

DOI: 10.19783/j.cnki.pspc.226429

基于小波去噪与随机森林的配电网高阻接地故障半监督识别方法

白浩¹, 潘姝慧¹, 邵向潮², 高健鸿³, 李巍¹, 雷金勇¹, 郭谋发³

(1. 南方电网科学研究院, 广东 广州 510663; 2. 广东电网有限责任公司东莞供电局, 广东 东莞 530600;
3. 福州大学电气工程与自动化学院, 福建 福州 350108)

摘要: 针对配电网高阻接地故障识别易受噪声干扰、无标签数据难以利用的难题, 提出一种基于小波去噪与随机森林的高阻接地故障半监督识别方法。区别于监督式学习方法仅利用标签数据, 基于协同训练方法能够充分利用有标签数据与无标签数据。首先, 使用小波阈值去噪算法消除零序电流中的噪声。其次, 采用波峰波谷故障启动算法判断线路是否发生故障或扰动事件。运用小波变换提取零序电流的小波系数作为故障特征。最后, 基于小波系数故障特征构建两个随机森林作为半监督分类器进行协同训练, 从而实现高阻接地故障的检测识别。仿真结果表明, 所提配电网高阻接地故障半监督识别方法可以充分挖掘配电网既有的故障案例中无标注数据蕴含的关键特征, 从而提高故障分类准确率, 具有较强的准确性和灵敏性。

关键词: 配电网; 高阻接地故障; 小波变换; 半监督学习; 随机森林

A high impedance grounding fault semi-supervised identification method based on wavelet denoising and random forest

BAI Hao¹, PAN Shuhui¹, SHAO Xiangchao², GAO Jianhong³, LI Wei¹, LEI Jinyong¹, GUO Moufa³

(1. Electric Power Research Institute, China Southern Grid, Guangzhou 510663, China; 2. Dongguan Power Supply Bureau, Guangdong Power Grid Co., Ltd., Dongguan 530600, China; 3. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China)

Abstract: To solve the problem that the identification of a high impedance grounding fault (HIF) in distribution networks is easily affected by noise, and the fact that it is difficult to use unlabeled data, a semi-supervised identification method of a high resistance grounding fault based on wavelet denoising and random forest is proposed. Different from supervised learning only using labeled data, the method can make full use of labeled and unlabeled data by collaborative training. First, the wavelet threshold denoising algorithm is used to filter the noise of zero-sequence currents. Secondly, the occurrence of an HIF can be detected by the peak and valley fault triggering algorithm. Then, applying wavelet transform to zero-sequence currents, the wavelet coefficients are extracted as fault features. Finally, two random forests are collaboratively trained with selected features to construct a semi-supervised classifier to detect the HIF. The simulation results show that the proposed method can use fully the key features in unlabeled data in the existed fault cases in distribution network to improve the accuracy of fault classification. It has strong reliability and flexibility.

This work is supported by the Science and Technology Project of China Southern Grid Co., Ltd. (No. GDKJXM 20198281).

Key words: distribution network; high impedance grounding fault; wavelet transform; semi-supervised learning; random forest

0 引言

高阻接地故障(high impedance fault, HIF)是配

电网中较难检测的一类单相接地故障, 其故障点过渡电阻阻值大, 故障信号极其微弱, 难以触发传统微机保护装置的启动阈值。同时, 此类故障长期存在将增加山林火灾风险, 进而威胁居民的人身安全。近年来, 多起电气火灾事件亦使得高阻接地故障检

基金项目: 南方电网公司科技项目资助(GDKJXM20198281)

测逐渐成为研究人员关注的热点、难点问题,亟待提出高阻接地故障准确识别的有效方法。

单相接地故障检测方法一般可分为暂态量方法、稳态量方法与人工智能方法。文献[1]提取暂态零模电流中的主谐振分量,结合暂态零模电压计算投影系数作为识别判据进而检测故障。经测量,实际树障试验中故障过渡电阻在故障初期可达数百欧至数万欧,相关电气暂态量难以检测,因此暂态量方法在高阻接地故障时难以奏效^[2-3]。文献[4]通过零序电流的功率变化量构建阈值以区分正常馈线与故障馈线。文献[5]基于南美洲电网的现场样本,结合小波变化与自适应降噪技术为当地电网设置合适的故障软阈值。为了避免保护装置频繁动作,稳态量方法的动作阈值一般较高且反映过渡电阻的能力有限^[6-7]。基于阈值的暂态量或稳态量方法可根据可靠系数、经验公式或自适应方式设置阈值,该类方法结构简单、物理意义明确,但在阈值设置中多依赖人为经验,无法充分利用现场既有故障案例中蕴含的关键信息。

在配电网故障诊断领域,人工智能方法主要表现为监督式学习,建立数据样本与样本标签之间的映射关系,实现故障检测与识别^[8-10]。文献[11]通过经验模态分解得到零序电流的模态分量作为故障特征,利用蚁群算法对模糊神经网络进行优化,最后利用该网络实现故障分类。文献[12]采用VMD得到故障相电流的第一层本征模态分量IMF₁,利用奇异值分解提取IMF₁的故障上升趋势,最后由支持向量机进行分类。文献[13]将故障相电流与相电压的小波系数能量作为故障特征,并定义不同馈线的能量矩阵用于构建半监督分类器。上述监督式学习算法均需要大量有标签数据构建样本集合用于分类器的训练。现有基于人工智能的配电网故障检测方法依赖人工标注样本,在标注过程将耗费大量人力,同时,人为标注的误判将影响所构建数据集质量,影响基于数据集的分类器模型性能。因此,需要一种能够同时利用有标签数据与无标签数据的半监督学习^[14]方法,有效减少人工标注过程的标签误判与资源浪费。

