

基于多特征融合和随机森林的串联电弧故障检测方法

郭敏^{1,2}, 郭小璇^{1,2}, 吴宁^{1,2}, 卢健斌^{1,2}, 陈卫东^{1,2}

(1. 广西电网有限责任公司电力科学研究院, 南宁 530023;

2. 广西电力装备智能控制与运维重点实验室, 南宁 535023)

摘要: 串联电弧故障在线路负载的限制下,其故障电流比较小,使得传统电气保护装置如低压断路器、熔断器等无法发生动作,因此,串联电弧故障的检测成为目前电弧故障检测研究的热点问题。为此,通过搭建电弧故障平台进行电弧实验,获得典型负载在正常工作和电弧故障时的电流数据;然后,在电弧多特征融合的基础上,结合特征的信息增益、增益率对电弧特征进行重要性排序,构建更加合适的串联电弧特征指标集;最后,在此基础上建立随机森林算法,并采用粒子群算法对参数进行优化。研究结果表明:与传统随机森林和BP神经网络相比,该方法有更高的识别率,其准确率和查全率都达到了99%,能够实现对串联故障电弧的有效诊断。

关键词: 串联电弧;故障识别;特征分析;特征选择;随机森林

Series Arc Fault Detection Method Based on Multi-feature Fusion and Random Forest

GUO Min^{1,2}, GUO Xiaoxuan^{1,2}, WU Ning^{1,2}, LU Jianbin^{1,2}, CHEN Weidong^{1,2}

(1. Electric Power Research Institute of Guangxi Power Grid Co., Ltd., Nanning 530023, China;

2. Guangxi Key Laboratory of Intelligent Control and Maintenance of Power Equipment,
Nanning 535023, China)

Abstract: Under the limitation of the line load, the fault current of the series arc fault is relatively small, which makes such traditional electrical protection devices as low-voltage circuit breakers, fuses cannot operate, therefore, the detection of series arc faults has become a hot issue in the current arc fault detection research. For this purpose, the arc experiment is performed by setting up arc fault platform and the current data of typical loads during normal operation and arc faults are obtained; Then, on the basis of arc multi-feature fusion, the information gain and the gain rate of the features are combined to sort the importance of the arc features so to construct a more suitable set of series arc features; Finally, on this basis, a random forest algorithm is set up and the particle swarm algorithm is used to optimize the parameters. The research results show that compared with the traditional random forest and BP neural network, the proposed method has a higher identification rate, and both its accuracy rate and completion rate reach 99%, which can achieve effective diagnosis of series fault arc.

Keywords: series arc fault; fault recognition; feature analysis; feature selection; random forest

0 引言

近年来,随着我国电气化程度的不断提高和电网的逐步智能化,极大的促进了工业水平的进步和提高了国民的生活质量,但也产生了一些隐患。根据全国商业消防与安全协会发布的火灾统计可知,2021年全国消防扑救火灾74.5万起,电气引发的

占28.4%。在引发电气火灾的起因中,电弧故障是主要原因^[1]。大多数时候,当并联电弧故障和接地电弧故障发生时,熔断器、断路器、剩余电流动作保护器等传统的电气保护装置能够容易检测出电弧故障起到保护效果,但在串联电弧故障发生时,由于电路电流的减小,导致熔断器、断路器等无法发生动作,最终引发电气火灾。因此,为了减少电器

收稿日期:2023-02-26

基金项目:南方电网公司科技项目(GXKJXM20220099)。

火灾带来的严重的经济损失和人身伤害,研究高准确率的串联型故障电弧诊断技术尤为重要。

目前,基于电弧数学模型、电弧物理现象和电流电压波形对串联故障电弧进行检测是主要的检测方法^[2-3]。相对于其他检测方法,基于电弧电流特征的电弧检测方法具有更好的可实现性,成为目前电弧故障检测研究的热点。文献[4-6]基于电流波形的频域特征对电弧故障进行检测;文献[7-8]基于电流波形的时域特征,分别通过周波分析法和电流信号的短时过零率对发生电弧时的电流信号进行分析,实现了对串联故障电弧的快速有效检测;文献[9-10]基于电流信号的小波特征,分别利用小波包能量熵和小波近似熵对串联电弧故障进行诊断。

上述的电弧故障检测方法虽然实现了串联电弧故障的诊断识别,但仅仅依靠单一特征容易受到诸多限制,并不足以实现故障电弧的有效检验:时域特征容易受到噪声的干扰导致稳定性较低,且并不适用于所有负载类型;电流频域特征的提取通常需要利用傅里叶变换进行频域变换,但忽视了时域特征;小波变换在一定程度上解决了傅里叶变换在局部时域信号分析方面的不足,但存在频谱混叠等问题。因此,为了解决单一特征不能全面地对电弧故障进行诊断的问题,文献[2]从不同角度提取了Lipschitz指数、Teager能量算子和近似熵这3个指标,将以往对单特征量的识别转换为对多特征空间位置分布关系的研究,实现对串联电弧故障的准确诊断。文献[11]利用时域、频域中的弧特征,通过分析故障电弧信号的电流间隙、三次谐波、谐波幅度等进行电弧的检测和识别。文献[12]以Fourier系数、Mel频率倒谱数据和小波特征作为深度神经网络的输入量,能够更全面地对电弧故障进行诊断。目前深度学习已广泛应用于电弧故障检测,以BP神经网络算法及其改进算法居多^[2,12-15],但BP神经网络存在容易局部极小化且收敛速度慢等问题。

