

DOI: 10.19783/j.cnki.pspc.211199

# 基于多粒度特征选择和模型融合的复合电能质量扰动 分类特征优化

阮梓航, 肖先勇, 胡文曦, 郑子萱, 汪颖

(四川大学电气工程学院, 四川 成都 610065)

**摘要:** 现代电力系统因其“双高”特性造成电能质量扰动模式愈加复杂, 对复合扰动的准确分类提出了挑战。传统电能质量扰动分类方法在特征提取阶段所提取的特征由人为确定, 难以判断所提取的特征对分类问题是否有效, 加之多重复合扰动特征相互耦合导致扰动特征的可分性确定困难。为此, 提出一种基于粒度的计算方法进行特征选择的模型。在提取的扰动特征集的基础上, 通过构建多粒度空间反映特征分布差异性, 进而挖掘各粒度下的最优特征子集以确定有效和冗余的分类特征, 达到优化分类效果的目的。在此基础上, 通过集成分类模型融合不同粒度空间最优扰动特征集所训练的同质弱分类器模型, 提出一种新的电能质量扰动多粒度集成分类方法。该方法克服了现有方法在进行多粒度分类时通过寻找最优单粒度空间特征而导致的其他粒度空间信息丢失的问题。实验表明, 多粒度特征选择算法可提取对分类有效的扰动特征, 集成分类模型可进一步改善模型的分类性能。

**关键词:** 电能质量; 复合扰动; 特征选择; 多粒度空间; 集成分类

## Multiple power quality disturbance classification feature optimization based on multi-granularity feature selection and model fusion

RUAN Zihang, XIAO Xianyong, HU Wenxi, ZHENG Zixuan, WANG Ying

(College of Electrical Engineering, Sichuan University, Chengdu 610065, China)

**Abstract:** Modern power systems with “double high” characteristics make power quality disturbance patterns more complex, and the accurate classification of multiple power quality disturbances becomes more difficult. In the feature extraction stage of traditional power quality disturbance classification methods, the extracted features are determined artificially. Thus, it is difficult to judge whether the extracted features are adequate for classification problems, and the coupling of multiple feature distribution will affect the separability of disturbance features. Therefore, this paper proposes a feature selection method based on granular computing to optimize the performance of the classification. Based on the original feature set, a multi-granularity space is constructed to reflect the difference in feature distribution. Then the optimal feature subsets under each granularity are mined to determine the effective and redundant classification features. The homogeneous base classifiers trained by optimal feature subsets corresponding to different granularity spaces are fused by the ensemble model. A new multi-granularity ensemble classification model for power quality disturbance is proposed. This method overcomes the problem of the existing techniques by searching for the optimal valuable information of a single granularity space in a multi-granularity classification, leading to other granularity spaces losing the useful information. The simulation results show that the multi-granularity feature selection algorithm can extract useful features for classification, and an integrated model can improve the classification performance of the model.

This work is supported by the National Natural Science Foundation of China (No. 51807126).

**Key words:** power quality; multiple disturbance; feature selection; multi-granularity space; integrated classification

## 0 引言

随着电网分布式电源与电力电子装置比例升高,

“双高”电力系统中电能质量扰动(Power Quality Disturbance, PQD)事件呈现频率高、模式复杂等特点<sup>[1-3]</sup>。如何准确辨识单一扰动构成的复合 PQD 事件, 是实现复杂扰动源追溯和治理的必要前提。

现有复合 PQD 分类方法主要采用机器学习方法。文献[4-5]利用基于卷积神经网络的深度学习方

基金项目: 国家自然科学基金项目资助(51807126); 中央高校基本科研业务费专项资金资助

法实现了电能质量的复合扰动分类,具有良好的学习和拟合能力,但该类方法训练时过度依赖数据导致模型计算量大、结构复杂,在训练样本数量较少的情况下学习能力不佳,而通过长期监测记录到的复合电能质量扰动事件数量远不足以满足深度学习的数据体量要求。其他机器学习方法在多数情况下可保证较好的学习和拟合能力,且在电能质量扰动分类领域广泛应用。例如,文献[6]设计了一种自适应滤波器来提取失真信号频域中的单频分量特征,改进了经验小波变换的特征提取效果;文献[7-9]分别利用变分分解、变频窗函数以及布莱克曼窗使 S 变换获得了更好的时频域分析能力。这些方法大多未考虑所提取的特征本身对分类性能的影响,冗余的特征会大量占用计算机资源并降低模型的性能,因此需要对提取的特征进行特征选择,现在针对 PQD 扰动的特征选择的方法已有研究<sup>[10-11]</sup>,但对多标签数据集不同样本间特征分布的差异性以及类域的交叉重叠等模糊性、不确定性以及不完整问题对特征选择效果的影响考虑较少。因此,在特征选择阶段利用粒度计算的原理通过将类域转换为不同的知识粒度,并从不同知识粒度的角度观察和分析同一问题的影响,可了解不同粒度空间下特征对扰动类别可分性的不同贡献程度,以此作为判别特征好坏的依据<sup>[12-16]</sup>。

粒度计算所面临的多粒度空间特征最优子集挖掘决定了后续分类器设计同样要求在多个粒度空间中构建,传统的基于粒度计算的分类方法<sup>[15]</sup>大多是基于最优粒度空间的选择,但该类方法会丢失其他粒度空间的有用信息,因此需要对每个粒度空间进行融合。在现有模型融合方法中, Bagging 模型采用投票或取平均值的方式集成多个同质弱学习器的分类结果,可降低分类模型的方差水平,提高鲁棒性<sup>[17]</sup>; Boosting 模型通常对多个同质弱学习器进行顺序拟合,着重于对分类性能较弱的基分类模型进行加权组合,可降低分类模型的偏置水平,提升准确率<sup>[18]</sup>,但面临过拟合的问题。Stacking 集成分类模型可突出分类性能优秀的基分类器模型,并采用非线性组合优化的方法进一步降低偏置水平,提升模型的泛化能力,使其具有更好的模型融合效果<sup>[19]</sup>,但传统 Stacking 模型仅在相同特征维度下基于异质弱分类器组合优化以实现集成分类<sup>[20]</sup>,不适用于本文多粒度空间中样本特征维度不同的情况下的组合优化。

为解决上述问题,本文提出一种基于多粒度空间的特征选择算法。该方法通过构建不同粒度特征

空间反映特征分布的差异性,进而为特征选择提供决策依据,得到各粒度空间下的最优扰动特征集,实现特征选择。在此基础上,考虑融合各粒度空间各自的分类模型,提出电能质量扰动多粒度集成分类方法。该模型适用于在训练样本较小的情况下,利用基于粒度计算的特征选择算法和模型融合方法对机器学习模型进行分类效果进行优化,并通过实验验证了本文方法的准确性与可靠性。

