

DOI: 10.7667/PSPC160899

基于多标签随机森林的电能质量复合扰动分类方法

瞿合祚¹, 刘恒², 李晓明¹, 黄建明¹

(1. 武汉大学电气工程学院, 湖北 武汉 430072; 2. 湖北省电力公司孝感供电公司, 湖北 孝感 432000)

摘要: 提出一种多标签随机森林(Multi-label Random Forest, ML-RF)分类算法, 并将其应用于电能质量复合扰动分类。ML-RF 是基于多标签决策树(Multi-label Decision Tree, ML-DT)的集成学习算法, 利用子决策树的组合来增强分类器的整体性能。首先对电能质量扰动信号进行平稳小波变换, 计算各层分解系数的小波能量熵作为分类特征向量。然后使用 Bootstrap 自助法和子空间采样构造不同的训练集训练子决策树。最后组合子决策树得到 ML-RF 分类器, 并对复合电能质量扰动进行分类。仿真结果表明, 在不同噪声情况下, 该方法均能有效进行复合扰动的分类, 具有较好的噪声鲁棒性, 是复合电能质量扰动分类的一种可行方法。

关键词: 电能质量; 复合扰动; 多标签分类; 随机森林; 决策树

Recognition of multiple power quality disturbances using multi-label random forest

QU Hezuo¹, LIU Heng², LI Xiaoming¹, HUANG Jianming¹

(1. School of Electrical Engineering, Wuhan University, Wuhan 430072, China;

2. Xiaogan Power Supply Company, Hubei Electric Power Company, Xiaogan 432000, China)

Abstract: Multi-label Random Forest (ML-RF) is proposed and applied to the recognition of multiple power quality disturbances. ML-RF is an ensemble learning algorithm based on the Multi-label Decision Tree (ML-DT), by assembling sub-classification decision tree to enhance the overall performance. Firstly, all power quality disturbances are decomposed by steady wavelet transform, and the energy entropy of the wavelet coefficients are extracted as eigenvectors. Then, it uses training sets that are constructed by Bootstrap re-sampling method and subspace sampling method to train sub-DTs. Finally, it combines the sub-DTs by majority voting to predict the type of multiple power quality disturbances. The simulation results show that the ML-RF can effectively recognize the multiple power quality disturbances under different noise conditions, and it is a feasible method with noise robustness.

This work is supported by National Natural Science Foundation of China (No. 51277134).

Key words: power quality; multiple disturbances; multi-label classification; random forests; decision tree

0 引言

随着新能源的不断发展和电力电子技术的广泛运用, 同时电力系统的负荷结构趋于复杂, 大量非线性冲击性负荷使得电能质量扰动事件频发, 对系统的安全稳定运行产生极大威胁。如何准确地识别电能质量扰动的类别是提高电能质量的关键一步。

电能质量扰动分类问题的建模过程通常分为特征提取和模式分类两步。电能质量扰动信号是典型的时变信号, 为了更充分地分析其特点, 通常采用

时频分析进行扰动特征提取, 主要有 S 变换^[1-2]、小波变换^[3]和短时傅立叶变换^[4]。短时傅立叶变换通过加窗使其具有时频分析能力, 能采用快速傅立叶变换计算, 物理意义明确, 但其时频分辨率固定, 无法兼顾高频和低频信号的分辨率需求。小波变换支持多分辨率分析, 在电能质量扰动特征提取中应用广泛, 能对高频信号提供高的时域分辨率, 对低频信号提供高的频域分辨率。S 变换继承和发展了短时傅立叶变换和小波变换的局部化优点, 采用与频率有关的可变高斯窗函数, 具有较好的时频特性, 但无法准确地检测瞬态脉冲信号。

电能质量扰动的模式分类问题通常不是简单的多分类问题, 复杂的电网环境使得电能质量扰动

趋于复杂,在实际生产中的扰动不会只是理想的单一扰动^[5],比如在电压暂升期间可能出现瞬态脉冲,发生电压闪变的同时可能伴随着谐波扰动。电能质量扰动分类应当视为多标签分类问题,对此已有学者提出二分类-支持向量机(Binary-SVM)^[6],多标签 k -近邻法(ML-KNN)^[7],多标签径向基神经网络(ML-RBF)^[8],排位支持向量机(Rank-SVM)^[9]等等方法。文献[8]采用均值聚类确定每一类样本的聚类中心,将其作为径向基函数神经网络基函数的中心进行训练得到 ML-RBF 分类器,但是该分类器无法有效解释推理过程和依据。文献[7]基于 k -近邻分类器和贝叶斯准则提出 ML-KNN,在多标签分类中效果较好,但其需要存储所有的训练样本,当训练样本增加时需要更多的计算机内存。文献[9]基于 Rank-SVM 和 Morlet 小波核函数提出的分类方法能消除高频干扰,具有良好的抗噪性能,为电能质量复合扰动多标签分类的研究提供了新思路。

