Anti Interference Identification Method for Abnormal Data of Transformer Oil Spectrum Based on Gaussian Distribution Algorithm*

CHEN Gang[1,2*], SHI Yingbin[1,2], HE Dandong[1,2]
(1.Electric Power Research Institute of State Grid Xinjiang Electric Power Co., Ltd., Urumqi Xinjiang 830013, China;
2.Xinjiang Key Laboratory of Extreme Environment Operation and Detection Technology of Power Transmission & Transformation Equipment, Urumqi Xinjiang 830013, China)

Abstract: The anti-interference recognition method for abnormal data of transformer oil chromatogram based on Gaussian distribution algorithm is studied, and the anti-interference recognition effect of abnormal data of transformer oil chromatogram is improved. The exponential weighted moving average method based on sliding windows is used, the historical dataset of transformer oil chromatography is processed, and a residual dataset of oil chromatography data is established. The Gaussian distribution algorithm is used, the Gaussian distribution probability value of the residual data of the oil chromatogram is calculated. The particle swarm optimization algorithm is used to determine the probability threshold of Gaussian distribution anomalies, when the probability value of Gaussian distribution in oil chromatography residual data is lower than the threshold, it indicates that the data is abnormal, and anti-interference identification of transformer oil chromatography abnormal data is completed. The experimental results show that this method can effectively establish a residual dataset for transformer oil chromatogram data, effectively determine the probability threshold of Gaussian distribution anomalies and calculate the Gaussian distribution probability of residual oil chromatogram data, completing anti-interference identification of transformers from chromatographic anomaly data. Under different noises, the Mathews coefficient of the method for anti-interference identification of abnormal data is relatively high, indicating that the accuracy of the method for anti-interference identification of abnormal data is relatively high.

Key words: Gaussian distribution algorithm; transformer; abnormal oil chromatogram data; anti interference; identification method; sliding window

EEACC: 7220; 8350 **doi**: 10.3969/j.issn.1005-9490.2026.03.023

基于高斯分布算法的变压器油色谱异常数据抗干扰识别方法*

陈 刚[1,2*], 石迎彬[1,2], 何丹东[1,2]
(1.国网新疆电力有限公司电力科学研究院, 新疆 乌鲁木齐 830013;
2.新疆输变电设备极端环境运行与检测技术重点实验室, 新疆 乌鲁木齐 830013)

摘 要: 研究基于高斯分布算法的变压器油色谱异常数据抗干扰识别方法, 提升变压器油色谱异常数据抗干扰识别效果。利用基于滑动窗口的指数加权移动平均法, 处理变压器油色谱历史数据集, 建立油色谱数据的残差数据集; 利用高斯分布算法, 计算油色谱残差数据的高斯分布概率值; 通过改进粒子群算法确定高斯分布异常概率阈值, 当油色谱残差数据高斯分布概率值低于阈值时, 则说明该数据为异常数据, 完成变压器油色谱异常数据抗干扰识别。实验结果表明: 所提方法可有效建立变压器油色谱数据的残差数据集, 有效确定高斯分布异常概率阈值, 并计算油色谱残差数据的高斯分布概率, 完成变压器油色谱异常数据抗干扰识别; 在不同噪声下, 所提方法异常数据抗干扰识别的马修斯系数均较高, 即异常数据抗干扰识别精度较高。

关键词: 高斯分布算法; 变压器; 油色谱异常数据; 抗干扰; 识别方法; 滑动窗口
中图分类号: TP391 **文献标识码**: A **文章编号**: 1005-9490(2026)03-0638-06

电力系统内的主要部件之一是变压器, 变压器的运行状态, 直接影响电力系统供电的安全稳定性[1-3]。变压器油色谱数据是描绘变压器运行状态的主要依据, 精准分析变压器油色谱数据, 可及时发现变压器内部是否存在故障问题, 帮助维修人员制定合理的维修方案, 降低故障发生概率[4], 提升电力系统运行的安全可靠性。因为变压器运行环境较为恶劣, 所以导致采集的变压器油色谱数据内, 经常出现异常数据, 降低变压器故障诊断精度。为此, 需要研究变压器油色谱异常数据识别方法, 提升变压