## 1 复合电能质量扰动的特点及特征提取

PQD 扰动并不总是单一扰动,也可能出现多种单一扰动同时存在的情况,即复合扰动。受篇幅所限,本文以电网中常见的各类 PQD 扰动类型为分析对象,具体包含电压暂降(C1)、电压暂升(C2)、暂态振荡(C3)、谐波(C4)、电压中断(C5)、电压闪变(C6) 6 种单一扰动模型,7 种双重复合扰动模型和 3 种三重复合扰动模型,共 16 种单一和复合扰动模型。

由于本文方法需要对初始特征进行优化,因此在初始特征提取阶段,可不依赖专家经验,提取尽可能多的常见初始特征。本文对各类扰动信号采用 S 变换<sup>[21-22]</sup>,分别得到时域最大值( $T_{max}$ )、最小值( $T_{min}$ )、平均值( $T_{mean}$ )、标准差( $T_{std}$ )、方均根值曲线( $T_{rms}$ )和频域最大值( $F_{max}$ )、最小值( $F_{min}$ )、平均值( $F_{mean}$ )、标准差( $F_{std}$ )、方均根值曲线( $F_{rms}$ );由于 Hilbert 边际谱曲线具有更高的频域分辨率,本文采用希尔伯特黄变换<sup>[23]</sup>获取扰动信号的 Hilbert 边际谱曲线( $F_{HMS}$ )共 11 种 PQD 扰动特征曲线,再从上述 11 种特征曲线中提取出以下 7 类特征参数。

1)  $F_1$ — $F_5$  分别为时频域特征曲线的最大值、最小值、平均值、标准差和均方根值,以反映基频扰动的幅值变化程度,可有效区分为暂升、暂降、中断、电压波动等扰动类型。2)  $F_6$  为频域特征曲线频率大于 100 Hz 频段的标准差,通过描述扰动特征在高频段数据与均值的偏离程度以衡量是否发生幅值波动,可有效区分谐波和振荡扰动类型。3)  $F_7$  为总谐波畸变率,即谐波与基波均方根值之比,对谐波扰动类型具有较高的区分度。

综上,本文构建 60 维时频域特征作为初始特征集合,如表 1 所示。需要说明的是,该集合仅是根据常见特征构造而来,实际应用中可提取更多特征,且冗余特征会在优化过程中被删除而不会造成分类性能下降。由于本文主要针对特征优化,特征提取方法不是本文的研究工作,其详细计算流程不再赘述,各参数的计算过程见相关文献<sup>[1,18,24-25]</sup>。

表 1 PQD 时频域特征

Table 1 PQD time-frequency domain features

参数	$F_1$	$F_2$	$F_3$	$F_4$	$F_5$	$F_6$	$F_7$
$T_{\max}$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	—	—
$T_{\min}$	$f_6$	$f_7$	$f_8$	$f_9$	$f_{10}$	—	—
$T_{\text{mean}}$	$f_{11}$	$f_{12}$	—	$f_{13}$	$f_{14}$	—	—
$T_{\text{std}}$	$f_{15}$	$f_{16}$	$f_{17}$	—	$f_{18}$	—	—
$T_{\text{rms}}$	$f_{19}$	$f_{20}$	$f_{21}$	$f_{22}$	—	—	—
$F_{\max}$	$f_{23}$	$f_{24}$	$f_{25}$	$f_{26}$	$f_{27}$	$f_{28}$	$f_{29}$
$F_{\min}$	$f_{30}$	$f_{31}$	$f_{32}$	$f_{33}$	$f_{34}$	$f_{35}$	$f_{36}$
$F_{\text{mean}}$	$f_{37}$	$f_{38}$	—	$f_{39}$	$f_{40}$	$f_{41}$	$f_{42}$
$F_{\text{std}}$	$f_{43}$	$f_{44}$	$f_{45}$	—	$f_{46}$	—	$f_{47}$
$F_{\text{rms}}$	$f_{48}$	$f_{49}$	$f_{50}$	$f_{51}$	—	$f_{52}$	$f_{53}$
$F_{\text{HMS}}$	$f_{54}$	$f_{55}$	$f_{56}$	$f_{57}$	$f_{58}$	$f_{59}$	$f_{60}$

## 2 多粒度特征空间下多标签特征选择方法

### 2.1 变精度模糊邻域粗糙集的基本原理

针对复合 PQD 的高维特征集合, 本节基于粗糙集的属性约简的粒度计算方法实现多粒度特征空间下的特征优化。一个多标签决策系统(Multi-Label Decision System, MLDS)如式(1)所示。

$$MLDS = \langle U, C \cup D \rangle \quad (1)$$

式中:  $U$  为非空样本集合,  $U = \{x_1, x_2, \dots, x_n\}$ ,  $n$  为样本数;  $C$  为条件属性集合,  $C = \{C_1, C_2, \dots, C_p\}$ ,  $p$  为特征标记序号;  $D$  为标签集合,  $D = \{d_1, d_2, \dots, d_m\}$ ,  $m$  为扰动标签序号。

为构建多标签决策系统, 首先引入模糊相似关系度量样本间的相似程度。样本  $x_i$  和  $x_j$  之间的模糊相似度定义, 如式(2)所示。

$$r_{ij} = \begin{cases} 1 - |x_i - x_j|, & |x_i - x_j| \leq 1 - \delta \\ 0, & |x_i - x_j| > 1 - \delta \end{cases} \quad (2)$$

式中:  $0 \leq r_{ij} \leq 1$ ;  $|x_i - x_j|$  为样本间距离, 可采用欧氏距离计算<sup>[13]</sup>。两两样本间的模糊相似度  $r_{ij}$  可构成一个  $n \times n$  维的模糊相似关系矩阵  $R_C$ , 该矩阵表征了条件属性集合  $C$  下的样本间的相似程度,  $r_{ij}$  越大表示样本  $x_i$  与样本  $x_j$  之间的关联程度越高。

其次需要根据样本间的相似关系, 挖掘出特征与分类标签之间的关联规则。为克服传统粗糙集模型无法评价特征重要程度的缺陷, 引入邻域粗糙集的概念, 将决策属性与条件属性之间的等价关系转换为基于样本间距离的邻域关系, 实现对特征重要度的描述。基于样本间距离的邻域关系由模糊相似关系矩阵  $R_C$  定义, 对于所有  $B \in C$ , 模糊邻域半径  $\delta$  控制下的样本间的模糊相似关系如式(3)所示。

$$[x_i]_B^\delta(x_j) = \begin{cases} 0, & R_B(x_i, x_j) < \delta \\ R_C(x_i, x_j), & R_B(x_i, x_j) \geq \delta \end{cases} \quad (3)$$

