

DOI: 10.7667/PSPC180414

基于不完全 S 变换与梯度提升树的电能质量复合扰动识别

许立武, 李开成, 罗奕, 肖贤贵, 张婵, 蔡得龙

(华中科技大学电气与电子工程学院强电磁工程与新技术国家重点实验室, 湖北 武汉 430074)

摘要: 针对电能质量复合扰动中特征选择困难和分类准确率不高的问题, 提出基于不完全 S 变换和梯度提升树的特征选择和分类器构建方法。首先通过选取特定频率的不完全 S 变换得到扰动的时频矩阵。再从时频矩阵中提取 53 种原始特征量, 并基于梯度提升树对各个特征的重要性进行度量, 选取重要特征。最后根据选取的特征集训练和构建梯度提升树, 得到扰动分类器。仿真实验表明, 对于包括 8 种复合扰动在内的共 17 种扰动类型, 该方法的分类准确率高于 CART 决策树、随机森林(RF)、多层感知机(MLP)等现有方法。不同噪声条件下的分类结果表明, 该方法具有良好的抗噪性能和算法鲁棒性, 展现出良好的应用前景。

关键词: 电能质量; 扰动识别; 梯度提升树; 不完全 S 变换; 特征选择

Classification of complex power quality disturbances based on incomplete S-transform and gradient boosting decision tree

XU Liwu, LI Kaicheng, LUO Yi, XIAO Xiangui, ZHANG Chan, CAI Delong

(State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: In order to solve the problem of feature selection and classification accuracy in complex power quality disturbances recognition, a method of feature selection and classifier construction based on incomplete S-transform and Gradient Boosting Decision Tree (GBDT) is proposed. Firstly, the time-frequency matrixes of the disturbances are obtained by the incomplete S-transform at the specific frequencies. Then the original feature set of 53 features is extracted from the time-frequency matrix, and the importance level of each feature is measured by GBDT to select optimal features. Finally, the disturbance classifier is trained and constructed by GBDT based on the selected feature set. The simulation results show that the classification accuracy of the proposed method is higher than the CART, the Random Forest (RF), the Multilayer Perceptron (MLP), and so on, for a total of 17 disturbance classes including 8 classes of combined disturbances. Moreover, the classification results in different noise conditions indicate that the method has good anti-noise performance and algorithm robustness, which shows good prospects for application.

This work is supported by National Natural Science Foundation of China (No. 51277080).

Key words: power quality; disturbances classification; gradient boosting decision tree (GBDT); incomplete S-transform; feature selection

0 引言

随着新能源发电和各种非线性负载的广泛使用, 电力系统中的电能质量扰动(Power Quality Disturbances, PQDs)事件日益增多。电能质量扰动通常表现为电压或电流的幅值、相位以及频率的突然变化或变形。这些扰动会带来一系列的问题, 如电

气设备损坏、能耗增加、电磁干扰增强, 甚至危害电力系统的安全稳定运行^[1]。录波器记录了大量的 PQDs 数据, 靠人工的观察来判断电能质量扰动类型是不现实的^[2]。因此, 需要研究能够自动判别电能质量扰动类型且高效准确的分类技术。

电能质量扰动的自动识别一般分为以下 3 个步骤: 信号分析与特征提取、特征选择、扰动分类。在信号分析和特征提取阶段, 常用的信号处理方法包括短时傅里叶变换(STFT)^[3]、小波变换(WT)^[4]、

希尔伯特-黄变换(HHT)^[5]、S 变换^[6-8]等。STFT 利用时-频窗口来定位信号的暂态特征,然而不变的窗口大小导致时-频分辨率固定;小波变换将扰动信号分解得到多层小波系数,然而小波变换对噪声很敏感;HHT 可以分析非平稳信号,然而它无法有效地构建频谱;相比其他方法,S 变换使用窗口大小随频率变化的高斯窗,具有良好的时频分析能力,得到了广泛应用^[6-8]。

特征选择可以移除冗余特征,提高准确率并减少计算时间^[9]。常用的特征选择方法包括粒子群算法(Particle Swam Optimization, PSO)、遗传算法(Genetic Algorithm, GA)等智能优化算法^[10-11]以及基于决策树的基尼系数等方法^[12]。利用智能算法进行特征选择计算量较大,并且选出的特征往往需要与特定的分类器配合使用。而基于决策树的方法则通过得到特征的重要性排序很方便地实现重要特征的选取。

在扰动分类阶段,多种分类技术都有应用^[13],但其中应用最广泛的是人工神经网络(Artificial Neural Networks, ANN)^[14-16]、支持向量机(Support Vector Machine, SVM)^[17-19]、决策树(Decision Tree, DT)^[7-8,12]等。ANN 通过学习权重可以学习到数据的映射关系,但网络结构的选择缺乏理论指导,且需要较长的训练时间;SVM 适用于具有大量特征的分类问题,然而解决多分类问题存在困难且对大规模训练样本难以实施;DT 可以构建出 if-then 的决策规则,结构简单且可解释性强。然而,DT 构建的规则常常过拟合,需要通过剪枝等方法提高泛化能力。