项目来源: 自治区重点研发计划项目(2022B01047-2)
收稿日期: 2024-03-06 **修改日期**: 2024-05-09

器故障诊断精度[5]。目前相关学者开始对该课题开展研究,例如,杨海强等[6]通过最小截平方和建立变压器油色谱异常数据识别模型;利用改进奇异值分解,求解该模型的最小二乘问题,得到变压器油色谱异常数据识别结果。该方法可快速精准地识别变压器油色谱异常数据。刘岩等[7]针对变压器油色谱历史数据,构造高斯混合模型,通过期望极大算法估计该模型参数,利用该模型对变压器油色谱数据进行异常识别。该方法可有效识别变压器油色谱异常数据。在正负样本分布不均衡时,上述方法均无法精准识别变压器油色谱异常数据,且上述方法的抗干扰能力较差。

高斯分布算法能够完成正负样本分布不均衡时的异常数据识别,可提升异常数据识别精度[8],且该算法具备较优的抗干扰能力。针对上述方法存在的问题,研究基于高斯分布算法的变压器油色谱异常数据抗干扰识别方法,该方法的技术思路为:

①利用基于滑动窗口的指数加权移动平均法,处理变压器油色谱历史数据集,建立变压器油色谱数据的残差数据集;

②利用高斯分布算法,计算变压器油色谱残差数据的高斯分布概率值,识别基于高斯分布算法的油色谱异常数据;

③通过改进粒子群算法确定异常数据抗干扰识别的最佳阈值,当变压器油色谱残差数据高斯分布概率值低于阈值时,则说明该数据为异常数据,完成变压器油色谱异常数据抗干扰识别。

该方法可以提升变压器油色谱异常数据抗干扰识别精度,确保变压器稳定运行。

1　变压器油色谱异常数据抗干扰识别方法

1.1　构建变压器油色谱数据的残差数据集

利用基于滑动窗口的指数加权移动平均法(Exponentially Weighted Moving Average, EWMA),处理变压器油色谱历史数据集,得到变压器油色谱历史数据和估计值的残差[9],进而建立变压器油色谱数据的残差数据集。

变压器油色谱历史数据集用 $X=\{(t_i,y_i),i=1,2,\cdots,N\}$ 描述,其中,第 i 个相对时间用 t_i 描述,变压器油色谱历史数据采样时间段用 N 描述,变压器油色谱数据监测值用 y_i 描述,y_i 在时间间隔是 ω 的滑动窗口中的数据用 $S_\omega(y_i)=\{y_{i-\omega},y_{i-\omega+1},y_{i-1}\}$ 描述,$i=\omega,\omega+1,\cdots,N$。因为变压器不同类型的油色谱数据间存在一定的时间相关性,所以需要利用 EWMA 分析 $S_\omega(y_i)$ 时间窗数据,同时求解和 y_i 的残差。

$S_\omega(y_i)$ 的指数加权移动平均值用 k_i 描述,公式如下:

$$k_i=\gamma y_{i-1}+k_{i-1}-\gamma k_{i-1} \tag{1}$$

式中:加权下降速率用 γ 描述。

初始化 $k_{i-\omega-1}$,使 $k_{i-\omega-1}=0$,则式(1)可展开成:

$$k_i=(1-\gamma)(y_{i-1}+\gamma y_{i-2}+\cdots+\gamma_{\omega-1}y_{i-\omega}) \tag{2}$$

式中:$\omega=\dfrac{1}{1-\gamma}$;因为 $k_{i-\omega-1}=0$,所以初期的 k_i 非常小,降低残差数据计算精度,为此,调整式(1)获取:

$$k_i=\frac{(1-\gamma)(y_{i-1}+\gamma y_{i-2}+\cdots+\gamma_{\omega-1}y_{i-\omega})}{1-\gamma_\omega} \tag{3}$$