模糊邻域粗糙集模型为多标签扰动的特征选择提供了一种有效的方法, 但因乐观粗糙集模型近似集决策条件过于宽松, 悲观粗糙集模型近似集决策条件过于严格, 导致对含噪声数据的特征选择和分类效果不佳。为防止因为决策条件过于严格或宽松而导致的数据缺失和误分类问题, 通过引入变精度系数  $\alpha$ , 构建变精度模糊邻域粗糙集(Variable Precision Fuzzy Neighborhood Rough Set, VPFNRS)模型, 通过降低近似集决策条件的要求, 进而从数据集中提取出更多的有用信息。各标签所对应决策空间中 VPFNRS 的下近似集  $R_B^{\delta, \alpha}(\tilde{D}_i)$  和上近似集  $\bar{R}_B^{\delta, \beta}(\tilde{D}_i)$  分别如式(4)和式(5)所示。

$$R_B^{\delta, \alpha}(\tilde{D}_i) = \{x \in U_i \mid I([x]_B^\delta, \tilde{D}_i) \geq \alpha\}, 0.5 \leq \alpha \leq 1 \quad (4)$$

$$\bar{R}_B^{\delta, \beta}(\tilde{D}_i) = \{x \in U_i \mid I([x]_B^\delta, \tilde{D}_i) > \beta\}, 0 \leq \beta < 0.5 \quad (5)$$

式中:  $i = \{1, 2\}$ ;  $\tilde{D}_1$  和  $\tilde{D}_2$  为样本  $x_i$  在多标签空间中相对于某决策类的隶属度;  $U_1$  和  $U_2$  为多标签决策系统中的两类互斥决策类样本集合, 分别如式(6)一式(9)所示。

$$\tilde{D}_1 = \bigvee_{j=1}^m \frac{|[x_i]_C \cap U_1^{d_j}|}{|[x_i]_C|} \quad (6)$$

$$\tilde{D}_2 = \bigwedge_{j=1}^m \frac{|[x_i]_C \cap U_2^{d_j}|}{|[x_i]_C|} \quad (7)$$

$$U_1 = \bigcup_{j=1}^m U_1^{d_j} \quad (8)$$

$$U_2 = \bigcup_{j=1}^m U_2^{d_j} - U_1 \quad (9)$$

式中:  $m$  是标签集合的维数;  $U_1^{d_j}$  和  $U_2^{d_j}$  是标签  $d_j$  定义的两类互斥决策类样本集合;  $[x_i]_C$  是模糊相似关系矩阵  $R_C$  中由样本  $x_i$  产生的模糊相似关系。由以上可知, 变精度模糊邻域粗糙集的粒度空间大小是由邻域半径  $\delta$  和变精度系数  $\alpha$  共同控制的。

根据式(4)一式(9)可构建出多标签决策空间下的 VPFNRS 模型, 其下近似集  $R_B^{\delta, \alpha}(D)$  和上近似集  $\bar{R}_B^{\delta, \beta}(D)$  分别如式(10)和式(11)所示。

$$R_B^{\delta, \alpha}(D) = \{R_B^{\delta, \alpha}(\tilde{D}_1), R_B^{\delta, \alpha}(\tilde{D}_2)\} \quad (10)$$

$$\bar{R}_B^{\delta, \beta}(D) = \{\bar{R}_B^{\delta, \beta}(\tilde{D}_1), \bar{R}_B^{\delta, \beta}(\tilde{D}_2)\} \quad (11)$$

对于所有  $B \in C$ , 标签集合  $D$  对条件属性集合  $B$  的依赖度的计算公式如式(12)所示。

$$\gamma_B^{\delta, \alpha}(D) = \frac{|\text{POS}_B^{\delta, \alpha}(D)|}{|U|} = \frac{|\bigcup_{i=1}^r R_B^{\delta, \alpha}(\tilde{D}_i)|}{|U|} \quad (12)$$

式中,  $0 \leq \gamma_B(D) \leq 1$ , 依赖度  $\gamma_B(D)$  越大, 表明条件属性集  $B$  能够更好地表征标签特征。对于所有  $c \in C - B$ , 特征  $c$  对标签  $D$  的重要度计算公式为

$$imp(c, B, D) = \gamma_{B \cup c}^{\delta, \alpha}(D) - \gamma_B^{\delta, \alpha}(D) \quad (13)$$

从而将重要度大于某一阈值的特征作为该粒度空间下的按特征重要度排序的最优特征子集。

### 2.2 考虑多粒度特征空间分布差异的特征选择算法

针对复合 PQD 扰动的多标签分类问题, 基于 VPFNRS 特征选择算法, 利用不同的模糊邻域半径  $\delta_r \in [\delta_1, \delta_R]$  且  $1 \leq r \leq R$  和变精度系数  $\alpha_s \in [\alpha_1, \alpha_S]$  且  $1 \leq s \leq S$  构建粒度不同的  $R \times S$  个粒度空间, 在各粒度空间下选择出对应的最优扰动特征子集, 算法的主要流程如下:

- 1) 对  $\delta_r \in [\delta_1, \delta_R]$  以  $\Delta\delta$  为步长取值;
- 2) 对  $\alpha_s \in [\alpha_1, \alpha_S]$  以  $\Delta\alpha$  为步长取值, 基于当前取值构建一个粒度空间  $(\delta_r, \alpha_s)$ ;
- 3) 对步骤2)构造的粒度空间  $(\delta_r, \alpha_s)$ , 计算样本间的模糊相似关系矩阵  $R_C$ , 置  $Pool = \Phi$ ;

4) 取一扰动特征  $c \in C - Pool$ ,  $Pool = Pool \cup c$ ;  
 5) 根据式(3)计算该粒度空间下的模糊相似关系  $R_{Pool}$ ;

6) 根据式(4)一式(12)计算多标签扰动特征集  $Pool$  相对于扰动标签的依赖度;

7) 根据式(13)计算重要度  $\gamma_{Pool}(D)imp(c, Pool, D)$ , 将满足  $imp(c, Pool, D) > 0$  所对应的属性  $c$  按重要度降序加入  $opt^{(\delta, \alpha)}$  集中;

- 8) 如果  $C - Pool \neq \Phi$ , 回到步骤 4);
- 9) 若  $\delta_r \leq \delta_R$  且  $\alpha_s < \alpha_S$ , 则回到步骤 2); 若  $\delta_r < \delta_R$  且  $\alpha_s = \alpha_S$ , 则回到步骤 1); 如果  $\delta_r = \delta_R$  且  $\alpha_s = \alpha_S$ , 则停止。

## 3 基于模型融合的复合电能质量扰动集成分类

### 3.1 Stacking 集成分类的基本原理

由于本文的特征选择算法在不同粒度空间下生成的最优特征子集仅对某些 PQD 扰动类型具有较好的表征能力, 为了获取对全样本空间的 PQD 扰动的识别能力, 采用基于同质弱分类器不同特征维度的 Stacking 集成分类模型。模型融合是通过特定方式将多个不同的基分类器模型进行组合, 得到多模型的综合分类系统, 以提升分类模型的性能, Stacking 模型融合是一种分层的模型集成策略, 既能有效提高模型分类性能, 又不会导致模型过拟合。算法流程如图 1 所示。

