

基于改进 MFCC 和 RF 的配电网故障电弧声纹识别方法

陈凯¹, 吴田¹, 杨莹², 何清³, 黎鹏¹, 普子恒¹

(1. 三峡大学电气与新能源学院, 湖北宜昌 443002; 2. 国网湖北送变电工程有限公司, 武汉 430063; 3. 国网湖北电力科学研究院, 武汉 430015)

摘要: 针对电力系统中配电网线路中的电弧故障检测易发生故障误判的问题, 提出了一种基于改进梅尔倒谱系数和随机森林算法的故障电弧声纹识别模型。该模型首先对故障电弧声信号进行分帧加窗处理, 提取其梅尔倒谱系数, 使用费舍尔比计算特征各维度对区分故障的贡献度, 根据贡献度设计出合适的权值对特征向量进行加权降维处理, 得到区分性更强的特征向量。最后采用随机森林算法对故障电弧声纹信号进行识别。为了验证该模型的有效性, 文中搭建故障电弧试验平台采集故障电弧声信号, 对故障电弧燃烧发展过程的声信号时频域进行分析, 确定了声信号特征与电弧燃烧程度之间存在相关性, 验证了以声信号作为故障电弧检测参量的可行性, 并对不同燃烧程度的电弧声信号使用故障电弧声纹识别模型进行计算分析。结果表明, 基于改进梅尔倒谱系数和随机森林算法的故障电弧声纹识别模型对故障电弧声信号的识别结果符合预期, 该系统可为配电网电弧故障检测提供一种新方法。

关键词: 配电网; 故障电弧; 声纹识别; 梅尔倒谱系数; 随机森林算法

Fault Arc Voiceprint Recognition Method in Distribution Network Based on Improved MFCC and RF

CHEN Kai¹, WU Tian¹, YANG Ying², HE Qing³, LI Peng¹, PU Ziheng¹

(1. College of Electrical Engineering and New Energy, China Three Gorges University, Hubei Yichang 443002, China; 2. State Grid Hubei Transmission & Transformation Engineering Co., Wuhan 430063, China; 3. State Grid Hubei Electric Power Research Institute, Wuhan 430015, China)

Abstract: To address the issue of frequent false positives in arc fault detection within power system distribution network lines, a model based on improved Mel-frequency cepstral coefficients (MFCC) and random forest algorithm for arc fault sound pattern recognition is proposed. The model first processes the arc fault sound signals by framing and windowing, then extracts their MFCCs. Using the Fisher ratio, the contribution of each feature dimension to fault differentiation is calculated. Based on these contributions, appropriate weights are designed to perform weighted dimensionality reduction on the feature vectors, yielding more discriminative feature vectors. Finally, the random forest algorithm is employed to recognize the arc fault sound patterns. To validate the effectiveness of this model, an arc fault test platform was constructed to collect arc fault sound signals. Analysis of the time-frequency domain of these signals during the arc fault combustion process established a correlation between the sound signal features and the degree of arc combustion, confirming the feasibility of using sound signals as parameters for arc fault detection. The model was then applied to sound signals from different combustion stages of the arc fault for calculation and analysis. Results indicate that the fault arc sound pattern recognition model, based on improved MFCC and random forest algorithm, achieves expected recognition outcomes. This system offers a novel method for detecting arc faults in distribution networks.

Key words: distribution network; arc fault; voiceprint recognition; Mel-frequency cepstrum coefficient; random forest algorithm

0 引言

随着新型电力系统的大力发展, 如何提高配电网的安全可靠性成为了目前的研究热点之一^[1-5]。电弧故障是配电网中常见的故障之一, 故障电弧产生

时会释放巨大的热量, 易引燃周围可燃物品, 发生电气火灾^[6-10], 影响电气设备的安全运行, 造成不必要的人身伤害和财产损失, 快速有效地检测识别

基金项目: 国家自然科学基金资助项目(51807110)。

Project Supported by National Natural Science Foundation of China(51807110).

故障电弧对提高电力系统的可靠性有一定的研究意义。

传统的故障电弧检测方法主要是基于故障电弧的电压和电流信号时频域阈值^[11-15]，其属于接触式采集，对于不同拓扑结构的线路，普适性较差，易发生故障误判等问题。与电压电流信号检测方法相比，声信号采集设备安装简单、成本低，具有非接触式采集的优势，不易受到不同拓扑结构的线路影响。目前配电网电弧故障检测方法主要采取检测故障电弧的电流电压信号，而故障电弧在发生时，往往会伴随产生着巨大的声、光、热等能量，这些物理现象中包含着丰富的信息，它们与电弧燃烧的发展有着一定的联系，融合故障电弧弧声的特征信息，将电弧弧声信号作为故障电弧是否发生的评判参数，对故障进行检测有一定的可行性。

声纹识别技术是一种通过提取声音信号中的特征来识别待测语音的技术^[16-19]。基本原理是输入声音样本，提取出该声音样本中特定的特征向量，进行模式识别模型训练，建立起相应的数据库。在需要进行语音识别时，将待测数据与已有模型库进行匹配，从而得出识别结果。

声纹识别技术主要分为特征提取和模式识别两部分，利用声纹识别技术对故障电弧进行故障检测最为关键的一步是对于故障声音信号的特征提取。不同的特征向量组，对于识别结果的识别速度和识别率有较大的影响，常用的声音信号特征向量主要有线性预测倒谱系数(linear prediction cepstrum coefficient, LPCC)和梅尔频率倒谱系数(mel frequency cepstrum coefficient, MFCC)和伽马通频率