为了解决现有方法在特征选取上的困难和分类准确率上的不足,本文提出一种基于不完全 S 变换与梯度提升树的特征选取和扰动分类方法。首先选取重要的频率进行不完全 S 变换并提取原始特征;再根据梯度提升树度量各个特征的重要性,进行重要特征的选取;最后在选取的重要特征集上构建梯度提升树,实现复合扰动的分类。仿真结果表明,利用该方法构建的扰动分类器,在不同的噪声环境下均具有较高的分类准确率,显示出良好的抗噪能力和识别能力。

1 S 变换

S 变换通过一个可以缩放和移动的高斯窗来提供可变的时频分辨率。

电能质量扰动信号的离散形式可表示为 $h(kT)$, 其中, $k = 0, 1, 2, \dots, N-1$, T 是采样间隔。它的快速傅里叶变换可表示为式(1)。

$$X\left[\frac{n}{NT}\right] = \frac{1}{N} \sum_{k=0}^{N-1} h(kT) e^{-i2\pi nk/N} \quad (1)$$

然后, $h(kT)$ 的离散 S 变换可由式(2)得到。

$$S(jT, n/NT) = \begin{cases} \sum_{m=0}^{N-1} X\left(\frac{m+n}{NT}\right) G(m, n) e^{i2\pi mj/N}, n \neq 0 \\ \frac{1}{N} \sum_{m=0}^{N-1} X\left(\frac{m}{NT}\right), n = 0 \end{cases} \quad (2)$$

式中: $j, m = 0, 1, 2, \dots, N-1$; $n = 0, 1, 2, \dots, N/2$; $G(m, n)$ 为高斯窗函数, 其定义为

$$G(m, n) = e^{-2\pi^2 m^2 \lambda / n^2} \quad (3)$$

参数 λ 为窗函数调整参数, 用来改善 S 变换在不同频率下的时频分辨率。 λ 值越小则 S 变换时间分辨率越高, λ 值越大则 S 变换的频率分辨率越高。

2 基于不完全 S 变换的特征提取

2.1 扰动信号类型

本文考虑正常信号(C0)、8 种单扰动和 8 种复合扰动在内的共 17 种扰动信号类型。8 种单扰动包括: 电压暂降(C1)、电压暂升(C2)、电压中断(C3)、暂态振荡(C4)、电压切口(C5)、电压尖峰(C6)、闪变(C7)、谐波(C8)。8 种复合扰动包括: 暂降+谐波(C9)、暂升+谐波(C10)、闪变+谐波(C11)、中断+谐波(C12)、闪变+暂降(C13)、闪变+暂升(C14)、暂降+振荡(C15)、暂升+振荡(C16)。扰动信号根据文献[4,8,16]的数学模型生成。

2.2 不完全 S 变换特征提取技术

对于由 N 个离散采样点构成的扰动信号, 经过 S 变换后, 将会得到一个 $N/2 \times N$ 的二维复数矩阵。该复数矩阵的幅值矩阵(记作 STA)能够反映信号的时间-频率特性, 其中, STA 的行反映某个频率下对应的时域信息, 而列反映某个时刻下对应的频域信息。

通过对扰动信号和 STA 矩阵进行分析, 发现“有用信息”往往集中在某些频率位置上, 因此可仅在特定频率对信号进行不完全 S 变换, 大幅减少计算时间^[7]。

特定频率选择 $m \times 50$ Hz 和 Nyquist 频率, 其中 $m = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$ 。 $m=1$ 时为基频, 可反映暂降、暂升, 中断和闪变的特性; $m=3, 5, 7, 9$ 时为奇次谐波主要成分所在频率, 可以反映谐波的特性; $m = \{1, 3, 5, 7, 9, 11, 13, 15, 17, 19\}$ 的频率范围包含了振荡中心频率(300~900 Hz), 可反映暂态振荡的特性; Nyquist 频率可反映切口和尖峰的周期特性。

为了在基频上有更好的时间分辨率, 方便识别暂降、暂升、中断、和闪变扰动, $m=1$ 时窗函数的

调整参数 λ 取 0.2, 而其他频率下 λ 取 5 来获得更好的频率分辨率。

对扰动信号进行不完全 S 变换得到 STA 矩阵后, 从中可以提取时频变化特征, 其中对应频率变化的特征向量包括: 各行最大值(r_{\max})、各行最小值(r_{\min})、各行均值(r_{mean})、各行标准差(r_{std})和各行均方根值(r_{rms})等; 对应时间变化的特征向量包括: 各列最大值(c_{\max})、各列最小值(c_{\min})、各列均值(c_{mean})、各列标准差(c_{std})和各列均方根值(c_{rms})等。