随机森林(Random Forest, RF)是一种基于决策树的集成学习算法^[10],其分类决策结果由构成森林的子决策树分类结果的众数决定。随机森林能在运算量没有显著提高的前提下提高了预测精度,对多元共线性不敏感,可以很好地预测多达几千个解释变量的作用。

本文提出一种基于多标签决策树的电能质量扰动多标签分类方法,多标签随机森林算法(ML-RF)。首先采用离散平稳小波变换分解扰动信号得到各层分解系数,并提取各层小波能量熵的统计参数作为分类特征;然后阐述构造多标签分类决策树的方法,并训练多棵子决策树构成随机森林分类器,最后对比随机森林和 Binary-SVM, ML-KNN, ML-RBF 对电能质量单一扰动和复合扰动的分类结果,验证了 ML-RF 算法的有效性。

1 特征提取

1.1 离散平稳小波变换

小波变换能同时满足低频和高频信号对时域频域不同的分辨率要求^[11],广泛应用于信号处理领域。根据应用场合和需求不同小波变换有许多变化,离散平稳小波变换是一种非正交的小波变换,如图1所示。与常用的采用 Mallat 算法^[12]进行小波分解与重构不同,计算离散平稳小波变换时并不对信号本身进行下抽样处理,而是对滤波器组进行内插补零(上采样),使其每级分解系数和原时间序列长度保持一致,这种处理方法可以有效地避免因信号下抽样造成重构信号发生振荡,也被称作多孔算法^[12]。对低通滤波器系数 $h[k]$ 和高通滤波器系数 $g[k]$ 进行

上采样操作如式(1)和式(2)所示。

$$h_{j+1}[k] = h_j[k] \uparrow 2 = \begin{cases} h_j[k/2] & k = 2n, n \in N \\ 0 & k = 2n+1, n \in N \end{cases} \quad (1)$$

$$g_{j+1}[k] = g_j[k] \uparrow 2 = \begin{cases} g_j[k/2] & k = 2n, n \in N \\ 0 & k = 2n+1, n \in N \end{cases} \quad (2)$$

相应地,离散平稳小波变换的分解公式如式(3)所示,其中 d_j 和 c_j 分别为每层的高频系数和低频系数, c_0 表示原信号。

$$\begin{cases} c_{j+1}[k] = h_j[-k] * c_j[k] \\ d_{j+1}[k] = g_j[-k] * c_j[k] \end{cases} \quad (3)$$

小波重构是与分解相反的逆过程,离散平稳小波变换的重构公式如式(4)所示。

$$c_j[k] = (h_j[k] * c_{j+1}[k] + g_j[k] * d_{j+1}[k]) / 2 \quad (4)$$

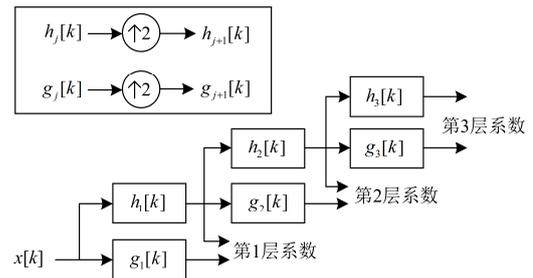


图1 平稳小波分解原理

Fig. 1 Algorithm of stationary wavelet decomposition

1.2 改进小波能量熵

小波变换可以同时时域和频域中定位分析非平稳时变信号,在其基础上引入熵的概念进而定义小波熵^[13-14],采用小波熵能对被分析信号在各频段上的分布特性进行统计分析。小波能量熵以小波变换的各层分解系数为基准,对信号不同频段的能量进行划分,最终以一个定量的熵值来反映信号能量在频域上的分布复杂度。在信号处理中运用较为广泛的是香农熵,设第 j 层小波分解系数序列为 S_j , $s_j[i]$ 表示该层第 i 个小波系数,则第 j 层小波分解系数的香农能量熵的定义如式(5)所示。

$$SE(S_j) = - \sum_{i=1}^N s_j^2[i] \ln s_j^2[i] \quad (5)$$

单一的小波能量熵反映的是某一频段小波系数的能量,没有考虑到能量随时间的变化情况,无法反映对应频段信号能量变化的局部信息。为了对非平稳时变信号采用小波能量熵进行分析,本文提出一种改进小波能量熵的计算方法,即通过对小波分解系数进行加窗处理再计算香农熵。设 g 为宽度为 N_w ($N_w \in N$) 的滑动窗,窗中心为 $\tau \in N$,则改进香农能量熵的计算如式(6)。在时轴上移动窗函数并对