倒谱系数(gammatone frequency cepstrum coefficient, GFCC)等。目前关于故障电弧声纹识别的研究成果较少，有学者将声纹识别技术用于变压器故障检测中并取得了一定的效果，文[20]运用长短期记忆网络(long short-term memory, LSTM)建立模型，通过提取变压器正常状态、过载和放电状态的声信号的梅尔倒谱系数，对3种状态进行了分类识别，结果表明该方法具有较高的准确性。文[21]运用鲸鱼优化算法(whale optimization algorithm, WOA)对随机森林算法(random forest algorithm, RF)进行优化并建立模型，通过提取变压器正常状态、典型机械故障的GFCC特征向量，进行分类识别，结果表明该方法具有较好的识别结果和优良的抗噪性和鲁棒性。文[22]运用顺序最小优化(sequential minimal optimization, SMO)对支持向量机(support vector machine, SVM)进行优化并建立模型，通过提取不同绝缘类型缺陷声信号的MFCC特征向量，对不同绝缘缺陷类型进行分类识别，结果表明，该方法具有较高的准确性和敏感性。其中MFCC是目前较为主流的声纹特征，但MFCC特征是基于人声信号特征进行设计的，故障电弧弧声与人声存在较大的差异，直接将MFCC应用于故障电弧弧声识别很难准确提取到较为有效的特征信号，导致识别效果有限，故仍需要对其进行研究。

文中针对配电网电弧故障检测易发生故障误判的问题，提出了一种故障电弧声纹识别模型，搭建了故障电弧试验平台，采集故障电弧声信号，使用目前应用最广泛的MFCC特征向量并针对故障电弧声信号与人声信号的差异性对其进行了基于费舍尔

(Fisher)比的加权降维优化, 使用随机森林算法(random forest algorithm, RF)对故障电弧声信号进行分类识别, 并与不同分类模型进行比较, 对不同信噪比下的识别率进行抗噪性测试, 验证该模型的有效性 with 抗噪性。

1 声纹识别模型及改进

1.1 声纹识别框架

文中基于 MFCC 特征向量, 针对故障电弧放电声信号提出了一种故障电弧声纹识别模型, 模型流程图见图 1。具体流程为先对故障电弧放电声信号进行预处理, 然后进行 MFCC 特征向量提取, 根据费舍尔(Fisher)比对提取的 MFCC 特征向量进行加权降维优化, 最后应用 RF 算法进行训练识别。

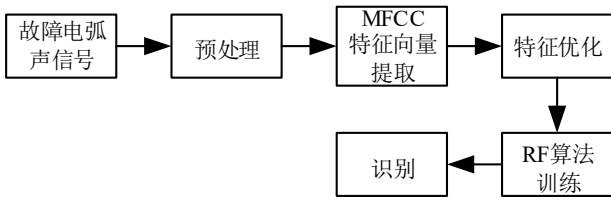


图 1 模型流程图

Fig. 1 Model workflow diagram

1.2 声信号预处理

因采集的声音样本长度过长, 在整段音频信号中很难提取特征值, 需要对信号进行预处理操作。根据信号拥有的短时平稳性, 先要对信号进行分帧、加窗操作。文中采用帧长 100 ms, 帧移为 50 ms 的汉明窗进行分帧加窗操作, 其中汉明窗的公式为

$$W(n) = \begin{cases} 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, & 0 \leq n \leq N \\ 0, & \text{其它} \end{cases} \quad (1)$$

式(1)中: N 为窗口长度。

1.3 MFCC 特征向量提取

MFCC 是一种基于人耳听觉感知特性提出的特征向量^[23], 是目前应用最为广泛的声纹识别特征之一。MFCC 特征向量能够从原始的声信号中提取频谱包络, 即能够描述信号频谱中频率分量的强度随频率变化的情况, 在声纹识别中频谱包络是区分语音特征的重要输入特征。MFCC 是基于 Mel 频率的倒谱系数, Mel 频率与频率的转换公式为

$$f_{mel} = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right) \quad (2)$$

式(2)中: f_{mel} 为 Mel 频率; f 为频率。梅尔频率与频率的关系曲线图见图 2。

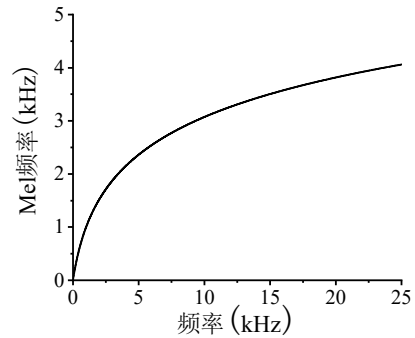


图 2 频率关系曲线

Fig. 2 Frequency response curve

在求取 MFCC 特征向量时, 需要对每一帧的信号分别进行 FFT 变换、Mel 滤波、对数变换、离散余弦变换, 而后组成特征向量组。其中 Mel 滤波由若干个三角带通滤波器组成的滤波器组实现, 每个三角带通滤波器的传递函数可表示为

$$H_i(k) = \begin{cases} 0, & k < f(i-1) \\ \frac{k - f(i-1)}{f(i) - f(i-1)}, & f(i-1) \leq k \leq f(i) \\ \frac{f(i+1) - k}{f(i+1) - f(i)}, & f(i) \leq k \leq f(i+1) \\ 0, & k > f(i+1) \end{cases} \quad (3)$$