取 STA 矩阵以及各特征向量的最大值与最小值的和 ($\max + \min$)、最大值和最小值的差 ($\max - \min$)、均值(mean)、标准差(std)、均方根(rms)构成特征 $F1 - F51$, 具体如表 1 所示。另外, 为了更好地反映信号的总体变化特征, 本文引入 STA 矩阵的高阶统计量偏度(Skewness)和峭度(Kurtosis)构成特征 $F52$ 和特征 $F53$ 。特征 $F1 - F53$ 组成原始特征集合, 各特征统计量计算方法参照文献[11-12]。

表 1 特征提取($F1 - F51$)

Table 1 Feature extraction ($F1 - F51$)					
	max+ min	max- min	mean	std	rms
STA	F1	F2	F3	F4	F5
c_{\max}	F6	F7	F8	F9	F10
c_{\min}	F11	F12	F13	F14	F15
c_{mean}	F16	F17	—	F18	F19
c_{std}	F20	F21	F22	F23	F24
c_{rms}	F25	F26	F27	F28	—
r_{\max}	F29	F30	F31	F32	F33
r_{\min}	F34	F35	F36	F37	F38
r_{mean}	F39	F40	—	F41	F42
r_{std}	F43	F44	F45	F46	F47
r_{rms}	F48	F49	F50	F51	—

3 基于梯度提升树的特征选择和分类器构建

3.1 梯度提升树原理

梯度提升树(Gradient Boosting Decision Tree, GBDT) 是一种以 CART(Classification and Regression Tree)中的回归树作为基学习器的提升算法, 通过基学习器的加权组合得到集成学习器^[20-21]。

GBDT 采用加法模型和前向分步算法实现学习的迭代过程, 在每一步中, 通过损失函数最小化来确定下一棵决策树。针对每一步的优化问题难以求解的难题, Freidman 提出了梯度提升方法: 即利用损失函数的负梯度在当前模型的值, 作为提升树的残差近似值, 拟合下一棵决策树^[20-21]。

GBDT 通常具有三种正则化方式, 来防止对于

训练数据的过拟合。第一种是 Shrinkage 策略, 通过设定学习率(learning rate)来减少每一棵回归树对预测的贡献。第二种是子采样(subsample), 在拟合一棵新的回归树时, 通过无放回的子抽样得到样本集的子集。第三种是对基学习器即 CART 回归树进行剪枝处理。除此之外, 树模型的复杂度也可作为正则项加进优化目标里, 进一步提高泛化能力^[21]。因此, 相比其他分类算法, GBDT 的泛化能力通常更强, 可获得更高的分类准确率。

3.2 特征选择方法

GBDT 可以计算得到各个特征的重要度, 实现特征重要性的排序。因此, 在得到原始特征集 $F1 - F53$ 之后, 可用 GBDT 进行特征选择, 来降低计算复杂度, 更有效地实现扰动的分类。

利用 GBDT 进行特征的重要度量是基于各个特征用于决策树节点分裂的次数和每次分裂后模型的结构增益, 累计求和后对所有的树进行平均^[20-21]。

特征 j 的重要度是通过该特征在所有树中的重要度的平均值来衡量:

$$\hat{I}_j = \frac{1}{M} \sum_{m=1}^M \hat{I}_j(T_m) \quad (4)$$

式中, M 为决策树的数量。特征 j 在单棵树中的重要度计算公式为

$$\hat{I}_j(T) = \sum_{t=1}^{L-1} \hat{i}_t l(v_t = j) \quad (5)$$

式中: 树的非叶子节点数量为 $L - 1$; v_t 是与节点 t 相关的分割特征; \hat{i}_t 是节点 t 分裂之后模型的结构增益。

3.3 分类器构建

利用 GBDT 算法构建扰动分类器, 是一个学习由输入空间到标记空间映射的过程。其中, 各个样本的特征向量构成输入空间, 对应的扰动类别标签构成标记空间。图 1 显示了基于梯度提升树的扰动分类器的构建过程。当得到一个新样本的特征向量后, 学得分类器可以对该样本的扰动类别进行预测, 进而实现对电能质量扰动的识别。

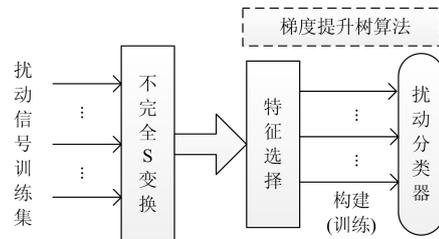


图 1 基于梯度提升树的扰动分类器构建
Fig. 1 Construction of classifier for PQDs based on GBDT algorithm

