

DOI: 10.19783/j.cnki.pspc.190569

基于分段改进 S 变换和随机森林的复合 电能质量扰动识别方法

王仁明, 汪宏阳, 张贇宁, 王凌云

(三峡大学电气与新能源学院, 湖北 宜昌 443002)

摘要: 各类分布式设备和智能设备接入电力系统, 使得电力系统对电能的波动越来越敏感, 这导致对电能质量扰动(PQD)的识别和处理变得越来越重要。通过将分段改进 S 变换(SMST)和随机森林(RF)算法相结合, 提出了一种用于复杂噪声环境下 PQD 识别的新方法。首先, 基于检测误差和峰度对 SMST 的不同频段进行分别调参, 并使用 SMST 提取待检测信号的 75 种时频特征, 构成原始特征集。然后, 改进分类回归树(CART)的节点分裂过程, 加入了离散值处理策略并使用 Gini 指数的下降作为新的节点分裂规则。同时, 在下次节点分裂前, 将基尼指数下降值为零的特征从特征集中删除。最后, 使用改进的 CART 算法构建了 RF 分类器并对复合 PQD 信号进行分类。实验证明, 在不同的信噪比条件下, 新方法均能有效识别多数单一 PQD 信号和常见的双重复合 PQD 信号。虽然新方法在运行效率方面仍有一定的改进空间, 但其在不同层面上的改进均能有效提升 PQD 识别精度, 且平均分类精度明显高于各类传统 PQD 识别方法。

关键词: 电能质量; 扰动分类; 分段改进 S 变换; Gini 指数下降; 随机森林

Composite power quality disturbance recognition based on segmented modified S-transform and random forest

WANG Renming, WANG Hongyang, ZHANG Yunning, WANG Lingyun

(College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China)

Abstract: Various types of distributed equipment and intelligent equipment are connected to the power system. It makes the power system more and more sensitive to power fluctuations, which has led to the identification and processing of Power Quality Disturbances (PQD) become increasingly important. By combining the Segmented Modified S-Transform (SMST) and the Random Forest (RF) algorithm, a new method for PQD identification under complex noise conditions is proposed. Firstly, different frequency bands of SMST are tuned based on various detection errors and kurtosis, and 75 time-frequency features are extracted from the signal using SMST to form the original feature set. Then, the node splitting process of Classification Regression Tree (CART) is improved. The discrete value processing strategy is added and the drop of Gini index is used as the new node splitting rule. Moreover, before the next node splitting, the feature whose Gini index drops to zero is removed. Finally, RF classifier is constructed with modified CART algorithm and used to classify the complex PQD signals. Experiments show that under the condition of different SNR, the new method can effectively identify most single PQD signals and common dual-compounded PQD signals. Although the new method still has some room for improvement in terms of efficiency, its improvement at different aspects can effectively benefit the accuracy of PQD recognition, and its average classification accuracy is significantly higher than traditional PQD recognition methods based on S-transform.

This work is supported by National Natural Science Foundation of China (No. 61603212 and No. 51407104).

Key words: power quality; disturbance classification; segmented modified S-transform; Gini index decline; random

0 引言

近年来, 随着分布式电源和微电网的发展, 大

量的电力电子装置和智能设备接入了电力系统, 其带来的各类非线性、波动性、冲击性负载是诸如设备损坏、敏感负荷工作失常和大小停电事故的主要诱因。这类电能质量不佳导致的安全事故已对电力系统的安全稳定运行造成了极大威胁^[1-2]。电力系统中的电能质量扰动(Power Quality Disturbance, PQD)信

基金项目: 国家自然科学基金项目资助(61603212;
51407104)

号包括电压波动(电压暂升、电压暂降和中断)、电压下陷、谐波、暂态振荡等^[3],具有持续时间短、发生频繁、种类及复合形式多样等特点,分析和识别 PQD 现象是开展针对性的治理和提高电能质量的前提。

识别 PQD 信号的关键问题是如何从海量的电能质量数据中提取特征并准确地进行分类。特征提取通常先采用时频变换的方法处理原始 PQD 信号,再提取信号的时域和频域特征。常用的时频分析方法包括短时傅里叶变换(Short Time Fourier Transform, STFT)^[4]、小波变换(Wavelet Transform, WT)^[5]、希尔伯特-黄变换(Hilbert-Huang Transform, HHT)^[6]以及 S 变换(S-transform, ST)^[7]等。STFT 通过对傅里叶变换进行加窗实现,具有明确的物理意义,但其窗函数相对固定,故而时频分辨率单一。WT 支持多分辨率分析,具有可变的时频分辨率,但其分辨率变换尺度与频率的关系固定,且其变换结果容易受到噪声影响。HHT 分解后可能会产生虚假模态和模态混叠等问题,不利于对 PQD 的时频特征进行准确分析。而 ST 采用了可变高斯窗函数,高斯窗宽随频率变化,抗噪性能较好且能产生大量的时频信息用于 PQD 识别。此外,其变换结果中低频部分时间分辨率高和高频部分频率分辨率高的特性与 PQD 识别的需求基本相符,因而 ST 与其各种改进形式广泛用于 PQD 信号的处理^[8]。

模式分类步骤一般采用基于机器学习理论的分类算法构建分类器,并基于 PQD 特征对信号进行分类,实现 PQD 信号的识别。常用的分类算法有:贝叶斯分类器(Bayes Classifier, BC)^[9]、K 最近邻(K-nearest Neighbor, KNN)^[10]、概率神经网络(Probabilistic Neural Network, PNN)^[11]、多标签学习(Multi-label Learning, MLL)^[12]、支持向量机(Support Vector Machine, SVM)^[13]、决策树(Decision Tree, DT)^[14]等。相较于其他分类器而言,DT 具有结构简单且分类规则易于理解、可解释性强,分类效率与准确率较高等优点。现有文献中,用于 PQD 识别的 DT 的构建多基于统计实验或经验确定的信号特征集和阈值^[15],这类方法在特定 PQD 集合和场景下分类效果良好,但用于复合扰动,不同信噪比 PQD 的识别效果则不甚理想。此外,人工统计实验的方法在面对不断变化的扰动类型和场景需求时,DT 分类器的构建难度和工作量十分巨大,这些缺点极大地限制了 DT 分类器在 PQD 识别领域的应用。

随机森林(Random Forest, RF)是将 bagging 集成学习理论与随机子空间方法相结合的一种基于

DT 的机器学习算法^[16]。其模型包含多个由 bagging 方法训练得到的 DT,当输入待分类样本时,最终的分类结果由所有单个决策树的输出结果投票决定。相比于使用同样特征样本集构建的 DT 分类器,RF 的运算量没有显著增加,但由于在分类器构建过程中加入了随机抽样,其过拟合风险被降低。同时,泛化性能增强、分类精度更高、噪声鲁棒性更强,其综合性能优势更为明显。