式(3)中: i 为滤波器数量; $f(i)$ 为三角带通滤

波器中心频率。

信号经过三角滤波器得到的能量谱的表达式为

$$s(i) = \ln\left(\sum_{k=1}^N |x(k)|^2 H_i(k)\right) \quad (4)$$

式(4)中： N 为滤波器数量； $x(k)$ 为经过 FFT 变换的信号。

最后对式(4)的得到的能量的对数进行离散余弦变换，计算出 MFCC 特征向量，其可表示为

$$c(n) = \sum_{i=1}^N s(i) \cos\left(\frac{(\pi n(i-0.5))}{N}\right), 1 \leq n \leq N \quad (5)$$

式(5)中： n 为 MFCC 倒谱系数个数。

1.4 特征优化

在进行分类模型训练时，对于一个 MFCC 特征向量组，不同维倒谱系数特征分量对分类的贡献度也有所不同，部分维度的特征向量对分类贡献度较小，可能会影响甚至降低识别率。在特征选取中常用的方法有增减分量法和费舍尔(Fisher)比分析某一维度对分类的贡献度。

增减分量法计算量较大并且需要多次的模型训练和评估，实现难度较大，并且需要考虑特征之间的相关性，而 MFCC 特征向量在提取过程中使用了离散余弦变换，在频域上可近似认为每一维特征相互独立，故增减分量法不适用于 MFCC 特征向量特征优化。

Fisher 比计算量较小且不需要多次的模型训练和评估，不需要考虑特征之间的相关性，能通过计算特征在不同类别之间的分离程度选择最优特征，在声纹识别中，使用 Fisher 比能够使语音的声纹特征更可能地聚集，减少噪声信号的干扰，有利于模型分类结果，改善故障误判的情况，降低识别时

间^[24]。以下为优化过程。

1)提取特征向量后得到一个 $a \times b$ 维 MFCC 特征矩阵

$$M = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1b} \\ m_{21} & m_{22} & \dots & m_{2b} \\ \vdots & \ddots & \vdots & \vdots \\ m_{a1} & m_{a2} & \dots & m_{ab} \end{bmatrix} \quad (6)$$

式(6)中： a 为分帧的帧数； b 为每一帧的 MFCC 特征的个数。

2)计算第 j 维的特征分量的均值与第 l 类样本在第 j 维上特征分量的均值：

$$\mu_j = \frac{1}{a} \sum_{i=1}^a m_{ij} \quad (7)$$

$$\mu_j^l = \frac{1}{n_l} \sum_{i=1}^{n_l} m_{ij}^l \quad (8)$$

式(7)、(8)中： μ_j 为第 j 维特征分量的均值； m_{ij} 为第 i 行、第 j 列的特征向量； μ_j^l 为第 l 类样本在第 j 维特征分量的均值； n_l 为第 l 类样本的样本数； m_{ij}^l 为第 l 类样本第 i 行、第 j 列的特征向量。

3)计算类间方差与类内方差：

$$S_{between}^2(j) = \sum_{l=1}^L (\mu_j^l - \mu_j)^2 \quad (9)$$

$$S_{within}^2(j) = \sum_{l=1}^L \left[\frac{1}{n_l} \sum_{x \in l} (x_j^l - \mu_j^l)^2 \right] \quad (10)$$

式(9)、(10)中： $S_{between}^2(j)$ 为第 j 维的类间方差； $S_{within}^2(j)$ 为第 j 维的类内方差； L 为样本总类数； x_j^l 为第 l 类的第 j 维的特征向量。

4)计算第 j 维的 Fisher 比

$$F(j) = \frac{s_{between}^2(j)}{s_{within}^2(j)} \quad (11)$$

通常情况，Fisher 比越大说明该分量在同类信号之间方差较小，在不同类信号之间方差越大，即 Fisher 比越大说明该维特征向量区分不同类别信号能力越强，在分类识别中贡献较大，根据不同维度 Fisher 比的大小，对维数 Fisher 比值过小的特征向量舍去以达到降维的目的。

5)将每一维的 Fisher 比作为加权重值对原特征向量组进行加权处理：

$$M^j = [m_{1j}, m_{2j}, \dots, m_{aj}]^T \quad (12)$$

$$M_1^j = F(j) \cdot M^j \quad (13)$$

将每一维的特征向量进行加权后得到的加权特征向量组为

$$M_f = [M_1^1, M_1^2, \dots, M_1^b] \quad (14)$$

1.5 分类识别算法

随机森林算法^[25](random forest algorithm, RF)是一种以决策树为估计器的集成算法，在预测的过程中采用了多个决策树作为估计器，综合多个决策器的分类预测结果，提高了模型的精度。相较于传统的分类模型，在处理高维数据时速度较快、抗过拟合能力强的优势，在故障电弧声纹识别中，故障电弧声信号特征干扰因素较多，使用随机森林算法具有较强的优势^[26-27]。

随机森林算法对故障电弧声纹识别流程如下：

1)使用 bootstrap 重采样方法从 MFCC 特征样本集中随机抽取 n 个子样本集， n 个子样本集分别组成 n 个决策树。

2)在每个决策树的训练过程中，从原始样本集中的 X 个特征输入中随机不放回地选取 x 个特征根据 Gini 指数最小原则从上述特征中选择一个最优的特征进行分裂生长，重复上述过程。

3)最后，通过集合 n 个决策树的分类结果，得出最后的分类结果。

随机森林在进行分类过程中在决策树的基础上引入了随机选择的原则，通过随机选择子样本集和随机选择特征输入量，对特征中的异常值具有一定的鲁棒性，在声纹识别的过程中，有利于减少噪声的干扰。