本文提出一种基于小波去噪与随机森林的配电网高阻接地故障半监督识别方法。首先,使用小波阈值去噪算法过滤故障信号噪声;其次,运用波峰波谷故障启动算法判断线路是否发生HIF;然后,采用小波变换提取样本的小波分量,并使用随机森林筛选出有效特征;最后,基于协同训练方法构建半监督分类器。在PSCAD/EMTDC上进行方法的

仿真验证。结果表明,所提方法能够降低噪声干扰,减少人工标注过程中的标注样本数量,充分利用标签数据与无标签数据进行训练,具有较强的准确性、灵敏性与较高的分类准确率。

1 故障启动算法

中低压配电网中常发生单相高阻接地故障与其他不对称扰动事件,这类不对称事件均会引起线路电流的周期性突变,高阻接地故障常持续8个周期以上,扰动事件持续6个周期左右^[13]。在实际配电网中,系统可能存在轻微的三相不对称,则对应零序电流为较小的交流量。为了避免启动算法误动作,经大量仿真试验,本文将算法启动阈值设置为0.1 A,以躲过正常工况时零序电流幅值。在实际工程中将根据现场具体工况对该阈值进行重新整定。一般而言,不对称事件持续的一个工频周期内至少存在两个波峰波谷点。基于上述现象,本文提出一种通过检测零序电流连续波峰波谷点的故障启动算法。当8个周期内出现连续的12个波峰波谷点时,则启动后续的故障识别算法。

首先,将采集的线路零序电流按式(1)进行零均值归一化处理。

$$x^* = \frac{x - \bar{x}}{\sigma} \quad (1)$$

式中: x^* 表示归一化处理后的样本; x 表示输入样本; \bar{x} 和 σ 分别表示样本的均值与方差。

其次,求取零序电流信号相邻点的差分值,设第 $n-1$ 、 n 、 $n+1$ 个点的幅值为 y_{n-1} 、 y_n 、 y_{n+1} ,差分值分别为 $dy[n-1]$ 、 $dy[n]$ 、 $dy[n+1]$ 。当满足 $dy[n-1] \geq 0$ 、 $dy[n] \geq 0$ 、 $dy[n+1] \leq 0$ 且 y_{n-1} 、 y_n 、 $y_{n+1} \geq \delta$ 时,则标记 n 点为可能波峰点(k_n, y_{kn})。其中 δ 为峰值的最小阈值,用于消除信号小范围波动造成的干扰,设为0.1 A。然后,设第 $m-1$ 、 m 、 $m+1$ 个点的幅值为 y_{m-1} 、 y_m 、 y_{m+1} ,差分值为 $dy[m-1]$ 、 $dy[m]$ 、 $dy[m+1]$ 。当满足 $dy[m-1] \leq 0$ 、 $dy[m] \leq 0$ 、 $dy[m+1] \geq 0$ 且 y_{m-1} 、 y_m 、 $y_{m+1} \leq -\delta$ 时,标记 m 点为可能波谷点(p_m, y_{pm})。

最后,设 g 、 i 、 h 为3个连续的可能波峰点, e 、 j 、 s 为3个连续的可能波谷点, i 、 j 为相邻的可能波峰点、可能波谷点,求取相邻可能波峰点横纵坐标 k_g 、 k_h 与 k_j 的差值,相邻可能波谷点横纵坐标 p_e 、 p_s 与 p_j 的差值以及相邻波峰波谷点的差值。当可能波峰点、可能波谷点的差值满足式(2)时,则确定 i 点为波峰点, j 点为波谷点。

$$\begin{cases} |k_g - k_i - N_T| \leq \frac{N_T}{5}, & |p_e - p_j - N_T| \leq \frac{N_T}{5} \\ |k_h - k_i - N_T| \leq \frac{N_T}{5}, & |p_s - p_j - N_T| \leq \frac{N_T}{5} \\ |k_i - p_j - \frac{N_T}{2}| \leq \frac{N_T}{10} \end{cases} \quad (2)$$

式中: k_g 、 k_h 与 k_i 为可能波峰点的采样点横坐标; p_e 、 p_s 与 p_j 为可能波谷点的采样点横坐标; N_T 为一周波的采样点数; $N_T/5$ 和 $N_T/10$ 表示偏差裕度。

2 小波变换与去噪算法

2.1 离散小波变换

离散小波变换 (discrete wavelet transform, DWT) 是一种离散信号二进制处理方法, 能够得到信号的多尺度频带, 适合分析时变的非线性暂态信号^[15]。相较于连续小波变换, DWT 具备多尺度分析的优势。对比希尔伯特-黄变换^[16], DWT 在低频部分的冗余信息更少。对比短时傅里叶变换^[17]与 Wigner-Ville 分布^[18], DWT 无需权衡时间分辨率与频率分辨率。配电网发生高阻接地故障与不对称扰动时, 线路零序电流产生暂态突变, 适合使用 DWT 进行分析。

DWT 通过多层低通与高通滤波器, 分别得到原信号的多级近似分量与细节分量, 如图 1 所示。其中 LF 表示低通滤波器, HF 表示高通滤波器, A_i 为近似分量, 表示原信号的高尺度低频分量, D_i 表示细节分量, 表示原信号的低尺度高频信息, $i=1,2,L,6$ 。文献^[19]指出, 高阻接地故障与扰动事件在 0~312.5 Hz 具有较明显的特征区别。如图 2 所示, 分别对高阻接地故障零序电流波形与非高阻接地故障零序电流波形进行离散小波变换, 获得多级小波分量。根据对比可知, 对应 0~312.5 Hz 的第 4 层细节分量 D_4 具有显著差异。进而, 为了充分利用高阻接地故障带来的丰富谐波分量且避免基波分量的影响, 本文在采样频率为 10 kHz 的条件下, 选用第 5 层与第 6 层的细节分量 D_5 (156.25~312.5 Hz)、 D_6 (78.125~156.25 Hz) 作为故障特征。