本文利用 ST 能产生大量时频信息的特点和随机森林噪声鲁棒性强的优势,提出了一种基于改进 S 变换(Modified S-transform, MST)^[17]和随机森林算法的电能质量复合扰动识别方法。首先,基于 PQD 信号的定义和频率分布特征将 MST 的处理频域进行分段,在引入峰度参数结合检测误差进行参数调整后,形成了分段改进 S 变换(Segmented Modified S-transform, SMST)。其次,仿真生成 PQD 信号,使用 SMST 处理原始信号后,基于其时频分布特点提取总计 75 种特征用于分类。然后,对 RF 算法中子决策树的生成过程进行了改进,在分类回归树(Classification and Regression Tree, CART)算法的基础上使用 Gini 指数的下降 $Gini_{dec}$ 作为分裂指标,并在节点分裂过程中剔除该值为零的特征,同时在原 CART 算法的实现过程中增加了离散值处理环节。最后,基于提出的方法构建了 RF 分类器,实验验证了新方法中各改进环节的有效性,并比较了本文所提出的方法与基于 ST 的传统 PQD 识别方法在不同信噪比、相同特征样本集下的分类结果。

实验结果表明,新方法对比文中列举的其他分类算法,在识别不同信噪比环境下的电能质量扰动信号时,特别是在信号源的信噪比较低或含随机噪声的情况下,优势更加明显,体现出了良好的抗噪能力。

1 基于分段改进 S 变换的 PQD 特征选择

1.1 S 变换和改进 S 变换

ST 和广义 S 变换(Generalized S-transform, GST)已被证明具有良好的抗噪性与时频特征表现能力^[18-19],能够满足复杂噪声环境下特征提取的需求。

文献[17]通过引入窗宽与频率平方根的倒数成正比的高斯窗得到了改进 S 变换,其离散形式如下。

$$\begin{cases} S(jT, \frac{n}{NT}) = \sum_{m=0}^{N-1} H(\frac{m+n}{NT}) e^{-\frac{2\pi^2 m^2 g}{n}} e^{i2\pi \frac{mj}{N}} & n \neq 0 \\ S(jT, 0) = \frac{1}{N} \sum_{m=0}^{N-1} h(\frac{m}{NT}) & n = 0 \end{cases} \quad (1)$$

式中:变量 f 离散化为 $\frac{n}{NT}$; 变量 τ 离散化为 jT ;

$H(\frac{n}{NT})$ 为信号 $h(kT)$ 的离散傅里叶变换; T 为采样周期; N 为总采样点数; i 为虚数单位; g 为窗宽调节因子, 当 $0 < g < 1$ 时, 窗宽随频率平方根呈反比加快变化, 时间分辨率更高, 当 $g > 1$ 时, 高斯窗变化速度减慢, 频率分辨率得以提高。

1.2 分段改进 S 变换

ST 和 MST 中窗宽调节因子 g 的大小对时频分析的时间和频率分辨率等有直接影响, 同时也间接影响后续分类精度, 但受限于海森堡测不准原理, 时间和频率分辨率两者不可兼得。

SMST 基于上述限制要求, 结合 PQD 信号频率分布的大体规律, 将分析频域分为三段, 并针对不同的分析需求分别整定窗宽调节因子: 在 $0 \sim 100$ Hz 区间主要分析如电压暂降、中断、凹陷等基频幅值变化较为明显的低频类扰动, 对应调节因子 g_1 ; $100 \sim 650$ Hz 区间主要包含电能质量关注的奇次谐波频率, 对应调节因子 g_2 ; 650 Hz 以上主要研究频率较高的暂态振荡等扰动对应因子 g_3 。

SMST 的离散形式如式(2)所示。

$$\begin{cases} S(jT, \frac{n}{NT}) = \sum_{m=0}^{N-1} H(\frac{m+n}{NT}) e^{-\frac{2\pi^2 m^2 g_1}{n}} e^{i2\pi \frac{mj}{N}} & 0 \leq n \leq 100NT \\ S(jT, \frac{n}{NT}) = \sum_{m=0}^{N-1} H(\frac{m+n}{NT}) e^{-\frac{2\pi^2 m^2 g_2}{n}} e^{i2\pi \frac{mj}{N}} & 100NT \leq n \leq 650NT \\ S(jT, \frac{n}{NT}) = \sum_{m=0}^{N-1} H(\frac{m+n}{NT}) e^{-\frac{2\pi^2 m^2 g_3}{n}} e^{i2\pi \frac{mj}{N}} & 650NT \leq n \leq n_{\max} \\ S(jT, 0) = \frac{1}{N} \sum_{m=0}^{N-1} h(\frac{m}{NT}), & n = 0 \end{cases} \quad (2)$$

式中: g_1 、 g_2 、 g_3 分别为低、中、高频段的调节因子; n_{\max} 为最大检测频率点; 其他变量含义与 MST 相同。

1.3 窗宽调节因子的整定

窗宽调节因子基于检测误差最小化原则进行整定, 图 1、图 2 分别是窗宽调节因子 g 取 0.1 和 1 时的电压暂升信号经 MST 处理后得到的时频图, 而图 3、图 4 则分别是窗宽调节因子 g 取 0.1 和 10 时谐波信号经 MST 处理后得到的时频图。

从图 1—图 4 可初步总结出窗宽调节因子的变化对检测效果的影响规律, 如图 5 所示, 即: 窗宽调节因子降低会导致时频分析时的频域分辨率降低, 时域分辨率提高; 窗宽调节因子提高会导致时频分析时的频域分辨率提高, 时域分辨率降低。综合来看, 窗宽调节因子的变化分别对主要含频域或时域信号特征的 PQD 信号造成不同的影响。

由于不同 PQD 信号的信号特征通常并不相同, 结合上述调节因子的变化对检测效果的影响规律

分析可知, 若想达到最佳效果, 则需基于各个 PQD 信号的不同特征分别整定 g 的值。于是本文在 MST 的基础上, 提出按频率分段整定其窗宽调节因子 g , 形成 SMST 以获得最佳分类效果。

从图 1—图 4 可以看出: 调节因子 g 取不同值时, 不同类型扰动分量峰部的尖度(即峰度, Kurtosis)有明显变化, 而扰动分量与周围区域相比幅值变化越明显, 其与非扰动发生段的区分度越高, 分类效果就越好。因此, 在不同 PQD 信号检测要求的基础上加入了峰度值, 不同类型扰动的误差检测项目设置如表 1 所示。

