

DOI:10.13296/j.1001-1609.hva.2024.07.019

基于多分类支持向量机的变压器在线监测数据错误模式识别

何宁辉¹, 吴旭涛¹, 张佩¹, 沙伟燕¹, 周秀¹, 丁培¹,
杨擎柱², 程养春²

(1. 国网宁夏电力有限公司电力科学研究院, 银川 750011; 2. 华北电力大学高电压与电磁兼容北京市重点实验室, 北京 102206)

摘要: 针对变压器在线油中溶解气体在线监测数据质量问题,统计了200多台监测装置的2020全年数据,总结了3种主要数据错误模式;提出了数据错误模式识别策略和特征参数,构建了多分类支持向量机进行错误数据识别与分类;并利用核主成分分析法和排列组合遍历寻优法对特征向量进行了降维优化。所构建的多分类支持向量机分类器对于H₂错误数据识别准确率达到97.5%,对于其他气体达到90%以上。应用所构建的分类器对2020全年数据进行了统计,其中H₂的错误数据达到27.14%,C₂H₂的错误数据达到1.75%。

关键词: 错误数据; 模式识别; 支持向量机; 在线监测; 变压器; 油中溶解气体分析

Error Pattern Identification of On-line Monitoring Data of Transformer Based on Multi-classification SVM

HE Ninghui¹, WU Xutao¹, ZHANG Pei¹, SHA Weiyan¹, ZHOU Xiu¹,
DING Pei¹, YANG Qingzhu², CHENG Yangchun²

(1. State Grid Ningxia EPRI, Yinchuan 750011, China; 2. Beijing Key Laboratory of High Voltage and EMC, North China Electric Power University, Beijing 102206, China)

Abstract: In view of online monitoring data quality of dissolved gas in transformer oil, the annual data of more than 200 monitoring devices in 2020 are counted and three main data error patterns are summarized. The identification strategy and characteristic parameters of data errors pattern are proposed, and a multi-classification support vector machine is constructed to identify and classify the data errors. The Kernel principal component analysis and permutation and combination traversal method are used to reduce dimension of the feature vector. The identification accuracy of the constructed multi-classification support vector machine classifier for H₂ error data is 97.5%, and for other gasses is more than 90%. The constructed classifier is used to make statistics on the annual data of 2020, in which the error data of H₂ and C₂H₂ reach 27.14% and 1.75% respectively.

Key words: error data; pattern identification; support vector machines; online monitoring; transformer; dissolved gas analysis

0 引言

电力变压器油中溶解气体在线监测技术是目前现场变压器能够采用的主要在线监测手段,得到

了广泛应用。其中绝大部分监测装置(下文简称装置)基于气相色谱分析原理^[1-3]。监测结果被用于变压器故障预警、状态评估、缺陷发展趋势分析和变压器寿命预测等^[4-8]。该类装置是实现变压器智能

收稿日期:2024-02-21; 修回日期:2024-05-07

基金项目:国家电网科学技术项目(5229DK19004Z)。

Project Supported by Science and Technology Project of SGCC(5229DK19004Z).

化运维的重要支撑。特别是在国家倡导的“碳达峰、碳中和”目标下,在线监测装置是实现电力设备和电力网络双向互动,进而实现安全、高效、灵活消纳高比例新能源电力的重要基础。装置的数据质量问题也逐渐得到重视。

由于存在产品质量瑕疵、数据传输通道不畅、现场干扰、装置日常维护不到位等情况,装置输出的数据质量不高^[9]。装置输出的错误数据严重影响了实时报警的准确度和可信度,还对基于历史数据的信息挖掘结果造成巨大影响,严重削弱了装置的应用效果。

根据变电站在线监测装置维护人员的经验,基于气相色谱法的装置经常会因载气欠压、色谱柱老化而造成输出结果偏离正确值;由断路器或者刀闸动作所产生的强电磁脉冲对装置内部的电路造成瞬态干扰,导致装置输出结果出现严重偏差;由数据传输通道中的偶然错误也会导致数据出现明显错误或者溢出^[10]。因此,非常有必要对装置输出的数据进行分析,剔除错误数据;并且识别错误类型,依此有针对性地提升装置性能和维护水平,进而提升现有装置的数据质量。

对于装置的数据质量问题的研究,仍处于起步阶段,缺乏明确的定义和统一的衡量标准。数据序列中存在的多种错误模式之间缺乏明确的界限,并且往往相互交织在一起。数据序列的错误现象还与对应的电力设备状态量的特点密切相关,存在显著的特异性。这些情况给数据错误模式的识别带来巨大难度。亟需开展在线监测装置数据错误模式识别方法的研究。

但是,对于异常数据识别方法的研究已经有相当多的成果。诸如聚类算法、马尔可夫模型、时间序列分析方法已经应用到电网异常负荷预测与检测中^[11-19]。利用智能算法识别电力设备状态量数据序列中的异常现象已经得到广泛应用,这给我们识别在线监测装置的错误数据提供了参考。其中多分类支持向量机(support vector machine, SVM)能够有效识别各种异常模式,算法结构简单,自学习泛化能力强^[16],非常适合文中监测装置错误数据的识别。