变压器油色谱数据残差值用 Δy_i 描述,公式如下:

$$\Delta y_i=y_i-k_i \tag{4}$$

标准化处理 Δy_i,可避免值域较大残差值,覆盖值域较小残差值[10-12],标准化处理公式如下:

$$g_i=\frac{\Delta y_i-\mu}{\sigma} \tag{5}$$

式中:标准化处理后的残差值用 g_i 描述;X 的残差数据集 $X'=\{(t_\omega,g_\omega),(t_{\omega+1},g_{\omega+1}),\cdots,(t_N,g_N)\}$,其平均数用 μ 描述;标准差用 σ 描述;X' 服从高斯分布。

1.2　基于高斯分布算法的油色谱异常数据识别

因为 X' 服从高斯分布,所以可以利用高斯分布算法,对 X' 进行异常数据抗干扰识别。

高斯分布用于呈现变压器油色谱残差数据 x' 出现的概率分布。通过计算 x' 的均值 θ 和方差 λ^2,估计 x' 的概率分布。利用高斯分布算法抗干扰识别变压器油色谱异常数据的原理为:计算 x' 的高斯分布概率 $p(x';\theta,\lambda^2)$,对比分析 $p(x';\theta,\lambda^2)$ 与高斯分布概率阈值 ε,当 $p(x')\geqslant\varepsilon$ 时,说明 x' 为正常数据,当 $p(x')<\varepsilon$ 时,说明 x' 为异常数据。

利用高斯分布算法抗干扰识别变压器油色谱异常数据的具体步骤如下:

步骤 1　计算 x' 的均值 θ 和方差 λ^2,公式如下:

$$\theta=\frac{\sum_{j=1}^{m}x'_j}{m} \tag{6}$$

$$\lambda^2=\frac{\sum_{j=1}^{m}(x'_j-\theta)^2}{m} \tag{7}$$

式中:x' 的维度用 m 描述;第 j 维变压器油色谱数据的残差数据用 x'_j 描述。

步骤 2　求解 x' 的高斯分布概率 $p(x';\theta,\lambda^2)$,公式如下:

$$p(x';\theta,\lambda^2)=\frac{e^{-\frac{(x'-\theta)^2}{2\lambda^2}}}{\theta\sqrt{2\pi}} \tag{8}$$

步骤 3　确定高斯分布概率阈值 ε,利用召回率 h 与精度 q,求解 $F1$ 分数,最高 $F1$ 对应的值便是阈值 ε,计算公式如下:

$$\begin{aligned} h&=\frac{TP}{TP+FP}\\ q&=\frac{TP}{TP+FN}\\ F1&=\frac{2hq}{h+q} \end{aligned} \tag{9}$$

式中:正确识别的正常样本数用 TP 描述;错误识别的正常正样本数用 FP 描述;错误识别的正常负样本数用 FN 描述。

如果 $p(x')<\varepsilon$,则该变压器油色谱数据 x' 为异常数据。

1.3　确定异常数据抗干扰识别的最佳阈值

为提升变压器油色谱异常数据抗干扰识别精度,利用改进粒子群算法,确定最佳高斯分布概率阈值。

令变压器油色谱异常数据抗干扰识别的高斯分布概率阈值 ε 数量为 η,在 $p(x';\theta,\lambda^2)$ 内代入已知标签的 a 组变压器油色谱残差数据样本,如果符合 $p(x';\theta,\lambda^2)\geqslant\varepsilon_l$,那么说明包含该组残差数据的样本是正例[13],否则是反例。a 组变压器油色谱残差数据样本完成抗干扰识别后,和实际情况进行对比分析,同时求解 $F1$ 分数,结束后可获取 η 个 $F1$ 分数,令 $F1$ 达到最大值的 ε 便是最佳 ε。

利用粒子群算法寻优 ε 的流程如图 1 所示。

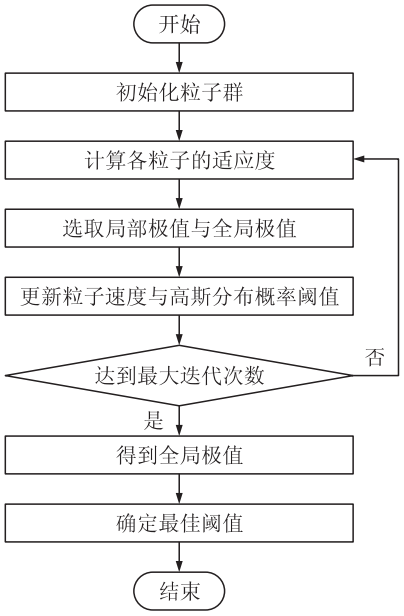


图 1　粒子群寻优最佳阈值的流程图

利用粒子群算法寻优 ε 的具体步骤如下:

步骤 1　确定粒子数量 η,即变压器油色谱异常数据抗干扰识别的高斯分布概率阈值 ε 数量为 η,构建粒子种群,即初始高斯分布概率阈值。

步骤 2　设置迭代次数 τ 以及它的最大值 τ_{max}。

步骤 3　以 $F1$ 为粒子适应度,利用式(9)计算各粒子的适应度,在初始高斯分布概率阈值内,选取最大 $F1$ 对应的粒子作为个体极值 $U_{best}(F1_{max})$,相应的粒子值是全局极值 G_{best}。

步骤 4　更新粒子速度 $v_{\tau+1}(l)$ 与高斯分布概率阈值 $\varepsilon_{\tau+1}(l)$,公式如下:

$$\begin{aligned} v_{\tau+1}(l)=&wv_\tau(l)+c_1r_1[U_{best}(F1_{max})-z_\tau(l)]+\\ &c_2r_2(G_{best}-z_\tau(l)) \end{aligned} \tag{10}$$

$$\varepsilon_{\tau+1}(l)=\varepsilon_\tau(l)+v_{\tau+1}(l) \tag{11}$$

式中:迭代 τ、τ+1 次后,第 l 个粒子的速度用 $v_\tau(l)$、$v_{\tau+1}(l)$ 描述;学习因子用 c_1、c_2 描述;随机数用 r_1、r_2 描述;权值用 w 描述;高斯分布概率阈值变化率用 $z_\tau(l)$ 描述。

w 的计算公式如下:

$$w=w_{max}-\frac{\tau(w_{max}-w_{min})}{\tau_{max}} \tag{12}$$

式中:w 的最大、最小值用 w_{max}、w_{min} 描述。

步骤 5　对比分析更新后粒子的 $F1_{\tau+1}(l)$ 分数,和上一代粒子的 $F1_\tau(l)$ 分数,如果 $F1_{\tau+1}(l)>F1_\tau(l)$,那么 $F1$ 内的 $F1_\tau(l)$ 调整成 $F1_{\tau+1}(l)$,$U_{best}(F1_{max})$ 内相应的值也调整成 $F1_{\tau+1}(l)$ 相应的 $\varepsilon_{\tau+1}(l)$,否则,保持原样[14],对比分析结束后,调整 G_{best} 为 $F1_{max}$ 相应的变压器油色谱异常数据抗干扰识别的高斯分布概率阈值 $\varepsilon_{\tau+1}(l)$。

步骤 6　反复操作步骤 4 与步骤 5,分析算法是否达到 τ_{max},若达到 τ_{max},则输出 G_{best} 对应的最佳高斯分布概率阈值 ε,若未达到 τ_{max},则返回步骤 3,令 $\tau=\tau+1$。

当初始粒子生成较差时,算法会收敛至错误结果[15],为此,需要对粒子群算法进行改进,改进的粒子群算法,将寻优最佳高斯分布概率阈值问题,变更成寻优最佳高斯分布概率阈值变化率问题,变更步骤如下:

步骤 1　令初始阈值 ε_0 为 1,初始变化率 $z\in(0,1)$,设置两个迭代次数 τ_1、τ_2。

步骤 2　以 $\varepsilon_0\times z$ 方式,调整高斯分布概率阈值,得到 ε_l,求解当下的 $F1$ 分数,再次调整阈值,求解 $F1$ 分数,调整 τ_1 次后,获取 τ_1 个 $F1$ 分数,存储最大 $F1$ 分数,以及相应的 z。

