故障前、后负荷波动对故障区段定位准确率的影响

Table 7 Influence of load fluctuation on

fault section location accuracy

| | , |
|----------|---------------|
| 负荷波动 / % | 故障区段定位准确率 / % |
| ±5 | 93.10 |
| ± 3 | 98.20 |
| ± 2 | 98.85 |
| ±1 | 98.80 |

由表7可知,若故障前、后负荷波动在±3%之 内,则故障区段定位准确率基本不受影响,若负荷波 动超过上述范围,则故障区段定位准确率会随着故 障前、后负荷波动范围的增加而下降。

5.6 误判结果识别方法验证

1)利用类别置信度识别误判。 通过采用不同的随机数种子,对样本集 Ω。进行 10次训练集与测试集划分,重复10次模型训练与预测过程,并统计每次测试集中的误判线路情况,得到 10组训练测试集误判结果如图4所示。图中,坐标 (X, Y, N_p) 表示第10(X-1)+Y条线路的故障样本数 量为 N_p ,则1—117为线路编号,118表示未故障(0 状态),119、120为空。



图4 误判结果分布示意图

Fig.4 Schematic diagram of misjudgment distribution

由图4可知,误判样本主要集中在线路1、7、8、 13、52—54、57、60、67以及未故障(0)这11种情况 下,误判样本类别置信度分布如图5所示。图中,N^r, 为置信度小于阈值0.98的样本数量。



图 5 误判样本类别置信度分布图

Fig.5 Shematic diagram of confidence distribution of misclassified sample categories

由图5可知:通过类别置信度阈值判别,能够将 大部分误判样本划入存疑区域;同时分析正判样 本在置信度阈值上下的分布情况,98.7%的正判样 本类别置信度在阈值0.98以上,因此类别置信度阈 值设置为0.98能够有效提取可疑的故障区段定位 样本。

2)标注易错线路消除误判影响。

由图4确定的10条易误判线路中,线路1、7、8、 13为相邻的上下级线路,线路52—54、57、60、67同 样也为相邻的上下级线路,发生在这2个区域(区域 标注见附录A图A3)的10条线路中的误判占全部误 判的89.6%,可以确定误判的发生具有一定的区域 聚集性。通过易错区域提前标注结合实际类别置信 度,可以有效降低误判的影响:将落入易错标注线路 集且类别置信度低于0.98的判断结果设置为可疑, 然后基于故障投诉电话或其他故障区段定位方法进 行进一步的确定,或在易错线路配置其他精确量测 设备进行故障区段的准确定位。

6 结论

中性点不接地配电网单相接地故障前、后,所有 线路电流稳态有效值变化量与故障区段之间存在一 定的映射关系,同时具备通过仿真建模获取大量样 本的条件,由此提出了基于LightGBM算法的中性点 不接地配电网单相接地故障区段定位方法。该方法 从全网角度预测最可能的故障线路,并通过大数据 样本训练得到故障区段定位模型,该模型的本质是 多分类模型,避免了基于机理方法的阈值设置,为传 统机理方法较难解决的问题提供了思路;在应用研 究中,引入类别权重消除了训练样本类别不平衡问 题,采用特殊值填充法较好地解决了数据缺失的问 题;提出了置信度辅助判别法与易错线路标注法,有 效提高了故障选线结果的可靠性。本文方法对于过 渡电阻在800Ω以下的高阻接地故障具有可靠的识 别定位效果。实际配电网扩建时需要增加训练样 本,若线路有投退运行的需要,则需要将相应的负荷 开关状态加入特征集中。

本文研究未考虑间歇性故障及电弧的不稳定燃 烧引起的电流稳态有效值不稳定的情况,以及分布 式电源接入情况,这将作为今后研究的关注点。

附录见本刊网络版(http://www.epae.cn)。

参考文献:

- [1]张保会,尹项根. 电力系统继电保护[M]. 2版. 北京:中国电力出版社,2010:51-57.
- [2] 白浩,李鹏,袁智勇,等.人工智能在配电网高阻接地故障检测 中的应用及展望[J].南方电网技术,2019,13(2):34-44.
 BAI Hao,LI Peng,YUAN Zhiyong, et al. Application and prospect of artificial intelligence in high impedance fault detection of distribution network[J]. Southern Power System Technology,2019,13(2):34-44.
- [3] 唐金锐,尹项根,张哲,等. 配电网故障自动定位技术研究综述
 [J]. 电力自动化设备,2013,33(5):7-13.
 TANG Jinrui, YIN Xianggen, ZHANG Zhe, et al. Survey of fault location technology for distribution networks
 [J]. Electric Power Automation Equipment,2013,33(5):7-13.
- [4] 庞清乐.小电流接地故障选线与定位技术[M].北京:电子工业出版社,2010:1-10.
- [5] 刘志文,董旭桂,邹林,等.基于零序电流衰减周期分量的高阻 接地故障区段定位[J].电力系统自动化,2020,44(7):161-168. LIU Zhiwen, DONG Xuzhu, ZOU Lin, et al. Section location method for high impedance grounding fault based on declining periodic component of zero sequence current[J]. Automation of Electric Power Systems,2020,44(7):161-168.
- [6] 童晓阳,张绍迅.基于灰色关联度的配电网故障区段定位与类型识别方法[J].电力系统自动化,2019,43(4):113-118.
 TONG Xiaoyang,ZHANG Shaoxun. Grey relational degree based fault section location and type recognition method for distribution network[J]. Automation of Electric Power Systems, 2019, 43(4):113-118.
- [7] 郭壮志,陈涛,徐其兴,等. 配电网故障区段定位的互补松弛约 束新模型与算法[J]. 电力自动化设备,2020,40(5):129-137. GUO Zhuangzhi,CHEN Tao,XU Qixing,et al. Novel fault section location model for distribution network with complementary

relaxation constraints and its algorithm[J]. Electric Power Automation Equipment, 2020, 40(5): 129-137.

- [8] 王英英,罗毅,涂光瑜.基于粗糙集与决策树的配电网故障诊断方法[J].高电压技术,2008,34(4):794-798.
 WANG Yingying,LUO Yi,TU Guangyu. Fault diagnosis method for distribution networks based on the rough sets and decision tree theory[J]. High Voltage Engineering, 2008, 34(4): 794-798.
- [9] THUKARAM D, KHINCHA H P, VIJAYNARASIMHA H P. Artificial neural network and support vector machine approach for locating faults in radial distribution systems[J]. IEEE Transactions on Power Delivery, 2005, 20(2):710-721.
- [10] 孙雅明,吕航. Petri 网和冗余纠错技术结合的配网故障区段定 位新方法[J]. 中国电机工程学报,2004,24(10):61-67. SUN Yaming,LÜ Hang. A new approach of the fault section locating for distribution systems based on Petri nets in combination with redundant correcting technique[J]. Proceedings of CSEE,2004,24(10):61-67.
- [11] 宋国兵,李广,于叶云,等. 基于相电流突变量的配电网单相接 地故障区段定位[J]. 电力系统自动化,2011,35(21):84-90. SONG Guobing,LI Guang,YU Yeyun, et al. Single-phase earth fault section location based on phase current fault component in distribution network[J]. Automation of Electric Power Systems,2011,35(21):84-90.
- [12] GUOLIN K, QI M, THOMAS F, et al. LightGBM: a highly efficient gradient boosting decision tree[J]. Advances in Neural Information Processing Systems, 2017, 30: 3146-3154.
- [13] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45 (1):5-32.
- [14] CORTES C, VAPNIK V. Support-vector networks [J]. Machine Learning, 1995, 20(3):273-297.
- [15] CHEN T, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM, 2016:785-794.