加窗处理后的小波分解系数进行计算得到随时间变换的香农熵。通过对小波变换各层分解系数计算改进小波香农熵可获得信号不同频段分量的能量在时域分布的情况。

$$SE(S_j, \tau) = SE(S_j[k]g[k - \tau]) = -\sum_{i=1}^N (s_j[i]g[i - \tau])^2 \ln(s_j[i]g[i - \tau])^2 \quad (6)$$

2 多标签分类决策树

决策树是构成随机森林的基本单位, 是一种简单但是广泛使用的分类器^[15]。决策树有两大优点: (1) 决策树模型可读性好, 具有描述性, 有助于人工分析; (2) 效率高, 决策树只需要一次构建, 反复使用, 每一次预测的最大计算次数不超过决策树的深度。为了引入随机森林算法解决复合电能质量扰动分类, 首先需要构造支持多标签分类的决策树, 即多标签分类决策树(ML-DT)。本文构造的 ML-DT 基于经典的 C4.5 决策树算法, C4.5 算法自顶向下逐步生成一棵决策树, 在构造子节点时使子节点所含样本集的杂度最小化。本文采用熵量化样本集的杂度, ML-DT 和 C4.5-DT 的主要差别在于前者对样本集的熵进行推广, 使之能应用于多标签数据集^[16]。

设训练样本集为 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, 其中 $x_i \in \mathbf{R}^q = (x_i^{(1)}, \dots, x_i^{(q)})^T$, $y_i = (y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(L)})$, q 为样本特征维数, L 为分类标签数, m 为训练样本数, $y_i^{(l)} \in \{-1, +1\}$, $y_i^{(l)} = +1$ 表示样本 i 属于标签 l , 反之则不属于。

在单标签分类中, 单标签样本集 D_{SL} 的熵采用式(7)计算, 其中 p 表示样本集 D 中属于标签的样本比例。在熵的计算中, 定义 $0 \log 0 = 0$, 可见当 D_{SL} 中全部样本均属于标签或者均不属于标签时样本集的熵为零。熵是衡量样本集不确定性的量, 熵越小说明该样本集的杂度越小, 越适合作为叶节点。

$$ENT(D_{SL}) = -p \log_2 p - (1-p) \log_2 (1-p) \quad (7)$$

在 ML-DT 算法中对样本集的熵的定义进行了推广, 采用式(8)计算多标签样本集 D_{ML} 的熵。

$$ENT(D_{ML}) = \sum_{l=1}^L -p_l \log_2 p_l - (1-p_l) \log_2 (1-p_l) \quad (8)$$

式中, p_l 表示样本集 D_{ML} 中属于标签 l 的样本所占的比例。

生成决策树 $h(x, D)$ 主要分为以下 5 步。

Step1: 创建决策树根节点, 该节点对应训练样本集 D_F ;

Step2: 根据当前节点对应的样本集 D_F , 选取使信息增益最大的分割参数 fea 和 val , 其中信息增

益的定义如式(9)所示。

$$IG(D_F, fea, val) = ENT(D_F) - \sum_{\rho \in \{-, +\}} \frac{|D^\rho|}{|D_F|} ENT(D^\rho) \quad (9)$$

式中, D^ρ 表示按分割参数分割 D_F 得到的两个子集 D^+ 和 D^- , 分别为 $D^+ = \{(x_i, y_i) | x_i^{(fea)} \leq val, 1 \leq i \leq |D_F|\}$, $D^- = \{(x_i, y_i) | x_i^{(fea)} > val, 1 \leq i \leq |D_F|\}$;

Step3: 按照使当前节点信息增益 IG 最大的最优分割参数将 D_F 分割为两个子集 D^+ 和 D^- , 并对两个子集分别按步骤(2)继续操作;

Step4: 决策树停止生长准则, 通常有两种。一是某节点对应的样本集中的样本数小于一定的数量时当前节点成为叶节点; 二是某节点对应样本集中得到的最大信息增益小于一定的阈值, 则当前节点成为叶节点 D_{leaf} ;

