

基于 PCA 和 Fast-MCD 的气体在线监测异常数据

王子凌¹, 汪科¹, 柴卫健¹, 李业欣², 邹国平², 安斯光²

(1. 国网浙江省电力有限公司电力科学研究院, 杭州 310014; 2. 中国计量大学机电工程学院, 杭州 310018)

摘要: 油中溶解气体在线监测装置对变压器运行状态的准确监测是保证其稳定运行的关键, 监测装置若出现故障会导致在线监测数据异常, 严重影响在线监测的效果。针对这一问题, 提出了一种基于 PCA(主成分分析)和 Fast-MCD(快速最小协方差行列式)的变压器油中溶解气体异常数据识别方法。首先, 利用 PCA 算法对油中溶解气体时间序列进行降维, 有效消除了冗余特征信息。然后, 结合稳健统计理论, 对降维后的数据进行 Fast-MCD 稳健分析, 实现异常值识别。最后, 分析识别出的异常值, 判断故障来源。算例研究结果表明, 该异常识别方法能够有效识别出油中溶解气体在线监测装置故障, 性能优于其他常规方法, 准确率高达 99.1%。

关键词: 异常检测; 油中溶解气体分析; Fast-MCD; 主成分分析

文章编号: xxx-xxxx(XXXX)XX-0001-10

Identification of Abnormal data in online dissolved gas monitoring in oil based on PCA and Fast-MCD

WANG Ziling¹, WANG Ke¹, CHAI Weijian¹, LI Yexin², ZOU Guoping², AN Siguang²

(1. State Grid Zhejiang Electric Power Co., Ltd. Research Institute, Hangzhou 310014, China; 2. School of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China)

Abstract: To detect the operational status of transformers by online monitoring of dissolved gases in oil is critical for ensuring its stable operation. However, faults in monitoring devices can result in abnormal data, severely impacting the effect of online monitoring. To address this issue, a method for identifying abnormal data in dissolved gas monitoring based on principal component analysis (PCA) and fast minimum covariance determinant (Fast-MCD) is proposed. First, the PCA is used to reduce the dimensionality of time-series data for dissolved gas in oil, effectively eliminating redundant features. Next, robust statistical theory is applied to the reduced data using Fast-MCD to identify anomalies. Finally, the identified anomalies are analyzed to determine the source of the fault. Case studies demonstrate that the proposed method effectively identifies faults in online dissolved gas monitoring devices, outperforming conventional methods with an accuracy of 99.1%.

Key words: anomaly detection; DGA; Fast-MCD; PCA

0 引言

电力变压器是电力系统中最关键的设备之一, 其故障将严重影响电力系统的安全稳定运行。随着在线监测技术的发展, 变压器油中 DGA(溶解气体分析)已成为诊断变压器故障的重要手段之一。通过油中溶解气体在线监测装置(以下简称“在线监测装置”)对绝缘油中特征气体组分进行监测分析, 可实时掌握充油设备的运行状况^[1]。但随着使用年限的增加, 在线监测装置面临的问题越来越多^[2-4]。一方面, 在线监测装置数量众多且维护不到位, 易出现故障; 另一方面, 各制造厂商产品质量参差不齐, 导致软硬件故障频发。这些异常会使数据失真, 干扰变压器故障诊断, 降低异常识别的准确率^[5]。因此, 及时发现异常数据并查明原因, 对在线监测系统的稳定运行至关重要。

现阶段, 在线监测装置异常数据的识别主要采用阈值法, 即数据绝对值或增长值超过设定的阈值时判定为异常^[6-7]。但是阈值法存在两个主要弊端: 一是监测误差较大且数据波动较剧烈时易引发误报警; 二是每种特征气体对应一个阈值, 不同气体之间的数据相互独立, 无法体现特征气体数据之间的内在联系, 无法识别未超阈值的潜在故障。

为了解决阈值法在识别 DGA 异常数据遇到的问题, 国内外学者结合现有的异常数据识别方法进行了深入研究。高树国等人提出一种基于 Fast-MCD(快速最小协方差行列式)稳健分析的异常值检测方法^[8], 采用 H₂、CO 和总烃的三特征分析法, 但该方法中的马氏距离计算受样本集影响较大, 当数据波动幅度较大时,

收稿日期: 2024-08-13

基金项目: 国网浙江省电力有限公司科技项目(B311DS230008)

作者简介: 王子凌(1987), 男, 工学硕士, 高级工程师, 主要从事配电网运维大数据分析方面的研究; 安斯光(1981), 女, 教授, 硕士生导师, 主要从事电气设备故障诊断、智能优化算法和电磁场数值计算方面的研究。