因此,本文对原始实验数据集进行时域、频域分析,并且为了避免信息的冗余,应用信息增益和增益率对电弧特征集进行降维,得到一个相对合适的电弧特征子集用于分类器的训练与测试。在分类器方面,基于随机森林算法,并且为了降低随机森林的复杂度和提高其性能,利用粒子群算法对随机森林进行了优化。

1 串联电弧故障实验

1.1 电弧故障平台搭建

串联故障电弧实验平台主要由空气开关、电弧

故障发生器、数据采集卡、电压电流互感器以及各类负载组成,其中电弧故障发生器由一个石墨圆棒和一个尖头铜棒组成,数据采集卡的采样频率为10 kHz,电源为220 V 50 Hz的交流电,该平台的具体结构见图1。

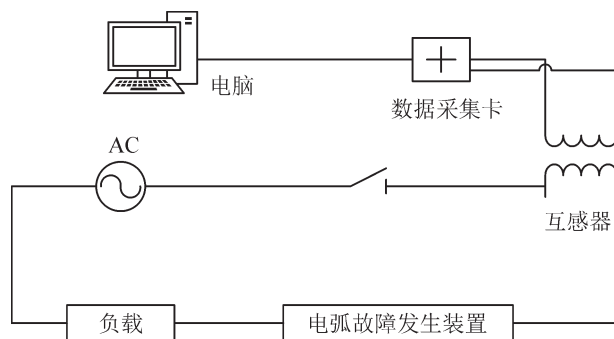


图1 串联电弧故障实验电路图
Fig. 1 Experimental circuit diagram of series arc fault

1.2 电弧故障数据采集

在串联电弧故障实验中,不同类型的负载在发生电弧故障时的电流波形不尽相同,在正常工作时的电流波形也并不一致,回路中电弧电流受负载类型的影响。本文对一些典型的单个负载进行实验,首先,将各类负载一一接入回路,采集其正常工作时的电流波形,然后通过分离石墨圆棒和尖头铜棒,模拟产生故障电弧,采集其出现故障电弧时的电流波形,表1为实验接入的负载类型,可分为电容-电感类负载、电阻类负载和开关类负载。

表1 实验所用负载类型
Table 1 Type of load for experiment

负载类型	负载
电容-电感类	荧光灯(10 W、20 W、45 W)、空调、调光灯、真空吸尘器
电阻类	吹风机(1 500 W)、烧水壶
开关类	手机、电脑、吹风机(800 W)

1.3 数据预处理与分析

首先,需要对正常工作的电流和故障电弧电流进行分类,分别标签为1和0。以采样点数200为一个周期,并对每个周期的电流波形进行归一化处理。为了简单起见,取5个周期对不同负载的正常状态和故障状态时电流波形进行分析。

图2为电容-电感类负载(荧光灯、空调、调光灯)正常工作和发生电弧故障时的电流波形,由图2可知,在发生电弧故障时,电容-电感类负载的电流特征并不一致,各有各的特点:1)调光灯、荧光灯10 W会出现高频毛刺,但空调的高频毛刺现象并不明显。2)3种负载的电流波形的周期性都被破坏,

电流波动的幅值区间增大,并且在部分区间电流变化速度加快,与正常工作时的电流波形有明显的区别。

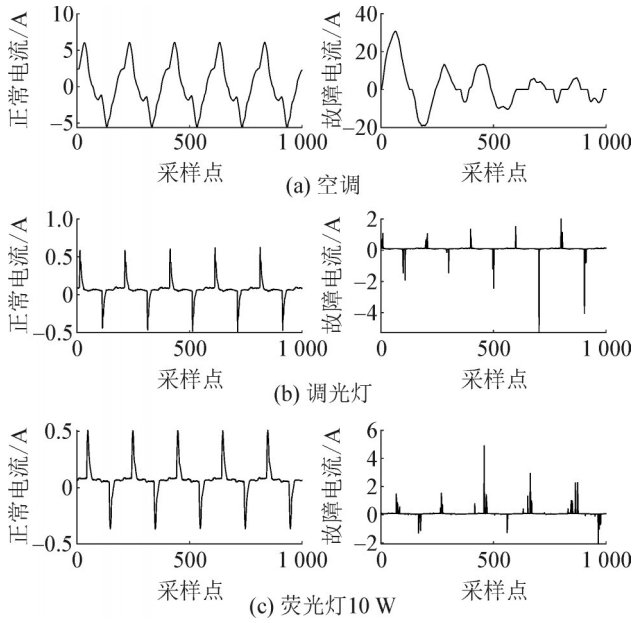


图2 电容-电感类负载电流波形
Fig. 2 Capacitive-inductive load current waveform

图3为开关类负载(手机、电脑)正常工作和发生电弧故障时的电流波形,由图3可知,与电容-电感类负载类似,开关类负载在正常工作时其电流波形在零点附近也会出现“平肩部”,电流波形发生畸变,高频毛刺现象明显。

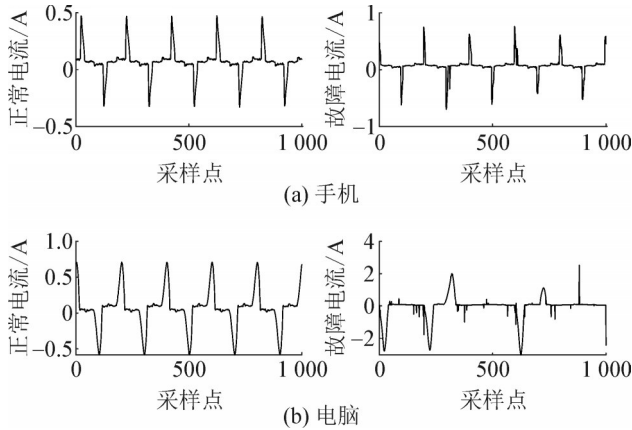


图3 开关类负载电流波形
Fig. 3 Current waveform of switching load

图4为电阻类负载(吹风机1500W、烧水壶)正常工作和发生电弧故障时的电流波形,由图4可知,在发生电流故障时,电阻类负载也会出现高频毛刺,但在时域上出现的零休现象相较于其他负载来说更加明显,并且其电流波动的幅度也没有其他负载种类大。