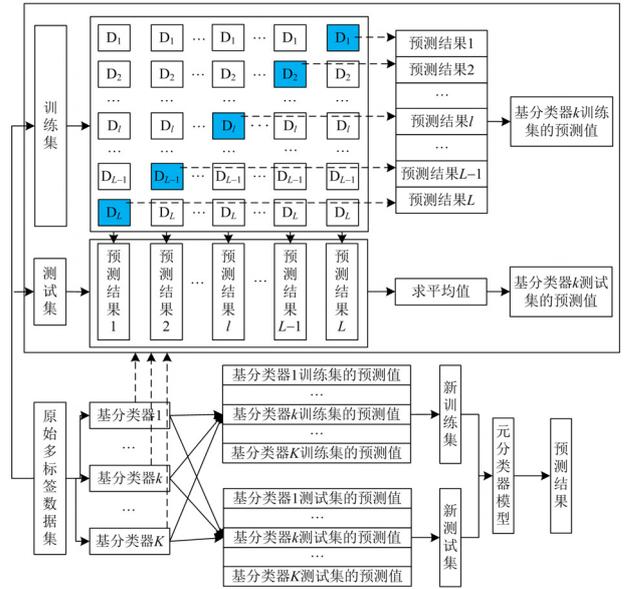


图 1 Stacking 集成分类器模型  
 Fig. 1 Stacking ensemble classifier model

1) 对所有同质基分类器进行  $L$  折交叉验证, 将训练集划分为  $L$  份, 分别利用其中  $L$  种不同组合的  $L-1$  份训练集对基分类器  $k$  进行训练, 再利用与训练集对应的第  $l$  折训练集  $D_l$  以及测试集进行测试, 得到各基分类器模型  $k$  的第  $l$  折训练集  $D_l$  与测试集的预测结果。

2) 将各基分类器模型  $k$  的第  $l$  折训练集  $D_l$  的预测结果组合作为元分类器模型的新训练集输入, 将各基分类器模型  $k$  的测试集的预测结果相加并求平均值作为元分类器模型的新测试集输入, 得到最终融合分类器模型, 并用于分类测试。

不同的基分类器模型具有不同的误差特性, Stacking 模型融合的目的旨在弱化单个基分类器模型的误差特性对整个集成分类模型分类性能的影响, 以提高多模型融合分类系统的分类性能。

### 3.2 基于模型融合的复合电能质量扰动分类辨识

XGboost 是一种基于 Boosting 思想的集成树模型, 通过不断对上一轮迭代的残差进行拟合, 将多个弱分类器模型融合为一个集成的强分类器模型<sup>[26]</sup>, 相对传统方法在具有更好分类效果的同时, 也具有更强的抗过拟合能力。因此本文选择 XGboost 作为分类器模型。XGboost 第  $t$  轮迭代生成的集成树模型如式(14)所示。

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (14)$$

式中:  $t$  为树的数量;  $\hat{y}_i^{(t)}$  为第  $t$  次迭代后第  $i$  个样本的预测结果;  $\hat{y}_i^{(t-1)}$  为第  $t-1$  次迭代后第  $i$  个样本

的预测结果:  $f_t(x_i)$  为第  $t$  棵树的模型。

XGboost 模型的目标函数包含模型的损失函数和抑制模型过拟合的正则化项  $\Omega$  两部分, 目标函数的计算如式(15)所示, 式中第一部分为模型的损失函数, 第二部分为模型树的复杂度之和, 可通过控制正则化项以抑制模型过拟合, 其计算方法如式(16)所示。

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^L \Omega(f_k) \quad (15)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|W\|^2 \quad (16)$$

式中,  $\gamma$  和  $\lambda$  代表对模型的惩罚系数。

模型损失函数越小, 代表模型的分类性能越好, 以第  $t$  轮迭代生成的集成树模型为例, XGBoost 模型在保持上一轮模型不变的情况下, 利用新的树模型  $f_t(x_i)$  来拟合  $t-1$  次预测的残差, 从而提升当前模型的分类性能。

为了获取对全样本空间的 PQD 扰动的识别能力, 突出各粒度空间最优特征子集的分类优势, 本文基于 VPFNRS-Stacking 思想提出了一种三层的复合 PQD 分类器架构。首先对原始多标签数据集进行  $L$  折划分取出训练集和测试集, 经特征选择能够得到  $R \times S$  个粒度空间的最优扰动特征子集, 为确保  $L$  折交叉验证中训练集数据不出现重复训练的情况, 需要将原始多标签数据集划分为  $L = R \times S$  个大小相等的子集, 对其他基分类器模型也进行同样的操作, 最后合并每个基分类器的分类结果, 输入集成分类器模型中的第二层元分类器模型进行训练, 为在第二层集成分类器中充分挖掘各粒度空间数据结构和空间的信息, 其元分类器通常选择对不同类型数据分类性能差异较大的模型, 不同分类性能的模型有助于从不同角度观测数据, 挖掘信息。最后将不同分类器的分类结果合并, 作为集成分类器第三层元分类器模型的输入, 第三层模型为防止过拟合, 需采用泛化能力较强的分类器, 以实现多粒度集成分类。基于 VPFNRS-Stacking 的多粒度集成分类器模型如图 2 所示, 该模型主要由三层分类器构成。

集成分类模型第一层为各个粒度空间基分类器模型, 由  $L$  折交叉验证所划分的训练集  $D = \{(x, y)\}$  按各个粒度空间所选择的特征向量构建基分类器模型的训练数据集  $D_l^{\delta_r, \alpha_s} = \{(x_l^{\delta_r, \alpha_s}, y) | 1 \leq r \leq R, 1 \leq s \leq S\}$  进行训练和测试, 获取各个粒度空间下的基分类器模型; 其中  $x_l$  为第  $l$  折训练集对应的特征向量;  $y$  为第  $l$  折训练集对应的多标签扰动标签集;  $x_l^{\delta_r, \alpha_s}$  为粒度空间  $(\delta_r, \alpha_s)$  按最优扰动特征向