识别效果不佳。李振杰在高树国研究基础上,提出了一种基于单变量 Fast-MCD 稳健分析的异常值检测方法^[9],该方法对单一特征气体波动的辨识灵敏度更高,能识别出更多的异常值点,但同样受样本集影响较大。荣智海等人提出了基于 KPCA(核主成分分析)的异常值识别方法^[10],通过统计量方法快速识别异常数据,但该方法适用的数据具有局限性,核函数的参数会影响特征提取效果,不同的样本数据需要选择合适的参数。黄成云等人采用设置滑动窗口和基于密度模式的空间数据聚类算法来识别异常数据^[11],并根据聚类簇半径将数据甄别为正常数据、零值数据、缺失值数据和异常数据四类。刘云鹏等人利用变分模态分解去除趋势项,保留残余项,并利用 3σ 准则进行异常识别^[12],该算法能够准确识别存在趋势变动的非平稳监测序列。沈晓东等人利用小波变换模极大值与 Lipschitz 指数的关系进行异常数据识别,并将其分类为快速渐变和阶跃两种类型^[13]。

相较于传统阈值法,现有的 DGA 异常数据识别方法虽提升了准确率,但因异常数据类型复杂,目前尚无一种方法能够解决所有 DGA 异常数据识别问题,且识别准确率仍有提升空间。基于此,本文提出一种基于 PCA(主成分分析)和 Fast-MCD 的变压器油中溶解气体在线监测异常数据识别方法,以提高 DGA 异常数据识别的准确率,为变压器及在线监测装置状态分析提供可靠手段。

1 基于 PCA 与 Fast-MCD 的 DGA 数据异常识别

1.1 PCA 降维

PCA 是多元统计分析的重要方法,其核心是将高维数据线性映射到较低维空间,并最大化投影维度上的数据差异,以较少的维度展示原数据的主要特性^[14-16]。DGA 异常数据识别问题面临数据维数高和变量耦合的挑战,这给异常值识别工作带来了干扰。PCA 可以有效减小数据的维度,降低模型的复杂度,使降维后的数据保留样本主要信息,离群点特征更加明显。

假设在线监测装置有 n 种特征气体,共采集了 m 天的历史数据,则 DGA 数据 X 可以写成如下的矩阵形式:

$$X = [x_1, x_2, \dots, x_n] = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}_{m \times n} \quad (1)$$

式中: x_1, x_2, \dots, x_n 为每种特征气体所有时间序列组成的列向量。

PCA 将 n 维的 X 利用空间特征 V 降至 k 维 ($k \leq n$) 的主成分矩阵 Y ^[17]。

$$Y = XV \quad (2)$$

具体步骤如下:

1) 首先对油中溶解气体时间序列 X 进行 Z-score 归一化处理,得到标准化矩阵 Z 。

$$Z_{ij} = \frac{x_{ij} - x_j}{S_j}, i = 1, 2, \dots, m; j = 1, 2, \dots, n \quad (3)$$

式中: x_{ij} 为第 j 个特征气体第 i 天的气体含量; x_j 和 S_j 分别为第 j 个特征气体含量的样本均值和样本标准差。

2) 计算 Z 的协方差矩阵 C , 并采用奇异值分解求出协方差矩阵 C 的特征值和特征向量。

3) 将特征值由大到小排列,使 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, 对应的特征向量 v_1, v_2, \dots, v_n 做出相应的变动。

4) 计算方差贡献率 T_i 和累计方差贡献率 D_i , 采用特征值表示方差大小^[18]。选取 $D_i \geq 90\%$ 对应的前 k 个特征向量构成空间特征矩阵 V , 则 $Y = XV$ 为降到 k 维后的数据。

$$T_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \times 100\% \quad (4)$$

$$D_i = \sum_{j=1}^i T_j \quad (5)$$

$$V = [v_1, v_2, \dots, v_k] = \begin{bmatrix} v_{11} & \cdots & v_{1k} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nk} \end{bmatrix}_{n \times k} \quad (6)$$

1.2 Fast-MCD 稳健分析

为降低异常点的屏蔽作用,提高识别结果的稳健性,采用 Fast-MCD 稳健分析算法,利用 C-step 迭代和马氏距离概念构造出稳健的统计量^[19],通过生成初始估计统计量、选择性迭代和嵌套扩展等过程,连续交换点,实现异常数据识别。

首先从 n 个样本数据中随机抽取 q 个样本,计算均值 T_1 和协方差 S_1 ;然后计算所有样本到均值 T_1 的马氏距离 $d_1(i)$,选出最小的前 q 个距离对应的样本;再求出上述 q 个样本的均值和协方差,经迭代获得稳健均值 T_{MCD} 和稳健协方差 S_{MCD} 。该迭代过程称为 C-step (Concentration-step) 迭代^[20],每次迭代专注距离最小的 q 个观测值。与 S_1 相比, S_2 行列式更小,数据更加集中。

Fast-MCD 算法具体步骤如下:

1) 从 m 行 k 维的样本数据 $Y = [Y_1, Y_2, \dots, Y_k]$ 中抽取任意 $h \times m$ 个样本。其中, m 是样本数量, k 是特征气体数量, h 是抽取系数,一般默认 h 为 $[(m+k+1)/2]$,也可设定为 $[(m+k+1)/2] \leq h \leq m$,此时异常值占原样本的 $(m-h+1)/m$ 倍^[21]。考虑到 DGA 数据存在异常时,工作人员会及时进行消缺处理,导致故障样本较少,故令 $h=0.85$ 。若遇到特殊情况,气体长期异常的话,则令 h 值减小。

2) 计算均值向量 $T_1 = [\mu_1, \mu_2, \dots, \mu_k]$ 。其中, μ_i 为第 i 种特征气体的均值;协方差矩阵 $S_1 = (s_{ij})_{k \times k}$, s_{ij} 为 Y_i 和 Y_j 的协方差。

3) 计算所有样本到均值 T_1 的马氏距离 $d_1(i) = \sqrt{(y_i - T_1)^T S_1^{-1} (y_i - T_1)}$, $i = 1, \dots, m$, 其中, $y_i = [y_{i1}, y_{i2}, \dots, y_{ik}]$ 为第 i 行特征气体向量。

4) 将马氏距离 $d_1(i)$ 按由小到大排序,选择最小的前 h 个距离对应的样本,求上述 h 个样本的均值 T_2 和协方差 S_2 。若 $\det(S_2) \neq 0$,则计算所有样本到均值 T_2 的马氏距离 $d_2(i)$ 。

5) 进行迭代,直到均值和协方差趋于稳定。此时,得到稳健均值 T_{MCD} 和稳健协方差 S_{MCD} 。

6) 根据步骤 5 得到的 T_{MCD} 和 S_{MCD} ,计算所有样本的稳健马氏距离 $d(i)$ 。马氏距离近似服从卡方分布^[22],自由度为 k_1 ,置信度为 α ,当 $d(i) > \sqrt{\chi_{k_1, \alpha}^2}$ 时,判定第 i 个数据为异常值,否则为正常值。筛出所有含有异常数据的样本点,构造异常样本点序列 O 。

1.3 算法流程

本文提出的 DGA 异常数据识别算法流程如图 1 所示。首先输入油中溶解气体时间序列,对输入气体数据进行 PCA,基于方差累计贡献率对数据进行降维;接着进行 Fast-MCD 稳健分析,利用 C-step 迭代与马氏距离迭代计算稳健统计量。若气体数据超过稳健统计量,则判定为异常值。将这些异常值构成异常样本序列 O 。最后人工查验这些异常值,分析异常情况,判断故障来源。

2 案例分析

为证明所提方法的有效性和正确性,选取杭州供电公司、宁波供电公司、丽水供电公司所辖多个变电站的 DGA 数据,对所提方法进行验证。

2.1 案例一

收集杭州供电公司某变电站 110 kV 1 号主变压器 2018-01-01—07-03 共 184 组 DGA 数据。选取 H_2 、 CO 、 CO_2 、 CH_4 、 C_2H_4 和 C_2H_6 共 6 种特征气体,其时间序列如图 2 所示。02-24—03-08 期间,特征气体数据出现阶跃波动,其中 H_2 向上阶跃,其余气体向下阶跃。03-09—05-12 期间, H_2 的值缓慢下降至零,并连续 32 天保持零值。

2.1.1 算法识别结果分析

首先,采用 PCA 方法对数据集降维。通过合理选择主成分个数,既能获得代表性特征气体,又能去除噪声干扰^[23]。设定截断阈值为 90%。经过 PCA 处理后,累计贡献率变化如图 3 所示,前 4 个主成分覆盖了 90%