基于 GBDT 构建扰动分类器的算法框架如下:

(1) 输入: 训练集 $T = \{(\bar{x}_1, y_1), (\bar{x}_2, y_2), \dots, (\bar{x}_N, y_N)\}$, $\bar{x}_i \in R^n, y_i \in \{1, 2, \dots, K\}$, K 为扰动类别数;

(2) 损失函数: 多分类交叉熵(cross-entropy)损失函数

$$L(y, f(\bar{x})) = - \sum_{k=1}^K y_k \log(p_k) \quad (6)$$

$$p_k(\bar{x}) = \exp(f_k(\bar{x})) / \sum_{l=1}^K \exp(f_l(\bar{x})) \quad (7)$$

其中, 如果样本输出类别为 k , 则 $y_k=1$, 否则 $y_k=0$;

(3) 输出: 提升树 $f_M(\bar{x})$, M 为决策树的数量;

(4) 算法流程

Step1: 初始化一棵只有根节点的树(常数值 c), 根节点的类别是使得损失函数最小的值

$$f_0(\bar{x}) = \arg \min_c \sum_{i=1}^N L(y_i, c) \quad (8)$$

式中, N 为训练样本的个数;

Step2: 迭代过程, 对 $m=1, 2, \dots, M$

① 计算在当前模型下损失函数的负梯度值, 并将它作为残差的估计值。

对于 $i=1, 2, \dots, m$, 计算

$$r_{mi} = - \left[\frac{\partial L(y_i, f(\bar{x}_i))}{\partial f(\bar{x}_i)} \right]_{f=f_{m-1}} \quad (9)$$

② 对 r_{mi} 拟合一颗回归树, 得到第 m 棵树的叶节点区域 R_{mj} , $j=1, 2, \dots, J_m$ 。

③ 计算在每个区域 R_{mj} 上的输出类别, 使损失函数最小化。

对 $j=1, 2, \dots, J_m$, 计算

$$c_{mj} = \arg \min_c \sum_{\bar{x}_i \in R_{mj}} L(y_i, f_{m-1}(\bar{x}_i) + c) \quad (10)$$

④ 更新梯度提升树

$$f_m(\bar{x}) = f_{m-1}(\bar{x}) + \sum_{j=1}^{J_m} c_{mj} I(\bar{x} \in R_{mj}) \quad (11)$$

Step3: 得到最终模型

$$f_M(\bar{x}) = \sum_{m=1}^M \sum_{j=1}^{J_m} I(\bar{x} \in R_{mj}) \quad (12)$$

4 仿真实验与对比分析

利用 Matlab 由随机参数生成各类扰动信号, 其中采样率为 6 400 Hz, 基频取 50 Hz, 采样周期为 10 个周期(0.2 s), 噪声为高斯白噪声。

GBDT 的算法参数设置如下: (1) 树的数目(M)取 50; (2) 树的最大深度(max_depth)取 5; (3) 学习率(learning rate)取 0.3; (4) 子采样率(subsample)取 0.8。

4.1 不完全 S 变换

首先生成信噪比在 20~50 dB 之间, 具有均匀分布随机值的各类扰动信号, 训练集每类信号 800 组, 测试集每类信号 200 组。然后对各扰动信号进行不完全 S 变换, 分别得到 STA 矩阵。图 2—图 4 给出了电压切口(C5), 闪变+谐波(C11), 暂降+振荡(C15)三种电能质量扰动事件的原始信号以及对应 STA 矩阵的三维时-频图。

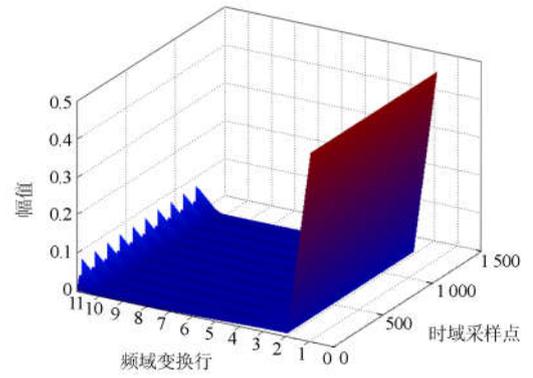
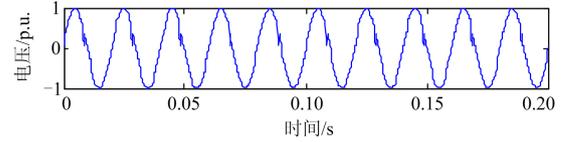