Step5: 当所有节点均无法继续分割为子节点时, 则决策树分类器 $h(x, D)$ 训练完毕。

采用决策树对未知样本 x 进行分类决策的过程从根节点开始, 按照每个非叶结点对应的最优分割参数, 选择样本对应的输出分支, 直到进入叶节点, 该未知样本的分类标签由叶结点对应的样本集 D_{leaf} 确定, 叶节点样本集中大多数样本均属标签 j , 则寻历至该节点的未知样本也被认为属于标签 j 。如式(10)所示, 其中 p_j 表示 D_{leaf} 样本集中属于标签 j 的样本所占的比例。

$$y = \{y^{(j)} | p_j > 0.5, 1 \leq j \leq L\} \quad (10)$$

3 多标签随机森林算法

随机森林算法基于集成学习从相同的数据集中构建一组存在细微差异的子决策树, 分类时所有决策树投票表决, 以降低单个模型的随机波动所产生的影响^[17-18]。为了使不同的子决策树具有一定的多样性, 随机森林使用 Bootstrap 自助法对原始样本集进行有放回均匀随机抽样获得自助样本子集, 然后使用子空间采样根据特征的不同子集构造不同的子决策树。通过引入上述两个随机采样过程使得生成的集成模型的多样性被显著增强, 且每个子决策树之间是相互独立的, 还可以减小每棵树的训练时间。通过采用支持多标签数据集的 ML-DT, 本文构造了多标签随机森林算法 ML-RF 用于复合电能质量扰动分类。

3.1 生成随机森林

构造随机森林的过程就是生成大量子决策树的过程, 可分为以下 4 步。

Step1: 对原始训练样本集 D 进行放回均匀抽样

方法随机生成自助样本集,通常自助样本集中会含有重复样本,同时会遗漏部分原始数据,这充分保证各个子决策树的差异性,使得算法不容易出现过拟合。本文中生成的自助样本集为 D_{boot} ,其样本数与 D 中样本数相同,均为 m 。

Step2: 对 D_{boot} 中的样本特征进行子空间采样。从 q 个原始特征中随机选取 $q_s (q_s \geq L)$ 个特征构成新的样本特征,得到用以训练子决策树的 D_T 样本集。

Step3: 利用 D_T 样本集采用完全分裂方式建立子决策树 $h(x, D_T)$ 。完全分裂方式的停止生长准则有两条,一是样本集中的样本数为 1,无法继续分裂;二是样本集的熵为 0,所有样本均指向同一个标签。第一条准则是第二条准则的特例,即样本集中仅含一个样本。

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

3.2 随机森林分类

设随机森林为 $H(x, D, n_{\text{RF}}, m, q_s) = \{h_i(x, D_{T_i}) | i=1, 2, \dots, n_{\text{RF}}\}$,采用随机森林对多标签电能质量扰动样本 x 进行分类的过程,即依次将未知样本的特征向量按照子决策树对应的决策特征进行提取,再使用子决策树进行分类得到结果 $h_i(x, D_{T_i}) (1 \leq i \leq n_{\text{RF}})$,其中 $h_i(\cdot) = \{h_{i_1}, h_{i_2}, \dots, h_{i_L}\}$, $h_{i_j} = 1$ 表示该分类器认为样本 x 属于标签 j ,反之 $h_{i_j} = 0$ 。最终的分类结果由子决策树的结果投票决定,即少数服从多数。如

式(11)所示。

$$y = \{y^{(j)} | p_{H_j} > 0.5, 1 \leq j \leq L\} \quad (11)$$

式(11)中的 p_{H_j} 表示将样本分类为标签 j 的子决策树占决策森林的比例,如式(12)。

$$p_{H_j} = \frac{1}{n_{\text{RF}}} \sum_{i=1}^{n_{\text{RF}}} h_{i_j}(x, D_{T_i}) \quad (12)$$

通常在生成决策树后还会对决策树进行剪枝,但是随机森林中的子树通常是不需要剪枝的,因为通过引入 Bootstrap 和子空间采样已经保证集成模型的随机性,较好地避免过拟合问题,因此不需要进行剪枝。

4 仿真分析

4.1 电能质量扰动模型

尽管电能质量扰动情况复杂,其基本的扰动信号主要有:电压暂升,电压中断,电压暂降,谐波,暂态振荡,瞬态冲击,闪变等。在 Matlab2015 环境中,参考文献[19-20],建立标准信号与电能质量基本扰动信号模型,如表 1 所示。基波频率 f_0 为 50 Hz,采样频率 f_s 为 12.8 kHz, T 为工频周期, $u(t)$ 为单位阶跃函数。为适应不同幅值的信号,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么值。谐波信号主要考虑(含)13 次以下的谐波(650 Hz),暂态振荡频率范围为 700~1500 Hz。

