

# 基于 Stacking 模型融合的串联故障电弧检测

王毅<sup>1</sup>, 陈进<sup>1</sup>, 李松浓<sup>2</sup>, 陈涛<sup>2</sup>, 戴莲丹<sup>3</sup>, 宣姝<sup>3</sup>

(1.重庆邮电大学 通信与信息工程学院, 重庆 400065;

2.国网重庆市电力公司电力科学研究院, 重庆 400014; 3.国网重庆市电力公司营销服务中心, 重庆 400023)

**摘要:** 针对低压交流配电网中由于电弧燃烧程度不同、电流畸变程度不同而导致漏检、错检问题, 提出一种基于 Stacking 模型融合的时域故障电弧检测方法。从回路电流中提取时域特征, 将时域特征组成特征矩阵对机器学习算法决策树和集成学习算法随机森林等进行参数寻优。最后, 将集成学习算法代替机器学习算法作为基学习器通过 Stacking 模型融合构建低压交流故障检测模型。实验共采集 6 种电器的并联电流共计 96 970 组, 结果表明, 相较于非集成算法和其他集成算法, 所提方法具有更高的准确率、精确度和 F1 指标, 其模型更为稳健。

**关键词:** 故障电弧; 电流采集; 电弧检测; 集成机器学习; Stacking 模型融合

中图分类号: TM501.2

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.201209

中文引用格式: 王毅, 陈进, 李松浓, 等. 基于 Stacking 模型融合的串联故障电弧检测[J]. 电子技术应用, 2021, 47(11): 53-57.

英文引用格式: Wang Yi, Chen Jin, Li Songnong, et al. Series fault arc detection based on Stacking model fusion[J]. Application of Electronic Technique, 2021, 47(11): 53-57.

## Series fault arc detection based on Stacking model fusion

Wang Yi<sup>1</sup>, Chen Jin<sup>1</sup>, Li Songnong<sup>2</sup>, Chen Tao<sup>2</sup>, Dai Liandan<sup>3</sup>, Xuan Shu<sup>3</sup>

(1.School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;

2.Chongqing Electric Power Research Institute, Chongqing 400014, China;

3.State Grid Chongqing Electric Power Company Marketing Service Center, Chongqing 400023, China)

**Abstract:** Aiming at the problems of missed detection and false detection due to different arc combustion levels and current distortion levels in low-voltage AC distribution networks, a time-domain arc fault detection method based on Stacking model fusion is proposed. The time domain features are extracted from the loop current, and the time domain features are formed into a feature matrix to optimize the parameters of the machine learning algorithm decision tree and the integrated learning algorithm random forest. Finally, the integrated learning algorithm is used as the base learner instead of the machine learning algorithm to build a low-voltage AC fault detection model through Stacking model fusion. The experiment collected a total of 96 970 groups of parallel currents of six electrical appliances. The results show that compared with non-integrated algorithms and other integrated algorithms, the proposed method has higher accuracy, precision and F1 index, and its model is more robust.

**Key words:** arc fault; current acquisition; arc detection; integrated machine learning; Stacking model fusion

### 0 引言

电弧是一种由于绝缘介质被电压击穿而导致的持续放电现象, 通常伴随着电极部分挥发<sup>[1-2]</sup>。在生活中由于输电线长时间带载、过载产生热量或者由于外力导致的绝缘层老化、破损, 会产生电弧现象<sup>[3-4]</sup>。这类电弧是人们不希望产生的, 称为故障电弧<sup>[5-6]</sup>。

近年来, 相关学者从不同角度对故障电弧的检测进行了广泛研究, 文献[7]通过快速傅里叶变换(FFT)得到电流的幅度谱, 再对其进行稀疏表示来进行故障检测, 在对稀疏表示的  $L_p$  范数进行调整时, 发现对不同的数