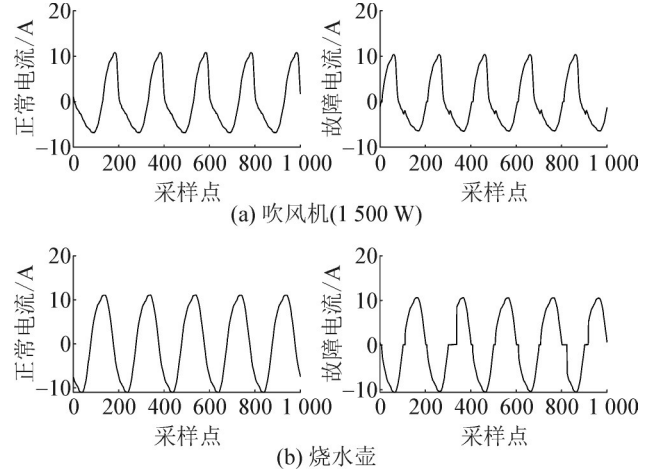


图4 电阻类负载电流波形
Fig. 4 Current waveform of resistive load

2 电弧故障特征的提取与选择

对于电流的时域、频域特征在电弧故障发生前后的变化比较显著的特点,本文从这两个方面进行特征提取,并且为了降低评估器的复杂度和提升其泛化性能,需要在提取的特征集的基础上进行特征选择,减少信息的冗余。

2.1 时域特征的提取

本文提取的时域特征主要包括电流变化速率、电流平均值的绝对值、短时过零率、零休时间、样本熵、香农熵。

在发生电弧故障前后,电流瞬时值的变化速率一般会发生变化,因此,以相邻样本点的数据做差,通过计算其最大值和最小值来表示电流变化速率^[16],其表达式为

$$\begin{cases} V_{\max} = \max[l_k - l_{k-1}] \\ V_{\min} = \min[l_k - l_{k-1}] \end{cases} \quad (1)$$

式中, l_k 为第 k 个采样点的电流瞬时值。

电弧故障发生时,周期内的电流信号波形通常会发生畸变,而电流平均值的绝对值可以反映一个周期内电流波形的不规则程度^[16]。设每个周期的采样点数为 N ,其表达式为

$$I_{\text{avg}} = \left| \frac{1}{N} \sum_{k=0}^{N-1} l_k \right| \quad (2)$$

短时过零率是相邻离散信号的符号改变次数与窗宽的比值,发生电弧故障时,随机性增强,短时过零率增大^[8]。将窗宽固定为一个周波,用时窗内过零次数代替短时过零率,其表达式为

$$R(n) = \sum_{m=-\infty}^{\infty} \{|\text{sgn}[s(m)] - \text{sgn}[s(m-1)]|w(n-m)\} \quad (3)$$

式中: $s(n)$ 是电流信号波形, $n \leq N$; N 为电流信号 $s(n)$ 的长度。 $\text{sgn}(x)$ 为符号函数,公式为

$$\text{sgn}(x) = \begin{cases} +1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases} \quad (4)$$

而 $w(x)$ 为垂直窗口函数,其表达式为

$$w(x) = \begin{cases} 1, & 0 \leq x \leq X \\ 0, & x < 0 \text{ 或 } x > X \end{cases} \quad (5)$$

零休时间是指每个周期内电流波形在零点附近电流瞬时值连续等于零的时间,通常以“平肩部”表现出来。串联电弧故障下的零休时间一般会有所变化。将周期内连续处于零点附近的采样点数表示零休时间,其表达式为

$$t_{\text{zero}} = \sum_{k=1}^{N-1} \text{dif}_k \quad (6)$$

$$\begin{cases} \text{dif}_k = 1, & |l_k| < 0.3 A \text{ 且 } |l_k - l_{k-1}| < 0.15 A \\ \text{dif}_k = 0, & \text{其他} \end{cases}$$

式中, $\text{dif}_k = 1$ 表示该采样点处于零休状态, $\text{dif}_k = 0$ 表示该采样点不处于零休状态。

样本熵和香农熵能充分反映电弧故障发生后复杂度的变化。样本熵计算的具体步骤可参照文献[17]。香农熵的表达式为

$$H(x) = - \sum_{i=1}^n p_i \log p_i \quad (7)$$

式中, $\{p_i\}$ 为离散信号的概率分布, $p_i > 0, i = 1, 2, \dots, n$ 。

2.2 频域特征的提取

本文提取的频域特征主要包括谐波、谐波因素、总谐波畸变率、功率谱熵、子代能量比、频率质心。

快速傅里叶变换(FFT)是离散傅里叶变换的快速算法,可用来求解电流信号序列的频谱。对每个周期的采样信号进行快速傅里叶变换,再对其求模,可得到每个周期的各谐波分量的幅值,若用 D_k 表示各谐波分量的幅值,则 D_0 和 D_1 分别为直流分量和基波的幅值。一般情况下,对于一个周期长度为 N 的采样信号来说,只需求 $0 \sim N/2-1$ 次谐波分量即可。