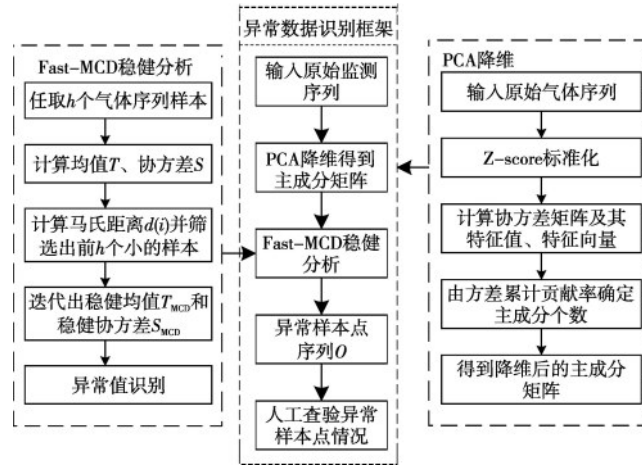


图1 DGA异常数据识别算法流程

Fig.1 Flowchart of abnormal data identification using DGA

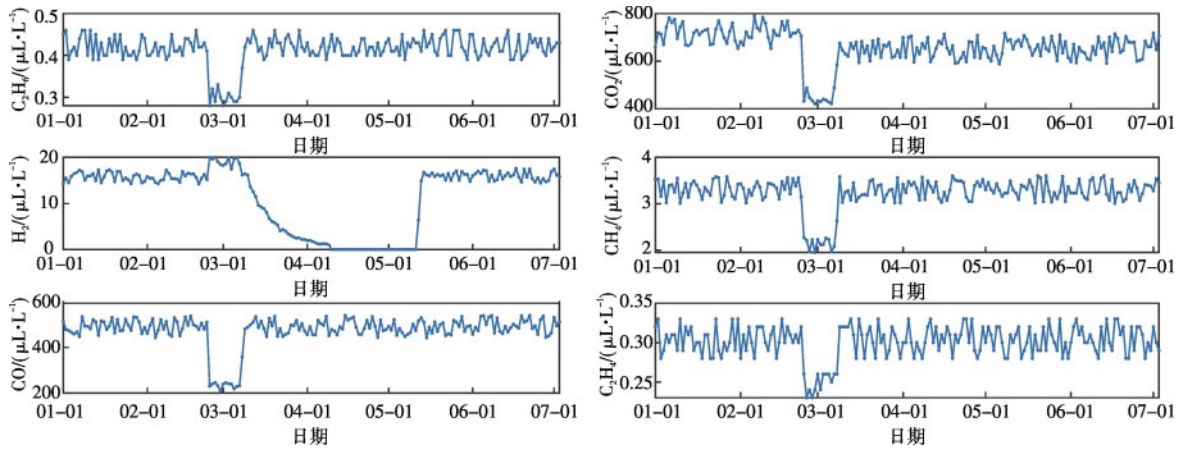


图2 特征气体时间序列

Fig.2 Time series of characteristic gases

以上原始输入参数的方差。主成分1和主成分2的散点图如图4所示,特征气体数据可分成三类:菱形为正常数据,方形为第一段阶跃异常,圆形为第二段H₂降为0的异常状态。PCA的前两个主成分包含着大量的特征气体信息,能够有效区分其状态。

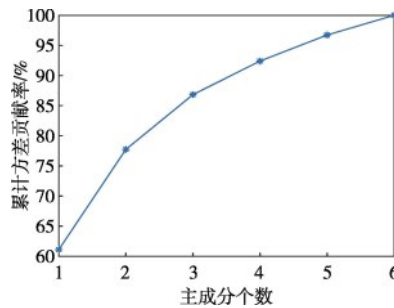


图3 累计方差贡献率

Fig.3 Cumulative variance contribution rates

提取前4个主成分构造数据集,运用Fast-MCD模型对数据进行训练,根据稳健均值 T_{MCD} 和稳健协方差 S_{MCD} 计算每个样本的稳健马氏距离 $d(i)$,其中参数设置为: $m=184, k=6, h=0.85$ 。当置信度 $\alpha=99\%$ 且自由度 $k_1=4$ 时,卡方分布的临界值 $\sqrt{\chi_{k_1, \alpha}^2} = 3.88$ 。因此,当 $d(i) > 3.88$ 时,该样本点被判定为异常值,否则为正常值。PCA结合Fast-MCD算法的异常值识别结果如图5所示。

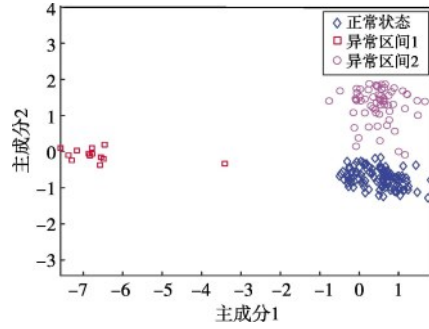


图4 主成分1和主成分2散点图

Fig.4 Scatter plot of principal component 1 and principal component 2

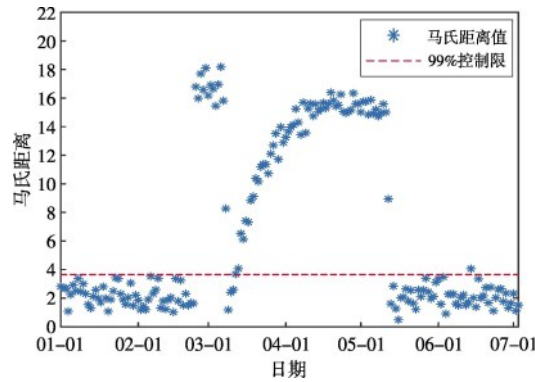


图5 PCA-FMCD 识别结果

Fig.5 The identification results from PCA-FMCD

从图5可以看出,共识别出78个异常点,包括两段异常样本区间和一个孤立异常点。对其进行故障来源分析:第一段异常区间(02-24—03-08)涉及全体气体的阶跃变化,规律性一致,排除变压器故障的可能性,推断是在线监测装置故障。第二段异常区间(03-12—05-12) H_2 缓慢下降至零,因变压器故障无此特征,推断是在线监测装置故障。孤立异常点(06-14) CO_2 较前一天上升了20%,是由现场干扰等偶然因素引起的短时扰动^[24]。