据采用不同的范数可以提高分类的准确性。文献[8]分别分析了小波熵算法、基于支持向量机(SVM)和逻辑回归的分类模型, 综合以上算法可知相对于简单方法, 该组合算法具有更好的检测功能。文献[9]集中讨论了电弧电流中的40~100 kHz的高频谐波部分, 通过FFT和阈值设置, 可以有效识别故障电弧。文献[10]提取电流中的9个时频域特征, 再使用主成分分析将9个特征在维度上缩减为3个参数, 最后使用SVM建立了负荷识别和串联电弧检测的综合模型。

上述故障识别均从故障的高频分量入手, 通常需要

高采样频率和高计算量。本文通过实验采集大量电流,各种信息混合情况下传统机器学习算法表现不佳,本文提出一种 Stacking 模型融合的时域故障电弧检测方法,经验证该算法的识别率能达到 98% 以上。

### 1 信号采集与分析

#### 1.1 电流信息采集平台

本文根据 GBT 31143-2014《电弧故障保护电器的一般要求》选用故障电弧发生器,根据各公开文献对家用电器的大致分类选用多个典型负载采集电流信息,实验采集原理如图 1 所示。

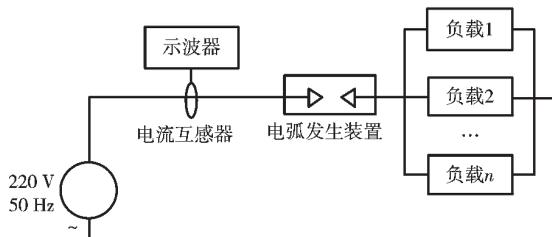


图 1 电流信息采集系统

本文使用白炽灯、电风扇、电烙铁、调光灯、LED 灯和笔记本作为典型负载进行测量,使用电流互感器和 picoscope 高性能示波器作为电流采集工具,采样频率为 10 kHz。在采集故障电流时,根据 GB14.287.4-2014《电器火灾监控系统第 4 部分:故障电弧探测器》规定电弧持续时间不超过 0.42 ms 或者不超过额定电流值 5% 的微小电弧不作为电弧统计。

如图 1 所示,采集正常电流时不接入故障电弧发生器,利用示波器与电流互感器相连测量电流;采集故障电流时串联接入故障电弧发生器,实验开始前旋转故障电弧发生器移动电极的螺杆使紫铜锥棒与固定电极石墨圆棒接触,此时电路闭合,实验开始后缓缓旋转螺杆将移动电极拉开使故障电弧发生器内产生连续电弧,保存此段电流作为故障电流。使用上述方法采集 6 种不同类型用电器单独和混合电流波形。

采集不同类型用电器混合工作状态下的正常和故障电流,用电器功率信息如表 1 所示,采集样本信息如表 2 所示。

表 1 典型负载

序号	负载	功率/W	类型
1	白炽灯	100	电阻型
2	电风扇	50	电容电感型
3	电烙铁	30	电容电感型
4	调光灯	25	开关电源型
5	LED	5	开关电源型
6	笔记本	90	开关电源型

选取表 1 中所示的负载作为典型负载进行测量。采集的电流信息包括各种电器单独工作以及不同类型用电器同时工作时正常和故障电流信息。采集电器和采集到的电流样本情况如表 2 所示。

表 2 电流信息采集情况

序号	负载	正常样本	故障样本
1	白炽灯	5 120	5 923
2	调光灯	3 520	5 210
3	电风扇	5 120	9 020
4	LED 灯	3 200	4 540
5	笔记本	3 520	4 782
6	电烙铁	3 520	4 940
7	电风扇+LED	1 920	2 680
8	电脑+LED	2 240	3 603
9	白炽灯+笔记本	1 920	2 375
10	电风扇+电烙铁	1 920	3 310
11	电烙铁+白炽灯	2 240	1 843
12	电风扇+白炽灯+LED	1 920	2 365
13	电烙铁+电风扇+LED	1 920	2 990
14	笔记本+电风扇+电烙铁+白炽灯	2 560	2 749