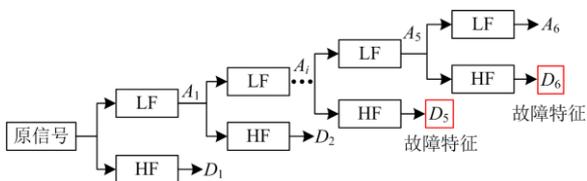


图 1 离散小波变换示意图

Fig. 1 Diagram of discrete wavelet transform

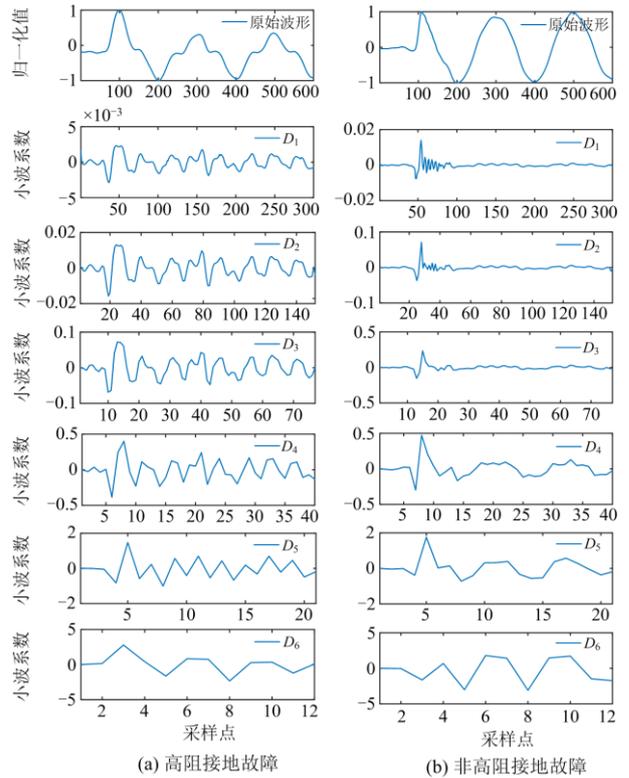


图 2 不同事件的零序电流波形及其多级小波分量

Fig. 2 Zero-sequence current waveforms of different events and their multi-level wavelet components

2.2 小波阈值去噪算法

高阻接地故障电流幅值微弱, 易受噪声干扰。为保证故障识别算法的准确性, 需要进行噪声过滤。设备采集零序电流信号由原始信号与噪声信号叠加组成, 经过小波变换得到近似分量与细节分量, 各分量仍由原始信号与噪声信号组成, 且噪声多为高频分量, 位于小波变换的细节分量中。原始信号的模值占比多大于噪声信号, 且实际噪声由多个来源复合而成, 可看作不同独立概率分布的变量之和。根据中心极限定理, 噪声分布将趋近于高斯分布。因此, 根据高斯分布的 3σ 原则, 噪声将主要分布于 $[-3\sigma, 3\sigma]$ 的区间内。本文使用极大极小阈值 λ , 如式(3)所示。

$$\lambda = \begin{cases} 0.3936 + 0.1829 \left(\frac{\ln N}{\ln 2} \right), & N > 32 \\ 0, & N \leq 32 \end{cases} \quad (3)$$

式中, N 为采样点数。

对各层分量在 $[-3\sigma, 3\sigma]$ 内的数据进行处理, 将各层去噪后的分量重构得到去噪信号。其中, σ 表示噪声的近似方差, 如式(4)所示。

$$\sigma = \text{median}(s_j(k))/0.6745 \quad (k = 0, 1, L, N-1) \quad (4)$$

式中： $s_j(k)$ 表示某一层分量的数据点；median 表示取得数据的中值。

在小波阈值去噪算法的应用中，不仅需要考虑噪声的分布范围，还需要考虑方差 σ 对阈值的影响，此处将各层分量的 σ 与各层的极大极小阈值相乘来得到新阈值。阈值的计算方法可分为硬阈值法与软阈值法。硬阈值法根据计算得到阈值 λ ，将需要去噪的小波分量中模值小于 λ 的点置零，大于 λ 的点保留；软阈值去噪法将分量中模值大于 λ 的点减去 λ ，模值小于 λ 的点置零。相较于硬阈值法，软阈值法能够更加平滑地去除噪声，保证波形的还原度，因此本文选用软阈值法进行去噪。

3 半监督学习算法

3.1 随机森林算法

随机森林由多棵决策树构成，可通过信息熵或信息增益构建样本的分类标准，同时适用于处理离散与连续数据^[20]。基于信息熵的随机森林构建过程^[21]如下所述。

1) 将需要进行分类与评估的样本保存为样本库 S ，并计算其基尼指数 $G(S)$ ，如式(5)所示。

$$G(S) = \sum_{i=1}^n \sum_{i' \neq i} p_i p_{i'} = 1 - \sum_{i=1}^n p_i^2 \quad (5)$$

式中： p_i 表示标签为 i 的样本在样本库 S 中的占比； $p_{i'}$ 表示标签为非 i 的样本在样本库 S 中的占比； n 表示样本库的总标签数； $G(S)$ 越小则表示标签为 i 的样本在样本库的占比越大。实际应用中使用 0 作为高阻接地故障标签，1 作为非高阻接地故障标签。