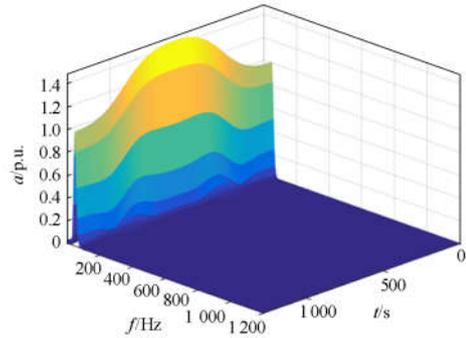


图 1 电压暂升信号时频图($g=0.1$)

Fig. 1 Spectrogram of the voltage swell signal with $g=0.1$

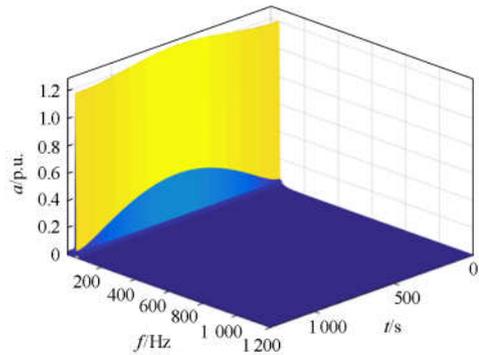


图 2 电压暂升信号时频图($g=1$)

Fig. 2 Spectrogram of the voltage swell signal with $g=1$

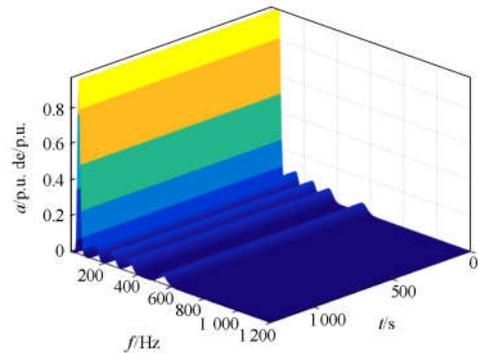


图 3 谐波信号时频图($g=0.1$)

Fig. 3 Spectrogram of harmonic signal $g=0.1$

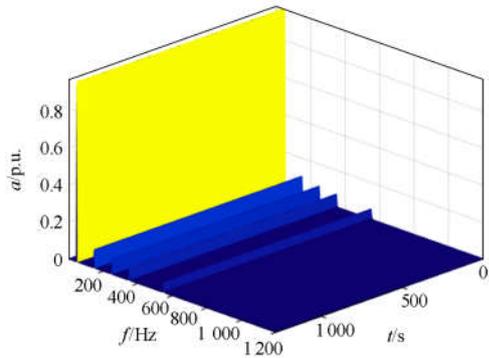


图 4 谐波信号时频图($g=10$)

Fig. 4 Spectrogram of harmonic signal with $g=10$

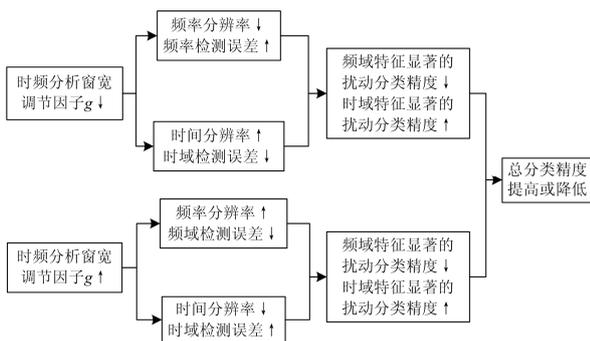


图 5 窗宽调节因子变化造成的影响

Fig. 5 Effect of the change of window width adjustment factor

表 1 不同类型扰动的检测项目

Table 1 Different types of disturbance detection items

类型	幅值	起始时间	结束时间	频率	峰度
	误差	误差	误差	误差	
电压暂降	√	√	√	—	√
电压暂升	√	√	√	—	√
电压中断	√	√	√	—	√
电压闪变	√	—	—	√	√
谐波	√	—	—	√	√
暂态振荡	√	√	√	√	√
暂态脉冲	√	√	√	—	√
电压尖峰	√	√	√	—	—
电压缺口	√	√	√	—	—

表 1 中，“√”表示在本类型扰动的参数整定中考虑该项目，“—”表示不考虑该检测项目或没有意义。

使用如表 3 中所示的，不含噪声的标准扰动信号进行参数整定实验，其中扰动的幅值参数取定义内最小值，同时频率参数取中位值，而各类检测误差则定义为 SMST 的参数检测值与设定的实际参数之差。

参数整定时，时轴起始值分别设置为各窗宽调

节因子的仿真步长值，其中 g_1 使用步长 0.000 2， g_2 使用步长 1， g_3 使用步长 0.005，分别计算表 1 中检测项目的值。

由于后续分类处理主要以变换后所得的模时频矩阵为基础，因此对所有类型的 PQD 信号来说，窗宽调节因子的整定优先保证幅值误差最小。在此基础上，检测时有起止时间检测要求的 PQD 信号，保证起止时间误差较小；有频率检测要求的 PQD 信号保证频率检测误差较小；对于皆有要求的暂态振荡则需要综合考虑。在 g 取值区间内其他各类误差变化不明显时，才考虑峰度值变化作为主要指标，否则仅作为次要因素参考。综上，初步将调节因子整定为 $g_1=0.010$ 、 $g_2=14$ 、 $g_3=0.250$ 。

由于环境噪声可能会对检测结果产生影响，另在前文的 PQD 最小参数模型基础上分别加入了 20 dB、30 dB、40 dB、50 dB 的高斯白噪声重新进行实验，总结噪声对各检测项目的影响，在参数整定时，根据其影响区间，保留一定的误差噪声裕度，最终将调节因子整定为 $g_1=0.009 8$ 、 $g_2=14$ 、 $g_3=0.255$ 。

1.4 基于 SMST 的 PQD 信号特征选择

利用原始 PQD 信号进行 ST 后能产生大量时频信息的特点，基于变换后产生的模时频矩阵(Modular Time-frequency Matrix, MTFM)进行特征提取，选择的 PQD 信号特征总结如表 2 所示， F_n 表示该特征的标号为 n 。

如表 2 所示，除基于 MTFM 的数学特征的提取的特征 $F_1—F_{37}$ 外，受 SMST 分频思想启发，还加入了局部频域特征 $F_{38}—F_{69}$ 。其中，表 2 中 PQD 信号局部频域特征的频率分段点设置与 1.2 节中 SMST 的频率分段点相同。

同时，由于如电压尖峰、电压缺口、电压暂降和中断等部分 PQD 扰动信号的频率特征不明显，但在 50 Hz 左右的频率最大幅值曲线形状有明显变化，因此增加特征 F_{70} 和 F_{71} 。同时考虑到噪声与谐波 PQD 信号在谐波频率点附近最大幅值的峰度有较大区别，进一步补充特征 $F_{72}—F_{75}$ ，具体如表 2 所示。