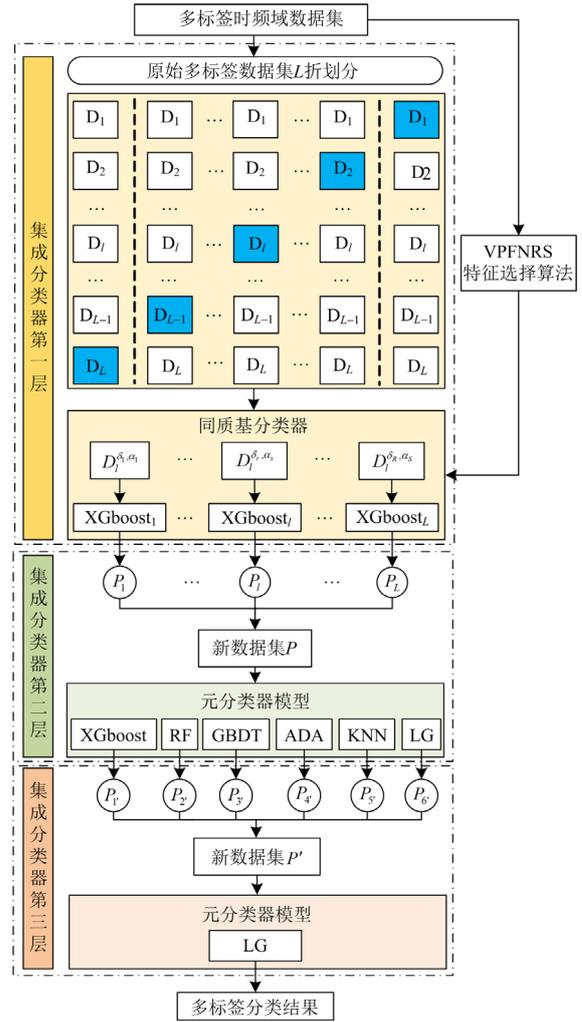


图 2 VPFNRS-Stacking 多粒度集成分类器模型

Fig. 2 VPFNRS-Stacking multi-granularity ensemble classifier model

量构建的扰动特征集;  $L = R \times S$  为选取的粒度空间总数。

集成分类模型第二层和第三层均为多粒度元分类器, 第一层各个粒度空间基分类器的预测输出为  $P_l$ , 将  $L$  个基分类器的预测输出构成新的数据集  $P = \{P_1, \dots, P_L, y\}$ , 将数据集  $P$  作为集成分类模型第二层元分类器的输入, 第二层  $M$  个分类器的预测输出分别为  $P_m$ , 又构成的新数据集为  $P' = \{P'_1, \dots, P'_M, y\}$ , 将新数据集  $P'$  作为第三层模型的输入, 进行训练和测试, 最终融合成多粒度集成分类器模型, 以达到融合各个基分类模型分类优势并提升集成分类模型分类性能的目的。

#### 4 实验分析

本文分别采用仿真生成数据与电网实测数据进

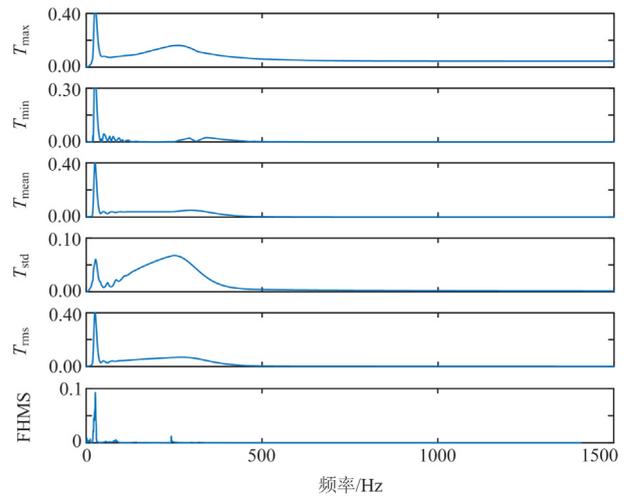
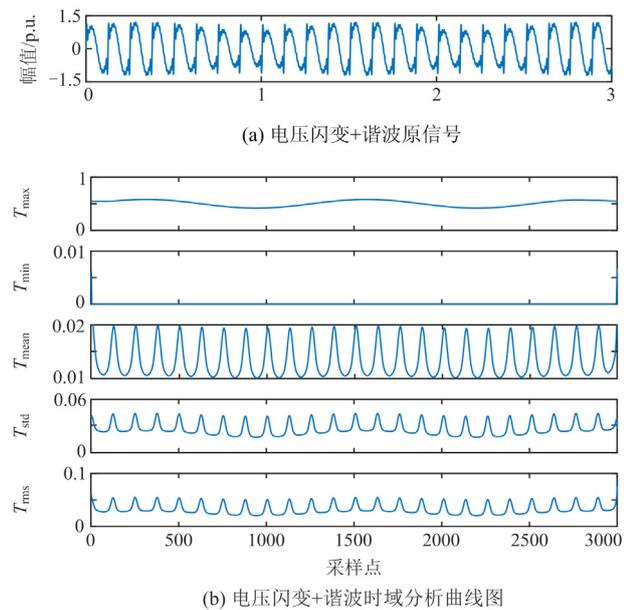
行验证, 仿真生成数据参照文献[1]的方法以及 IEEE 标准<sup>[26]</sup>, 其中包含电网中常见的 16 类单一和复合扰动模型, 扰动模型均采用归一化幅值曲线。每类扰动生成 100 个样本, 其中 50 个作为训练样本, 50 个作为测试样本。为了验证模型的抗噪性能, 对每类扰动信号添加 30 dB、40 dB、50 dB 高斯白噪声, 作抗噪性能对比实验。信号基波频率为 50 Hz, 采样频率为 6.4 kHz, 采样周期为 24 个周期(0.48 s)。电网实测数据共包括 24 个样本。本文对特征选择和分类模型均采用如下评价指标: 平均分类准确率(Average Precision, AP)、汉明损失(Hamming Loss, HL)、一类错误(One-error, OE)、覆盖率(Coverage, C)、排位损失(Ranking Loss, RL)和召回率(Recall, R), 其中模型的汉明损失、一类错误、覆盖率、排位损失越小说明模型的分类能力越强。

#### 4.1 时频域扰动特征量的提取

本文利用 Matlab 仿真 16 种单一和复合扰动的 PQD 扰动模型, 对各类扰动模型分别采用 S 变换和希尔伯特黄变换, 根据第 1 节的方法提取出 11 种时频域扰动特征曲线, 以电压闪变+谐波双重复合 PQD 扰动为例, 扰动信号的时频域特征曲线如图 3 中(b)和(c)所示。

为进一步验证采用粒度计算方法进行特征选择的必要性, 以特征  $f_1$  与  $f_{16}$  构成的特征空间为例, 图 4 给出了若干单一扰动、复合扰动样本的分布图。

由图 4(a)可知, 在单标签分类中一个样本仅对应一种扰动类型, 受扰动时频域特征交叉和重叠的影响, 在面对复杂的复合扰动分类问题时难以确定作为决策边界的超平面, 导致可分性较差, 因此现



(c) 电压闪变+谐波频域分析曲线图

图 3 电压闪变+谐波时频域分析

Fig. 3 Voltage flicker with harmonic time-frequency domain analysis

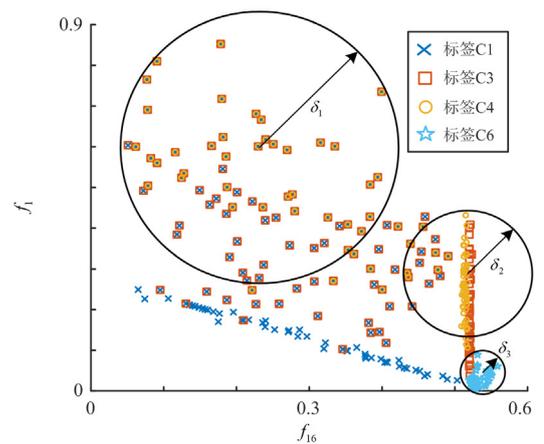
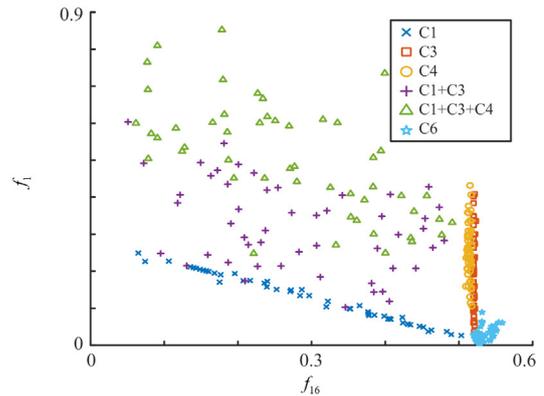