步骤 3　变更 z 的值,返回步骤 2,重新搜索最

大 $F1$ 分数,以及相应的 z,并与上次 $F1$ 分数进行对比,如果高于上次,那么存储全部数据,反之,保持原样。

步骤 4　反复操作步骤 3,共操作 τ_2 次,最终获取的 z,就是最佳高斯分布概率阈值变化率。

步骤 5　利用最佳高斯分布概率阈值变化率,计算变压器油色谱异常数据抗干扰识别的最佳高斯分布概率阈值,提升算法的稳定性。

2　实验分析

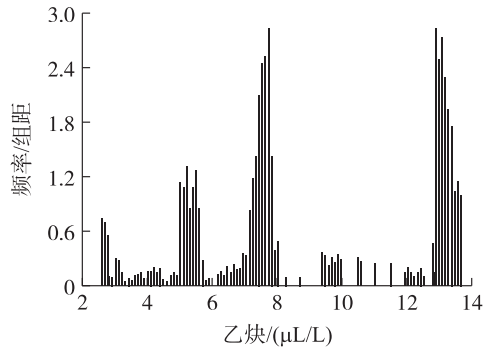
以某 110 kV 油浸有载变压器为实验对象,其型号为 SZ-50000/110,该变压器的相关参数见表 1。

表 1　变压器相关参数

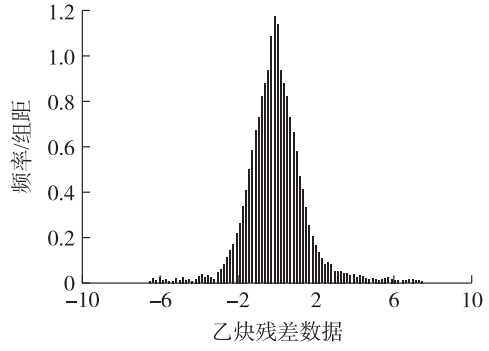
名称	标准数值	
额定值	高压绕组额定电压/kV	110
	低压绕组额定电压/kV	10.5
	额定频率/Hz	50
	高压绕组额定容量/MVA	50
	低压绕组额定容量/MVA	50
	相数	3
	调压方式	有载
	调压位置	高压中性点
	调压范围/kV	±8×1.25%
绝缘水平	高压线端雷电全波冲击电压峰值/kV	480
	低压线端雷电全波冲击电压峰值/kV	75
	中性点端子雷电全波冲击电压峰值/kV	325
	高压线端雷电截波冲击电压峰值/kV	530
	低压线端雷电截波冲击电压峰值/kV	86
温升限值/K	顶层油	56
	平均绕组	66
	热点绕组	79
	油箱表面	76

利用文献[6]基于 SVD 的改进 LTS 气动数据异常检测方法和本文方法分别计算变压器油色谱数据的残差数据,并以频率直方图的形式呈现,以变压器油色谱数据中的乙炔数据为例,两种方法下乙炔残差数据的频率直方图见图 2。

在同一变压器油色谱的异常数据工况下,由图 2(a)可知,基于 SVD 的方法乙炔残差数据频率直方图没有任何规律,说明乙炔残差数据分布比较杂乱,不服从高斯分布,无法直接利用残差数据进行变压器油色谱异常数据抗干扰识别。由图 2(b)可知,本文方法可有效计算乙炔的残差数据,且计算获取的乙炔残差数据服从高斯分布,能够利用本文方法进行变压器油色谱异常数据抗干扰识别。