2) 样本共具有 K 个特征样本，在算法实际应用中， K 个特征样本是将 D_5 、 D_6 当作一个特征，经过随机森林筛选后，保留下有效的点作为输入分类的故障特征。设其某一特征为 k ，据此划分标签为 i 的样本 S_i^k ，并计算特征 k 的基尼指数 $G(S, k)$ ，如式(6)所示。

$$G(S, k) = \sum_{i=1}^n \frac{|S_i^k|}{|S|} G(S_i^k) \quad (6)$$

式中， $|S|$ 与 $|S_i^k|$ 表示对应的样本数量。

3) 有放回地抽取 m 个样本，随机选择 h 个特征 ($h \ll K$)，由 h 个特征中基尼指数最小的特征构建决策树枝，遍历特征后，生成 m 棵决策树。

4) 重复步骤 2)、3)，满足单棵决策树枝下的标签一致或遍历 K 个特征，则单棵决策树构建成功。当所有决策树构建完毕时，建成随机森林，根据森林中所有决策树的分类投票结果决定输入样本的标

签，随机森林构建过程如图 3 所示。

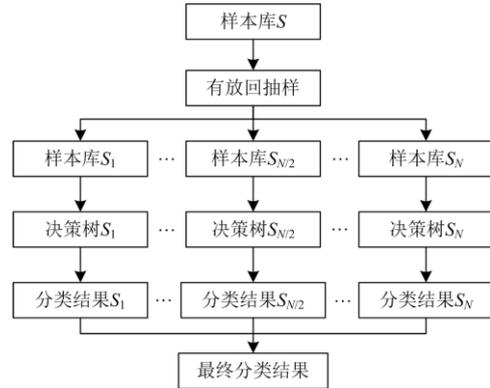


图 3 随机森林构建过程

Fig. 3 Random forest construction process

随机森林是一种集成算法，能够实现高度并行化训练与故障特征的重要性划分，相较普通单一分类器算法具有更高的准确性。下文采用协同训练方法构造半监督分类器，根据提取的故障特征进行分类。

3.2 协同训练算法

配电网中大量录波数据缺乏明确的标签，只能依靠人为观察并标注数据，这将消耗大量人力物力，也不利于充分利用采集的无标签样本。为有效利用无标签样本，本文采用协同训练方案^[22]，由两个随机森林分类器 (random forest 1 与 random forest 2，以下简称为 RF_1 与 RF_2) 构成半监督分类器。协同训练过程如图 4 所示，具体步骤如下所述。

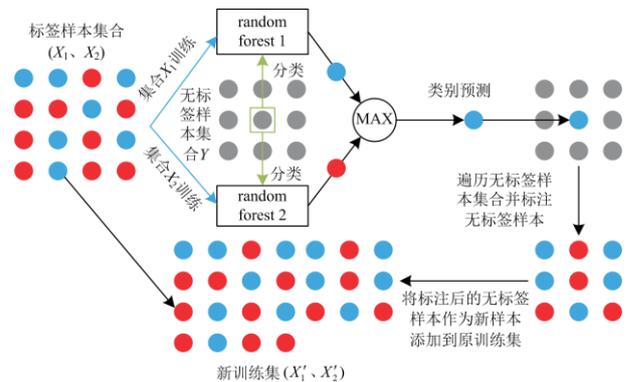


图 4 协同训练过程

Fig. 4 Diagram of collaborative training

1) 利用小波阈值去噪对有标签的零序电流数据进行处理，提取相关小波系数用于训练随机森林，构建两个均等的标签样本集合 X_1 、 X_2 ，分别训练两个不同的随机森林模型。

2) 用训练好的随机森林模型对无标签样本集合 Y 中的无标签样本进行标注，并将标注后的样本作为新样本加入原训练集中构建新训练集，即利用

RF_1 对 Y 中元素进行标记, 并将标记后的数据添加至 X_2 作为新训练集 X'_2 , 用 RF_2 对 Y 中元素进行标记, 并将标记后的数据添加至 X_1 作为新训练集 X'_1 。

3) 利用新训练集再次训练两个随机森林模型, 并重复上述操作, 直到所有无标签样本均被两个随机森林标注上同一标签时, 则完成对无标签数据的学习。

4) 训练完成后, 两个随机森林得到确定的权值, 当两个随机森林的分类结果不同时, 选择权值占比较大的分类结果作为最终结果。

4 高阻接地故障识别算法流程

配电网高阻接地故障协同训练智能识别算法的检测流程如图 5 所示, 具体步骤如下。

1) 信号去噪: 实时采集线路上的零序电流, 长度为 8 个周波, 使用小波阈值去噪算法过滤信号中附带的噪声。

2) 故障启动: 当波峰波谷算法检测到出现 12 个波峰波谷点, 判定为疑似故障发生, 启动后续的故障识别算法。

3) 特征提取: 使用小波变换得到去噪后零序电流的 6 层小波分量, 并保留第 5 层与第 6 层的小波分量 D_5 、 D_6 , 将这两个分量拼接构成故障特征。