表 1 电能质量标准信号与扰动信号模型

Table 1 Power quality standard signal and disturbance signal model

信号类型	信号模型	参数
标准信号(C ₀)	$v(t) = \sin(\omega t)$	$\omega = 2\pi f_0$
电压暂升(C ₁)	$v(t) = \{1 + k[u(t_2) - u(t_1)]\} \sin(\omega t)$	$0.1 \leq k \leq 0.9, 0.5T \leq t_2 - t_1 \leq 30T$
电压中断(C ₂)	$v(t) = \{1 - k[u(t_2) - u(t_1)]\} \sin(\omega t)$	$0.9 < k \leq 1, 0.5T \leq t_2 - t_1 \leq 30T$
电压暂降(C ₃)	$v(t) = \{1 - k[u(t_2) - u(t_1)]\} \sin(\omega t)$	$0.1 \leq k \leq 0.9, 0.5T \leq t_2 - t_1 \leq 30T$
谐波(C ₄)	$v(t) = \sin(\omega t) + \sum_{k=2}^{13} a_k \sin(k\omega t)$	$0.01 < a_k < 0.2$
暂态振荡(C ₅)	$v(t) = \sin(\omega t) + a e^{-\lambda(t-t_1)} \sin(\beta\omega t) u(t-t_1)$	$0.05 < a < 1, 15 < \lambda < 130, 14 \leq \beta < 30$
瞬态冲击(C ₆)	$v(t) = \sin(\omega t) + a[u(t_1) - u(t_2)]$	$a > 0.8, 1 \text{ ms} < t_2 - t_1 < 3 \text{ ms}$,
电压闪变(C ₇)	$v(t) = [1 + a \sin(\beta\omega t)] \sin(\omega t)$	$0.05 < a < 0.2, 0.1 < \beta < 0.5$

4.2 特征提取

不同扰动信号在时频分布上的差异主要体现在在时频平面上不同时频段能量分布的差异,各时频区能量分布的均匀性差异反映扰动分量状态的差别。基于平稳小波分析建立的改进小波能量熵测度可以对时频平面上能量分布的均匀性进行定量描述,并对扰动分量的不确定性和复杂度加以定量度量,所以对电能质量扰动具有较好的特征提取和分

类能力。常用的小波基有 harr 小波、dbN 小波、coifN 小波和 symN 小波等,由于 coifN 小波相对其他小波基具有更长的支撑长度,更有利于信号能量的集中和各层小波能量熵特征的提取,然而过大的支撑长度会增加计算时间。信号的离散小波变换实质上是对信号进行多尺度分析,综合考虑信号采样频率、小波基性质和不同类型扰动所在频段的差异这些因素,本文采用 coif4 小波对扰动信号进行 6 层小波

分解, 每层分解系数对应着不同频段, 如图 2 所示。分别计算 $d_1 \sim d_6$ 各层系数的改进香农熵 $SE_{d1}(k) \sim SE_{d6}(k)$, 通过大量实验和特征规律的总结, 提取下述特征量作为电能质量扰动分类的原始特征量:

(1) $F_1 \sim F_6$ 为各层平稳小波变换高频系数的改进小波香农熵 $SE_{d1}(k)$ 的最大值;

(2) $F_7 \sim F_{12}$ 为各层平稳小波变换高频系数的改进小波香农熵 $SE_{d1}(k)$ 的最小值;

(3) $F_{13} \sim F_{18}$ 为各层平稳小波变换高频系数的改进小波香农熵 $SE_{d1}(k)$ 的平均值;

(4) $F_{19} \sim F_{24}$ 为各层平稳小波变换高频系数的改进小波香农熵 $SE_{d1}(k)$ 的标准差;

(5) $F_{25} \sim F_{28}$ 依次为第 6 层平稳小波变换低频系数的改进小波香农熵的 $SE_{c6}(k)$ 的最大值、最小值、平均值和标准差;

(6) 在分解到一定尺度时, 可以将第 6 层低频分量近似地认为是基波信号, 采用半个周期的滑动有效值计算方法计算第 6 层低频分量的有效值曲线, 如式(13)所示, 提取其中的最大值和最小值作为特征量 F_{29} 和 F_{30} 。

$$RMS_{c6}(k) = \sqrt{\frac{1}{128} \sum_{i=k-128}^k C_6(k)^2}, \quad k > 128 \quad (13)$$