谐波因素(H_k)为各次谐波分量有效值与基波分量有效值的比值,表示各次谐波相对基波的能量权重;总谐波畸变率(THD)为全部谐波含量均方根值与基波均方根值之比,可以描述电弧故障发生前后畸变波形偏离正弦波形的程度,用百分数表示^[16];频率质心(w_c)的作用类似于物理上的质量之心,反映了电流波形频率发布的中心。谐波因素、总谐波畸变率和频率质心的表达式分别为

$$H_k = \frac{D_k}{D_1}, k = 2, 3, \dots, N/2 - 1 \quad (8)$$

$$\text{THD} = \sqrt{\sum_{k=2}^{N/2-1} \left(\frac{D_k}{D_1}\right)^2} \quad (9)$$

$$w_c = \frac{\sum_{k=0}^{N/2-1} D_k^2 k}{\sum_{k=0}^{N/2-1} D_k^2} \quad (10)$$

式中: D_k 为第 k 次谐波分量的幅值; D_0 和 D_1 分别为直流分量和基波的幅值。

子代能量比为频域中各子代能量与总能量的比值。基于奈奎斯特频率,将频域划为4个子代区间,用谐波次数表示为 $s_0 \in [0, \frac{N}{16})$, $s_1 \in [\frac{N}{16}, \frac{N}{8})$, $s_2 \in [\frac{N}{8}, \frac{N}{4})$, $s_3 \in [\frac{N}{4}, \frac{N}{2})$,则子代能量比 SR_i 的表达式为

$$SR_i = \frac{\sum_{k \in s_i} D_k^2}{\sum_{n=0}^{N/2-1} D_n^2} \quad (11)$$

功率谱熵是依据时域信号通过傅里叶变换,在时域变换到频域的过程中遵循能量守恒原则的一种能量描述^[17-18],其表达式为

$$H_p = - \sum_{i=1}^n q_i \lg q_i$$

$$q_i = \frac{S_i}{\sum_{i=1}^n S_i} \quad (12)$$

$$S_i = |y_i|^2$$

式中: y_i 为时域信号的傅里叶变换; S_i 为频率谱序列 y_i 在对应频率处的功率谱值; q_i 为第 i 个功率谱在整个功率谱中所占的百分比。

2.3 特征分析

在发生电弧故障前后,不同负载种类的特征取值变化一般不同,而且即便相同的负载类型也可能存在差异。因此,为了研究负载种类对特征取值的影响,对冰箱(电容-电感类)、电脑(开关类)、烧水壶(电阻类)这3种典型负载发生电弧故障前后的时域和频域特征取值进行分析。

2.3.1 时域特征分析

3种典型负载在发生电弧故障前后的零休时间(T_{zero})、电流平均值的绝对值(I_{avg})、香农熵(H)以及电流变化速率最大值(V_{max})的变化曲线见图5。图中,后10个周期表示发生电弧故障,左纵轴表示冰箱和电脑相应的时域特征,右纵轴表示烧水壶相应的时域特征。结合电流正常工作和故障时的波形

特征,可以得到以下结论:

1)在发生电弧故障时,不同负载的时域特征取值变化各有特点:3种典型负载的电流平均值的绝对值和电流变化速率最大值都普遍变大,香农熵普遍变小;烧水壶的零休时间显著增大,但由于电脑、冰箱这两类负载在正常工作时也存在“平肩部”,因此其零休时间的变化是上下波动的;

2)3种典型负载在正常工作状态时,其时域特征的取值一般不同,在发生电弧故障后,其时域特征的变化范围也有差异。也就是说,负载种类对时域特征的取值有很大影响;

3)对于给定负载,发生电弧故障后的时域特征取值区间有可能会与正常工作时的取值区间发生重叠。因此,单一的时域特征不适合直接作为区分电路状态的依据;

4)对于不同的时域特征,在发生电弧故障后,其波动的幅度存在差别。

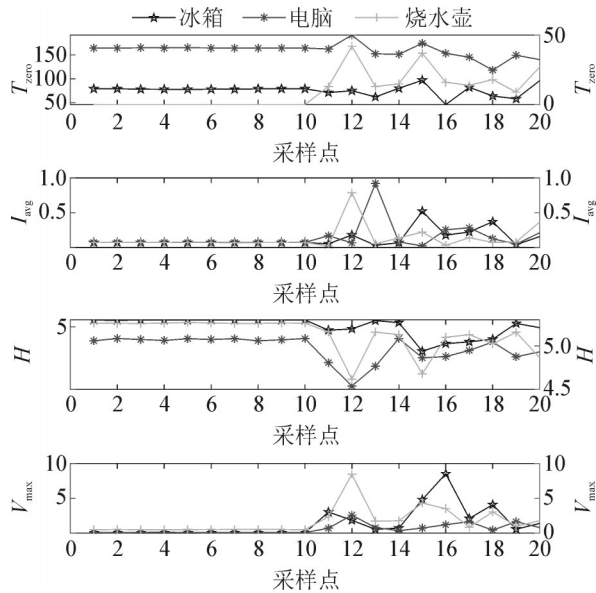


图5 不同负载下的时域特征图
Fig. 5 Time domain characteristic diagram under different loads

2.3.2 频域特征分析

3种典型负载在第10个周期发生电弧故障后总谐波畸变率(THD)、频率质心(W_c)、子代能量比(SR_2)以及谐波因素(H_2 、 H_3)的变化曲线见图6,结合电流正常工作和故障时的波形特征,可以得到以下结论:

1)在发生电弧故障时,不同负载的频域特征取值变化各有特点:对于电脑、冰箱这两类负载,其频域特征波动明显,而对于烧水壶这类负载来说,其总谐波畸变率、频率质心和子代能量比的波动并不明显,并且不同负载的频域特征值也存在差异;

2)对于给定负载,发生电弧故障后的频域特征取值区间也有可能会与正常工作时的取值区间发生重叠。因此,无法通过单一频域特征对电路状态进行区分。

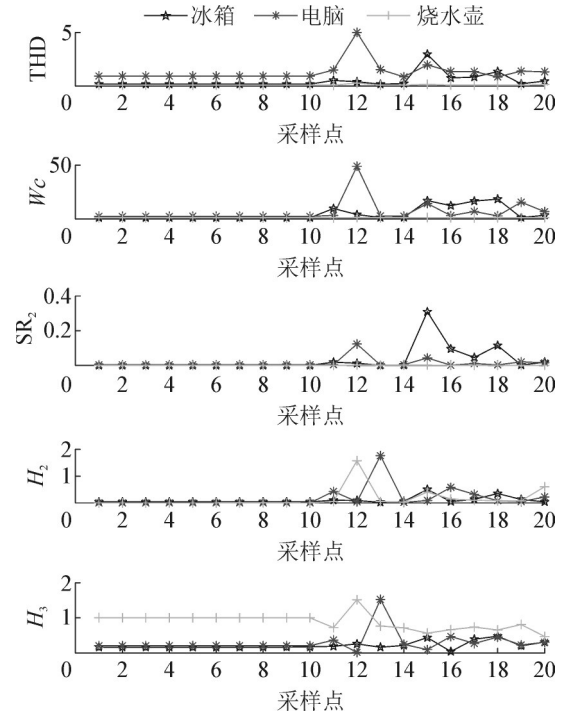


图6 不同负载下的频域特征
Fig. 6 Frequency domain characteristics under different loads

2.4 特征选择

在进行一个分类任务时,经过一系列方法的处理得到原始特征集。但是,在原始特征集中,除了存在一些对分类任务是有用、关键的特征外,还会存在一些对分类任务没有什么用的特征,称为“无关特征”。“无关特征”会造成诊断模型结构冗余,收敛速度较慢等问题。特征选择是特征工程的一个重要环节,其目标就是基于某种方法从原始特征集寻找最优特征子集。一方面,特征选择能够去除“无关特征”,在提高分类模型泛化性能的同时减少运行时间。另一方面,特征选择是为了防止分类模型出现过拟合问题:能够很好的对训练集数据进行分类,但在测试集中的表现却不尽人意。

“信息熵”是度量样本集合纯度最常用的一种指标^[19]。假定当前样本集合 D 的第 k 类样本所占的比例为 $p_k(k=1,2,\dots,n)$,则 D 的信息熵定义为

$$Ent(D) = -\sum_{k=1}^n p_k \log_2 p_k \quad (13)$$

若某个特征 a 有 V 个不同的取值,假定为 $\{a^1, a^2, \dots, a^V\}$,如果使用该特征来划分样本集 D ,那么 V 个 a 的取值可将 D 划分为 V 个样本子集,也就

是产生 V 个分支节点,则第 V 个分支节点所包含的样本为 D 中在 a 上取值为 a^v 的样本,记为 D^v 。 D^v 的信息熵可用式(13)计算得到。用特征 a 对样本集合 D 进行划分后的“信息增益”定义为 D 的经验熵与特征 a 给定的条件下 D 的经验熵之差公式为

$$Gain(D,a)=Ent(D)-\sum_{v=1}^V\frac{|D^v|}{D}Ent(D^v)$$

(14)

信息增益表示一个特征对数据样本的分类能力。通常一个特征的信息增益越小,用该特征进行节点的划分会导致“纯度降低”,不利于数据样本的分类。

事实上,信息增益准则对特征值较多的特征情有独钟。为了避免这种情况,需要另一种计算方式对信息增益进行矫正,这就是增益率。与信息增益不同的是,增益率准则是偏向于可取值数目较少的特征。增益率的定义如下

$$Gain_ratio(D,a)=\frac{Gain(D,a)}{IV(a)}$$

$$IV(a)=-\sum_{v=1}^V\frac{|D^v|}{D}\log_2\frac{|D^v|}{D}$$

(15)

一般的,单个特征评估器不同,评估特征重要性的角度也不同。信息增益准则对特征值较多的特征情有独钟,而增益率准则与此相反。因此,C4.5 算法并没有直接在信息增益准则和增益率准则中选其一,而是通过对两种准则的某种结合方法来确定最大的候选划分特征。本文将基于信息增益和增益率的这两个单个属性评估器结合起来,对特征进行更全面、更准确的评价。具体步骤为:首先,将单个属性评估器结果归一化到[0, 1]区间,然后,将这两个评估器对每个特征的评分分别进行平均,最后得到每个特征的得分。

2.5 特征选择结果

在进行特征选择前,为了消除每种负载的每个特征的数值大小和量纲不同对结果的影响,需要对每种负载的特征进行归一化。特征选择结果见表 2。表 2 给出了每个特征的信息增益(Info Gain)、增益率(Gain Ratio)、综合得分(Score)以及按综合得分降序的排名(Rank)。