4) 特征分类: 使用训练好的半监督分类器对故障特征进行分类以区分高阻接地故障与非高阻接地故障。

5 仿真算例分析

为验证算法准确性, 基于 PSCAD/EMTDC 仿真软件建立谐振接地配电网, 如图 6 所示, 采样率

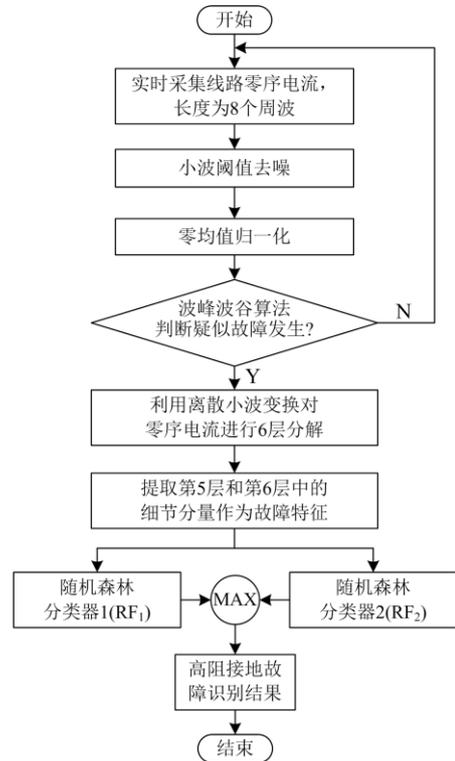


图 5 识别算法流程图

Fig. 5 Flow chart of identification method

为 10 kHz, 线路参数如表 1 所示。此处补偿度取 8%, 消弧线圈电感值为 0.4223 H。选用并联 Emanuel 模型进行接地电弧建模^[23], 如图 7 所示。该模型设置可变电阻实现故障电流幅值半周期不对称性, 采用两个反并联二极管与直流电源串联来模拟故障电流的零休特性, 同时, 通过调节直流电源幅值、电阻的大小以及开关通断时间来模拟多重故障重叠的效果。该模型的参数设置如表 2 所示。

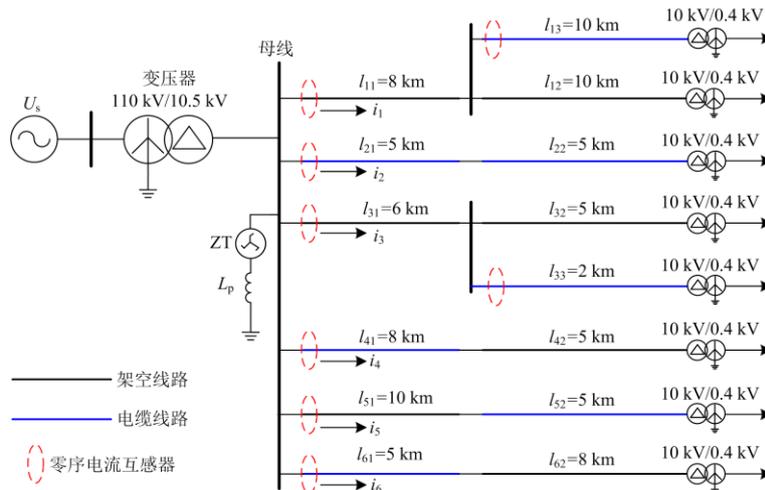


图 6 10 kV 谐振接地配电网

Fig. 6 10 kV resonant ground distribution network

表 1 配电网线路参数

Table 1 Line parameters of distribution network

线路	相序	线路电阻/ (Ω/km)	线路电感/ (mH/km)	线路电容/ (μF/km)
架空线	正序	0.1700	1.2000	0.0097
	零序	0.2300	5.4800	0.0060
电缆线	正序	0.2700	0.2550	0.3390
	零序	2.7000	1.0190	0.2800

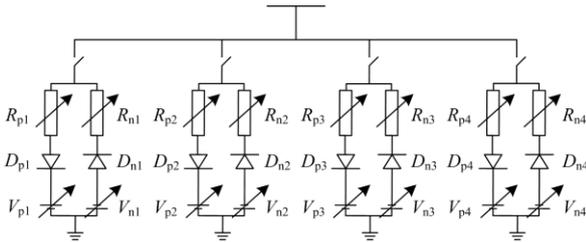


图 7 Emanuel 模型

Fig. 7 Emanuel model

表 2 并联 Emanuel 模型参数设置

Table 2 Parameter of parallel Emanuel model

可变电阻	阻值/Ω	直流电源	电压/kV
$R_{p1}, R_{p4}, R_{n1}, R_{n4}$	500~600	$V_{p1}, V_{p2}, V_{p3}, V_{p4}$	3.95~4.35
$R_{p2}, R_{p3}, R_{n3}, R_{n2}$	400~500	$V_{n1}, V_{n2}, V_{n3}, V_{n4}$	2.74~3.46

本文对高阻接地故障、负荷投切、电容器投切等不对称事件进行仿真并采集对应的零序电流，仿真事件参数如表 3 所示。按照表 3 采集对应事件的线路零序电流作为故障数据集，添加噪声构成信噪比为 3~30 dB 的噪声数据集，再经由小波阈值去噪算法过滤噪声构成去噪数据集，将上述噪声数据集与去噪数据集经由波峰波谷启动算法进行测试，测试结果如图 8 所示。

表 3 仿真事件参数

Table 3 Parameter of simulation events

不对称事件	数量	位置	样本类型
负荷投切	85	$l_{12}, l_{13}, l_{21}, l_{33}, l_{52}$	不对称事件
高阻接地故障	170	$l_{12}, l_{13}, l_{22}, l_{32}, l_{52}$	发生前 1 周波+
电容器投切	85	$l_{12}, l_{13}, l_{21}, l_{33}, l_{52}$	后 7 周波

当信噪比降低至 15 dB 以下时，噪声数据集的启动准确率明显下降，而去噪数据集保持着较高的准确率，可见，小波阈值去噪算法能够有效提高启动算法的准确性。以 10 dB 噪声波形为例，部分标记波峰波谷的波形如图 9 所示。其中，图 9(a)~图 9(c)分别表示高阻接地故障、负荷投切与电容器投切发生时线路的零序电流，图中红点表示算法检测到的波峰与波谷点。纵向对比原始波形、信噪比为 10 dB 的噪声波形与小波阈值去噪后的波形可知，

该启动算法在 3 种条件下均能在 8 个周期内检测到 12 个波峰波谷点，符合启动条件。然而，受噪声影响，10 dB 噪声波形所检测到的波峰波谷点会发生部分缺失，而经小波阈值去噪后的波形则不会存在未检测波峰波谷点的情形。因此，为提高启动算法的准确性，采集的零序电流需要先经过小波阈值去噪算法过滤噪声，再经由波峰波谷启动算法进行判别是否发生故障或扰动，若启动成功，则采用后续的故障识别算法做进一步判断。