综上所述，最终确定了需要提取的共计 75 种特征，即： $F_1—F_{75}$ ，具体如表 2 所示，其中特征的计算参考文献[10,20]设计。

2 基于随机森林算法的分类

2.1 改进的节点分裂指标和离散值处理方法

RF 算法为每个训练子集分别建立一棵决策树而不剪枝，每棵决策树任其生长，最终生成 n 棵 DT 形成“森林”。分类时，将测试集分别输入每个 DT，最终的分类结果由森林中所有 DT 投票决定。

表 2 PQD 信号扰动特征
Table 2 Disturbance features of PQD signals

特征总结	特征标号	特征描述
PQD 信号 时域特征	F ₁ —F ₂₅	MTFM 信号基频幅值最大最小值之差、最大最小值之和、均值、标准差、均方根值对应特征: F ₁ —F ₅ ; 各列均值最大最小值之差、最大最小值之和、均值、标准差、均方根值对应特征: F ₆ —F ₁₀ ; 标准差最大最小值之差、最大最小值之和、均值、标准差、均方根值对应特征: F ₁₁ —F ₁₅ ; 均方根最大最小值之差、最大最小值之和、均值、标准差、均方根值对应特征: F ₁₆ —F ₂₀ ; 最大值的最大最小值之差、最大最小值之和、均值、标准差、均方根值, 分别对应特征: F ₂₁ —F ₂₅ 。
PQD 信号 频域特征	F ₂₆ —F ₃₇	MTFM 各行最大值的最大值、标准差、均方根、谐波总含量对应特征: F ₂₆ —F ₂₉ ; 均值的最大值、标准差、均方根、谐波总含量对应特征: F ₃₀ —F ₃₃ ; 均方差的最大值、标准差、均方根、谐波总含量对应特征: F ₃₀ —F ₃₃ 。
PQD 信号局部 频域特征	F ₃₈ —F ₆₉	MTFM 各行极差的低频范围对应段的最大值、最小值、均值、标准差、均方根、谐波畸总变率、峰度对应特征: F ₃₈ —F ₄₃ ; 标准差曲线的低频范围对应段的最大值、最小值、均值、标准差、均方根、谐波畸总变率、峰度对应特征: F ₄₄ —F ₄₉ ; 各行最大值的低频范围对应段的最大值、均值、标准差、均方根对应特征: F ₅₀ —F ₅₃ ; 均方根的中频范围对应段的最大值、均值、标准差、均方根对应特征: F ₅₄ —F ₅₇ ; 各行极差的高频对应段的最大值、均值、标准差、均方根对应特征: F ₅₈ —F ₆₁ ; 标准差的高频对应段的最大值、均值、标准差、均方根对应特征: F ₆₂ —F ₆₅ ; 最大值的高频对应段的最大值、均值、标准差、均方根对应特征: F ₆₆ —F ₆₉ 。
PQD 信号 补充特征	F ₇₀ —F ₇₅	50 Hz 左右频率最大幅值峰度 F ₇₀ , 50 Hz 左右频率最大幅值偏度 F ₇₁ , 3、5、7、11 次谐波对应频率附近最大幅值峰度 F ₇₂ —F ₇₅ 。

随机森林中的子决策树生成流程如图 6 所示。对每个 bootstrap 抽样生成的训练子集都训练一棵决策树。生成子树时, 对训练子数据集的特征空间再进行一次随机选择。生成特征子集后, 基于选定的节点分裂算法开始决策树的节点分裂, 直到达到建树停止条件。

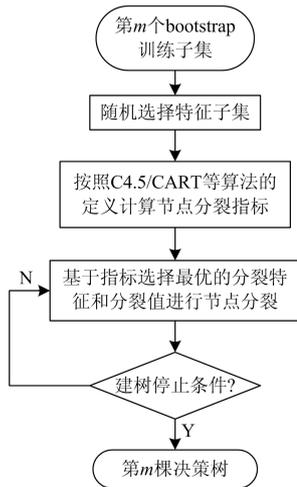


图 6 随机森林子决策树生成流程图

Fig. 6 Flow chart of random forest subdecision tree generation

本文选择了 CART 节点分裂算法进行 RF 的分类器的构建。CART 算法采用 Gini 指数作为节点分裂规则^[21], 单一样本集的 Gini 指数计算过程如下。

$$Gini(S) = 1 - \sum_{i=1}^M P_i^2 \quad (3)$$

式中, P_i 为类别 C_i 在样本集 S 中出现的概率。若 S 被划分为两个子集 S_1 和 S_2 , 则样本集划分后的 Gini 指数为

$$Gini_{\text{split}}(S) = \frac{|S_1|}{|S|} Gini(S_1) + \frac{|S_2|}{|S|} Gini(S_2) \quad (4)$$

在节点分裂时, 选择使得划分后 Gini 指数最小的特征作为最优划分特征, 并选择该特征的最小特征值作为最优划分点进行节点分裂。

由于提取的扰动特征维数较高, 存在一些冗余和噪声, 本文使用基于 Gini 指数改进的节点分裂规则, 其计算过程如下。

$$Gini_{\text{dec}}(S) = Gini(S) - Gini_{\text{split}}(S) \quad (5)$$

如此一来, $Gini_{\text{dec}}$ 越大, 则表明本特征在分类过程中起到的作用越大。而若某特征的 $Gini_{\text{dec}}$ 值为 0, 则表示数据集的一致性没有变化, 即该特征在分类过程中未起作用, 则在下次节点分裂前将该特征从子特征集中剔除, 可在一定程度上去除特征冗余, 减少噪声对识别效果的影响。

由于本文实现 CART 算法时, 采用了基于黄金分割法和二次插值法的单维寻优法来计算连续特征的 Gini 指数或 $Gini_{\text{dec}}$, 这种方法默认将处理的特征视为连续特征, 无法很好地处理离散量, 因此本文借鉴了 C4.5 算法对离散值的处理思路, 节点分裂过程中将取值种类低于阈值的特征视为离散特征, 使用逐个遍历的方法代替单维寻优法确定最优分裂节点, 以便更好地处理离散 PQD 信号特征。