随机森林算法流程图见图 3。

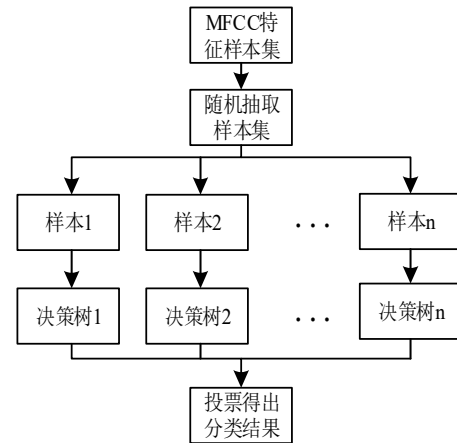


图 3 随机森林算法分类流程

Fig. 3 Random forest algorithm classification process

2 试验系统与方案

2.1 试验平台搭建

电弧声信号采集系统原理图见图 4，该系统由故障电弧试验平台、声学传感器、数据采集系统和计算机组成。故障电弧试验平台包括高压电源部分、故障电弧发生部分，其中高压电源部分包括 220 V 交流电、调压台以及保护电阻(10 kΩ, 10 kW)，声学传感器采用全指向型电容式麦克风(MPA416，灵敏度 47 mV/Pa，频率响应 20 Hz~20 kHz)采集电弧

声信号，数据采集系统采用容阻式分压器(分压比 1:1 000)、信号采集卡(NI-PCI-4474)和信号采集卡(TiePie HS5)用于采集声信号以及电压电流信号，传输至计算机进行分析。

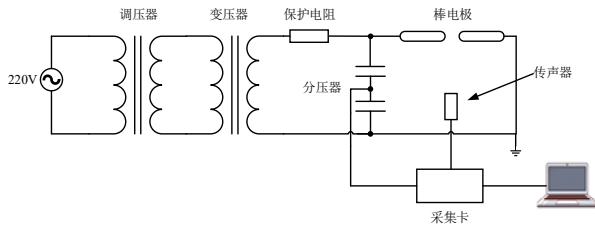


图 4 试验接线原理图

Fig. 4 Experimental wiring diagram

2.2 试验步骤

试验过程中，试验环境条件为常温常压，为了减少环境因素对声信号采集的影响，试验地点周围的空间尽量保证开阔。在保证试验设备安全的同时，能够采集到比较完整的声信号，传声器与放电电极之间的距离设置为 1 m。根据奈奎斯特定理可知，采样频率需大于信号最高频率的两倍，为了使采集的信号不失真，采样频率设置为 50 kHz，采样时间为 7 s。

在设置完毕试验条件后，启动电源，控制调压台，加压方法采用 GB/T 16927.1—2011《高电压试验技术第一部分：一般定义及试验要求》^[28]的放电试验方法，当所加电压低于 75% 预估放电电压时，以较快的速度进行升压；当电压升至 75% 预估放电电压之后，以每秒 2% 预估放电电压的速度进行升压，直至电极之间发生放电现象。为了减少随机放电影响和保证试验的可重复性，在电压达到 75% 预估放电电压时，在进行升压操作后，电极需在同一电压水平下保证耐压 1~2 min 后继续缓慢提升电压值直至发生放电现象。

3 结果分析

3.1 试验现象分析

在电极两端达到 15 kV 电压水平之前，电极之间没有发生很明显的电弧，此时只能听到较为细微的“嗡嗡”声，电极之间尚未形成电弧。在进行耐压 1 min 之后，继续进行升压操作后，直至达到 20 kV 电压水平时，电极之间开始有明显的电弧出现，此时能够听到较为强烈的“滋滋”声，电弧燃烧反复出现“燃弧—熄灭—重燃”的拉弧现象，其电压波形图见图 5。

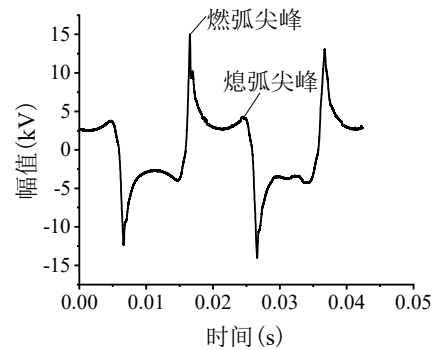
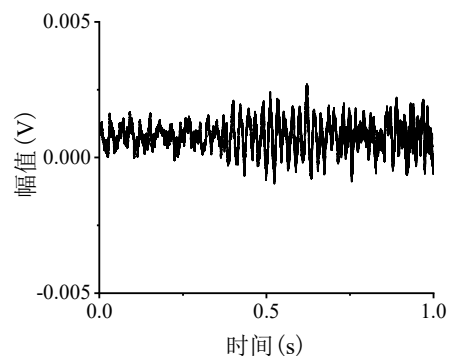