图 2 切口及其时频分析

Fig. 2 Notch and its time-frequency analysis

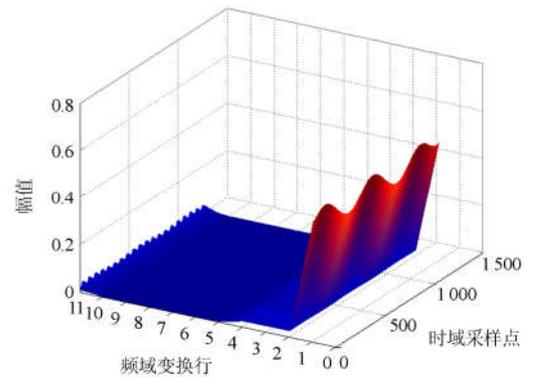
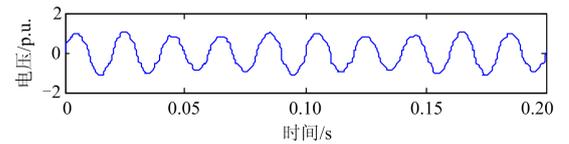
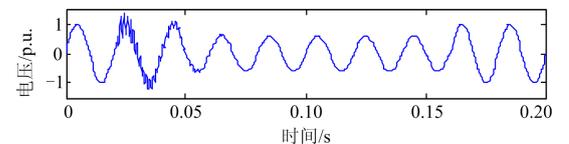


图 3 闪变+谐波及其时频分析

Fig. 3 Flicker with harmonic and its time-frequency analysis



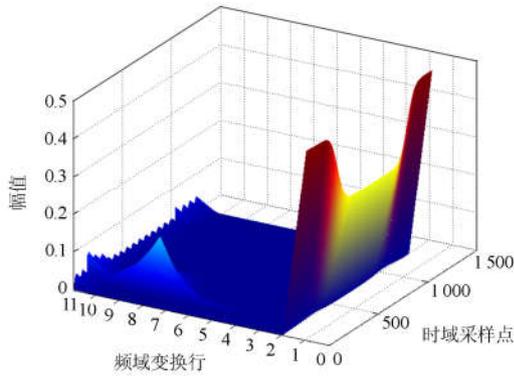


图 4 暂降+振荡及其时频分析

Fig. 4 Sag with oscillation and its time-frequency analysis

由图 2—图 4 可知：利用本文的不完全 S 变换方法得到的 STA 矩阵可以很好地反映各类扰动信号的特性，验证了该方法的有效性。同时，由于只需进行 11 次变换过程，相比于需要进行 $N/2$ 次变换过程(本文采样率和采样周期下 $N=1280$)的传统 S 变换，时间仅需原来的 $22/N$ 倍，大幅降低了计算量和计算时间，可以更好地满足实时性的要求。

4.2 重要特征的选取

得到各个样本的 STA 矩阵后，从中提取原始特征集($F1-F53$)。使用 GBDT 算法训练分类器后，得到各原始特征的重要性，前二十个重要特征归一化后的相对重要性如图 5 所示。

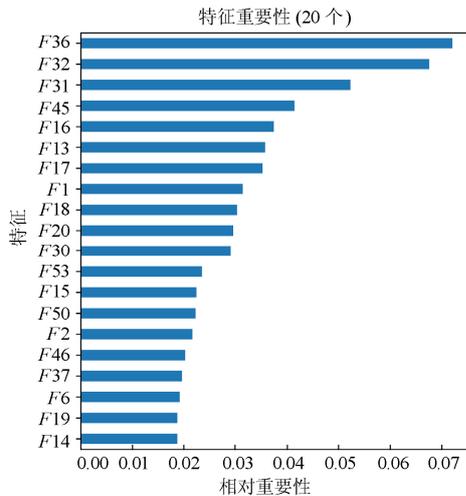


图 5 前二十个重要特征的相对重要性

Fig. 5 Relative importance of top 20 important features

为了得到选取的特征数目与测试集准确率的关系，从最重要的特征($F36$)开始，依次选取对应数目的重要特征，分别得到重要特征子集。然后使用 GBDT 算法重新训练分类器，得到特征数目与测试集准确率的对对应关系，如表 2 和图 6 所示。

表 2 测试集准确率与选取的特征数目的关系
Table 2 Relationship between the accuracy of test set and the number of selected features

特征数	1	2	3	4	5
准确率/%	43.41	87.50	92.32	95.97	96.76
特征数	6	7	8	9	10
准确率/%	96.97	97.26	97.32	97.35	97.41
特征数	11	12	13	14	15
准确率/%	97.35	97.65	97.65	97.82	97.65
特征数	16	17	18	19	20
准确率	97.79	97.71	97.74	97.91	97.76

当不进行特征选取，使用原始特征集 53 个特征时，测试集准确率为 97.79%。而从表 2 可以看出，当选取的特征数多于 3 个时，训练得到的分类器在测试集上的准确率均超过 95%，表明了该特征选取方法可以有效地提取重要特征，并大幅降低计算量。