2.2 改进随机森林分类器的构建

图 7 结合了离散值处理模块, 并使用 $Gini_{\text{dec}}$ 作为子 DT 生成策略的 RF 子树生成流程, 流程可总结为以下 4 步。

Step1: 使用 bagging 方法对原始样本集 S 进行有放回随机抽样, 生成子样本集 S_{bag} 作为每棵树的

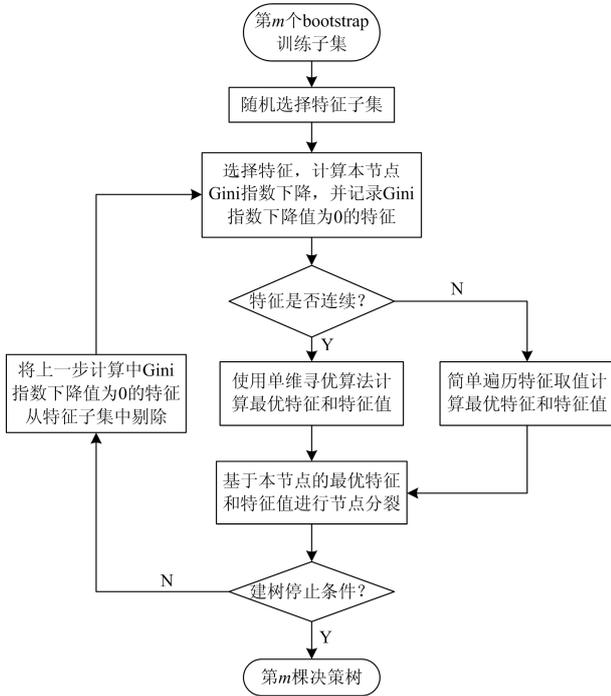


图7 改进的子决策树生成流程图

Fig. 7 Flowchart of the improved subdecision tree generation

训练数据, 其样本数与 S 中的样本数相同。由于每棵树的数据集是由原始数据集随机构成, 因此可能重复包含, 也可能不包含某些数据, 保证了各个子 DT 的差异性, 降低了算法的过拟合风险。

Step2: 使用随机选择输入变量(Forest-RI)方法, 对 S_{bag} 中的样本特征进行子空间采样。从 N_i 个原始特征中随机选取 N_{i_s} (默认值 $N_{i_s} = \sqrt{N_i}$) 个特征, 得到用于训练子决策树的 S_T 样本集。

Step3: 利用 S_T 样本集, 建立子 DT。 $h(x, S_T)$ 使用 $Gini_{\text{dec}}$ 作为节点分裂规则, 在节点分裂前, 将 $Gini_{\text{dec}}$ 值为 0 的特征从 S_T 对应的特征集中剔除。子 DT 停止生长准则与完全生长方式的要求基本相同: 一是当前所有的样本均指向同一标签或只剩单一样本; 二是当前所有样本中, 指向某标签的样本占大多数。在子 DT 生长过程中出现上述任意一种情况便停止子 DT 生长, 进入下一轮循环。

Step4: 重复上述 3 个步骤 n_{RF} 次, 生成的 n_{RF} 棵子决策树便构成了随机森林。

RF 算法引入的 bagging 方法和 Forest-RI 方法保证了集成模型的随机性, 能较好地避免过拟合问题^[16], 从而不需要进行剪枝。同时, 在子 DT 节点分裂时剔除 $Gini_{\text{dec}}$ 值为 0 的特征进一步降低了子 DT 间的相关性, 使得过拟合风险更低。

3 仿真分析

3.1 电能质量扰动模型

参考文献[21-23], 本文分析如下 18 类 PQD 信号。其中电压暂降(C1)、电压暂升(C2)、电压中断(C3)、闪变(C4)、电压缺口(C5)、电压尖峰(C6)、谐波(C7)、暂态振荡(C8)、暂态脉冲(C9)为 9 类单一扰动信号; 谐波+暂降(C10)、谐波+暂升(C11)、谐波+中断(C12)、谐波+闪变(C13)、暂降+振荡(C14)、暂升+振荡(C15)、中断+振荡(C16)、闪变+振荡(C17)、谐波+振荡(C18)为 9 类双重复合扰动信号, 并据此建立了标准 PQD 信号模型。

标准模型中, 电压幅值为归一化幅值, 实际应用中相当于以电压峰值为基准的标幺值。其中的单一 PQD 信号模型如表 3 所示, 复合 PQD 模型由对应的单一 PQD 信号复合而成。基波频率 f_0 为 50 Hz, 采样频率 f_s 为 6.4 kHz, T 为工频周期, 信号样本长度为 10 个周期, $u(t)$ 为单位阶跃函数。谐波信号主要考虑含 13 次谐波以下的谐波, 暂态振荡频率取值为 700~1 200 Hz^[24-25]。

3.2 噪声环境下新方法的表现

利用 Matlab 软件生成具有随机参数且分别含信噪比值为 50 dB、40 dB、30 dB、20 dB、20~50 dB 的 PQD 信号, 每类扰动各 400 组, 即每种信噪比条件下共 7 200 组, 并选取其中 60% 的样本构成训练集, 其余的构成测试集。

影响 RF 性能的参数主要有两个, 子 DT 棵数 n_{RF} 和节点子空间大小 N_{i_s} 。文献[16]指出 N_{i_s} 对算法性能有较大影响, 并给出了建议值, 由于本文从 PQD 信号中提取的特征较多, $N_{i_s} = 75$, 因此选择 $N_{i_s} = \sqrt{N_i} \approx 9$ 以平衡树模型的随机性与分类精度。 n_{RF} 主要影响算法的稳定性, 对分类精度的影响随着其值的增大不断减小, 但却会极大影响算法的运行速度, 本文中取 $n_{\text{RF}} = 40$ 以平衡算法的效率和性能。

以下各表中将电能质量扰动识别方法概括表示为: 特征提取方法+分类方法, “+”号前的一部分表示其使用的特征提取方法, 后一部分表示其使用的分类方法, 因此本文提出的新方法表示为 SMST+RFd, 代表使用分段改进 S 变换和经改进后的 RF 分类器。表 8 中 DTCART 和 DTC4.5 分别表示基于 CART 和 C4.5 节点分裂算法的 DT 算法。(20~50 dB)×rand 类别表示该类样本的信噪比皆为 20~50 dB 的一个随机值, 其他符号含义与上文相同。