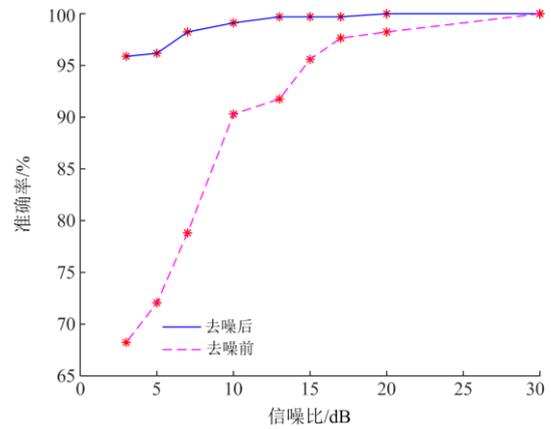


图 8 启动算法抗噪测试准确率

Fig. 8 Accuracy of start-up algorithm anti-noise test

故障识别算法启动后，选择启动算法中检测到的波峰波谷点前 1 个与后 2 个周波的波形作为一组故障样本。当 1 组长度为 8 个周波的零序电流检测到 12 个波峰波谷点时，则至少有 8 组故障样本可用于识别。将 8 组样本经小波变换得到的第 5 层与第 6 层细节分量拼接作为故障特征，输入两个随机森林构成的半监督分类器中进行分类。若 8 组波形中有 5 组以上波形判定为高阻接地故障，则确认发生高阻接地故障。

t 分布随机近邻嵌入(t-SNE)是常用的降维算法^[24]，其能够提取高维数据的特征。使用 t-SNE 对数据进行降维，能够提取数据样本的分布特征，直观展示数据相似度。可通过 t-SNE 将故障特征进行可视化，图 10(a)和图 10(b)分别表示未经处理的小波第 5 层与第 6 层细节分量与经随机森林筛选故障特征的可视化结果，其中，图标“0”表示高阻接地故障，图标“1”表示非高阻接地故障，可见，未经筛选的故障特征能够区分高阻接地故障与非高阻接地故障，但经随机森林筛选的故障特征比起未经筛选的特征具有更清晰的分界线。因此，由零序电流的第 5 层与第 6 层的小波细节分量(D_5 、 D_6)构成的故障特征经随机森林再次筛选后具备更好的区分性。

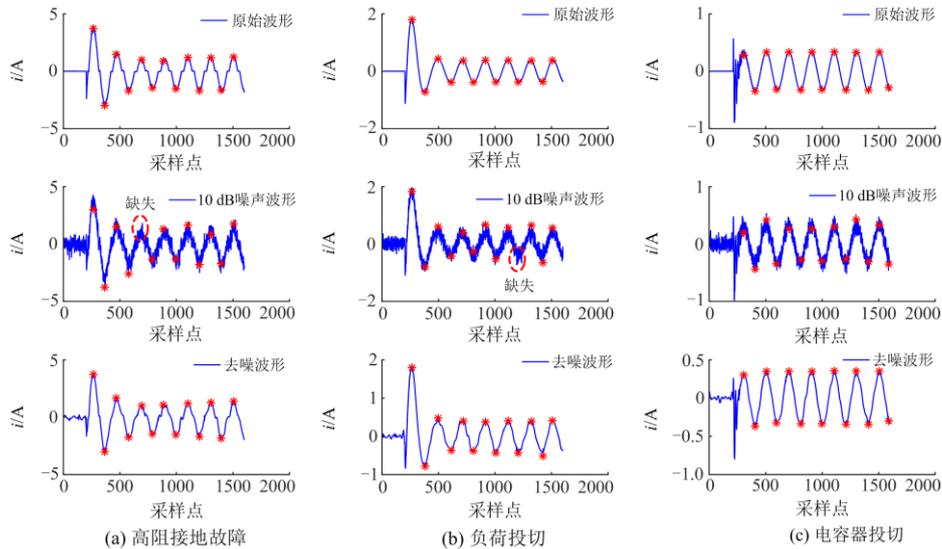


图9 仿真波形

Fig. 9 Simulation waveforms

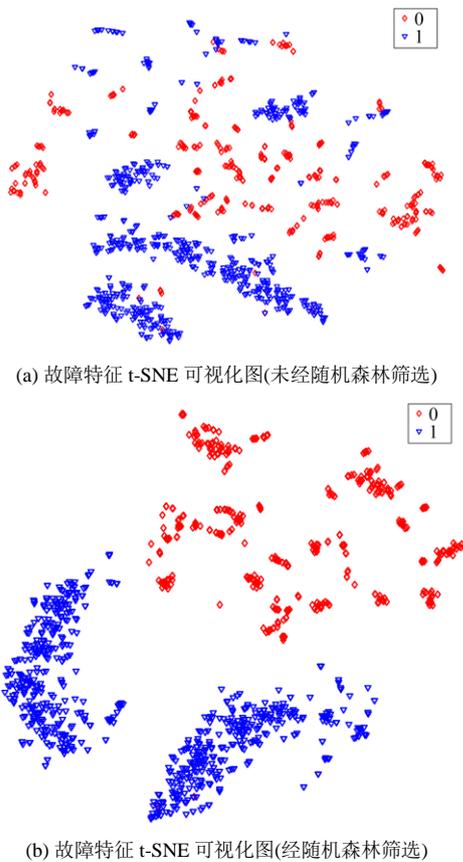


图10 t-SNE可视化的特征提取过程

Fig. 10 Feature extraction process for t-SNE visualization

为验证该半监督识别算法能够在更少的标签样本下获得比监督式学习更高的准确率与泛化能力, 选用文献[25]与文献[12]作为对比, 其中, 文献[25]