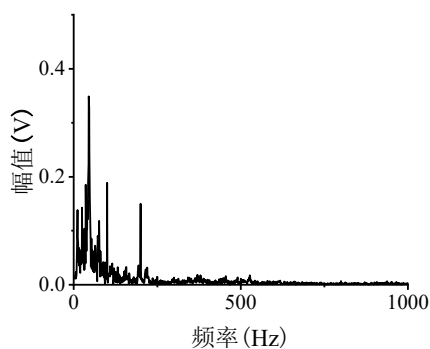
图 5 电弧电压波形图

Fig. 5 Arc voltage waveform diagram

未发现放电现象时，其声信号时频域波形图见图 6。时域上的波形图大部分以背景噪声为主，在频域上主要分布在低频部分。



(a) 时域波形图

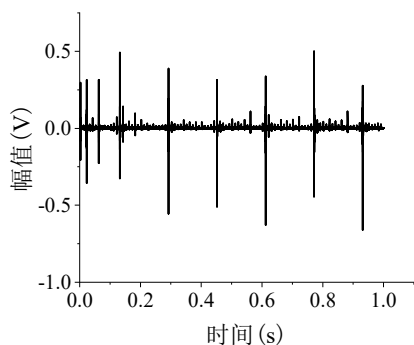


(b) 频域波形图

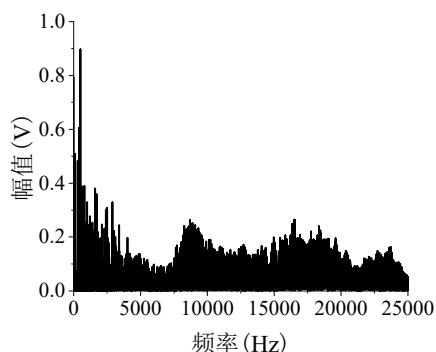
图 6 未放电声信号时频域波形图

Fig. 6 Non-discharge sound signal frequency domain waveform diagram

放电初期，其声信号时频域波形图见图 7，电弧时域上分布较为稀疏，频域上在 7.5~25 kHz 都有分布，最为明显的峰值分布在 8、16、19、23 kHz。



(a) 时域波形图

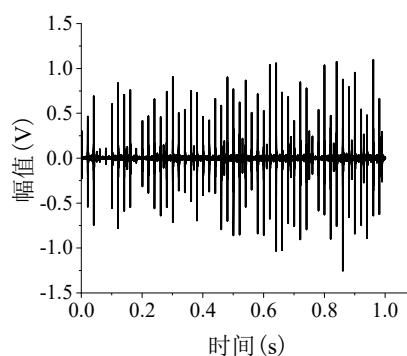


(b) 频域波形图

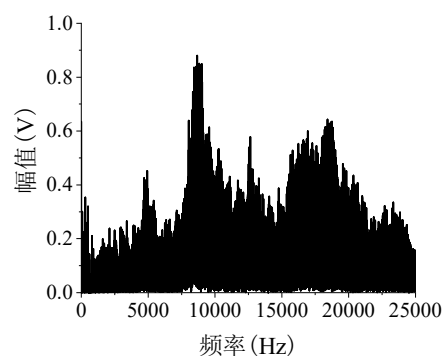
图 7 放电初期声信号时频域波形图

Fig. 7 Initial discharge sound signal frequency domain waveform

随着电弧燃烧程度的加剧，其声信号时频域波形图见图 8，时域上分布变得密集，频域上主要分布在 10~25 kHz，最为明显的峰值分布在 5、8、12、16、19、23 kHz，峰值幅值随燃烧程度加剧而增大，高频信号能量分布因声信号传播过程中衰减而有一定的减小。



(a) 时域波形图



(b) 频域波形图

图 8 剧烈燃烧声信号时频域波形图

Fig. 8 Intense burning sound signal frequency domain waveform

综上所述，在电弧放电初期到剧烈燃烧的过程中，声信号时域的幅值与分布密度随燃烧程度而增加，频域高频范围分布增强，整体幅值变高，在放电初期至剧烈燃烧之间，电弧产生的声信号强度和能量都有所增加，电弧弧声在剧烈燃烧之前就已经出现并具有一定的辨识度，故以声信号特征作为检测故障电弧发生的参量是具有可行性的。

3.2 识别结果

对试验采集的声信号以电弧燃烧程度分为未发生放电、放电初期、剧烈燃烧 3 类，每类取稳定声纹信号 1 s，进行 MFCC 特征提取，提取到的 MFCC 特征向量为 19 帧，每一帧的维数为 31 维，计算出其 Fisher 比。31 维 MFCC 特征向量的 Fisher 比值见图 9。由图 9 可知，MFCC 特征向量每一维对模型分类的贡献度有所不同，其中 1-12、17、20、29 维特征向量 Fisher 比值较大，为提升识别的准确度以及减少识别时间，根据其 Fisher 比进行特征优化使得 MFCC 特征向量声纹特征更加聚集并减少识别贡献度较小的冗余部分。

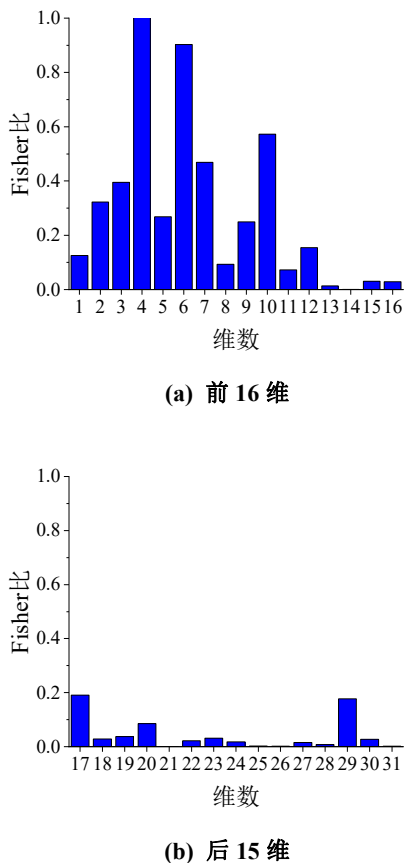
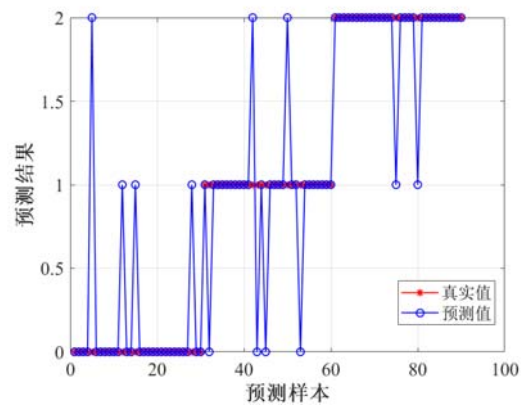