经变压器离线运维数据证实,工作人员在03-08更换了在线监测装置的CF卡,因CF卡存储功能损坏,造成全体气体数据连续阶跃波动。05-15更换了工控机主板,原因是主板短路或老化故障^[25]。

2.1.2 算法性能对比

为验证本文算法的有效性,将本文算法与多变量Fast-MCD算法^[8]、单变量Fast-MCD算法^[9]和KPCA算法^[10]的异常值识别结果进行对比。DGA异常数据的识别结果对比如图6所示。

由图6可知,单变量Fast-MCD算法虽然检测出了两段异常区间,但也错误识别了一些非异常值。以

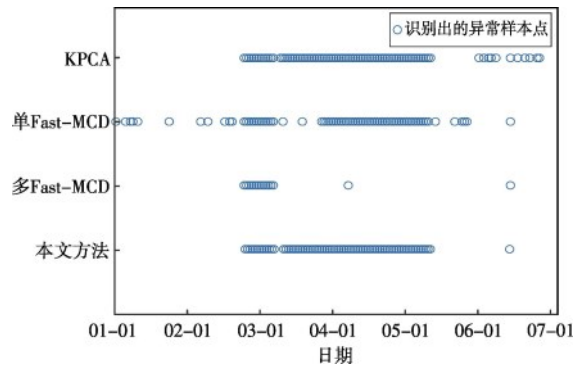


图6 多种方法识别结果

Fig.6 The identification results from multiple methods

CO₂为例,由于该算法只考虑单气体的波动,且CO₂气体含量基数较大,轻微的波动就可能致前期正常气体被误判为异常,如图7所示。多变量Fast-MCD算法只检测出一段异常区间,把H₂的长期零值误判为正常值。这是由于该算法没有经过PCA降维,多种特征气体成分的冗余对异常识别造成干扰,忽略了零值点。KPCA算法依赖于T²和SPE统计量,该方法需要两个统计量都超过阈值才能判定为异常,对样本数据的要求较高,因此可能会将正常数据误判为异常。此外,该算法需要划分训练集和数据集,限制了数据样本的选取。相比之下,本文算法能够准确识别出两段异常区间,识别效果有所提升。

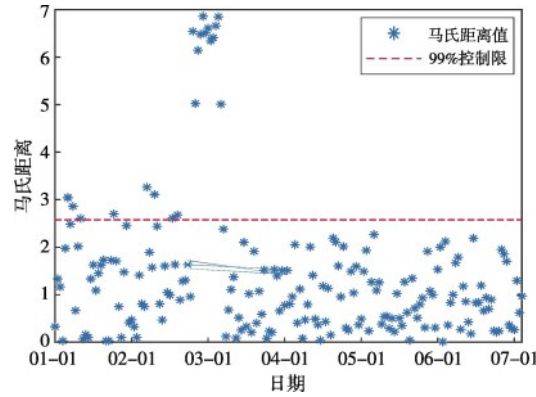


图7 单变量Fast-MCD中CO₂的识别结果

Fig.7 Identification results of CO₂ in univariate Fast-MCD

2.2 案例二

从宁波供电公司某变电站1号主变压器2020-06-17—12-31期间的197组DGA数据中,选取H₂、CO、CO₂、CH₄、C₂H₄和C₂H₆共6种特征气体,气体时间序列如图8所示。由图8可见,从11-20开始,CO₂和C₂H₆出现了明显的波动,CH₄、H₂和CO呈现出下降趋势;08-15,C₂H₆出现了离群点。

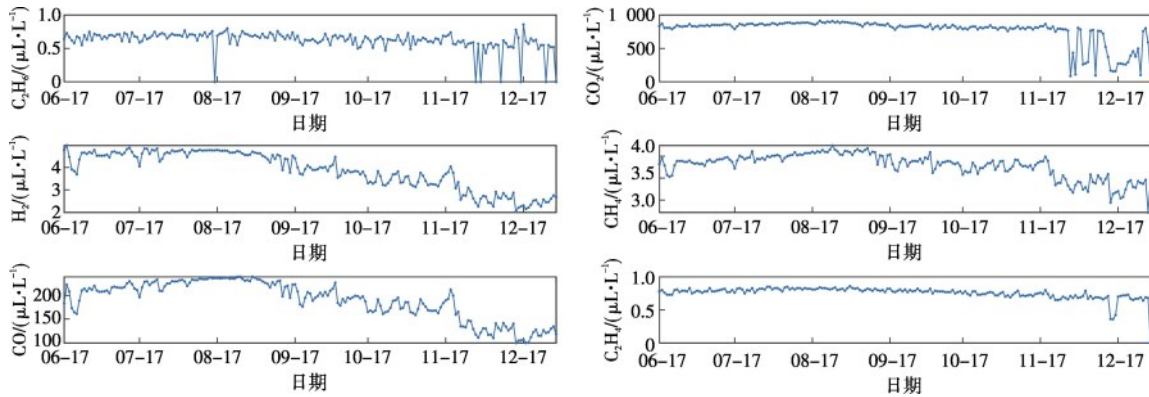


图8 某变电站特征气体时间序列

Fig.8 Time series of characteristic gases in a substation

通过PCA可知,前5个主成分超过99%的原始输入参数的方差如图9所示。因此,提取前5个主成分构造数据集,运用Fast-MCD模型对数据进行训练,根据稳健均值 T_{MCD} 和稳健协方差 S_{MCD} 计算每个样本的稳健马氏距离 $d(i)$,其中参数设置为: $m=197, k=6, h=0.85$ 。