表 3 PQD 信号标准模型
Table 3 Standard model of PQD signals

信号类型	信号模型	参数
电压暂降(C1)	$h(t) = [1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.9, T \leq t_2 - t_1 \leq 9T$
电压暂升(C2)	$h(t) = [1 + \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$	$0.1 \leq \alpha \leq 0.8, T \leq t_2 - t_1 \leq 9T$
电压中断(C3)	$h(t) = [1 - \alpha(u(t-t_1) - u(t-t_2))] \sin(\omega t)$	$0.9 \leq \alpha \leq 1.0, T \leq t_2 - t_1 \leq 9T$
电压闪变(C4)	$h(t) = (1 + \alpha \sin(\beta \omega t)) \sin(\omega t)$	$0.1 \leq \alpha \leq 0.2, 0.1 \leq \beta \leq 0.4$
电压缺口(C5)	$h(t) = \sin(\omega t) - \text{sgn}(\sin(\omega t)) \times \left\{ \sum_{n=0}^9 \kappa \times [u(t - (t_1 + 0.02n)) - u(t - (t_2 + 0.02n))] \right\}$	$0.1 \leq \kappa \leq 0.4, 0 \leq t_2, t_1 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$
电压尖峰(C6)	$h(t) = \sin(\omega t) + \text{sgn}(\sin(\omega t)) \times \left\{ \sum_{n=0}^9 \kappa \times [u(t - (t_1 + 0.02n)) - u(t - (t_2 + 0.02n))] \right\}$	$0.1 \leq \kappa \leq 0.4, 0 \leq t_2, t_1 \leq 0.5T$ $0.01T \leq t_2 - t_1 \leq 0.05T$
谐波(C7)	$h(t) = \alpha_1 \sin(\omega t) + \alpha_2 \sin(3\omega t) + \alpha_3 \sin(5\omega t) + \alpha_4 \sin(7\omega t) + \alpha_5 \sin(11\omega t)$	$0.005 \leq \alpha_{3,5,7,11} \leq 0.15, \sum \alpha_i^2 = 1$
暂态振荡(C8)	$h(t) = \sin(\omega t) + \alpha e^{-\frac{(t-t_1)}{\tau}} \cdot \sin(2\pi f_m t) [u(t-t_1) - u(t-t_2)]$	$0.1 \leq \alpha \leq 0.8, 8 \text{ ms} \leq \tau \leq 40 \text{ ms}$ $0.5Tn \leq t_2 - t_1 \leq 3Tn$ $700 \text{ Hz} \leq f_m \leq 1200 \text{ Hz}$
暂态脉冲(C9)	$h(t) = \sin(\omega t) + \alpha [u(t_1) - u(t_2)]$	$0.8 \leq \alpha \leq 2, 1 \text{ ms} \leq t_2 - t_1 \leq 3 \text{ ms}$

在不同信噪比条件下, 本文提出的算法分类精度具体如表 4 所示。

表 4 不同信噪比下的分类精度

Table 4 Accuracy under different SNR

扰动类型	分类精度					
	20 dB	30 dB	40 dB	50 dB	(20~50 dB)*rand	
单 一 扰 动	C1	0.963	0.975	0.994	0.969	0.975
	C2	0.975	1.000	0.981	1.000	0.963
	C3	0.981	1.000	1.000	0.988	0.981
	C4	1.000	1.000	1.000	1.000	0.981
	C5	0.850	0.975	0.975	1.000	0.963
	C6	0.931	0.938	0.938	0.994	0.969
	C7	1.000	1.000	1.000	1.000	0.994
	C8	0.938	1.000	1.000	1.000	1.000
	C9	1.000	1.000	1.000	1.000	1.000
双 重 复 合 扰 动	C10	0.956	0.981	0.975	0.925	0.969
	C11	0.931	0.981	0.975	1.000	0.994
	C12	0.975	0.988	0.988	0.988	0.988
	C13	0.994	0.994	1.000	1.000	1.000
	C14	0.844	0.969	0.963	0.963	0.963
	C15	0.913	1.000	0.988	0.975	0.969
	C16	0.925	0.975	0.975	0.994	0.963
	C17	0.913	1.000	0.994	1.000	0.975
	C18	0.938	1.000	1.000	1.000	0.988
总分类精度	0.946	0.988	0.986	0.989	0.980	

由表 4 可知, 对文中提出的 18 类扰动进行分类时, 除个别扰动类型由于定义较为接近分类精度稍低外, 本文所提方法的分类精度均较好。在信噪比为 30~50 dB 时, 总分类精度均大于 98.5%。同时, (20~50 dB)*rand 随机噪声条件下的总分类精度为 98.0%, 接近低噪声水平下的表现。在 20 dB 信噪比环境下, 几乎所有类别扰动识别的准确率都有所

下降, 但总分类精度仍为 94.6%。

3.3 方法改进效果验证

在保持信号处理环节使用 SMST 不变的情况下, 本文提出的新方法和仅使用原 RF 分类器算法时的分类精度对比如表 5 所示, 旨在验证本文中提出的随机森林部分改进的效果。

表 5 新方法和原 RF 算法的分类精度

Table 5 Accuracy of the new method and the original RF algorithm

识别方法	分类精度				
	20 dB	30 dB	40 dB	50 dB	(20~50 dB)*rand
SMST+RFd	0.945	0.989	0.990	0.987	0.978
SMST+RF	0.939	0.983	0.987	0.986	0.971

对比表 5 中的数据可以看出, 本文提出的方法在不同信噪比条件下分类精度的提升幅度并不相同, 最大提升为 0.7%, 最小仅为 0.1%, 但都强于基于原始 RF 的方法, 这说明在改用 $Gini_{dec}$ 和冗余特征剔除策略后, 对分类精度的改善的确有一定作用, 但由于只剔除 $Gini_{dec}$ 为 0 的特征, 导致该过程具有一定的随机性, 因此改进幅度有一定波动。

在均使用原始 RF 算法的情况下, 本文提出的 SMST 新方法和使用原 ST 方法时的分类精度对比如表 6 所示, 旨在验证本文中提出的 SMST 的效果。

表 6 SMST 和原 ST 方法的分类精度

Table 6 Accuracy of SMST and original ST methods

识别方法	分类精度				
	20 dB	30 dB	40 dB	50 dB	(20~50 dB)*rand
SMST+RF	0.948	0.989	0.991	0.986	0.977
ST+RF	0.898	0.960	0.965	0.972	0.923

从表 6 中的数据可以看出, 改用 SMST 后, 方法的整体分类精度提升较大, 特别在 20 dB 和 (20~50 dB)×rand 条件下, 与原始 ST 相比, 提升幅度分别达到了 5.0%和 5.4%。

在使用均使用 SMST 和原始 RF 算法的情况下, 本文提出的新方法和使用传统特征选择方案时的分类精度对比如表 7 所示, 旨在验证本文中提出的特征提取方法的效果。

表 7 新方法和传统特征选择方法的分类精度

Table 7 Accuracy of the new method and the traditional

识别方法	feature selection method				
	20 dB	30 dB	40 dB	50 dB	(20~50 dB)×rand
SMST+RF-75	0.948	0.989	0.991	0.986	0.977
SMST+RF-37	0.846	0.914	0.932	0.939	0.890

本文中提到的传统特征选择方案为上文提到的 F_1 — F_{37} , 不含受 SMST 特性启发加入的频域局部特征和补充特征的方案。

从表 7 中的数据可清楚地看出, 本文提出的特征提取方法在各个信噪比下对应的分类精度远好于传统方法, 其中在 20 dB 和(20~50 dB)×rand 处的精度提升分别达到 10.2%和 8.7%。