图 9 MFCC 各维度 Fisher 比值

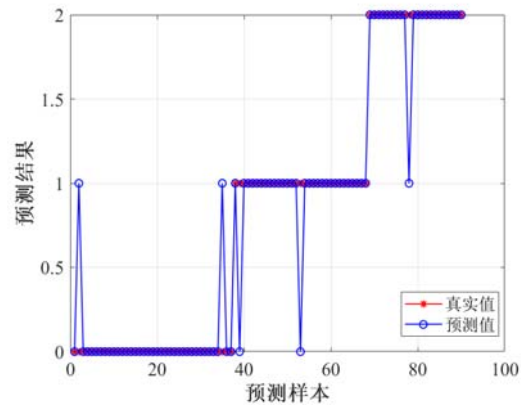
Fig. 9 MFCC dimension-wise fisher ratio

将未发生放电、放电初期、剧烈燃烧标签分别

定为 0、1、2，分别抽取 150 个样本进行训练和测试，并将这 450 个样本随机抽取 80% 的样本作为训练集，剩下的 20% 的样本作为验证集，使用随机森林算法对其进行分类识别，综合考虑模型的泛化性和运算效率，设置决策树个数为 50，叶子节点最小样本数为 1。识别结果见图 10。



(a) MFCC 特征向量识别结果



(b) 改进 MFCC 特征向量识别

图 10 RF 识别结果

Fig. 10 RF recognition result

使用 MFCC 特征向量进行识别，对未放电情况的识别率达到 86.7%，其中有 3 个样本被误判成放电初期，1 个样本被误判成剧烈燃烧；对于放电初期情况的识别率达到 80%，其中有 4 个样本被误判成未放电，2 个样本被误判成剧烈燃烧；对于剧烈燃烧情况的识别率达到 93.3%，其中 2 个样本被误

判成放电初期。总识别率为 86.7%，识别时间为 5.6 s。

使用改进 MFCC 特征向量进行识别时，对于未放电情况的识别率达到 94.6%，其中 2 个样本被误判成放电初期；对于放电初期情况的识别率达到 93.5%，其中有 2 个样本被误判成未放电；对于剧烈放电情况的识别率达到 95.5%，其中有 1 个样本被误判成放电初期。总识别率为 94.5%，识别时间为 1.7 s，相较于 MFCC 特征向量，识别率提高 7.8%，识别时间减少 3.9 s。

为了比对不同分类识别模型在故障电弧声纹识别中的识别能力，采用了多种不同的分类识别模型进行比较，包括使用径向基函数的支持向量机(RBF-SVM，惩罚系数参数 10.0，核参数 0.01)、卷积神经网络(CNN，初始学习率 0.001，学习率下降因子 0.1，最大迭代次数 500 次)、长短期记忆网络(LSTM，初始学习率 0.01，学习率下降因子 0.1，最大迭代次数 1 000 次)、偏最小二乘(PLS，保留主成分个数 6)、极限学习机(ELM，隐藏层节点个数 50)、反向传播神经网络(BP，隐藏层神经元 6 个，最大迭代次数 1 000 次)识别准确度结果见表 1。

表 1 不同模型识别率

分类模型	MFCC	改进 MFCC
RBF-SVM	77.8	86.7
CNN	74.4	87.8
LSTM	78.9	92.2
PLS	85.6	93.3
ELM	85.6	88.9
BP	78.8	82.2
RF	86.7	94.5

由表 1 可得，不同的分类模型识别率有所不同，

ELM、PLS、LSTM 以及本文使用的 RF 识别效果较好，其中 RF 对于 MFCC 特征向量的识别率为 86.7%，识别准确度超过 LSTM(78.9%)、PLS(85.6%)、ELM(85.6%)等分类模型；对于改进 MFCC 特征向量的识别率为 94.5%，识别准确度超过 LSTM(92.2%)、PLS(93.3%)、ELM(88.9%)等分类模型，具有更好的识别准确度。

考虑到实际情况中，声信号采集过程中易受到环境噪声干扰，对试验中得到声信号加入高斯白噪声。为了检验文中所提方法对于噪声的抗噪性能，引入平均信噪比(signal-to-noise ratio, SNR)对带噪声信号进行质量评估，SNR 决定了信号与噪声之间的相对强度。不同的信噪比代表着不同的信号质量。加入不同信噪比的高斯白噪声，MFCC 特征向量与改进 MFCC 特征向量使用 RF 的识别率见表 2。

表 2 不同信噪比的识别率

Table 2 Different signal-to-noise ratios recognition rates %

特征	信噪比/dB			
	-5	0	5	10
MFCC	24.3	38.6	63.4	86.7
改进 MFCC	58.3	72.3	81.6	94.5

由表 2 可得，改进 MFCC 特征向量在低信噪比(-5、0)的情况下，对比 MFCC 特征向量的识别率 24.3%、38.6%分别上升了 34%、33.7%，具有较好的抗噪性。