(a) 基于SVD的方法下乙炔残差数据频率直方图



(b) 本文方法下乙炔残差数据的频率直方图

图 2　两种方法下乙炔残差数据频率直方图对比结果

利用本文方法确定变压器油色谱异常数据抗干扰识别的高斯分布概率阈值,阈值确定结果见图 3。

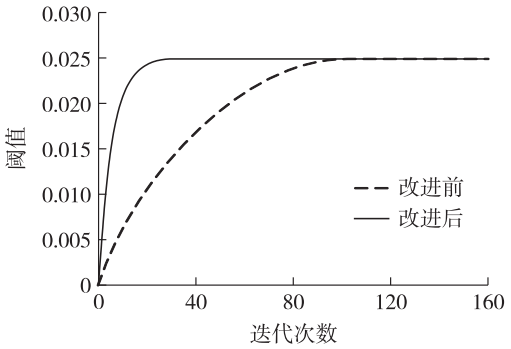


图 3　异常数据抗干扰识别高斯分布概率阈值确定结果

由图 3 可知,本文方法可有效确定变压器油色谱异常数据抗干扰识别的高斯分布概率阈值,本文方法改进前,当迭代次数达到 100 次时,完成收敛,阈值稳定在 0.025 左右;本文方法改进后,当迭代次数达到 30 次时,完成收敛,阈值稳定在 0.025 左右;说明本文方法改进后,可有效加快收敛速度,提升阈值确定效率,加快变压器油色谱异常数据抗干扰识别速度。结果表明,本文方法确定的高斯分布概率阈值是 0.025。

利用本文方法对该变压器油色谱异常数据进行抗干扰识别,以变压器油色谱数据中的氢气、乙炔与总烃为例,识别结果见表 2。

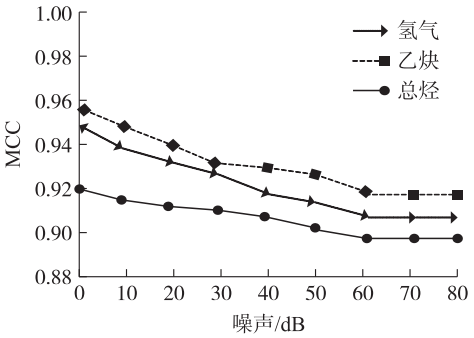
表 2 变压器油色谱异常数据抗干扰识别结果

样本编号	氢气高斯分布概率值	乙炔高斯分布概率值	总烃高斯分布概率值
1	0.031	0.028	0.036
2	0.045	0.038	0.029
3	0.053	0.033	0.034
4	0.034	0.008	0.043
5	0.029	0.043	0.049
6	0.033	0.046	0.011
7	0.042	0.031	0.053
8	0.039	0.039	0.047
9	0.015	0.053	0.022
10	0.042	0.055	0.058
11	0.046	0.044	0.047
12	0.035	0.041	0.016
13	0.021	0.039	0.046
14	0.037	0.028	0.039
15	0.043	0.037	0.032

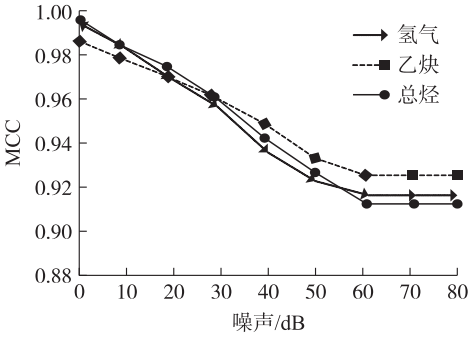
由表 2 可知,采用本文方法对该变压器油色谱异常数据进行抗干扰识别,样本编号为 9 与 13 的氢气数据高斯分布概率值低于阈值,其余样本的氢气数据高斯分布概率值均高于阈值,说明样本编号为 9 与 13 的氢气数据为异常数据;样本编号为 4 的乙炔数据高斯分布概率值低于阈值,其余样本的乙炔数据高斯分布概率值均高于阈值,说明样本编号为 4 的乙炔数据为异常数据;样本编号为 6、9 与 12 的总烃数据高斯分布概率值低于阈值,其余样本的总烃数据高斯分布概率值均高于阈值,说明样本编号为 6、9 与 12 的总烃数据为异常数据。结果表明,本文方法可有效抗干扰识别变压器油色谱异常数据。

利用马修斯系数(Matthews Correlation Coefficient,MCC)衡量基于 SVD 的方法和本文方法异常数据抗干扰识别精度,其值越接近 1,说明异常数据抗干扰识别精度越高。在变压器油色谱数据内添加不同噪声,分析两种方法在不同噪声时异常数据抗干扰识别效果,分析结果见图 4。