提取样本中的 $F_1 \sim F_{30}$ 形成原始特征集合 $\mathbf{F} = \{F_1, F_2, F_3, \dots, F_{30}\}$ 。

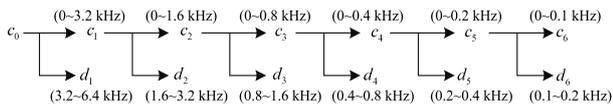


图 2 信号的平稳小波分解示意图

Fig. 2 Stationary wavelet decomposition of signals

4.3 分类效果仿真验证

本文参考文献[4], 考虑的电能质量扰动信号为表 1 中的 7 种单一扰动信号以及由它们复合而成的复合扰动信号, 其中双重复合扰动 11 种, 三重复合扰动 12 种, 四重复合扰动 7 种, 共 37 种不同类型的电能质量扰动信号。采用 Matlab 生成上述扰动信号, 每类扰动生成 200 个随机样本, 共 7400 个样本。每类扰动分别叠加信噪比(SNR)为 30 dB、40 dB 和 60 dB 的高斯白噪声, 选取 60% 的样本构成训练集, 其余的构成测试集。

首先分析随机森林算法参数对分类性能的影响, 考察采用不同算法参数的 ML-RF 分类器对 40 dB 下三重扰动数据的分类准确率, 其中 RF 规模 n_{RF} 从 1 棵到 100 棵, 步长 1 棵; RF 子空间大小 q_s 从 1 到 30, 步长为 1; Booststrap 样本集与原始训练集大小相同, 结果如图 3 所示。

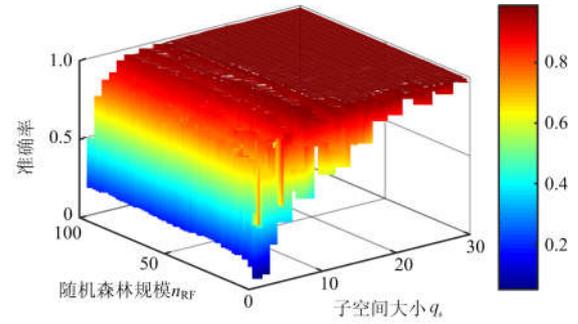


图 3 不同参数下的 ML-RF 分类准确率

Fig. 3 Classification accuracy of ML-RF with different parameters

由图 4 和图 5 可知, 当子空间大小 q_s 分别取不同的定值时, 随着森林规模 n_{RF} 的增长, 分类准确率均会提高, 当 n_{RF} 增长到一定值后继续增大随机森林的规模对准确率提高不再有显著作用, 这说明盲目选择过大的森林规模可能导致浪费计算资源; 当森林规模 n_{RF} 分别取不同的定值时, 随着子空间大小 q_s 的增大, 分类准确率会提高, 当 q_s 增大到一定值后继续增大子空间大小不会使准确率继续提高, 因此在使用随机森林算法时应当选择合适的 q_s , 过小可能导致准确率难以提高, 过大会导致子决策树的差异性减小, 同时增加计算量在后面的仿真中, 随机森林的规模 n_{RF} 取 60, 子空间大小 q_s 取 15。

为了进一步比较本文方法与其他多标签分类方法的性能, 在不同的噪声强度下, 分别采用 Binary-SVM, ML-RBF 和 ML-KNN 3 种多标签分类法(简记为 B-S, M-R 和 M-K)对电能质量扰动信号进行分类, 并将其分类结果与本文提出的 ML-RF(简记为 M-F)方法进行对比, 实验结果如表 2 所示。

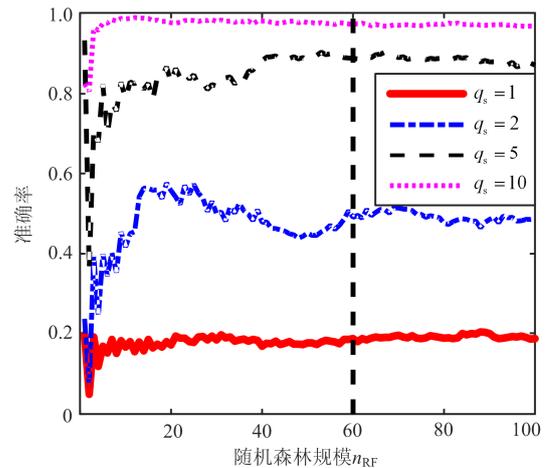
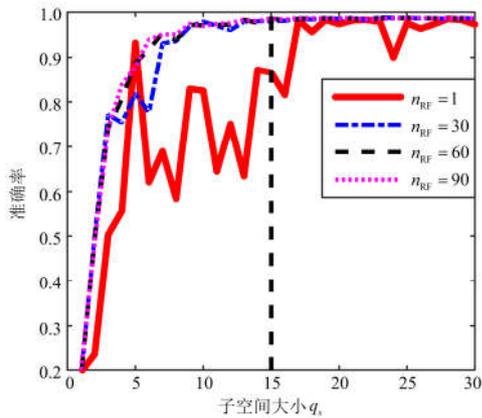