4 结论

文中提出了一种基于改进 MFCC 特征向量的故障电弧声纹识别模型，并搭建了故障电弧试验平台，对采集的声信号进行分析识别。结果表明：

1)通过对试验现象进行分析，随着电弧燃烧程

度的加剧，故障电弧声信号时域的幅值与分布密度增加，频域高频范围分布增强，整体幅值变高，故障电弧声出现早于剧烈燃烧并具有一定的辨识度。

2)模型对未放电以及早期放电和电弧剧烈燃烧3种情况的改进 MFCC 特征向量识别率均超过了90%，充分验证了模型的有效性。通过对比不同分类模型的识别率，可以得出 RF 相较于其他模型具有更好的识别性能。

3)对比 MFCC 特征向量与改进后 MFCC 特征向量，在加入不同信噪比的高斯白噪声，改进后 MFCC 特征向量在信噪比降低的情况下，分别上升 34%、33.7%，具有良好的抗噪性。

参考文献:

[1] 盛戈皞, 钱 勇, 罗林根, 等. 面向新型电力系统的电力设备运行维护关键技术及其应用展望[J]. 高电压技术, 2021, 47(9): 3072-3084.
SHENG Gehao, QIAN Yong, LUO Linggen, et al. Key technologies and application prospects for operation and maintenance of power equipment in new type power system[J]. High Voltage Engineering, 2021, 47(9): 3072-3084.

[2] 刘科研, 董伟杰, 肖仕武, 等. 基于电压数据 SVM 分类的有源配电网故障判别及定位[J]. 电网技术, 2021, 45(6): 2369-2379.
LIU Keyan, DONG Weijie, XIAO Shiwu, et al. Fault identification and location of active distribution network based on SVM classification of voltage data[J]. Power System Technology, 2021, 45(6): 2369-2379.

[3] 蒋 原, 李 擎, 冯 茜, 等. 基于 BP 神经网络的直流电网故障定位与保护方法[J]. 高压电器, 2020, 56(8): 23-28.
JIANG Yuan, LI Qing, FENG Qian, et al. Fault location and protection method for DC power grid based on BP neural network[J]. High Voltage Apparatus, 2020, 56(8): 23-28.

[4] 朱晓红, 杨伟荣, 张 蓉, 等. 基于 RNN-LSTM 神经网

络的小电流接地故障选线方法[J]. 高压电器, 2023, 59(7): 213-220.
ZHU Xiaohong, YANG Weirong, ZHANG Rong, et al. Line selection method of low current grounding fault based on RNN-LSTM neural network[J]. High Voltage Apparatus, 2023, 59(7): 213-220.

[5] 罗晗菁, 曾祥君, 喻 锟, 等. 基于多维波形差异度聚类分析的配电网故障区段定位方法[J]. 南方电网技术, 2024, 18(6): 58-68, 97.
LUO Hanjing, ZENG Xiangjun, YU Kun, et al. Fault section location method for distribution network based on Multi-Dimensional waveform difference clustering analysis[J]. Southern Power System Technology, 2024, 18(6): 58-68, 97.

[6] 秦泽为, 张蓬鹤, 宋如楠, 等. 新型电力系统下电弧故障诊断技术及发展趋势[J]. 电测与仪表, 2024, 61(2): 1-9.
QIN Yiwei, ZHANG Penghe, SONG Runan, et al. Arc fault diagnosis technology and trend of development in novel power system[J]. Electrical Measurement & Instrumentation, 2024, 61(2): 1-9.

[7] 熊 庆, 陈维江, 汲胜昌, 等. 低压直流系统故障电弧特性、检测和定位方法研究进展综述[J]. 中国电机工程学报, 2020, 40(18): 6015-6026.
XIONG Qing, CHEN Weijiang, JI Shengchang, et al. Review of research progress on characteristics, detection and localization approaches of fault arc in low voltage DC system[J]. Proceedings of the CSEE, 2020, 40(18): 6015-6026.

[8] 李松浓, 晏 尧, 向 菲, 等. 光伏直流系统故障电弧检测方法研究综述[J]. 电测与仪表, 2024, 61(2): 10-16.
LI Songnong, YAN Yao, XIANG Fei, et al. A comprehensive review on detection method for DC fault arc in photovoltaic system[J]. Electrical Measurement & Instrumentation, 2024, 61(2): 10-16.

[9] 何志鹏, 李伟林, 邓云坤, 等. 低压交流串联故障电弧辨识方法[J]. 电工技术学报, 2023, 38(10): 2806-2817.
HE Zhipeng, LI Weilin, DENG Yunkun, et al. The detection of series AC arc fault in low-voltage distribution system[J]. Transactions of China Electrotechnical Society, 2023, 38(10): 2806-2817.

[10] 贺 胜, 疏学明, 胡 俊, 等. 基于消防大数据的电气火灾风险预测预警方法[J]. 清华大学学报(自然科学版), 2024, 64(3): 478-491.