最优特征子集的选取本质上是个组合优化问题，当特征数目增多时，分类准确率既有可能因为假设空间的增大而上升，也有可能因为冗余特征的加入而下降^[11]。从图 6 可以看出，在特征数小于等于 14 个特征时，测试集准确率总体上随着特征数目的增加而增加，在 14 个特征时测试集准确率达到 97.82%。当选取的特征数继续增加时，仅当特征数目为 19 个时，测试集准确率增加到 97.91%，而其他情况下均小于 14 个特征时的结果。

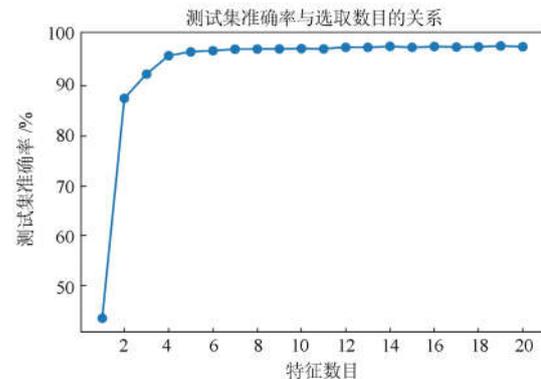


图 6 测试集准确率的变化曲线

Fig. 6 Curve of the test set accuracy

与 14 个特征相比，19 个特征时特征的数目增加了 35.71%，而准确率仅仅提高了 0.09%。考虑到计算复杂度与分类准确率的权衡，本文选取前 14 个重要特征构成重要特征集 $\{F36, F32, F31, F45, F16, F13, F17, F1, F18, F20, F30, F53, F15, F50\}$ 。该特征集上测试集的准确率比原始特征集高，且相比原始特征集(53 个特征)，特征的数目要少 73.6%左右，显著降低了分类器构建的计算复杂度。

4.3 不同信噪比下的扰动分类器构建

为了验证本文提出的方法在不同噪声环境中的表现, 首先利用 Matlab 仿真生成信噪比分别为 50 dB、40 dB、30 dB、20 dB 的扰动信号训练集和测试集。各个信噪比下, 训练集中每类样本 800 组, 测试集中每类样本 200 组, 分别利用 GBDT 算法训练和构建扰动分类器。特征集在特征选取步骤中已选出: $\{F36, F32, F31, F45, F16, F13, F17, F1, F18, F20, F30, F53, F15, F50\}$, 实验结果如表 3 所示。

由表 3 可知, 随着信噪比的降低, 由于噪声干扰越来越大, 分类器的分类准确率随之降低。然而, 在普通的噪声环境下(信噪比大于或等于 30 dB), 各分类器的分类准确率均高达 98%以上。即使是在极端噪声环境下(信噪比 20 dB 时), 分类器仍然具有高达 95.56%的分类准确率。这表明了本文基于梯度提升树的特征提取和分类器构建方法具有良好的抗噪能力和识别准确率。

表 3 不同信噪比下的分类性能

Table 3 Classification performances in different SNRs

扰动类型	分类准确率/%				
	50 dB	40 dB	30 dB	20 dB	
单 扰 动	C0	100.00	100.00	99.00	91.50
	C1	99.00	99.00	99.50	98.00
	C2	99.50	99.50	98.00	95.50
	C3	99.50	100.00	100.00	99.50
	C4	100.00	99.50	100.00	94.50
	C5	98.50	95.00	95.00	83.00
	C6	98.50	98.00	96.00	82.00
	C7	100.00	100.00	100.00	100.00
复 合 扰 动	C8	100.00	100.00	100.00	100.00
	C9	100.00	99.00	100.00	99.50
	C10	99.50	100.00	100.00	100.00
	C11	100.00	100.00	100.00	100.00
	C12	100.00	100.00	99.50	100.00
	C13	98.50	99.50	98.50	98.50
	C14	98.50	99.50	99.00	97.50
	C15	99.00	98.00	97.00	94.00
	C16	94.00	92.00	93.00	91.00
平均值	99.09	98.76	98.50	95.56	

以 20 dB 噪声环境为例, 图 7 显示了扰动分类器构建过程中, 训练集和测试集上分类错误率的收敛过程。当迭代次数(CART 回归树的数目)不断增加时, 训练集和测试集上的分类错误率均表现出明显的下降和收敛特性。梯度提升树构建完成时, 分类器在训练集上的分类错误率接近 0, 泛化误差(测试集上的分类错误率)也仅在 4.4%左右。这表明了分类器在高噪声环境中仍具有良好的泛化识别能力, 可以有效识别各种电能质量扰动事件。