使用S变换提取故障电流的特征(方法1), 文献[12]使用经验模态分解得到故障电流的固有模态分量(方法2), 两者均采用神经网络进行监督式分类学习。

为了进一步验证所提方法和对比方法的抗干扰性能, 采用340组无噪声零序电流波形, 并分别加入高斯白噪声, 形成信噪比为7 dB、10 dB、20 dB和30 dB的噪声波形, 噪声波形与原波形共1700组。设置不同场景分配训练集与测试集进行测试。

场景1: 选取238组无噪声波形及其对应噪声波形952组共1190组数据作为训练集, 剩下的510组作为测试集, 方法1和方法2训练集均带有标签, 本文方法所提算法将训练集中595组数据标签删除, 作为无标签数据, 与有标签数据共同训练, 分类结果如表4所示, 当标签训练集充足时, 方法1和方法2取得了较高的准确率, 本文方法在标签训练集减少一半后准确率仍高于另外两种方法。

场景2: 选取102组及其对应噪声波形408组, 共510组数据作为训练集, 剩下的1190组作为测试集, 方法1和方法2训练集均带有标签, 本文方法所提算法将训练集中595组数据标签删除, 作为无标签数据, 与有标签数据共同训练, 测试结果如表4所示。相对于场景1, 方法1和方法2随着训练集中标签数据的减少, 准确率明显下降, 本文方法依旧保持着与场景1接近的准确率。基于协同训练的半监督算法在更少的标签训练集下, 优于所对比的监督式算法, 这是由于协同训练通过两个分类器相互学习无标签数据, 从而获得更准确的分类效果。

表 4 实验对比结果

Table 4 Experimental comparison results

场景	方法	标签 训练集	无标签 训练集	测试集	准确率/%
场景 1	方法 1	1190	0	510	82.35
	方法 2	1190	0	510	88.24
	本文方法	595	595	510	99.41
场景 2	方法 1	510	0	1190	82.03
	方法 2	510	0	1190	82.92
	本文方法	255	255	1190	99.15
场景 3	本文方法	395	795	510	98.03
		155	355	1190	97.05

场景 3: 分别将场景 1 中 200 组标签数据和场景 2 中 100 组标签数据转变为无标签数据对本文算法进行测试, 如表 4 所示, 本文所提方法准确率均略微下降, 当使用过多的无标签数据训练, 则会增大分类器的学习难度, 导致准确率下降。

6 结语

针对配电网高阻接地故障标签数据少, 即有大量无标签数据难以利用的难题, 本文提出一种基于小波去噪与随机森林的高阻接地故障半监督识别方法。经过 PSCAD/EMTDC 仿真测试与现有方法对比, 得到以下结论:

1) 针对含噪声的零序电流, 本文提出波峰波谷启动算法能在信号信噪比为 15 dB 以上的条件下, 有效检测故障与扰动事件的发生。经小波阈值去噪后, 该算法能够获得更好的启动效果。

2) 通过小波变换与随机森林实现特征提取与特征筛选, 筛选后的故障特征经由随机森林组成的协同训练半监督分类器进行分类, 能够在少量有标签训练样本的条件下, 可靠区分高阻接地故障与扰动事件, 并具备一定的抗噪能力。

3) 相较于监督学习, 本文提出半监督学习方案, 能够充分利用有标签数据和无标签数据进行训练, 能够有效提高训练效率, 降低人为标注成本。

参考文献

[1] 薛永端, 李娟, 陈筱菁. 谐振接地系统高阻接地故障暂态选线与过渡电阻辨识[J]. 中国电机工程学报, 2017, 37(17): 5037-5048, 5223.
XUE Yongduan, LI Juan, CHEN Xiaoru. Faulty feeder selection and transition resistance identification of high impedance fault in resonant grounding system using transient signals[J]. Proceedings of the CSEE, 2017, 37(17): 5037-5048, 5223.

[2] 管廷龙, 薛永端, 徐丙垠. 基于故障相电压极化量的

谐振接地系统高阻故障方向检测方法[J]. 电力系统保护与控制, 2020, 48(23): 73-81.

GUAN Tinglong, XUE Yongduan, XU Bingyin. Method for detecting high-impedance fault direction in a resonant grounding system based on voltage polarization of the fault phase[J]. Power System Protection and Control, 2020, 48(23): 73-81.

[3] 梅睿, 程孟晗, 管廷龙, 等. 电流波形聚类的谐振接地系统高阻接地故障定位方法[J]. 供用电, 2020, 37(5): 36-41.

MEI Rui, CHENG Menghan, GUAN Tinglong, et al. High-resistance grounding fault location method for resonance grounding system based on current waveform clustering[J]. Distribution & Utilization, 2020, 37(5): 36-41.

[4] 龙毅, 欧阳金鑫, 熊小伏, 等. 基于零序功率变化量的配电网单相高阻接地保护[J]. 电工技术学报, 2019, 34(17): 3687-3695.

LONG Yi, OUYANG Jinxin, XIONG Xiaofu, et al. Protection principle of single-phase high resistance fault for distribution network based on zero-sequence power variation[J]. Transactions of China Electrotechnical Society, 2019, 34(17): 3687-3695.

[5] YE H G, SIM S, BRAVO R J. Wavelet and denoising techniques for real-time HIF detection in 12 kV distribution circuits[J]. IEEE Systems Journal, 2019, 13(4): 4365-4373.

[6] 叶远波, 汪胜和. 高阻接地故障时消弧线圈并联小电阻接地的控制方法研究[J]. 电力系统保护与控制, 2021, 49(19): 181-186.

YE Yuanbo, WANG Shenghe. Study on the control method of high impedance faults in the neutral via arc suppression coil paralleled with a low resistance grounded system[J]. Power System Protection and Control, 2021, 49(19): 181-186.

[7] 曾祥君, 黄慧, 喻锴. 基于柔性调控零序电压的配电网高阻接地及单相断线故障的选相方法[J]. 电力系统保护与控制, 2022, 50(3): 9-18.