图 4 不同分类方法下的时频域特征散点图

Fig. 4 Scatter diagram of different classification methods in time-frequency domain

有方法多采用多标签分类。在多标签分类中一个样本可对应多个扰动标签, 该方法通过将单标签分类同时针对多种扰动类型进行分类的方法转换为对扰动所含标签类型的预测, 简化了复杂的复合扰动分类问题。由图 4(b)可知, 在进行特征选择或分类时, 各扰动标签所对应的扰动时频域特征的分布状态具有差异性, 有些特征分布相对离散, 有些特征分布相对集中, 有些则交叉重叠具有模糊性, 在确定作为各扰动标签之间决策边界的超平面时, 若仅从单一粒度视角分析会影响对特征的区分能力, 例如采用  $\delta_3$  粒度对含有扰动标签 C6 的特征进行区分时, 具有较高的区分度, 但若采用该粒度对含有扰动标签 C1 的特征进行区分, 则区分效果不佳, 因此需

要同时采用多种合适的粒度构建区分能力不同的决策边界对扰动标签进行区分。

#### 4.2 特征选择实验

本文利用 VPFNRS 算法进行特征选择, 设置模糊邻域半径  $\delta = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ , 变精度系数  $\alpha = \{0.5, 0.6, 0.7, 0.8, 0.9\}$ , 构建 25 个粒度空间。每个粒度空间下都可得到一组最优扰动特征子集  $opt^{(\delta, \alpha)}$ 。为评估每个粒度空间的最优扰动特征子集对分类性能的影响, 以 XGboost 作为评估最优特征子集分类性能的分类器模型, 分类性能随粒度空间的变化曲面图如图 5 和图 6 所示。

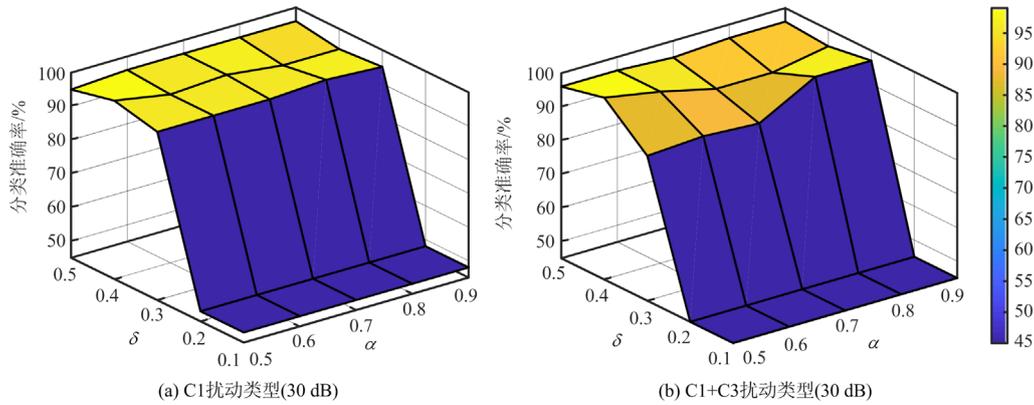


图 5 各粒度空间下扰动分类性能变化(30 dB)

Fig. 5 Variation of classification performance for disturbances in different granularity spaces (30 dB)

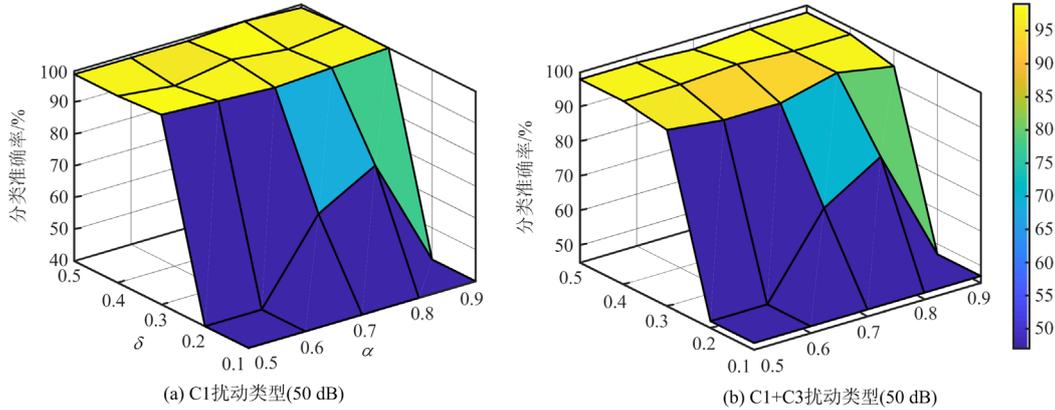


图 6 各粒度空间下扰动分类性能变化(50 dB)

Fig. 6 Variation of classification performance for disturbances in different granularity spaces (50 dB)

由图 5 和图 6 可以看出, 不同粒度空间对不同类别扰动的分类性能具有显著差异性, 扰动在部分粒度空间下分类性能可达 95% 以上。在加入 50 dB 高斯白噪声后, 部分粒度空间的分类准确率仍然能够保持在 95% 以上, 证明该特征选择算法选择的特征是有效的, 且对含噪声数据具有一定的耐受能力。

不同粒度空间对应的最优特征集所构建的分类器模型具有不同的分类表现。

本节对 VPFNRS-stacking、VPFNRS 最优粒度模型、随机森林(Random Forest, RF)、递归特征消除(Recursive Feature Elimination, RFE)和主成分分析法(Principal Component Analysis, PCA) 5 种特征

选择方法进行对比, 为了仅讨论特征选择对分类准确性的影响, 本文方法与现有方法均采用相同的 XGboost 分类器。VPFNRS 单粒度模型使用寻优算法找到最优粒度空间对应的最优扰动特征集进行实验, 各特征选择方法性能对比如表 2 所示。