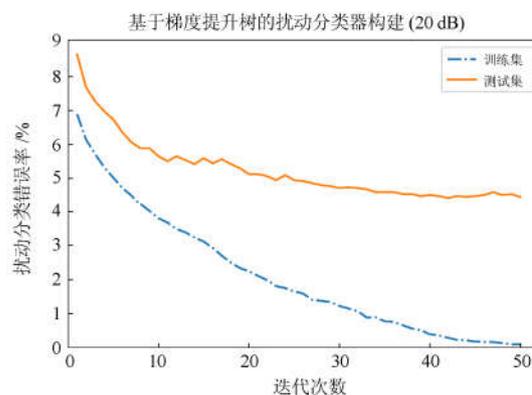


图 7 分类器分类错误率的收敛过程(20 dB)

Fig. 7 Convergence process of classification error rate (20 dB)

4.4 与其他扰动分类器的对比

为了验证本文基于梯度提升树的方法的有效性, 与基于 CART^[12]、KNN^[22]、朴素贝叶斯(Naïve Bayes)^[23]、随机森林(Random Forest)^[24]、多层感知机(MLP)^[25]的一系列分类方法进行比较, 并采用本文的不完全 S 变换方法和原始特征集, 对比结果如表 4 所示。由表 4 可知, 本文提出的方法在分类准确率上优于用来对比的各种算法, 可有效识别 17 种扰动, 并展现出良好的抗噪性能和算法鲁棒性。

表 4 不同分类算法性能比较

Table 4 Performance comparisons of different classification algorithms

分类算法	分类准确率/%			
	50 dB	40 dB	30 dB	20 dB
GBDT	99.09	98.76	98.50	95.56
CART	98.08	98.29	97.67	92.53
kNN	90.97	91.65	90.91	85.23
Naïve Bayes	94.97	95.20	94.35	89.41
Random Forest	98.88	98.70	98.26	95.46
MLP	96.41	96.14	94.91	92.61

5 结论

针对电能质量复合扰动存在的特征混叠现象而导致的特征选择困难和识别准确率不高的问题, 本文提出基于不完全 S 变换与梯度提升树(GBDT)的特征选择和扰动分类器构建方法:

(1) 提出了一种选取特定频率进行不完全 S 变换的方法, 大幅降低了 S 变换的计算量;

(2) 提出了一种利用 GBDT 算法来计算各个特征重要性的特征选择方法;

(3) 提出了基于 GBDT 算法的电能质量复合扰

动分类器构建方法。

仿真实验的结果表明,相比其他一些已有方法,该方法具有更高的分类准确率,能够有效识别包括8种复合扰动在内的17种电能质量扰动事件。且本文的方法在高噪声环境下对各类扰动仍具有很好的识别效果,表明了该方法具有良好的抗噪能力和算法鲁棒性。接下来准备将本文提出的方法应用于实际电能质量扰动信号,以验证本文方法的工程应用价值。

参考文献

[1] LI J, TENG Z, TANG Q, et al. Detection and classification of power quality disturbances using double resolution S-transform and DAG-SVMs[J]. IEEE Transactions on Instrumentation and Measurement, 2016, 65(10): 2302-2312.

[2] KHOKHAR S, ZIN A A B M, MOKHTAR A S B, et al. A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances[J]. Renewable and Sustainable Energy Reviews, 2015, 51: 1650-1663.

[3] 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.

[4] 任子晖, 刘昊岳, 徐进霞. 基于小波变换和改进 Prony 方法的电能质量扰动分析[J]. 电力系统保护与控制, 2016, 44(9): 122-128.
REN Zihui, LIU Haoyue, XU Jinxia. Power quality disturbance analysis based on wavelet transform and improved Prony method[J]. Power System Technology, 2016, 44(9): 122-128.

[5] 田振果, 傅成华, 吴浩, 等. 基于 HHT 的电能质量扰动定位与分类[J]. 电力系统保护与控制, 2015, 43(16): 36-42.
TIAN Zhenguo, FU Chenghua, WU Hao, et al. Power quality disturbance for location and classification based on HHT[J]. Power System Protection and Control, 2015, 43(16): 36-42.

[6] 郭俊文, 李开成, 何顺帆, 等. 基于改进不完全 S 变换与决策树的实时电能质量扰动分类[J]. 电力系统保护与控制, 2013, 41(22): 103-110.
GUO Junwen, LI Kaicheng, HE Shunfan, et al. A real time power quality disturbance classification based on improved incomplete S-transform and decision tree[J].

Power System Protection and Control, 2013, 41(22): 103-110.

[7] BISWAL M, DASH P K. Detection and characterization of multiple power quality disturbances with a fast S-transform and decision tree based classifier[J]. Digital Signal Processing, 2013, 23(4): 1071-1083.