图 4 不同 n_{RF} 下的 ML-RF 分类准确率

Fig. 4 Classification accuracy of ML-RF with different n_{RF}

图 5 不同 q_s 下的 ML-RF 分类准确率Fig. 5 Classification accuracy of ML-RF with different q_s .

由表 2 可以看出, ML-RF 算法在不同的噪声强

表 2 不同噪声条件下不同分类法的准确率

Table 2 Performance comparison of binary-SVM, ML-RBF, ML-KNN and ML-RF under different SNR values

SNR/dB	50				40				30			
分类器	B-S	M-R	M-K	M-F	B-S	M-R	M-K	M-F	B-S	M-R	M-K	M-F
单重扰动	0.955	0.959	0.968	0.996	0.964	0.966	0.977	0.993	0.966	0.968	0.977	0.991
二重扰动	0.866	0.927	0.818	0.997	0.890	0.922	0.849	0.997	0.889	0.925	0.856	0.995
三重扰动	0.814	0.827	0.610	0.984	0.821	0.844	0.596	0.988	0.813	0.834	0.590	0.980
四重扰动	0.759	0.830	0.634	0.968	0.804	0.838	0.630	0.980	0.800	0.843	0.643	0.989
总体	0.848	0.886	0.758	0.986	0.870	0.892	0.763	0.989	0.867	0.893	0.766	0.989

5 结论

为了更好地对复合电能质量扰动进行有效分类, 本文对单标签决策树进行推广得到支持多标签数据集的 ML-DT 算法, 以此为基础构建随机森林分类器 ML-RF, 运用该算法实现复合电能质量扰动分类。ML-RF 充分考虑多标签数据的特性, 在充分学习复合扰动分类特征的前提下, 较好地避免过拟合问题。在与现有其他分类方法 Binary-SVM, ML-RBF 和 ML-KNN 的对比中证明了该方法的有效性, 且 ML-RF 具有良好的抗噪性能。

参考文献

- [1] 肖助力, 龚仁喜, 陈双, 等. 基于改进 S 变换的电力系统谐波检测方法[J]. 电力系统保护与控制, 2015, 43(3): 84-90.
XIAO Zhuli, GONG Renxi, CHEN Shuang, et al. Detection of harmonics in power system based on modified S-transform[J]. Power System Protection and Control, 2015, 43(3): 84-90.
- [2] 何连杰, 史常凯, 闫卓, 等. 基于广义 S 变换能量相对熵的小电流接地系统故障区段定位方法[J]. 电工技术学报, 2017, 32(8): 274-280.

度和扰动复杂程度下的分类准确率均优于其他 3 种分类算法。信噪比 SNR 从 50 dB 减小至 30 dB, ML-RF 的分类准确率只有 0.5% 的降低, 因此 ML-RF 具有较好的噪声鲁棒性。ML-RF 适于进行电能质量多标签分类, 随着扰动复杂度的增加, 其分类准确率不会明显降低, 至少有 98.0% 的准确率。相较之下, Binary-SVM, ML-RBF 和 ML-KNN 对单重扰动分类效果较好, 但对多重扰动的分类准确率显著低于单重扰动的准确率。ML-RF 能通过训练子决策树充分学习样本训练集的分类特征, 同时 Bootstrap 采样和子空间采样技术既使得每棵子决策树具有自己擅长的分类领域, 同时避免随机森林出现过学习, 因此 ML-RF 分类方法具有较好的泛化能力, 适用于复合电能质量扰动的多标签分类。

- [3] 范小龙, 谢维成, 蒋文波, 等. 一种平稳小波变换改进阈值函数的电能质量扰动信号去噪方法[J]. 电工技术学报, 2016, 31(14): 219-226.
FAN Xiaolong, XIE Weicheng, JIANG Wenbo, et al. An improved threshold function method for power quality disturbance signal de-noising based on stationary wavelet transform[J]. Transactions of China Electrotechnical Society, 2016, 31(14): 219-226.
- [4] 徐健, 张语勃, 李彦斌, 等. 短时傅里叶变换和 S 变换用于检测电压暂降的对比研究[J]. 电力系统保护与控制, 2014, 42(24): 44-48.
XU Jian, ZHANG Yuqing, LI Yanbin, et al. Comparative study of STFT and S transform on detecting voltage sag[J]. Power System Protection and Control, 2014, 42(24): 44-48.
- [5] 张杨, 刘志刚. 一种基于时频域多特征量的电能质量混合扰动分类新方法[J]. 中国电机工程学报, 2012, 32(34): 83-90.