表 2 各特征选择方法的性能对比

Table 2 Performance comparison of each feature selection method

特征选择方法	AP	HL	OE	C	RL	R
VPFNRS-Stacking	0.915	0.018	0.081	1.069	0.036	0.96
VPFNRS(单粒度)	0.909	0.019	0.088	1.072	0.04	0.959
RF-ML	0.903	0.02	0.104	1.106	0.042	0.957
RFE-ML	0.876	0.025	0.124	1.199	0.057	0.946
PCA-ML	0.849	0.028	0.151	1.304	0.07	0.931
无特征选择	0.905	0.02	0.095	1.124	0.045	0.956

由表 2 可知, 多粒度 VPFNRS-Stacking 模型相较于单粒度 VPFNRS 模型得到的特征具有更好的分类性能。此外, VPFNRS 方法无论是单粒度模型还是多粒度模型的特征选择和分类效果也明显优于传统的 RF、RFE 和 PCA 特征选择方法, 且具有一定的抗噪能力, 并有效地改善了模型的性能。

### 4.3 复合电能质量扰动分类实验

基于 VPFNRS-Stacking 分类方法的复合 PQD 扰动分类在噪声 30 dB、40 dB 和 50 dB 情况下的测试结果如表 3 所示。

表 3 16 种单一和复合扰动分类准确率

Table 3 Classification accuracy of 16 single and complex disturbance

扰动类型	分类准确率/%		
	30 dB	40 dB	50 dB
C1	96.0	96.0	98.0
C2	100.0	100.0	100.0
C3	100.0	100.0	100.0
C4	100.0	100.0	98.0
C5	98.0	100.0	98.0
C6	100.0	100.0	100.0
C1+C3	100.0	96.0	100.0
C1+C4	82.0	86.0	86.0
C2+C4	96.0	92.0	92.0
C3+C4	76.0	80.0	78.0
C3+C6	100.0	100.0	100.0
C4+C5	86.0	80.0	90.0
C4+C6	92.0	82.0	92.0
C1+C3+C4	72.0	80.0	74.0
C2+C3+C4	84.0	90.0	84.0
C3+C4+C5	72.0	80.0	78.0
平均分类准确率/%	90.9	91.4	91.7

由表 3 可知, 本文方法对单一扰动类型的平均分类准确率可达 96%以上, 对双重扰动类型的平均分类准确率可达 90%以上, 对三重复合扰动类型的平均分类准确率可达 80%以上, 证明本文的复合 PQD 扰动分类模型是有效的, 且具有较强的抗噪性能。其中 C3+C4、C1+C3+C4、C3+C4+C5 三种复合电能质量扰动类型的分类准确率较低, 是由于本文利用普通且常见的特征参数构建初始特征集, 因此导致在扰动信号中所提取到的扰动特征本身对 C3+C4 复合扰动类型中非平稳的、非高斯的特征区分度不高, 进而影响最后的分类准确率。

为进一步对比本文方法与常用分类方法的性能, 采用深度卷积神经网络<sup>[3]</sup>(Deep Convolutional Neural Networks, DCNN)模型、两类分类支持向量机模型<sup>[1]</sup>和基于 XGboost 的排位分类模型进行对比。其中 DCNN 和 Binary-SVM 由于不涉及排位函数, 因此只使用平均分类准确率、汉明损失和召回率作为评价指标。各分类模型的性能对比如表 4 所示。

表 4 各分类模型的性能对比

Table 4 Performance comparison of each classification model

分类模型	AP	HL	OE	C	RL	R
VPFNRS-stacking	0.915	0.018	0.081	1.069	0.036	0.96
DCNN	0.643	0.358	—	—	—	0.643
Rank-XGboost	0.908	0.019	0.089	1.04	0.035	0.956
Binary-SVM	0.901	0.021	—	—	—	0.952

由表 4 可知, VPFNRS-Stacking 模型的性能均优于其他 3 种分类模型。在训练样本量较小的情况下, DCNN 的效果不佳。

为进一步证明本文分类模型的有效性, 根据某电网中实测数据共 24 个样本进行测试, 采样频率为 12.8 kHz, 1536 个采样点。使用相同规格的仿真数据生成 24×50 个扰动信号作为训练样本。电网实测数据的分类结果如表 5 所示。由表 5 可知, 本文的分类模型在实测数据中也具有较好的分类效果。

表 5 电网实测数据分析

Table 5 Real measurement data analysis of power grid

序号	标定扰动标签	分类结果
1	C1 5 组	C1 5 组
2	C2 3 组	C2 3 组
3	C4 4 组	C4 4 组
4	C1+C3 5 组	C1+C3 5 组
5	C1+C4 4 组	C1+C4 3 组, C1 1 组
6	C1+C3+C4 1 组	C1+C3 1 组
7	C2+C3+C4 2 组	C2+C3+C4 2 组

#### 4.4 集成分类模型对分类准确率的影响

为对比本文采用的多粒度 Stacking 集成分类模型与 Bagging 和 Boosting 集成分类模型的性能, 在 VPFNRS 各粒度空间所构建的同质弱分类器的基础上应用不同的集成分类策略进行实验, 实验结果如表 6 所示。由表 6 可见, 基于 Stacking 的集成分类策略相较于其他集成分类策略具有更好的模型融合效果。

表 6 各集成模型的性能对比

分类模型	AP	HL	OE	C	RL	R
VPFNRS-Stacking	0.915	0.018	0.081	1.069	0.036	0.96
VPFNRS-Bagging	0.903	0.021	0.099	1.094	0.043	0.959
VPFNRS-Boosting	0.9	0.023	0.101	1.096	0.045	0.955

## 5 结论

本文提出一种基于多粒度特征选择和模型融合的复合电能质量扰动分类特征优化方法, 并论证了该方法能够有效改善分类模型的性能, 对复合 PQD 分类具有较好的准确性和有效性, 结论如下:

1) 提出基于变精度模糊粗糙集模型的特征选择方法, 解决了多标签数据集不同样本间特征分布的差异性以及类域的交叉重叠等造成数据模糊而难以寻找最优特征区分面的问题, 提升了特征选择的效果。

2) 考虑到各粒度空间下不同扰动特征集对分类性能的影响, 为防止忽略各粒度空间中对分类有益的有效信息, 基于 Stacking 集成分类原理, 提出了三层复合电能质量扰动分类器架构, 构建了多粒度集成分类模型, 该模型融合了各个粒度空间下的基分类器的识别能力, 具有一定的自适应性。

本文在分析多标签数据时假设样本标签的重要度是一致的, 而目前尚缺乏针对标签重要度不等的多标签数据特征选择模型的相关研究; 此外, 初始特征集合的选择仍依赖于专家经验。因此, 下一步将分别从标签分布增强和初始特征选择的角度研究进一步的改进方法。