[8] KUMAR R, SINGH B, SHAHANI D T, et al. Recognition of power-quality disturbances using S-transform-based ANN classifier and rule-based decision tree[J]. IEEE Transactions on Industry Applications, 2015, 51(2): 1249-1258.

[9] XUE B, ZHANG M, BROWNE W N, et al. A survey on evolutionary computation approaches to feature selection[J]. IEEE Transactions on Evolutionary Computation, 2016, 20(4): 606-626.

[10] AHILA R, SADASIVAM V, MANIMALA K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances[J]. Applied Soft Computing, 2015, 32: 23-37.

[11] SINGH U, SINGH S N. Optimal feature selection via NSGA-II for power quality disturbances classification[J]. IEEE Transactions on Industrial Informatics, 2017, 14(7): 2994-3002.

[12] 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): 776-786.
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): 776-786.

[13] 刘志刚, 张巧革, 张杨. 电能质量复合扰动分类的研究进展[J]. 电力系统保护与控制, 2013, 41(13): 146-153.
LIU Zhigang, ZHANG Qiaoge, ZHANG Yang. Review of power quality mixed disturbances identification[J]. Power System Protection and Control, 2013, 41(13): 146-153.

[14] NAIK C A, KUNDU P. Power quality disturbance classification employing S-transform and three-module artificial neural network[J]. International Transactions on Electrical Energy Systems, 2014, 24(9): 1301-1322.

[15] 何巨龙, 王根平, 刘丹, 等. 基于提升小波和改进 BP 神经网络的配电网系统电能质量扰动定位与识别[J]. 电力系统保护与控制, 2017, 45(10): 69-76.
HE Julong, WANG Genping, LIU Dan, et al. Localization and identification of power quality disturbance in distribution network system based on lifting wavelet and

- improved BP neural network[J]. Power System Protection and Control, 2017, 45(10): 69-76.
- [16] 陈晓静, 李开成, 肖剑, 等. 一种实时电能质量扰动分类方法[J]. 电工技术学报, 2017, 32(3): 45-55.
CHEN Xiaojing, LI Kaicheng, XIAO Jian, et al. A method of real-time power quality disturbance classification[J]. Transactions of China Electrotechnical Society, 2017, 32(3): 45-55.
- [17] 黄南天, 徐殿国, 刘晓胜. 基于 S 变换与 SVM 的电能质量复合扰动识别[J]. 电工技术学报, 2011, 26(10): 23-30.
HUANG Nantian, XU Dianguo, LIU Xiaosheng. Identification of power quality complex disturbances based on S-transform and SVM[J]. Transactions of China Electrotechnical Society, 2011, 26(10): 23-30.
- [18] 张巧革, 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.
- [19] 王凌云, 李开成, 肖厦颖, 等. 基于稀疏分解的复合电能质量扰动分类[J]. 电测与仪表, 2018, 55(1): 14-20.
WANG Lingyun, LI Kaicheng, XIAO Xiaying, et al. Classification for multiple power quality disturbances based on sparse decomposition[J]. Electrical Measurement & Instrumentation 2018, 55(1): 14-20.
- [20] KE G, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C] // Advances in Neural Information Processing Systems, 2017: 3149-3157.
- [21] 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, ACM, 2016: 785-794.
- [22] PANIGRAHI B K, PANDI V R. Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm[J]. IET Generation, Transmission & Distribution, 2009, 3(3): 296-306.
- [23] JAMALI S, FARSA A R, GHAFARZADEH N. Identification of optimal features for fast and accurate classification of power quality disturbances[J]. Measurement, 2018, 116: 565-574.
- [24] 瞿合祚, 刘恒, 李晓明, 等. 基于多标签随机森林的电能质量复合扰动分类方法[J]. 电力系统保护与控制, 2017, 45(11): 1-7.
QU Hezuo, LIU Heng, LI Xiaoming, et al. Recognition of multiple power quality disturbances using multi-label random forest[J]. Power System Protection and Control, 2017, 45(11): 1-7.
- [25] FERREIRA D D, DE SEIXAS J M, CERQUEIRA A S. A method based on independent component analysis for single and multiple power quality disturbance classification[J]. Electric Power Systems Research, 2015, 119: 425-431.

收稿日期: 2018-04-13; 修回日期: 2018-06-07

作者简介:

许立武(1995—), 男, 硕士研究生, 研究方向为电能质量分析, 机器学习与深度学习、优化方法等在电力系统中的应用; E-mail: liwu_xu@hust.edu.cn

李开成(1962—), 男, 通信作者, 教授, 博士生导师, 研究方向为电子式互感器、电能质量分析及电磁测量与仪器; E-mail: likaicheng@hust.edu.cn

罗奕(1991—), 男, 博士研究生, 研究方向为电能质量分析、模式识别、智能仪器等。E-mail: yiluo2009@hust.edu.cn

(编辑 葛艳娜)