本文设定选取 15 个特征用于后续模型的训练,因此选择得分排名前 15 的特征作为模型的输入量,包括时域特征中的短时过零率 R ,零休时间 t_{zero} ,电流平均值的绝对值 $I_{|avg|}$,以及频域中的谐波 D_9 、 D_{15} 、 D_7 、 D_{11} 、 D_{13} 、 D_{33} ,谐波因素 H_9 、 H_{15} 、 H_{55} 、 H_{31} 、 H_{37} ,总谐波畸变率 THD。

表 2 特征选择结果
Table 2 Results of feature selection

Index	Info Gain	Gain Ratio	Score	Rank
R	0.830 8	0.211 0	2.000 0	1
D_{15}	0.436 9	0.058 7	0.718 3	2
H_9	0.407 6	0.055 8	0.665 5	3
D_9	0.405 4	0.054 6	0.656 6	4
$I_{ avg }$	0.391 6	0.052 9	0.629 9	5
H_{15}	0.389 6	0.052 5	0.625 3	6
THD	0.384 9	0.052 7	0.620 1	7
H_{55}	0.382 2	0.052 9	0.617 8	8
H_{31}	0.379 3	0.051 9	0.608 7	9
D_7	0.375 3	0.050 8	0.597 9	10
D_{11}	0.374 5	0.050 0	0.593 3	11
T_{zero}	0.066 4	0.129 1	0.587 9	12
H_{37}	0.368 4	0.050 5	0.587 5	13
D_{33}	0.369 7	0.049 8	0.585 6	14
D_{13}	0.367 6	0.049 1	0.579 8	15
H_{19}	0.363 4	0.049 8	0.577 5	16
H_{33}	0.362 5	0.049 4	0.574 1	17
H_{75}	0.358 4	0.050 4	0.573 9	18
H_{39}	0.359 9	0.049 4	0.570 8	19
H_{73}	0.352 6	0.049 3	0.560 9	20

3 多特征融合随机森林串联电弧故障识别方法

目前广泛应用于电弧故障检测的 BP 神经网络存在收敛速度慢和精度差等缺点,因此,为了解决该问题,本文基于随机森林算法,并结合粒子群算法对随机森林的参数进行优化,提出了随机森林粒子群优化算法识别串联电弧故障的检测流程。

3.1 串联电弧故障检测随机森林粒子群优化模型

随机森林是以决策树为估计器的 Bagging 算法,他将多个决策树结合起来,具有参数少、不易过拟合和泛化能力强的优点^[20]。与传统决策树不同的是,随机森林中的决策树的每一次划分并未用到当前节点的特征集合(假定有 M 个特征),而是从该节点的特征集中随机选取出一定数量(假定 m 个)的特征子集,再从该子集中选择一个最优特征进行划分。一般情况下, m 取 $\log_2 M$ 或 $\sqrt{M}$ 比较合适。与 Bagging 算法一样,随机森林算法也需要使用投票策略。投票策略是一种遵循少数服从多数原则的方法,即将多个弱分类器的结果进行汇总,然后取其分类类别的众数作为最终的分类结果来提高模型的泛化能力和分类准确率。随机森林的输出结果表达式为

$$R(x)=\frac{1}{L}\sum_{i=1}^Ly_i\tag{16}$$

式中： L 为随机森林中决策树的棵树， y_i 是每棵决策树的输出结果， $R(x)$ 为随机森林的输出结果。

在构建随机森林算法模型时，决策树棵树和最小叶子数是需要确定的两个重要参数。一般来说，决策树棵数越多、最小叶子数越小能够显著提升随机森林的分类准确率，但同时也增加了随机森林的复杂度，导致随机森林过拟合，降低了其性能。对此，本文利用粒子群优化算法^[21-23]，对随机森林的决策树棵数、最小叶子数进行优化，降低随机森林的复杂度。粒子群优化算法源于对鸟类觅食行为的观察研究，通过寻找最优种群收敛到全局最优，能够很好地对随机森林的决策树棵树和最小叶子数进行优化，从而提高随机森林模型分类的准确率。本文采用的随机森林粒子群优化算法的主要步骤为

- 1)确定算法的初始参数，随机设定决策树的棵树、最小叶子数，并默认随机特征个数 m 为 $\sqrt{M}$ ；
- 2)以分类准确率为适应度值，利用粒子群优化算法搜寻最优决策树棵树 L 和最优叶子数 $MINLS$ ；
- 3)将样本集划分为训练集和测试集；
- 4)对训练集采用 Bootstrap 生成 L 个采样集，每个采样集用于训练生成一棵决策树；
- 5)在决策树的每个节点分裂时，对该节点的特征集进行无放回的抽样，选取 m 个特征，只保留该 m 个特征对应的数据作为训练样本，若划分节点包含的样本数小于最小叶子数 $MINLS$ 时，决策树停止分裂；
- 6)训练生成 L 棵决策树，通过公式(16)得到模型最后的分类结果。

3.2 随机森林粒子群优化算法识别串联故障电弧的流程

首先，收集典型负载在正常工作时的和发生电弧故障时的电流波形，并提取电流的时域、频域特征，并对时域、频域特征进行归一化处理。然后，利用2.4节中的特征选择方法对得到的串联电弧特征指标进行选择，构建电弧特征指标集，作为分类模型的输入特征。最后，利用电弧特征指标集对分类模型进行训练和测试。