- HE Sheng, SHU Xueming, HU Jun, et al. Prediction and early-warning method of electrical fire risk based on fire-fighting big data[J]. Journal of Tsinghua University(Science and Technology), 2024, 64(03): 478-491.
- [11] YAO Xiu, LE V, LEE I. Unknown input observer-based series DC arc fault detection in DC microgrids[J]. IEEE Transactions on Power Electronics, 2022, 37(4): 4708-4718.
- [12] LE V, YAO Xiu, MILLER C, et al. Series DC arc fault detection based on ensemble machine learning[J]. IEEE Transactions on Power Electronics, 2020, 35(8): 7826-7839.
- [13] 王毅, 刘黎明, 李松浓, 等. 基于经验小波变换复合熵值与特征融合的故障电弧检测[J]. 电网技术, 2023, 47(5): 1912-1919.
- WANG Yi, LIU Liming, LI Songnong, et al. Arc fault detection based on empirical wavelet transform composite entropy and feature fusion[J]. Power System Technology, 2023, 47(5): 1912-1919.
- [14] WANG Yangkun, ZHANG Feng, ZHANG Xueheng, et al. Series AC arc fault detection method based on hybrid time and frequency analysis and fully connected neural network[J]. IEEE Transactions on Industrial Informatics, 2019, 15(12): 6210-6219.
- [15] 焦治杰, 李腾, 王莉娜, 等. 基于卷积神经网络的光伏系统直流串联电弧故障检测[J]. 电工电能新技术, 2019, 38(7): 29-34.
- JIAO Zhijie, LI Teng, WANG Lina, et al. DC series arc-fault detection of photovoltaic system based on convolutional neural network[J]. Advanced Technology of Electrical Engineering and Energy, 2019, 38(7): 29-34.
- [16] 季坤, 张晨晨, 丁国成, 等. 粒子群优化算法在电力变压器声纹识别中的应用[J]. 沈阳工业大学学报, 2023, 45(6): 643-648.
- JI Kun, ZHANG Chenchen, DING Guocheng, et al. Application of particle swarm optimization algorithm in power transformer voiceprint recognition[J]. Journal of Shenyang University of Technology, 2023, 45(6): 643-648.
- [17] CUI Yao, HUANG Xin, ZHANG Xin. Deep neural network based acoustic pattern recognition system for fault localization application[J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1): 01232.
- [18] WANG J, ZHAO Z, ZHU J, et al. Improved support vector machine for voiceprint diagnosis of typical faults in power transformers[J]. Machines, 2023, 11(5): 539.
- [19] 周梦茜, 唐志国, 王泽瑞, 等. 基于声纹识别系统的局部放电超声信号识别研究[J]. 高压电器, 2022, 58(9): 127-133.
- ZHOU Mengqian, TANG Zhiguo, WANG Zerui, et al. Study on ultrasonic signal recognition of partial discharge based on voiceprint recognition system[J]. High Voltage Apparatus, 2022, 58(9): 127-133.
- [20] 于达, 张玮, 王辉. 基于LSTM神经网络的油浸式变压器异常声纹诊断方法研究[J]. 智慧电力, 2023, 51(2): 45-52.
- YU Da, ZHANG Wei, WANG Hui. Abnormal voiceprint diagnosis method of oil-immersed transformer based on LSTM neural network[J]. Smart Power, 2023, 51(2): 45-52.
- [21] 耿琪深, 王丰华, 金霄. 基于Gammatone滤波器倒谱系数与鲸鱼算法优化随机森林的干式变压器机械故障声音诊断[J]. 电力自动化设备, 2020, 40(8): 191-196.
- GENG Qishen, WANG Fenghua, JIN Xiao. Mechanical fault sound diagnosis based on GFCC and random forest optimized by whale algorithm for dry type transformer[J]. Electric Power Automation Equipment, 2020, 40(8): 191-196.
- [22] YAO W, XU Y, QIAN Y, et al. A classification system for insulation defect identification of gas-insulated switchgear(GIS), based on voiceprint recognition technology[J]. Applied Sciences, 2020, 10(11): 3995.
- [23] 孙汉文, 李喆, 盛戈峰, 等. 基于机器学习与卷积神经网络的放电声音识别研究[J]. 高压电器, 2020, 56(9): 107-113.
- SUN Hanwen, LI Zhe, SHENG Gehao, et al. Study of discharge sound diagnosis based on machine learning and convolutional neural networks[J]. High Voltage Apparatus, 2020, 56(9): 107-113.
- [24] 董哲为. 基于Fisher融合的含噪声声纹特征提取与识别研究[D]. 西安: 长安大学, 2023.
- DONG Zhewei. Research on features extraction and recognition of noisy voiceprint based on fisher fusion algorithm[D]. Xi'an: Chang'an University, 2023.
- [25] 刘芹, 彭在兴, 王颂, 等. 基于随机森林算法的断路器分合闸线圈故障电流曲线识别[J]. 高压电器, 2019, 55(7): 93-100.
- LIU Qin, PENG Zaixing, WANG Song, et al. Fault current curves identification of circuit breaker opening/closing coil

based on random forest algorithm[J]. High Voltage Apparatus, 2019, 55(7): 93-100.

[26]赵洲峰, 赵志勇, 邹君文, 等. 基于声振信号组合特征的随机森林模型及其在绝缘子故障识别中的应用[J]. 广东电力, 2022, 35(12): 93-100.

ZHAO Zhoufeng, ZHAO Zhiyong, ZOU Junwen, et al. A random forest model based on combination features of acoustic-vibration signals and its application in insulator fault identification[J]. Guangdong Electric Power, 2022, 35(12): 93-100.

[27]王兰兰, 朱捷, 周正平, 等. 基于随机森林的滚动轴承故障辨识方法研究[J]. 机电工程, 2021, 38(12): 1599-1604.

WANG Lanlan, ZHU Jie, ZHOU Zhengping, et al. Fault identification method of rolling bearing based on random forest[J]. Journal of Mechanical & Electrical Engineering, 2021, 38(12): 1599-1604.

[28]高电压实验技术第 1 部分: 一般定义及实验要求: GB/T 16927.1—2011[S].2011.

High-voltage test techniques.Part 1: General definitions and test requirements: GB/T 16927.1—2011[S].2011.

陈凯(2000—), 男, 硕士研究生, 主要从事故障电弧检测与定位技术研究(E-mail: 673681917@qq.com)。

吴田(1983—), 男, 博士, 高级工程师, 主要从事电网智能运检和带电作业(E-mail: wutian_08@163.com)。