- [16] 张晨宇,王慧芳,叶晓君. 基于XGBoost算法的电力系统暂态 稳定评估[J]. 电力自动化设备,2019,39(3):77-83,89.
 ZHANG Chenyu, WANG Huifang, YE Xiaojun. Transient stability assessment of power system based on XGBoost algorithm[J]. Electric Power Automation Equipment, 2019, 39(3): 77-83,89.
- [17] SHEN S F, LIN D, WANG H F, et al. An adaptive protection scheme for distribution systems with DGS based on optimized Thevenin equivalent parameters estimation[J]. IEEE Transactions on Power Delivery, 2017, 32(1):411-419.
- [18] RUSSELL B D, AUCOIN B M, TALLEY T J. Detection of arcing faults on distribution Feeders[R]. College Station, USA: Texas A&M University, 1982.
- [19] 熊卫红,毛兴华,李景禄,等.小电阻接地方式对人身安全的影响及智能电阻接地方式研究[J].电力系统保护与控制,2019,47(14):166-172.

XIONG Weihong, MAO Xinghua, LI Jinglu, et al. Influence of small resistance grounding mode on personal safety and research of intelligent resistance grounding mode[J]. Power System Protection and Control, 2019, 47(14):166-172.

作者简介:



郑一斌(1996—),男,浙江衢州人,硕 士研究生,主要研究方向为配电网故障定位 (E-mail:yibin_zheng@zju.edu.cn);

王慧芳(1974—),女,浙江宁波人,副 教授,博士,通信作者,主要研究方向为电力 系统继电保护、电网状态检修、电力系统数 据挖掘(E-mail:huifangwang@zju.edu.cn);

张 磊(1986—),男,江苏扬州人,高 级工程师,主要研究方向为电力系统保护控

制、智能电网技术及应用(E-mail:zhang_l7@ecidi.com)。 (编辑 任思思)

Single-phase grounding fault section location in distribution network based on LightGBM algorithm

ZHENG Yibin¹, WANG Huifang¹, ZHANG Lei², JIANG Kuan¹, YANG Wenbin², ZHOU Caiquan²

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China;

2. POWERCHINA Huadong Engineering Corporation Limited, Hangzhou 311122, China)

Abstract: Limited by the configuration of measurement equipment, the fault section location in the isolated neutral distribution network widely distributed in China has been a problem for a long time. Faced with this issue, a fault section location method based on LightGBM (Light Gradient Boosting Machine) algorithm under the condition of limited measurement is proposed. The proposed method can accurately locate the fault section in the isolated neutral distribution network by regular RMS (Root-Mean-Square) values of phase currents without additional installation of PMU (Phasor Measurement Unit), μ PMU (micro-Phasor Measurement Unit), voltage transformer and other measuring devices. By analyzing the measurement information and grid structure of the original target distribution network, the LightGBM algorithm is used to establish offline fault section location model to quickly locate the fault section online. Specifically, the variations of the pre-fault and post-fault steady-state RMS values of line currents are taken as the core features. Otherwise, aiming at the possible deviation of model prediction results, the error-prone line labeling and category probability method are used to assist in identifying misjudgment situation which further improves the reliability of prediction results. The feasibility and efficiency of the proposed method are verified by using a modified IEEE 123-bus test system.

Key words: fault section location; distribution network; superposition principle; LightGBM algorithm; artificial intelligence



图 A1 中性点不接地系统发生单相接地故障前、后的稳态示意图 Fig.A1 Schematic diagram of steady state before and after a single-phase grounding fault in isolated neutral distribution network

图中, E_s 为系统侧电压; Z_s 为系统阻抗; Z_{11} ~ Z_{13} 为 1—3 号线路的线路阻抗; R_f 为故障过渡电阻。



(b) 单相接地故障区段定位在线应用部分

图 A2 方法流程图 Fig.A2 Flowchart of proposed algorithm



图 A3 修改后的 IEEE123 节点系统拓扑图 Fig.A3 Modified IEEE123 node system topology diagram

图中, 阴影区域为易错线路。