4 结果与分析

基于混淆矩阵常用的测试指标有准确率、查全率以及查准率，根据这些指标，可以对分类模型的

性能进行测试。混淆矩阵的构造见表3。准确率(Accuracy)的定义为分类正确的样本数与总样本数之比；查全率(Recall)的关注点在于故障样本的漏判，即有多少故障样本没有被检测出来，其定义为真负类数与负类样本数之比；查准率(Precision)的关注点在于故障样本的误判，即有多少故障样本被判错，其定义为真负类数与预测得到的负类总数之比；这3个指标的表达式分别为

$$Accuracy=\frac{TP+TN}{TN+FN+FP+TP}\tag{17}$$

$$Recall=\frac{TP}{TP+FP}\tag{18}$$

$$Precision=\frac{TP}{TP+FN}\tag{19}$$

表3 混淆矩阵
Table 3 Confusion matrix

预测类		真实类
0(故障)	1(正常)	
TP	FP	0(故障)
FN	TN	1(正常)

为了更好地显示本文随机森林粒子群优化模型对混合负载的诊断能力，与传统随机森林、BP神经网络进行比较。基于这3种算法构建的分类器的训练数据均采用本文所提取的特征集进行性能测试，其中，BP神经网络学习率为0.01，训练次数为500。这3种分类器的测试指标及其比较见表4。从表4可知，由于训练分类器时使用的数据为本文提取并且经过选择后的特征指标，因此分类器的各项测试指标较高，并且相较于传统随机森林和BP神经网络而言，本文方法对混合负载测试集具有更好准确率、查全率以及查准率。

表4 优化前后测试指标对比
Table 4 Comparison of test data before and after optimization

算法	准确率/%	查全率/%	查准率/%
传统随机森林	98.89	98.5	99.3
本文方法	99.59	99.4	99.8
BP神经网络	98.56	98.0	99.1

另外，为了更好地显示本文随机森林粒子群优化算法对各类负载的诊断能力，选取未经过训练的每种负载正常工作状态和电弧故障状态的数据各50组作为测试数据，与传统随机森林、BP神经网络进行比较，计算不同分类器对不同负载的准确率、查全率、查准率，如表5-7所示。

由表5-7可知，本文使用的随机森林粒子群优化模型与传统随机森林以及BP神经网络相比，对

大多数负载都有更好的准确率、查全率、查准率,可以更好的应用于串联故障电弧的有效识别。

表5 3种分类器对不同负载的准确率
Table 5 Accuracy rate of three classifiers for different loads

负载	随机森林/%	本文方法/%	BP神经网络/%
空调	100	100	99
冰箱	100	100	100
吹风机热风	98	100	96
电脑	100	100	100
调光灯	97	98	97
烧水壶	100	100	99
荧光灯 10 W	98	99	96
荧光灯 20 W	100	99	100
真空吸尘器	100	100	100

表6 3种分类器对不同负载的查全率
Table 6 Recall of three classifiers for different loads

负载	随机森林/%	本文方法/%	BP神经网络/%
空调	100	100	100
冰箱	100	100	100
吹风机热风	98	100	92
电脑	100	100	100
调光灯	98	98	96
烧水壶	100	100	98
荧光灯 10 W	96	98	96
荧光灯 20 W	100	98	100
真空吸尘器	100	100	100

表7 3种分类器对不同负载的查准率
Table 7 Accuracy rate of three classifiers for different loads

负载	随机森林/%	本文方法/%	BP神经网络/%
空调	100	100	98.4
冰箱	100	100	100
吹风机热风	98	100	100
电脑	100	100	100
调光灯	96.08	98	97.96
烧水壶	100	100	100
荧光灯 10 W	100	100	96
荧光灯 20 W	100	100	100
真空吸尘器	100	100	100

5 结语

1)采用了电流的时域、频域特征作为特征量输入随机森林粒子群优化模型,相较于单一的频域或时域特征,时域、频域特征的融合能够更全面地对串联电弧故障进行诊断识别。

2)在随机森林训练模型中,采用粒子群算法对

随机森林的参数进行优化,能够防止随机森林过拟合,提高模型的分类准确率。

3)提出的多特征融合随机森林粒子群优化算法相较于传统随机森林和BP神经网络具有更高的准确率、查全率和查准率,对串联电弧故障的检测具有一定的价值。

参考文献

[1] 赵长征. 电气火灾防治与调查技术[M]. 沈阳:辽宁大学出版社,2009.

[2] 刘晓明,赵洋,曹云东,等. 基于多特征融合的交流系统串联电弧故障诊断[J]. 电网技术,2014,38(3):795-801.
LIU Xiaoming, ZHAO Yang, CAO Yundong, et al. Multi-feature fusion based diagnosis of series arc faults in AC system[J]. Power System Technology, 2014,38(3): 795-801.

[3] 李佳威. 低压配电系统串联故障电弧诊断方法研究[J]. 电子制作,2014(12):195-195.
LI Jiawei. Research on arc diagnosis method of series fault in low-voltage distribution system[J]. Practical Electronics, 2014(12): 195-195.

[4] 廖水容,张认成,李夏河. 低压串联电弧故障电流高次谐波含有率试验[J]. 河南理工大学学报(自然科学版),2013,32(2):179-182.
LIAO Shuirong, ZHANG Rencheng, LI Xiahe. Experiment on current high order harmonic ratio for series low voltage arc fault[J]. Journal of Henan Polytechnic University (Natural Science), 2013,32(2): 179-182.

[5] 王尧,韦强强,葛磊蛟,等. 基于电弧电流高频分量的串联交流电弧故障检测方法[J]. 电力自动化设备,2017,37(7):191-197.
WANG Yao, WEI Qiangqiang, GE Leijiao, et al. Series AC arc fault detection based on high-frequency components of arc current[J]. Electric Power Automation Equipment, 2017,37(7): 191-197.