ZENG Xiangjun, HUANG Hui, YU Kun. Voltage phase selection method for high resistance grounding and a single-phase disconnection fault of a distribution network based on flexible control of zero-sequence voltage[J]. Power System Protection and Control, 2022, 50(3): 9-18.

[8] BAI Hao. AI in arcing-HIF detection: a brief review[J]. IET Smart Grid, 2020, 3(4): 435-444.

[9] JAMALI S, BAHMANYAR A, RANJBAR S. Hybrid classifier for fault location in active distribution networks[J]. Protection and Control of Modern Power Systems, 2020, 5(2): 174-182.

[10] 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.
- [11] 孙鹏, 曹雨晨, 刘洋, 等. 采用二进制蚁群模糊神经网络的配电网故障分类方法[J]. 高电压技术, 2016, 42(7): 2063-2072.
- SUN Peng, CAO Yuchen, LIU Yang, et al. Fault classification technique for power distribution network using binary ant colony algorithm and fuzzy neural network[J]. High Voltage Engineering, 2016, 42(7): 2063-2072.
- [12] LALA H, KARMAKAR S. Detection and experimental validation of high impedance arc fault in distribution system using empirical mode decomposition[J]. IEEE Systems Journal, 2020, 14(3): 3494-3505.
- [13] CHAITANYA B K, YADAV A, PAZOKI M. An intelligent detection of high-impedance faults for distribution lines integrated with distributed generators[J]. IEEE Systems Journal, 2019, 14(1): 870-879.
- [14] ABDELGAYED T S, MORSI W G, SIDHU T S. Fault detection and classification based on co-training of semi-supervised machine learning[J]. IEEE Transactions on Industrial Electronics, 2017, 65(2): 1595-1605.
- [15] GUO M F, YANG N C, YOU L X. Wavelet-transform based early detection method for short-circuit faults in power distribution networks[J]. International Journal of Electrical Power & Energy Systems, 2018, 99: 706-721.
- [16] 高伟, 杨耿杰, 郭谋发. 采用振动信号二维特征向量聚类的配电开关机械状态识别新方法[J]. 福州大学学报(自然科学版), 2017, 45(5): 674-680.
- GAO Wei, YANG Gengjie, GUO Moufa. A novel mechanical state identification method for distribution switch based on vibration signal 2-D feature vector with clustering algorithm[J]. Journal of Fuzhou University (Natural Science Edition), 2017, 45(5): 674-680.
- [17] LIMA É M, DOS SANTOS JUNQUEIRA C M, BRITO N S D, et al. High impedance fault detection method based on the short-time Fourier transform[J]. IET Generation, Transmission & Distribution, 2018, 12(11): 2577-2584.
- [18] CAI K, CAO W, AARNIOVUORI L, et al. Classification of power quality disturbances using Wigner-Ville distribution and deep convolutional neural networks[J]. IEEE Access, 2019, 7: 119099-119109.
- [19] LIN C, GUO M F. Discrete wavelet transform-based triggering method for single-phase earth fault in power distribution systems[J]. IEEE Transactions on Power Delivery, 2019, 34(5): 2058-2068.
- [20] 杨杰, 吴浩, 董星星. 基于电流故障分量特征和随机森林的输电线路故障类型识别[J]. 电力系统保护与控制, 2021, 49(13): 53-63.
- YANG Jie, WU Hao, DONG Xingxing. Transmission line fault type identification based on the characteristics of current fault components and random forest[J]. Power System Protection and Control, 2021, 49(13): 53-63.
- [21] 刘东奇, 曾祥君, 王耀南. 基于信息熵的智能配电变压器终端安全态势评估[J]. 南方电网技术, 2020, 14(1): 18-23.
- LIU Dongqi, ZENG Xiangjun, WANG Yaonan. Security situation assessment of intelligent distribution transformer terminal unit based on information entropy[J]. Southern Power System Technology, 2020, 14(1): 18-23.
- [22] ABDELGAYED T S, MORSI W G, SIDHU T S. Fault detection and classification based on co-training of semi supervised machine learning[J]. IEEE Transactions on Industrial Electronics, 2017, 65(2): 1595-1605.
- [23] 许晔, 郭谋发, 陈彬. 配电网单相接地电弧建模及仿真分析研究[J]. 电力系统保护与控制, 2015, 43(7): 57-64.
- XU Ye, GUO Moufa, CHEN Bin. Modeling and simulation analysis of arc in distribution network[J]. Power System Protection and Control, 2015, 43(7): 57-64.
- [24] 谷玉海, 韩秋实, 徐小力, 等. t分布随机近邻嵌入机械故障特征提取方法研究[J]. 机械科学与技术, 2016, 35(12): 1900-1905.
- GU Yuhai, HAN Qiushi, XU Xiaoli, et al. An early fault feature extraction method based on t-distribution stochastic neighbor embedding for large rotating machinery[J]. Mechanical Science and Technology for Aerospace Engineering, 2016, 35(12): 1900-1905.
- [25] MISHRA M, ROURAY P, ROURAY P K. A universal high impedance fault detection technique for distribution system using S-transform and pattern recognition[J]. Technology & Economics of Smart Grids & Sustainable Energy, 2016, 1(1): 1-9.

收稿日期: 2021-12-10; 修回日期: 2022-03-17

作者简介:

白浩(1987—), 男, 博士, 高级工程师, 研究方向为配电网快速故障处理、分布式能源控制、数字孪生与全景仿真等; E-mail: baihao@csg.cn

潘姝慧(1994—), 女, 硕士, 工程师, 研究方向为配电网运行及供电可靠性分析等;

邵向潮(1988—), 男, 硕士, 工程师, 研究方向为电力系统。

(编辑 周金梅)