当置信度为 $\alpha=99\%$ 且自由度 $k_1=4$ 时,卡方分布临界值为3.88。因此,当 $d(i)>3.88$ 时,判定数据为异常值,异常值的检测结果如图10所示。由图10可知,所提算法识别出了36个异常值点,包含11-20之后的波动区间和C₂H₆的离群点。

经变压器离线运维数据证实,工作人员在11-20更换了在线监测装置的载气瓶,由于载气故障导致所有气体在一段时间内连续波动。08-15没有变压器或在线装置的运维记录,因此可以确定C₂H₆的离群点是噪声干扰所致。

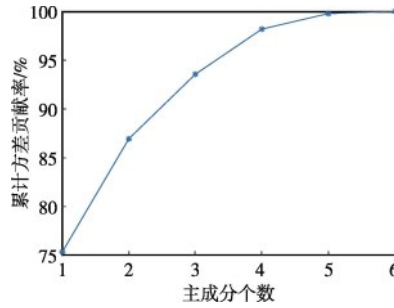


图9 累计方差贡献率

Fig.9 Cumulative variance contribution rates

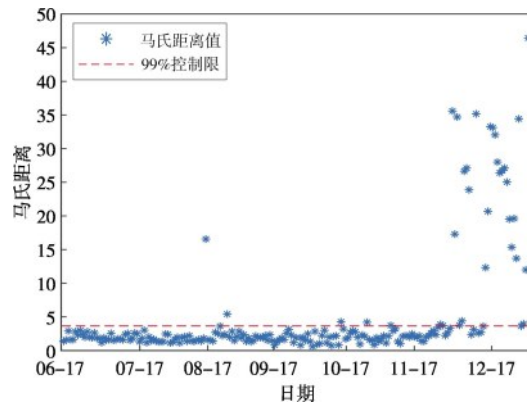


图10 PCA-FMCD 识别结果

Fig.10 Identification results using PCA-FMCD

2.3 案例三

为了进一步验证本文提出的算法对不同变压器 DGA 异常识别的准确性与稳定性,选取了杭州供电公司、宁波供电公司、丽水供电公司所辖多个变电站的 1 549 个 DGA 数据进行异常值识别。

2.3.1 评价指标

为全面、有效地评估所提算法的识别性能,选用基于混淆矩阵的分类评价指标体系^[26],混淆矩阵如表 1 所示。由表 1 可知,混淆矩阵通过 4 种指示方法反映真实值和预测值之间的相互关系。其中,TP 和 TN 分别表示真实值的正常数据和异常数据被模型正确识别的情况;FP 即假正,表示真实值为异常数据但被识别成正常数据的情况;FN 即假负,表示真实值为正常数据但被识别成异常数据的情况^[27]。

表 1 混淆矩阵

Table 1 Confusion matrix

混淆矩阵		真实值	
		正常值	异常值
预测值	正常值	TP	FP
	异常值	FN	TN

选取准确率、假正率和假负率作为评估指标,计算公式如式(7)–(9)所示。

$$Q_A = \frac{T_1 + T_2}{T_1 + T_2 + F_1 + F_2} \tag{7}$$

$$Q_{FPR} = \frac{F_1}{T_2 + F_1} \tag{8}$$

$$Q_{FNR} = \frac{F_2}{T_1 + F_2} \tag{9}$$

式中: Q_A 、 Q_{FPR} 、 Q_{FNR} 分别为准确率、假正率和假负率; T_1 和 T_2 分别代表 TP 和 TN; F_1 和 F_2 分别代表 FP 和 FN。

2.3.2 算法识别结果分析

采用本文提出的算法、多变量 Fast-MCD 算法^[8]、单变量 Fast-MCD 算法^[9]和 KPCA 算法^[10]对 1 549 个 DGA 数据进行异常数据识别,对比结果如表 2 所示。

表 2 不同算法的性能对比

Table 2 Performance comparison between different methods %

算法	评价指标		
	准确率 Q_A	假正率 Q_{FPR}	假负率 Q_{FNR}
本文算法	99.10	2.80	0.54
KPCA ^[10]	86.50	4.70	15.60
单变量 MCD ^[9]	90.45	16.40	8.23
多变量 MCD ^[8]	85.20	57.10	6.28