由图 4 可知,随着噪声的增长,对比基于 SVD 的方法,本文方法在进行不同类型异常数据抗干扰识别时的 MCC 值均不断下降,当噪声为 60 dB 时,三种类型油色谱异常数据抗干扰识别的 MCC 值均趋于稳定,氢气异常数据抗干扰识别的最低 MCC 值在 0.92 左右,乙炔异常数据抗干扰识别的最低 MCC 值在 0.93 左右,总烃异常数据抗干扰识别的最低 MCC 值在 0.918 左右,均与 1 较为接近。结果表明,在不同噪声下,本文方法异常数据抗干扰识别的 MCC 值均较高,即本文方法异常数据抗干扰识别精度较高,且具备较优的抗干扰性能。



(a) 基于SVD的方法下异常数据抗干扰识别效果



(b) 本文方法下异常数据抗干扰识别效果

图 4 两种方法下不同噪声时异常数据抗干扰识别效果

3 结论

针对传统方法无法快速发现变压器油色谱异常数据,以及异常数据识别抗干扰性能较差问题,研究基于高斯分布算法的变压器油色谱异常数据抗干扰识别方法,提升异常数据抗干扰识别能力。实验结果表明:本文方法可有效完成变压器油色谱异常数据抗干扰识别,且本文方法异常数据抗干扰识别精度较高,抗干扰性能较佳。

参考文献:

[1] 刘展程,王爽,唐波. 基于布谷鸟搜索算法和 DBN 模型的变压器故障识别[J]. 电力科学与技术学报,2022,37(2):3-11.

[2] 明菊兰,钱洲亥,牛莉,等. 基于最小二乘原理的电力变压器绝缘油色谱在线监测系统校准方法[J]. 武汉大学学报(理学版),2021,67(3):281-291.

[3] 胡啸宇,丁国成,张晨晨,等. 一起 500 kV 变压器局放超标及油色谱异常的分析与处理[J]. 变压器,2021,58(3):80-83.

[4] 李文志,朱娟,唐星昱,等. 一起换流变压器油中溶解气体异常增长的原因分析及处理[J]. 变压器,2022,59(10):66-69.

[5] 王文森,杨晓西,刘阳,等. 基于层次聚类分析的变压器油中溶解气体在线监测数据异常检测[J]. 高压电器,2023,59(1):142-147.

[6] 杨海强,黄俊,黎茂锋,等. 一种基于 SVD 的改进 LTS

气动数据异常检测方法[J]. 电光与控制,2021,28(7):78-82.
[7] 刘岩,李文文,周丽霞,等. 基于高斯混合模型的光伏发电出力中高比例异常数据检测方法研究[J]. 电测与仪表,2021,58(9):14-21.
[8] 官国宇,王运豪,别一鸣. 周期数据时段划分的双截断高斯混合模型及其 EM 算法[J]. 数理统计与管理,2022,41(1):108-123.
[9] 王亚京,包鹏赞,陈金香,等. 基于 FPGA 的旋转变压器信号解码设计[J]. 电子设计工程,2022,30(24):75-78,85.
[10] 杜修明,谢佳,吴杰,等. 一起现场安装缺陷引发的 500 kV 变压器油色谱异常分析及排查处理[J]. 变压器,2021,58(10):70-72.
[11] 雷萌,郭鹏,刘博嵩. 基于自适应 DBSCAN 算法的风电机组异常数据识别研究[J]. 动力工程学报,2021,41(10):859-865.
[12] 赵天辉,张耀,王建学. 基于空间密度聚类和异常数据域的负荷异常值识别方法[J]. 电力系统自动化,2021,45(10):97-105.
[13] 刘云鹏,王权,许自强,等. 基于多层架构的油中溶解气体数据清洗与异常识别方法研究[J]. 华北电力大学学报(自然科学版),2022,49(1):81-89.
[14] 刘文君,贺馨仪,王彬,等. 基于异常检测集成算法的油色谱在线监测数据质量评价体系[J]. 电网与清洁能源,2022,38(8):16-23.
[15] 张平,关丽红. 基于概率统计的多维关联数据动态挖掘仿真[J]. 计算机仿真,2022,39(3):402-406.

陈　刚(1975—),男,高级工程师,主要从事电气设备的绝缘介质理化技术检测、技术监督方面的工作,yanfu68423@163.com。