### 参考文献

- [1] 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.
- [2] 肖湘宁. 电能质量分析与控制[M]. 北京: 中国电力出版社, 2010.
- [3] 徐达, 洪文慧, 季天瑶, 等. 基于 S 变换和长短期记忆网络的电能质量复合扰动识别[J]. 广东电力, 2021, 34(7): 33-39.  
XU Da, HONG Wenhui, JI Tianyao, et al. Composite disturbance identification of power quality based on S transform and long short term memory network[J]. Guangdong Electric Power, 2021, 34(7): 33-39.
- [4] WANG S, CHEN H. A novel deep learning method for the classification of power quality disturbances classification using deep convolutional neural networks[J]. Applied Energy, 2019, 235: 1126-1140.
- [5] 郑炜, 林瑞全, 王俊, 等. 基于 GAF 与卷积神经网络的电能质量扰动分类[J]. 电力系统保护与控制, 2021, 49(11): 97-104.  
ZHENG Wei, LIN Ruiquan, WANG Jun, et al. Power quality disturbance classification based on GAF and a convolutional neural network[J]. Power System Protection and Control, 2021, 49(11): 97-104.
- [6] THIRUMALA K, PAL S, JAIN T, et al. A classification method for multiple power quality disturbances using EWT based adaptive filtering and multiclass SVM[J]. Neurocomputing, 2019, 334: 265-274.
- [7] 徐艳春, 高永康, 李振兴, 等. 基于 VMD 初始化 S 变换的混合动力系统电能质量扰动检测与分类[J]. 中国电机工程学报, 2019, 39(16): 4786-4798, 4980.  
XU Yanchun, GAO Yongkang, LI Zhenxing, et al. Power quality disturbance detection and classification of hybrid power system based on VMD initialization S-transform[J]. Proceedings of the CSEE, 2019, 39(16): 4786-4798, 4980.
- [8] TANG Qiu, QIU Wei, ZHOU Yicong. Classification of complex power quality disturbances using optimized S-Transform and Kernel SVM[J]. IEEE Transactions on Industrial Electronics, 2020, 67(11): 9715-9723.
- [9] 李建文, 秦刚, 李永刚, 等. 基于布莱克曼窗 S 变换与数据库查询的电能质量扰动识别与分类新方法[J]. 电网技术, 2020, 44(12): 4734-4743.  
LI Jianwen, QIN Gang, LI Yonggang, et al. Recognition and classification new method of power quality disturbances based on Blackman window S transform and database query[J]. Power System Technology, 2020, 44(12): 4734-4743.
- [10] 王燕. 电能质量扰动检测的研究综述[J]. 电力系统保护与控制, 2021, 49(13): 174-186.  
WANG Yan. Review of research development in power quality disturbance detection[J]. Power System Protection and Control, 2021, 49(13): 174-186.
- [11] 许立武, 李开成, 罗奕, 等. 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[J]. 电力系统保护与控制, 2019, 47(6): 24-31.
- XU Liwu, LI Kaicheng, LUO Yi, et al. Classification of complex power quality disturbances based on incomplete S-transform and Gradient boosting decision tree[J]. Power System Protection and Control, 2019, 47(6): 24-31.
- [12] 徐袭, 石敏. 一种基于粗糙集与小波变换的电能质量分类方法[J]. 电力自动化设备, 2005, 25(11): 15-18.
- XU Xi, SHI Min. Power quality classification based on rough set and wavelet transform[J]. Electric Power Automation Equipment, 2005, 25(11): 15-18.
- [13] WANG Changzhong, SHAO Mingwen, HE Qiang, et al. Feature subset selection based on fuzzy neighborhood rough sets[J]. Knowledge Based Systems, 2016, 111: 173-179.
- [14] WANG Changzhong, HUANG Yang, SHAO Mingwen, et al. Fuzzy rough set-based attribute reduction using distance measures[J]. Knowledge Based Systems, 2018, 164: 205-212.
- [15] CHEN Panpan, LIN Menglei, LIU Jinghua, et al. Multi-label attribute reduction based on variable precision fuzzy neighborhood rough set[J]. Digital Object Identifier, 2020, 8: 133565-133576.
- [16] 段洁, 胡清华, 张灵均, 等. 基于邻域粗糙集的多标记分类特征选择算法[J]. 计算机研究与发展, 2015, 52(1): 56-65.
- DUAN Jie, HU Qinghua, ZHANG Lingjun, et al. Feature selection for multi-label classification based on neighborhood rough sets[J]. Journal of Computer Research and Development, 2015, 52(1): 56-65.
- [17] RIBERIO M H D M, COELHO L D S. Ensemble approach based on bagging, boosting and stacking for short term prediction in agribusiness time series[J]. Applied Soft Computing Journal, 2020, 86.
- [18] 李长松, 刘凯, 肖先勇, 等. 基于条件互信息特征选择法和 Adaboost 算法的电能质量复合扰动分类[J]. 高电压技术, 2019, 45(2): 579-585.
- LI Changsong, LIU Kai, XIAO Xianyong, et al. Classification of multiple power quality disturbances based on conditional mutual information feature selection method and Adaboost algorithm[J]. High Voltage Engineering, 2019, 45(2): 579-585.
- [19] 汪颖, 陈春林, 肖先勇. 电压暂降源异质堆叠集成学习识别法[J]. 电力系统保护与控制, 2021, 49(15): 1-8.
- WANG Ying, CHEN Chunlin, XIAO Xianyong. Heterogeneous stacking integrated learning identification method for voltage sag sources[J]. Power System Protection and Control, 2021, 49(15): 1-8.
- [20] RADHAKRISHNAN P, RAMAIYAN K, VINAYAGAM A, et al. A Stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network[J]. Measurement, 2021, 175.
- [21] 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.
- [22] LI Jinsong, LIU Hao, WANG Dengke, et al. Classification of power quality disturbance based on s-transform and convolution neural network[J]. Frontiers in Energy Research, 2021, 9.
- [23] 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.
- [24] 黄南天, 王达, 刘座铭, 等. 复杂噪声环境下电能质量复合扰动特征选择[J]. 仪器仪表学报, 2018, 39(4): 82-90.
- HUANG Nantian, WANG Da, LIU Zuoming, et al. Feature selection of composite power quality disturbances under complex noise environment[J]. Chinese Journal of Scientific Instrument, 2018, 39(4): 82-90.
- [25] IEEE recommended practice for monitoring electric power quality: IEEE Std. 1159—2009[S]. New York: IEEE, 1995.
- [26] 陈振宇, 刘金波, 李晨, 等. 基于 LSTM 与 XGBoost 组合模型的超短期电力负荷预测[J]. 电网技术, 2020, 44(2): 614-620.
- CHEN Zhenyu, LIU Jinbo, LI Chen, et al. Ultra short-term power load forecasting based on combined LSTM-XGBoost model[J]. Power System Technology, 2020, 44(2): 614-620.

收稿日期: 2021-08-31; 修回日期: 2021-11-11

作者简介:

阮梓航(1997—), 男, 硕士研究生, 研究方向为电能质量与优质供电; E-mail: 1069612544@qq.com

肖先勇(1968—), 男, 博士, 教授, 研究方向为电能质量与优质供电; E-mail: xiaoxianyong@163.com

胡文曦(1993—), 男, 通信作者, 博士, 讲师, 研究方向为电力扰动数据分析与应用。E-mail: 408803186@qq.com

(编辑 魏小丽)