#### 1.2 典型负载电流信息分析

本文主要针对低压交流故障电弧进行检测,此类故障电弧的一般特性为:故障电弧信号中含有高频噪声;电弧电流上升速度快于非电弧电流;电弧电流存在“平肩部”<sup>[2]</sup>。分别取白炽灯、笔记本正常和故障电流的 5 个周期进行归一化后绘制图 2 所示电流波形。

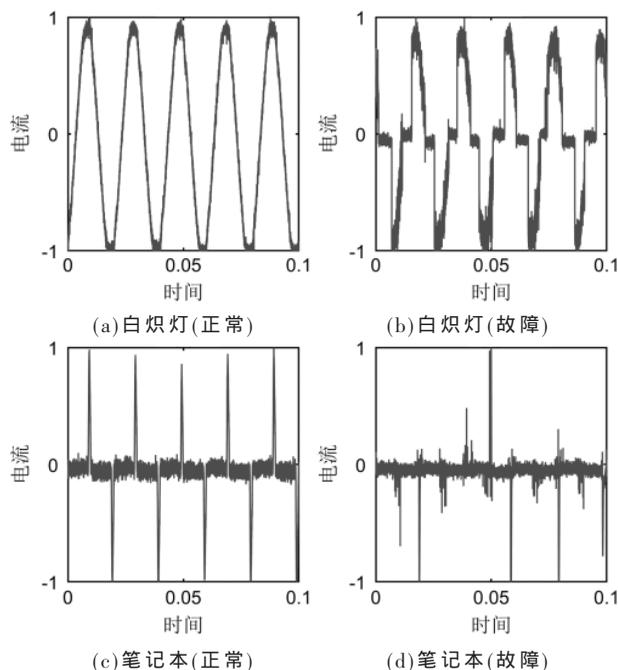


图 2 白炽灯和笔记本正常和电弧状态电流波形

## 2 机器学习算法

### 2.1 机器学习

机器学习是一门多领域交叉学科,涉及概率论、统计学等多门学科<sup>[11]</sup>。机器学习算法可分为三大类:监督学习、无监督学习和半监督学习。本文主要使用监督学习算法进行故障电弧检测,将所有正常电流的类标签设

为 1,故障电流数据类标签设为 2。

提取时域电流数据的平均值、中位数、方差以及极差进行故障识别。表 3 为求解上述特征的表达式。其中,  $N$  表示一个电流周期内的采样电流个数;  $x_i$  表示当前电流周期内的第  $i$  个电流样本。

表 3 时域特征表达式

特征	表达式
电流平均值	$x_{\text{mean}} = \frac{1}{N} \sum_{i=1}^N x_i$
电流中位数	$x_{\text{med}} = \frac{x_{(N/2)} + x_{(N/2+1)}}{2}$
电流方差	$x_{\text{var}} = \frac{1}{N} \sum_{i=1}^N (x_i - x_{\text{mean}})^2$
电流极差	$x_{\text{diff}} = \max(x_N) - \min(x_N)$

本文选取决策树进行故障识别,决策树是一种树形结构<sup>[12]</sup>,由节点和有向边组成。节点有两种类型:非叶子节点和叶子节点。非叶子节点表示一个特征或属性,叶子节点表示一个类。决策树结构如图 3 所示。

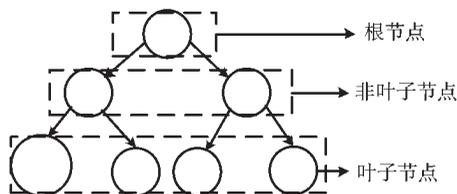


图 3 决策树结构

决策树的学习本质上是从训练集中归纳出一组分类规则,得到与数据集矛盾较小的树形结构。决策树学习的损失函数通常是正则化的极大似然函数,通常采用启发式方法,近似求解这一最优化问题。