- ZHANG Yang, LIU Zhigang. A new method for power quality mixed disturbance classification based on time-frequency domain multiple features[J]. Proceedings of the CSEE, 2012, 32(34): 83-90.
- [6] HSU C W, LIN C J. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415-425.
- [7] 管春, 周维维, 卢伟国. 基于多标签 RBF 神经网络的电能质量复合扰动分类方法[J]. 电工技术学报, 2011, 26(8): 198-204.
- GUAN Chun, ZHOU Luowei, LU Weiguo. Recognition of multiple power quality disturbances using multi-label RBF neural networks[J]. Transactions of China Electrotechnical Society, 2011, 26(8): 198-204.
- [8] ZHANG M L, ZHOU Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [9] 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.
- [10] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [11] 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 Protection and Control, 2016, 44(9): 122-128.
- [12] SHENSA M J. Generalized and parallelized a Troun and Mallat algorithms to design non-uniform filter-banks[J]. IEEE Transactions on Signal Processing, 1992, 40(10): 2464-2482.
- [13] 何正友, 蔡玉梅, 钱清泉. 小波熵理论及其在电力系统故障检测中的应用研究[J]. 中国电机工程学报, 2005, 25(5): 38-43.
- HE Zhengyou, CAI Yumei, QIAN Qingquan. A study of wavelet entropy theory and its application in electric power system fault detection[J]. Proceedings of the CSEE, 2005, 25(5): 38-43.
- [14] 郭凤仪, 李坤, 陈昌昱, 等. 基于小波近似熵的串联电弧故障识别方法[J]. 电工技术学报, 2016, 31(24): 164-172.
- GUO Fengyi, LI Kun, CHEN Changken, et al. Series arc fault identification method based on wavelet approximate entropy[J]. Transactions of China Electrotechnical Society, 2016, 31(24): 164-172.
- [15] QUINLAN J R. Induction on decision tree[J]. Machine Learning, 1986, 1(1): 81-106.
- [16] CLARE A, KING R D. Knowledge discovery in multi-label phenotype data[C] // Principles of Data Mining and Knowledge Discovery, Freiburg, Germany, September 3-5, 2001: 42-53.
- [17] 曹正凤. 随机森林算法优化研究[D]. 北京: 首都经济贸易大学, 2014.
- CAO Zhengfeng. Study on optimization of random forests algorithm[D]. Beijing: Capital University of Economics and Business, 2014.
- [18] 陈伟根, 龙震泽, 谢波, 等. 不同气隙尺寸的油纸绝缘气隙放电特征及发展阶段识别[J]. 电工技术学报, 2016, 31(10): 49-58.
- CHEN Weigen, LONG Zhenze, XIE Bo, et al. Characteristics and development stage recognition of air-gap discharge within oil-paper insulation considering effect of cavity size[J]. Transactions of China Electrotechnical Society, 2016, 31(10): 49-58.
- [19] 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.
- [20] 邓志勇, 彭晓亮, 陈俊, 等. ITD 算法在电能质量扰动信号特征量提取中的应用[J]. 电力系统保护与控制, 2016, 44(7): 32-38.
- DENG Zhiyong, PENG Xiaoliang, CHEN Jun, et al. Application of feature extraction of power quality disturbance signal using ITD[J]. Power System Protection and Control, 2016, 44(7): 32-38.

收稿日期: 2016-06-17; 修回日期: 2016-12-12

作者简介:

瞿合祚(1993—), 男, 硕士研究生, 研究方向为电力系统运行与控制; E-mail: 806475405@qq.com

刘恒(1980—), 男, 硕士研究生, 高级工程师, 研究方向为电力系统运行与控制; E-mail: 9853214@qq.com

李晓明(1955—), 男, 教授, 博士生导师, 主要研究方向为电力系统运行与控制、电能质量、智能电网等。E-mail: xmlm@whu.edu.cn

(编辑 周金梅)