综合上述表 5—表 7 中的结论, 可知特征选择过程对于分类器分类精度的影响最大, 合理的特征选择能大大提升 PQD 扰动识别算法的性能。

实验结果表明, 本文方法从特征提取、特征选择、分类器设计三个层面对传统 PQD 信号识别方法进行了改进, 且分类精度有不同程度地提升。总体来看, 在 20 dB 及(20~50 dB)×rand 随机噪声条件下的提升幅度较大。

3.4 分类效果对比

新方法和其他方法在不同噪声条件下的分类精度对比如表 8 所示。

表 8 新方法和其他方法的分类精度

Table 8 Accuracy of the new method and other methods

使用方法	分类精度				
	20 dB	30 dB	40 dB	50 dB	(20~50 dB)×rand
ST+PNN	0.484	0.669	0.751	0.802	0.674
ST+KNN	0.488	0.656	0.744	0.799	0.668
ST+DTCART	0.883	0.941	0.956	0.958	0.917
ST+DTC4.5	0.780	0.937	0.961	0.966	0.895
SMST+RFd	0.946	0.988	0.986	0.987	0.980

从表 8 可直观地看出, 新方法的分类精度表现要明显好于其他各方法, 除 20 dB 条件下分类精度为 94.6%相对较低外, 其他情况下均大于 98.0%, 接近 99.0%。

同时, 两种基于 DT 方法的表现明显好于基于 PNN 和 KNN 的方法, 这体现了基于 DT 及其衍生的分类算法在电能质量分析中的优势。但新方法相较 ST+DTCART 和 ST+DTC4.5, 分类精度的表现随信噪比的变化仍较小, 最高点到最低点的波动为 30~20 dB, 波动值仅为 4.2%。且其余各方法在 20 dB 和(20~50 dB)×rand 点处分类精度下降明显, 但新方法在 20 dB 处仍能保证 94.6%的分类精度, 且在 (20~50 dB)×rand 处, 新方法的分类精度高达 98.0%, 精度下降不太明显, 这与应用中噪声环境的客观要求相契合。

由于在每棵决策树的节点分裂阶段都需要遍历和判断该节点处是否有特征需要移除, 与基于 CART 决策树的原始 RF 算法相比, 新方法在总运行时间上存在一定劣势, 具体如图 8 所示。

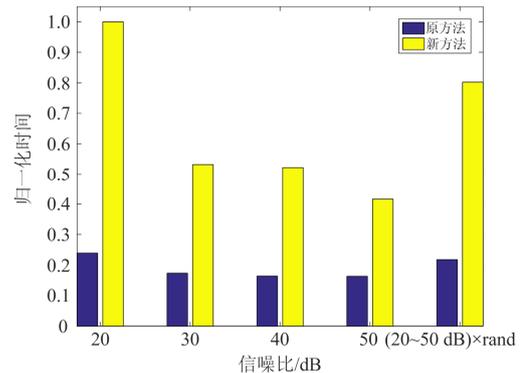


图 8 新方法和原方法的运行时间对比

Fig. 8 Comparison of the running time between the new and the original method

从图 8 可以看出, 新方法总是需要更长的运行时间, 且与原方法的差距随着信噪比的降低和噪声情况的复杂化逐渐加大。说明新方法虽然在分类精度和 OOB 误差方面的改进较大, 但在运行效率方面仍存在一定的改进空间。

综上所述, 本文提出的新方法在特征提取、特征选择、分类器设计三个不同层面上的改进均有一定效果, 且合理的特征选择对分类精度的影响最大。与基于 ST 的各类传统方法相比, 虽有运行效率和总运行时间方面的不足, 但新方法在不同噪声条件下的分类精度表现均较好, 特别是在 20 dB 和 20~50 dB 的随机噪声条件下, 精度下降更加不明显, 体现出了较好的噪声鲁棒性。

4 结论

针对现有的基于 DT 的 PQD 识别方法对复合扰

动和不同信噪比下的识别效果不理想, 以及人工统计调参的方法在扰动类型较多时难以实施的不足等问题, 本文提出了一种基于 SMST 和改进 RF 算法的复合电能质量扰动识别新方法, 主要工作包括以下几个方面。

1) 基于 18 种常见的单一和双重 PQD 模型, 优化了 SMST 的窗宽调节因子参数, 并提出了基于 SMST 且适用于 RF 分类器分类的特征提取方法。

2) 受 SMST 思路启发, 在传统特征选择方法的基础上, 提出并加入了一些 PQD 局部频域特征和补充特征用于 PQD 扰动识别。

3) 建立了基于改进 CART 的 RF 分类器。基于 CART 算法, 通过加入离散值处理策略, 将 RF 中子决策树的节点分裂指标修改为 Gini 指数下降 ($Gini_{dec}$), 并基于 $Gini_{dec}$ 在节点分裂时对样本特征集进行筛选, 优化了子决策树的结构, 在降低 RF 整体结构复杂度的同时, 提高了分类精度。

使用具有随机扰动参数的 PQD 信号特征样本集进行仿真验证, 结果表明, 相比传统基于 ST 的识别方法, 新方法具有更高的分类精度, 能有效识别包括 9 种复合扰动在内的共 18 种常见 PQD 扰动。且在低信噪比和随机噪声条件下识别效果仍很好, 分类精度下降随信噪比的降低不太明显, 体现了良好的噪声鲁棒性。后续将进一步研究特征选择 PQD 信号识别效果的影响, 提升该过程的可解释性, 并验证本文方法在实际应用中的价值。