## 2.2 决策树参数选择

本文使用网格搜索法对决策树的最大深度进行寻优。网格搜索法首先为想要调参的参数设定一组候选值,然后穷举参数的组合,再根据设定的评分机制找到最好的设置。本文设定决策树的最大深度从 1 到 20 依次递增,取表 2 中 14 种用电器的正常和故障电流信息各 1 500 组组成样本集,将样本集按照 5 份切分进行交叉验证,将 5 次测试集准确率的平均值作为评分标准,找到评分最高的组对应的最大深度作为最优值进行决策树建模。过程如图 4 所示。

图 4 为具有交叉验证的网格搜索法。上述方法在 Python3.8 环境下仿真可得仿真结果,如图 5 所示。

由图 5 可得出决策的最大深度最优值为 14,最优识别率为 94.11%。现考虑使用集成机器学习的方法对改模型进行改进,提高故障识别率。

## 3 集成学习及 Stacking 模型融合算法

### 3.1 Bagging 及 Boosting 集成机器学习框架

集成学习将训练集数据输入若干弱学习器,通过一

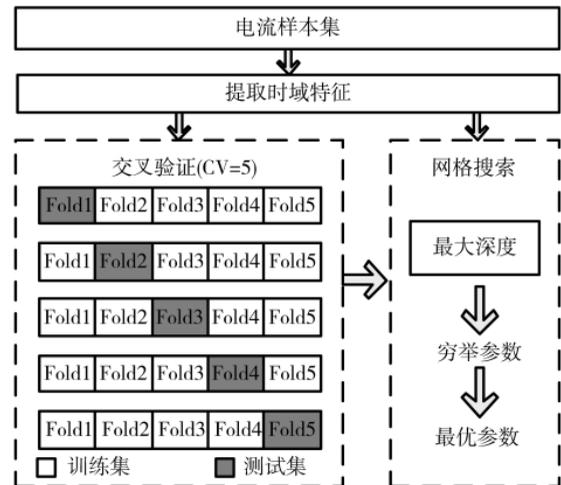


图 4 具有交叉验证的网格搜索法

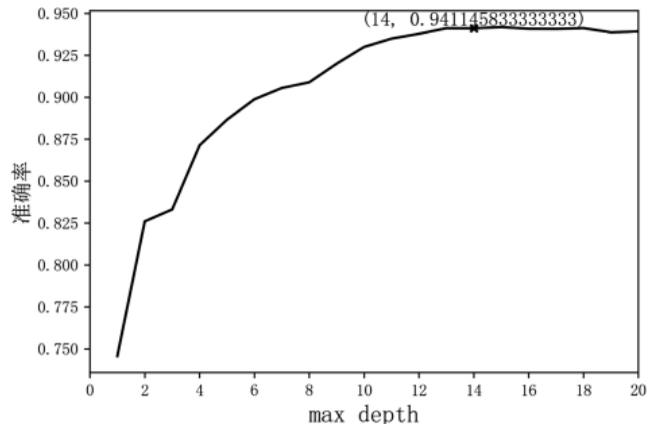


图 5 决策树参数寻优

定的结合策略,最终可以得到一个强学习器<sup>[13-14]</sup>。

集成方法可分为两类:第一类是串行集成方法,其中参与训练的弱学习器按照顺序生成(Boosting 框架)。序列方法的原理是利用弱学习器之间的依赖关系,通过对之前训练中错误标记的样本赋值较高的权重,可以提高整体的预测效果。另一类是并行集成方法,其中参与训练的弱学习器并行生成(Bagging 框架)。并行方法的原理是利用弱学习器之间的独立性,通过平均可以显著降低错误。

本文选择基于 bagging 框架的随机森林和极限树以及基于 boosting 框架的 Adaboost 和 XGboost 算法对决策树进行优化。

随机森林和 Adaboost 选择 14 作为每棵树的最大深度。其基学习器个数使用参数寻优的方法,以算法耗费的时间和准确率为评分机制,找到最佳参数。集成算法准确率如图 6 所示。

图 6 中随着树数量的增加,算法准确率也有所增加,其中 Adaboost 在树数量增加 64 后,准确率基本不上升但耗费时间会增多,故对于 Adaboost 来说,选择 64 棵树作为基学习器是合适的;随机森林在 100 棵树的条件下准确率较高,同时耗费的时间不多,故选择 100 棵树