由表 2 可知:相较于 KPCA、单变量 Fast-MCD、多变量 Fast-MCD,本文提出的算法准确率高达 99.1%,假正率为 2.8%,假负率为 0.54%,综合性能优于其他算法。尽管 KPCA 假正率较低,但其假负率高,容易把正常数据误判为异常数据。多变量 MCD 的准确率较低,且假正率高,容易把异常数据误判为正常数据。单变量 MCD 虽然准确率较高,但其假正率和假负率均较高。

3 结论

1)本文提出的算法能够准确稳定地识别不同样本数据中的异常数据,无需划分训练集与测试集,能够应用于大规模 DGA 数据的异常识别。

2)利用 PCA 对特征气体进行特征量的提取,实现对多维特征量的高效降维,有效消除了冗余特征信息,能够识别出单一多变量 Fast-MCD 算法识别不出的气体长期零值。

3)相较于传统方法,基于 PCA 与 Fast-MCD 的在线监测装置异常识别方法具有较高的准确率,其准确率达到了 99.1%。该研究可为在线监测装置异常数据的识别提供参考。

参考文献:

- [1] 郎雪淞,唐红,赵君娇,等. 载气发生器式油色谱在线监测装置技术及应用[J]. 东北电力技术,2019,40(6):4-6.
LANG Xuesong, TANG Hong, ZHAO Junjiao, et al. Technology and application of carrier gas generator oil chromatography on-line monitoring device[J]. Northeast Electric Power Technology, 2019, 40(6): 4-6.
- [2] 苏鑫. 电力变压器油色谱在线监测装置的故障识别[D]. 福州:福建工程学院,2019.
SU Xin. Fault identification of on-line oil chromatographic monitoring device for power transformer[D]. Fuzhou: Fujian University of Technology, 2019.
- [3] 张正东,宋博,甘强. 主变在线监测装置缺陷导致轻瓦斯报警事件的分析[J]. 浙江电力,2023,42(5):120-124.
ZHANG Zhengdong, SONG Bo, GAN Qiang. Analysis of a light gas alarm triggered by a fault of an online monitoring device of the main transformer[J]. Zhejiang Electric Power, 2023, 42(5): 120-124.
- [4] FARIA H, COSTA J S, OLIVAS J L M. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis[J]. Renewable and Sustainable Energy Reviews, 2015, 46: 201-209.
- [5] 荣智海. 基于油色谱监测数据的变压器差异化评价及分层诊断方法[D]. 北京:华北电力大学,2020.
RONG Zhihai. Differential evaluation and hierarchical diagnosis method of transformer based on dissolved gas monitoring [D]. Beijing: North China Electric Power University, 2020.
- [6] 李红雷,何清,钱之银. 变压器油中特征气体分析诊断及检测技术[M]. 北京:中国电力出版社,2020.
- [7] WANI S A, GUPTA D, FAROOQUE M U, et al. Multiple incipient fault classification approach for enhancing the accuracy of dissolved gas analysis (DGA)[J]. IET Science, Measurement & Technology, 2019, 13(7): 959-967.
- [8] 高树国,王学磊,李庆民,等. 基于 MCD 稳健统计分析的变压器油色谱异常值检测及分布特性[J]. 高电压技术,2014,40(11):3477-3482.
GAO Shuguo, WANG Xuelei, LI Qingmin, et al. Outliers detection and distribution characteristics of the transformer DGA