参考文献

- [1] 高健, 崔雪, 邹晨露, 等. 基于改进能量集中度的 S 变换与随机森林的电能质量扰动识别[J]. 电测与仪表, 2019, 56(1): 8-14.
GAO Jian, CUI Xue, ZOU Chenlu, et al. S-transform based on modified energy concentration and identification of power quality disturbance in random forest[J]. Electrical Measurement & Instrumentation, 2019, 56(1): 8-14.
- [2] SAQIB M A, SALEEM A Z. Power-quality issues and the need for reactive-power compensation in the grid integration of wind power[J]. Renewable and Sustainable Energy Reviews, 2015, 43: 51-64.
- [3] 王知芳, 杨秀, 潘爱强, 等. 基于改进集成聚类和 BP 神经网络的电压偏差预测[J]. 电工电能新技术, 2018, 37(5): 73-80.
WANG Zhifang, YANG Xiu, PAN Aiqiang, et al. Voltage deviation forecasting based on improved ensemble clustering and BP neural network[J]. Advanced Technology of Electrical Engineering and Energy, 2018, 37(5): 73-80.
- [4] 黄建明, 瞿合祚, 李晓明. 基于短时傅里叶变换及其谱峭度的电能质量混合扰动分类[J]. 电网技术, 2016, 40(10): 3184-3191.
HUANG Jianming, QU Hezuo, LI Xiaoming. Classification for hybrid power quality disturbance based on STFT and its spectral kurtosis[J]. Power System Technology, 2016, 40(10): 3184-3191.
- [5] 张艳, 殷礼胜, 马瑞卿, 等. 基于复小波变换和有效值算法的电压暂降检测方法[J]. 电测与仪表, 2017, 54(10): 74-79.
ZHANG Yan, YIN Lisheng, MA Ruiqing, et al. Voltage sag detection method based on complex wavelet transform and RMS algorithm[J]. Electrical Measurement & Instrumentation, 2017, 54(10): 74-79.
- [6] AFRONI M J, SUTANTO D, STIRLING D. Nonstationary power-quality waveforms using iterative Hilbert Huang transform and SAX algorithm[J]. IEEE Transactions on Power Delivery, 2013, 28(4): 2134-2144.
- [7] 杨剑锋, 姜爽, 石戈戈. 基于分段改进 S 变换的复合电能质量扰动识别[J]. 电力系统保护与控制, 2019, 47(9): 64-71.
YANG Jianfeng, JIANG Shuang, SHI Gege. Classification of composite power quality disturbances based on piecewise-modified S transform[J]. Power System Protection and Control, 2019, 47(9): 64-71.
- [8] 黄南天, 张卫辉, 蔡国伟, 等. 采用改进多分辨率快速 S 变换的电能质量扰动识别[J]. 电网技术, 2015, 39(5): 1412-1418.
HUANG Nantian, ZHANG Weihui, CAI Guowei, et al. Power quality disturbances classification with improved multiresolution fast S-transform[J]. Power System Technology, 2015, 39(5): 1412-1418.
- [9] WU Junfeng, LI Yuzhe, QUEVEDO D E, et al. Data-driven power control for state estimation: a Bayesian inference approach[J]. Automatica, 2015, 54: 332-339.
- [10] 周维维, 管春, 卢伟国. 多标签分类法在电能质量复合扰动分类中的应用[J]. 中国电机工程学报, 2011, 31(4): 45-50.
ZHOU Luwei, GUAN Chun, LU Weiguo. Application of multi-label classification method to categorization of multiple power quality disturbances[J]. Proceedings of the CSEE, 2011, 31(4): 45-50.
- [11] WANG Huihui, WANG Ping, LIU Tao. Power quality disturbance classification using the S-transform and probabilistic neural network[J]. Energies, 2017, 10(1): 107.
- [12] 张巧苹, 刘志刚, 朱玲, 等. 基于多标签 Rank-WSVM 的复合电能质量扰动分类[J]. 中国电机工程学报,

- 2013, 33(28): 114-120.
- ZHANG Qiaoge, LIU Zhigang, ZHU Ling, et al. Recognition of multiple power quality disturbances using multi-label wavelet support vector machine[J]. Proceedings of the CSEE, 2013, 33(28): 114-120.
- [13] 任子晖, 王琦. 基于优化 DDAGSVM 多类分类策略的电能质量扰动识别[J]. 电力系统保护与控制, 2018, 46(5): 82-88.
- REN Zihui, WANG Qi. Power quality disturbance recognition based on improved DDAGSVM multi-class classification strategy[J]. Power System Protection and Control, 2018, 46(5): 82-88.
- [14] 全慧敏, 戴瑜兴. 电能质量扰动的广义 S 变换分析和决策树算法分类[J]. 湖南大学学报: 自然科学版, 2009, 36(4): 43-47.
- QUAN Huimin, DAI Yuxing. Power quality disturbances classification based on generalized S-transform and decision tree[J]. Journal of Hunan University: Natural Sciences, 2009, 36(4): 43-47.
- [15] KUBENDRAN P, KUMAR A, LOGANATHAN A K. Detection and classification of complex power quality disturbances using S-transform amplitude matrix-based decision tree for different noise levels[J]. International Transactions on Electrical Energy Systems, 2016, 27(4): 1-12.
- [16] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [17] PINNEGAR C R, MANSIHA L. Time-local spectral analysis for non-stationary time series: the S-transform for noise signal[J]. Fluctuation and Noise Letters, 2003, 3(3): 357-364.
- [18] DHARMAPANDIT O, PATNAIK R K, DASH P K. A fast time-frequency response based differential spectral energy protection of AC microgrids including fault location[J]. Protection and Control of Modern Power Systems, 2017, 2(2): 331-358. DOI: 10.1186/s41601-017-0062-0.
- [19] 刘军, 黄纯, 江亚群, 等. 电能质量扰动分析中广义 S 变换算法的改进与实现[J]. 电力系统及其自动化学报, 2017, 29(3): 35-41.
- LIU Jun, HUANG Chun, JIANG Yaqun, et al. Improved generalized S-transform algorithm for power quality disturbances analysis and its implementation[J]. Proceedings of the CSU-EPSA, 2017, 29(3): 35-41.
- [20] 黄南天, 彭华, 蔡国伟, 等. 电能质量复合扰动特征选择与最优决策树构建[J]. 中国电机工程学报, 2017, 37(3): 776-785.
- HUANG Nantian, PENG Hua, CAI Guowei, et al. Feature selection and optimal decision tree construction of complex power quality disturbances[J]. Proceedings of the CSEE, 2017, 37(3): 776-785.
- [21] CUTLER D R, EDWARDS Jr T C, BEARD K H, et al. Random forests for classification in ecology[J]. Ecology, 2007, 88(11): 2783-2792.
- [22] RODRIGUEZ-GALIANO V F, GHIMIRE B, ROGAN J, et al. An assessment of the effectiveness of a random forest classifier for landcover classification[J]. ISPRS J Photogrammetry Remote Sensing, 2012, 67: 93-104.
- [23] LARIVIERE B, POEL D V D. Predicting customer retention and profitability by using random forests and regression forests techniques[J]. Expert Systems with Applications, 2005, 29: 472-484.
- [24] MURAT U, SELCUK Y, MUHSIN T G. An effective wavelet-based feature extraction method for classification of power quality disturbance signals[J]. Electric Power Systems Research, 2008, 3(2): 1747-1755.
- [25] 周祖宁. 基于 S 变换的电能质量扰动识别算法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
- ZHOU Zuning. Research on power quality disturbance recognition algorithm based on S-transform[D]. Harbin: Harbin Institute of Technology, 2017.

收稿日期: 2019-05-20; 修回日期: 2019-12-11

作者简介:

王仁明(1964—), 男, 通信作者, 教授, 硕士生导师, 研究方向为分布式发电分析与控制; E-mail: eermwang@ctgu.edu.cn

汪宏阳(1995—), 男, 硕士研究生, 研究方向为电能质量分析. E-mail: 894686764@qq.com

(编辑 许威)