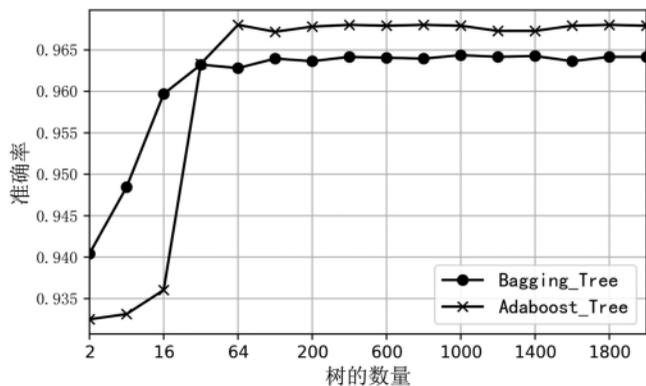


图6 集成算法准确率

作为随机森林的基学习器。

对于极限树和 XGboost 算法的最大深度和树的数量,首先在 1~1 000 以内按 200 划分为 5 份,在每份中分别按 10 划分寻找最优值。对于极限树和 XGboost,表 4 和表 5 分别表示网格搜索得到的最优值。

表 4 极限树参数寻优

最大深度	基学习器数		
	200	210	220
50	0.968 687 5	0.968 656 25	0.968 531 25
60	0.968 75	0.968 781 25	0.968 593 75
70	0.968 718 75	0.968 75	0.968 625

表 5 XGBoost 参数寻优

最大深度	基学习器数		
	110	120	130
8	0.967 062 5	0.966 875	0.966 906 25
10	0.967 625	0.967 656 25	0.967 562 5
12	0.966 562 5	0.966 281 25	0.966 656 25

极限树最大深度从 10 至 100 按间隔 10 递增,基学习器个数从 100 至 300 按 10 递增。表 4 列出了最大深度 50 至 60、基学习器 200 至 220 的准确率,从表中可看出最大深度为 60、基学习器 210 时准确率最高,故后续仿真中极限树的参数均以上述数值为准。

XGboost 最大深度从 8 至 10 间隔 2 递增,基学习器个数从 10 至 200 间隔 10 递增。表 5 列出了最大深度 8 至 12、基学习器个数 110 至 130 对应的准确率,可得最大深度为 10、基学习器 120 个时准确率最高。

### 3.2 Stacking 模型融合

将个体学习器结合在一起时使用的方法叫作结合策略。有一种结合策略是使用另外一个机器学习算法来将个体机器学习器的结果结合在一起,这个策略就是 Stacking<sup>[15]</sup>。

本文区别于常用的将机器学习算法作为基学习器的方法,将经过参数寻优的集成机器学习模型作为基学习器进行 Stacking 模型融合。在 Stacking 方法中次级学

习器一般使用线性算法,本文使用逻辑回归算法作为次级学习器。

图 7~图 10 分别表示在不同训练集下提升树、随机森林、极限树、XGBoost 以及 Stacking 集成算法的准确率、精确率、召回率以及 F1 指标。其中,Stacking 使用前 4 种集成机器学习算法作为初级学习器。

准确率是分类正确的样本个数占总样本个数的比例,是分类问题中最简单最常用的评价指标。如图 7 所示,本仿真中 4 种算法的准确率随着样本数量的增大而有所增加,但其识别率基本处于 95% 以上,在样本集大的情况下,准确率能达到 97% 以上,相较于非集成算法决策树而言,准确率有所提升。且 Stacking 模型融合算法准确率在集成学习的基础上又有所提升,在样本量较大的情况下,准确率能达到 98%~99%。