[6] LEE W J, SAHNI M, METHAPRAYOON K, et al. A novel approach for arcing fault detection for medium-/low-voltage switchgear[J]. IEEE Transactions on Industry Applications, 2009,45(4): 1475-1483.

[7] 朱森,孙鹏. 基于周波分析法的串联型故障电弧诊断技术[J]. 电气技术,2016(2):46-48.
ZHU Sen, SUN Peng. The series fault arc detection method based on cycle analysis[J]. Electrical Engineering, 2016(2): 46-48.

[8] 刘晓明,徐叶飞,刘婷,等. 基于电流信号短时过零率的电弧故障检测[J]. 电工技术学报,2015,30(13):125-133.
LIU Xiaoming, XU Yefei, LIU Ting, et al. The arc fault detection based on the current signal short time Zero crossing rate[J]. Transactions of China Electrotechnical Society, 2015,30(13): 125-133.

- [9] 刘晓明,王丽君,赵洋,等. 串联故障电弧检测方法的研究[J]. 电气开关,2014,52(1):13-18.
LIU Xiaoming, WANG Lijun, ZHAO Yang, et al. Research on the detection method of series arc fault[J]. Electric Switcher, 2014, 52(1): 13-18.
- [10] 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.
- [11] MÜLLER P, TENBOHLEN S, MAIER R, et al. Characteristics of series and parallel low current arc faults in the time and frequency domain[C]//2010 Proceedings of the 56th IEEE Holm Conference on Electrical Contacts. [S.l.]: IEEE, 2010: 1-7.
- [12] 龙官微,穆海宝,张大宁,等. 基于多特征融合神经网络的串联电弧故障识别技术[J]. 高电压技术,2021,47(2):463-471.
LONG Guanwei, MU Haibao, ZHANG Daning, et al. Series arc fault detection technology based on multi-feature fusion neural network[J]. High Voltage Engineering, 2021, 47(2): 463-471.
- [13] 张士文,张峰,王子骏,等. 一种基于小波变换能量与神经网络结合的串联型故障电弧辨识方法[J]. 电工技术学报,2014,29(6):290-295.
ZHANG Shiwen, ZHANG Feng, WANG Zijun, et al. Series arc fault identification method based on energy produced by wavelet transformation and neural network[J]. Transactions of China Electrotechnical Society, 2014, 29(6): 290-295.
- [14] 殷浩楠,竺红卫,丁鑫,等. 一种基于代价敏感学习的故障电弧识别方法[J]. 传感器与微系统,2018,37(4):13-16.
YIN Haonan, ZHU Hongwei, DING Xin, et al. An arc-fault recognition method based on cost-sensitive learning[J]. Transducer and Microsystem Technologies, 2018, 37(4): 13-16.
- [15] 吴春华,冯夏云,袁同浩,等. 基于BP神经网络的光伏故障电弧检测方法研究[J]. 太阳能学报,2016,37(11):2958-2964.
WU Chunhua, FENG Xiayun, YUAN Tonghao, et al. Photovoltaic arc fault detection method based on BP neural network[J]. Acta Energiae Solaris Sinica, 2016, 37(11): 2958-2964.
- [16] 殷浩楠. 低压故障电弧特征分析与模式识别[D]. 杭州:浙江大学,2017.
- [17] 崔芮华,李泽,佟德栓. 基于三维熵距和熵空间的航空电弧故障检测与分类技术[J]. 电工技术学报,2021,36(4):869-880.
CUI Ruihua, LI Ze, TONG Deshuan. Arc fault detection and classification based on three-dimensional entropy distance and entropy space in aviation power system[J]. Transactions of China Electrotechnical Society, 2021, 36(4): 869-880.
- [18] 谢平,陈矛,祝刚. 基于功率谱熵距的航空发电机转子转动故障诊断方法研究[J]. 电子测量技术,2018,41(21):99-102.
XIE Ping, CHEN Mao, ZHU Gang. Research on rotor rotation fault diagnosis method of aero-generator based on power spectrum entropy distance[J]. Electronic Measurement Technology, 2018, 41(21): 99-102.
- [19] 周志华. 机器学习[M]. 北京:清华大学出版社,2016:147-169.
- [20] 武炜杰. 随机森林算法的应用与优化方法研究[D]. 无锡:江南大学,2021.
- [21] 程学新. 粒子群优化加权随机森林算法研究[D]. 郑州:郑州大学,2017.
- [22] 唐俊. PSO算法原理及应用[J]. 计算机技术与发展,2010,20(2):213-216.
TANG Jun. Principle and application of PSO algorithm[J]. Computer Technology and Development, 2010, 20(2): 213-216.
- [23] 张丽萍. 粒子群优化算法的理论及实践[D]. 杭州:浙江大学,2005.
- 作者简介:
郭敏(1984—),男,高级工程师,主要研究方向为电力电子、电能质量和无功电压分析与控制技术。E-mail:329106373@qq.com。
郭小璇(1986—),女,高级工程师,通信作者,主要研究方向为需求侧管理及电能替代技术。E-mail:guo_xx.sy@gx.csg.cn。
吴宁(1991—),男,工程师,主要研究方向为综合能源技术应用研究。E-mail:wu_n.sy@gx.csg.cn。
卢健斌(1991—),男,工程师,主要研究方向为智慧用能技术。E-mail:lu_jb.sy@gx.csg.cn。
陈卫东(1983—),男,高级工程师,主要研究方向为微电网及电能质量的分析与控制技术。E-mail:chen_wd.sy@gx.csg.cn。