- data based on MCD robust statistics[J]. *High Voltage Engineering*, 2014, 40(11): 3477-3482.
- [9] 李振杰, 卞朝晖, 陈学民, 等. 变压器油色谱在线监测数据有效性评估[J]. *高压电器*, 2018, 54(6): 158-163.
LI Zhenjie, BIAN Zhaohui, CHEN Xuemin, et al. Validity evaluation of online monitored chromatographic data of transformer oil[J]. *High Voltage Apparatus*, 2018, 54(6): 158-163.
- [10] 荣智海, 齐波, 张鹏, 等. 基于核主成分分析的油色谱在线监测装置异常状态快速辨识[J]. *高电压技术*, 2019, 45(10): 3308-3316.
RONG Zhihai, QI Bo, ZHANG Peng, et al. Fast identification of abnormal state based on kernel principal component analysis for oil chromatography analysis on-line monitoring device[J]. *High Voltage Engineering*, 2019, 45(10): 3308-3316.
- [11] 黄云程, 高阿娜, 王翌, 等. 变压器油色谱在线监测数据甄别与清洗技术[J]. *电力科学与工程*, 2019, 35(6): 37-43.
HUANG Yuncheng, GAO Ana, WANG Yi, et al. Screening and cleaning technology of transformer oil chromatographic on-line monitoring data[J]. *Electric Power Science and Engineering*, 2019, 35(6): 37-43.
- [12] 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.
LIU Yunpeng, WANG Quan, XU Ziqiang, et al. Research on data cleaning and abnormal recognition method of dissolved gas in oil based on multi-layer architecture[J]. *Journal of North China Electric Power University (Natural Science Edition)*, 2022, 49(1): 81-89.
- [13] 沈晓东. 基于小波模极大值变压器油色谱在线异常数据识别[J]. *新型工业化*, 2018, 8(6): 26-31.
SHEN Xiaodong. Anomaly data identification based on wavelet modulus maximum transformer oil chromatography[J]. *The Journal of New Industrialization*, 2018, 8(6): 26-31.
- [14] 时海涛, 苏智剑. 齿轮测量中异常值的判定及剔除系统设计[J]. *机械制造与自动化*, 2020, 49(3): 53-56.
SHI Haitao, SU Zhijian. Design of judgment and elimination system for outliers in gear measurement[J]. *Machine Building & Automation*, 2020, 49(3): 53-56.
- [15] 周建平, 李聪, 万书亭, 等. 基于优化型SVM的高压断路器故障诊断方法研究[J]. *浙江电力*, 2019, 38(3): 14-19.
ZHOU Jianping, LI Cong, WAN Shuting, et al. Research on fault diagnosis of high voltage circuit breaker based on optimized SVM[J]. *Zhejiang Electric Power*, 2019, 38(3): 14-19.
- [16] PENG C, CHEN Y Y, KANG Z, et al. Robust principal component analysis: a factorization-based approach with linear complexity[J]. *Information Sciences*, 2020, 513: 581-599.
- [17] ZHAO B T, DONG X, GUO Y C, et al. PCA dimensionality reduction method for image classification[J]. *Neural Processing Letters*, 2022, 54(1): 347-368.
- [18] 刘伟鹏, 张国钢, 刘亚魁, 等. 基于主成分分析和支持向量机的高压断路器机械状态识别方法[J]. *高压电器*, 2020, 56(9): 267-272.
LIU Weipeng, ZHANG Guogang, LIU Yakui, et al. Mechanical status identification of high voltage circuit breakers based on principal component analysis and support vector machines[J]. *High Voltage Apparatus*, 2020, 56(9): 267-272.
- [19] LI X, DENG S R, LI L F, et al. Outlier detection based on robust mahalanobis distance and its application[J]. *Open Journal of Statistics*, 2019, 9(1): 15-26.
- [20] ROUSSEEUW P J, VAN DRIESSEN K. A fast algorithm for the minimum covariance determinant estimator[J]. *Technometrics*, 1999, 41(3): 212.
- [21] 张倩茹. 充分降维问题中基于MCD方法的稳健降维[D]. 昆明: 云南财经大学, 2021.
ZHANG Qianru. Robust dimension reduction based on MCD method in sufficient dimension reduction[D]. Kunming: Yunnan University of Finance and Economics, 2021.
- [22] 杨雅琴, 徐鹏, 吴细水. 基于Fast-MCD的自适应建模探索轨道不平顺劣化[J]. *清华大学学报(自然科学版)*, 2022, 62(3): 516-522.
YANG Yaqin, XU Peng, WU Xishui. Adaptive modeling method based on the Fast-MCD to analyze railway track irregularity deterioration[J]. *Journal of Tsinghua University (Science and Technology)*, 2022, 62(3): 516-522.
- [23] 李诗勇, 薛静, 吴冕之, 等. 基于粒子群优化的改进加权支持向量回归的变压器顶层油温预测[J]. *高压电器*, 2021, 57(12): 103-109.
LI Shiyong, XUE Jing, WU Mianzhi, et al. Prediction of transformer top-oil temperature with the improved weighted support vector regression based on particle swarm optimization[J]. *High Voltage Apparatus*, 2021, 57(12): 103-109.

- [24] 盛乐,李学锋,周波,等. 油色谱在线监测装置运维情况分析[J]. 国网技术学院学报,2020,23(3):1-6.
SHENG Le,LI Xuefeng,ZHOU Bo,et al. Analysis of operation and maintenance of online monitoring devices for oil chromatography[J]. Journal of State Grid Technology College,2020,23(3):1-6.
- [25] 童超,童军心,周梦垚,等. 油色谱在线监测装置异常情况特征分析[J]. 江西电力,2022,46(2):31-34.
- [26] 刘云鹏,许自强,和家慧,等. 基于条件式 Wasserstein 生成对抗网络的电力变压器故障样本增强技术[J]. 电网技术,2020,44(4):1505-1513.
LIU Yunpeng,XU Ziqiang,HE Jiahui,et al. Data augmentation method for power transformer fault diagnosis based on conditional Wasserstein generative adversarial network[J]. Power System Technology,2020,44(4):1505-1513.
- [27] 唐志国,曹智,何宁辉. 卷积神经网络迁移学习在局部放电类型诊断中的应用[J]. 高压电器,2022,58(4):158-164.
TANG Zhiguo,CAO Zhi,HE Ninghui. Application of convolutional neural network transfer learning in partial discharge type diagnosis[J]. High Voltage Apparatus,2022,58(4):158-164.