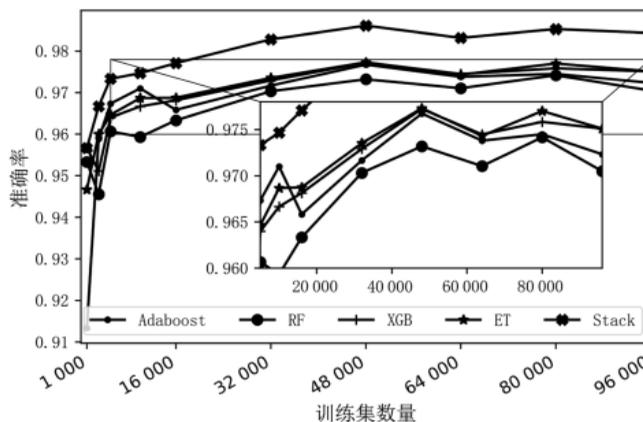


图 7 算法准确率

在本文仿真中,正常电流特征被标记为正例,故障电流特征被标记为反例。精确率指模型预测为正的样本中实际也为正的样本占被预测为正的样本的比例,体现了模型对负样本的区分能力,精确度越高,模型对负样本的区分能力越强。本文提出的故障电弧检测算法主要需要检测出电流中的负样本。如图 8 所示,Stacking 集成算法精确度不论在大样本还是小样本情况下都明显高于其他算法。

召回率指实际为正的样本中被预测为正的样本所

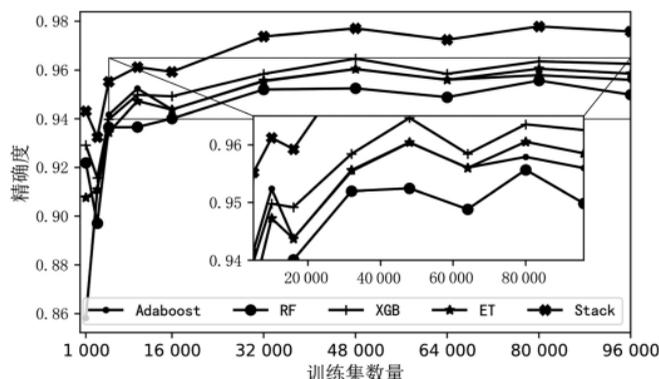


图 8 算法精确率

占实际为正的样本的比例,体现模型对负样本的识别能力,召回率越高,模型对正样本的识别能力越强。对于本文,召回率可以认为是正常样本被错判为故障样本的比例,召回率越高,其误判的可能性越低。如图9所示,在样本量较大的情况下 Stacking 算法在召回率上也有很好的表现。就此指标来说,Stacking 算法在误报率上也有很好的性能。

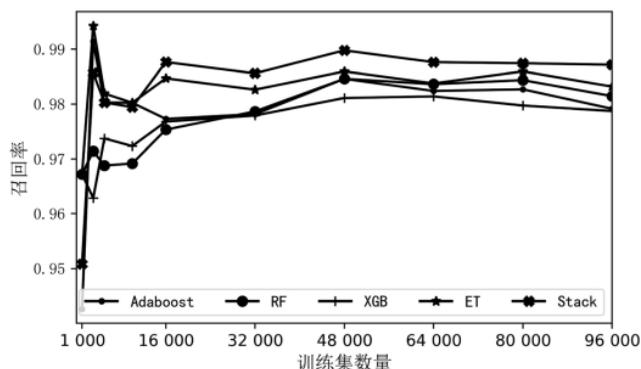


图9 算法召回率

在实际应用中,为了评价模型的综合性能,引入了F1指标对模型进行进一步评判。F1指标是精确率和召回率的调和平均值,F1指标越高,说明模型更稳健。在图10中可看出,在样本数量较大或较小的情况下,Stacking 算法的F1指标均高于其他算法,故 Stacking 模型融合算法具有更好的稳健性。

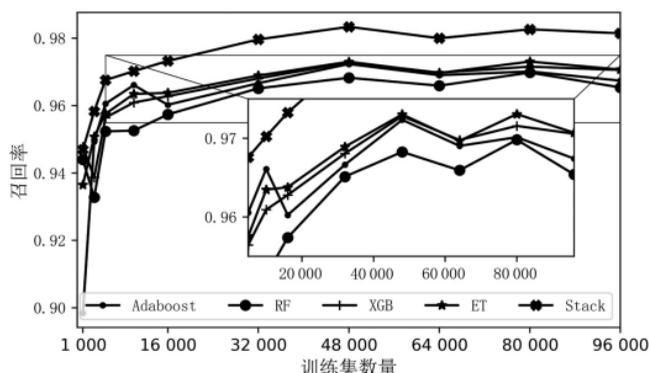


图10 算法F1指标

综上所述,Stacking 模型融合算法不论在准确性、精确性、召回率以及F1指标都有较好的表现,Stacking 算法更适用于故障电弧检测。

#### 4 结论

本文主要针对住宅和配电网情况下的低压交流故障电弧检测,在此供电条件下,使用符合标准的故障电弧发生器采集电流信息。由于时频转换算法需要耗费大量计算时间,故本文只使用时域特征组成特征矩阵,结合模型融合进行故障识别。经过大量实验数据的验证可知,Bagging 集成框架和 Boosting 集成框架对故障识别算法均有改进。但是将上述集成算法作为 Stacking 模型融

合的初级学习器进行故障检测可获得更好的性能。

#### 参考文献

- [1] UL standard for safety for arc-fault circuit-interrupters(UL 1699)[S].2008.
- [2] 刘官耕,杜松怀,苏娟,等.低压电弧故障防护技术研究与发展趋势[J].电网技术,2017,41(1):305-313.
- [3] 孙鹏,郑志成,闫荣妮,等.采用小波熵的串联型故障电弧检测方法[J].中国电机工程学报,2010,30(S1):232-236.
- [4] 张冠英,张晓亮,刘华,等.低压系统串联故障电弧在线检测方法[J].电工技术学报,2016,31(8):109-115.
- [5] 王尧,田明,牛峰,等.低压交流电弧故障检测方法研究综述[J].电器与能效管理技术,2018(10):8-13,44.
- [6] 赵怀军,秦海燕,刘凯,等.基于相关理论及零休特征融合的串联故障电弧检测方法[J].仪器仪表学报,2020,41(4):218-228.
- [7] QU N, WANG J H, LIU J H. An arc fault detection method based on current amplitude spectrum and sparse representation[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(10):3785-3792.
- [8] XIA K, LIU B Z, FU X L, et al. Wavelet entropy analysis and machine learning classification model of DC serial arc fault in electric vehicle power system[J]. IET Power Electronics, 2019, 12(15):3998-4004.
- [9] XIA K, HE Z H, YUAN Y, et al. An arc fault detection system for the household photovoltaic inverter according to the DC bus currents[C]//2015 18th International Conference on Electrical Machines and Systems, 2015.
- [10] JIANG J, WEN Z, ZHAO M X, et al. Series arc detection and complex load recognition based on principal component analysis and support vector machine[J]. IEEE Access, 2019, 7:47221-47229.
- [11] 徐艺.机器学习算法及其应用研究[D].长沙:湖南大学,2019.
- [12] 栾丽华,吉根林.决策树分类技术研究[J].计算机工程,2004,30(9):94-96,105.
- [13] 沈学华,周志华,吴建鑫,等.Boosting 和 Bagging 综述[J].计算机工程与应用,2000(12):31-32.
- [14] 徐继伟,杨云.集成学习方法:研究综述[J].云南大学学报(自然科学版),2018,40(6):1082-1092.
- [15] 史佳琪,张建华.基于多模型融合 Stacking 集成学习方式的负荷预测方法[J].中国电机工程学报,2019,39(14):4032-4041.

(收稿日期:2020-12-16)

#### 作者简介:

王毅(1981-),男,博士,副教授,主要研究方向:宽带电力线载波通信、智能电网用电信息采集系统。

陈进(1996-),通信作者,女,硕士研究生,主要研究方向:故障电弧检测,E-mail:1574429243@qq.com。



扫码下载电子文档

## 版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所