文中利用SVM方法,针对变压器油中溶解气体在线监测装置,开展了错误数据模式识别方法的研究,总结了数据错误的模式,提出了识别策略,优化了特征向量,构建了多分类支持向量机分类器并检验了其准确度。

1 数据错误模式的种类

为了统计变压器油中溶解气体在线监测装置

数据发生错误的情况,文中收集了中国4个网省公司200多台基于气相色谱原理的监测装置的历史数据。这些历史数据的形式是每日一测的时间序列,持续时间为2020年度全年。数据中包括氢气、甲烷、乙烷、乙烯、一氧化碳、二氧化碳和乙炔等7种气体的体积分数。其中氢气等前6种气体的错误模式比较相似,而乙炔的错误模式相对特殊。我们总结到突变、抖动、固化等3种主要错误模式。

1.1 数据突变

数据发生突变的典型现象见图1,在一段数据序列中,大部分数据处于正常数值范围,个别数据显著偏离之前近期的数据。有时这种偏离现象持续时间很短,成为孤立的“离群”数据;有时这种偏离一直持续,形成新的“基线”。前者往往源自装置受到瞬时干扰,后者往往源自载气欠压或者装置的传感部件出现系统性偏差。

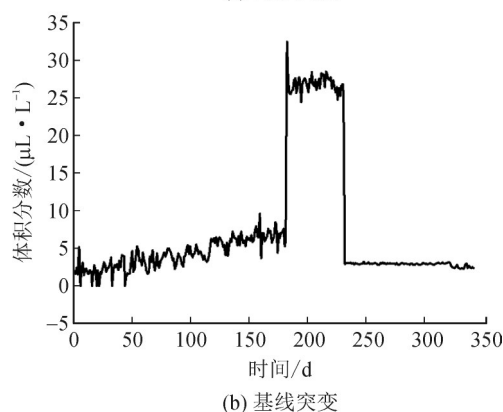
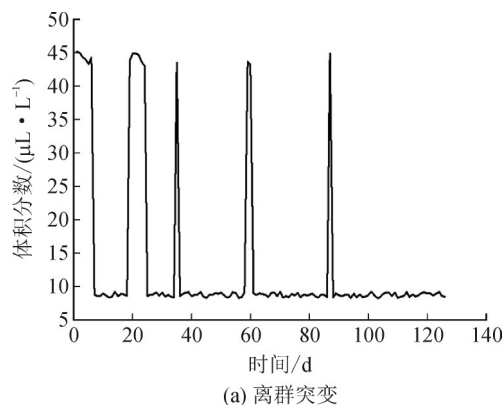


图1 数据突变现象

Fig. 1 Data mutation

1.2 数据抖动

数据发生抖动的典型现象见图2。在一段数据序列中,数据发生超出正常范围的剧烈波动,出现多次交替突增和突减。鉴于变压器油中溶解气体的消散过程比较缓慢,气体体积分数的下降过程漫长而缓慢,显然这种剧烈抖动的数据不符合油中溶解气体的自然变化规律。发生这种情况的原因可能是装置的电路元件损坏,产生了巨大的随机噪

声;也可能是数据上传通道受到干扰,上传的数据断断续续,在零值与正常值之间来回跳转。

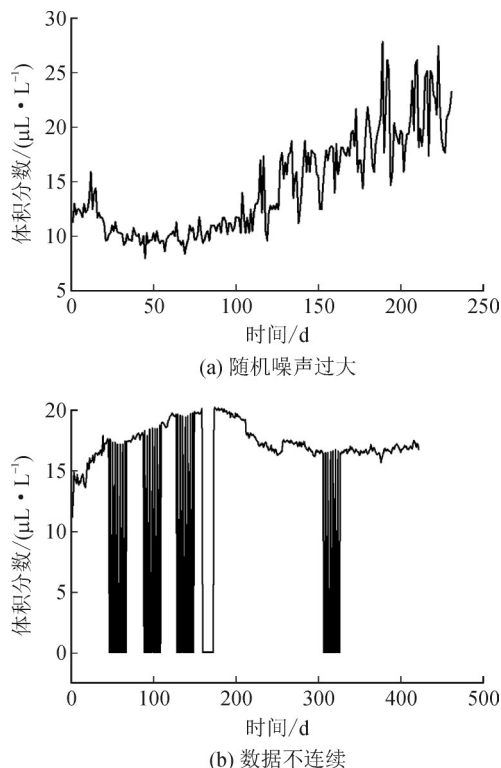


图2 数据抖动现象

Fig. 2 Data jitter

1.3 数据固化

数据发生固化的典型现象见图3。在一段数据序列中,数据固定在某一数值上,缺少必要的变化。正常情况下,油中溶解气体的体积分数会因气体的产生和扩散过程中的各种因素的变化而变化;即使气体体积分数稳定不变,监测结果也会因各种随机干扰而表现出某种概率分布特征。因此,装置输出的数据固定不变是一种异常现象。其缘由可能是装置内部控制系统出现错误,持续上传之前的某一数据;或者数据交换机发生错误,持续发送某一虚假数据。

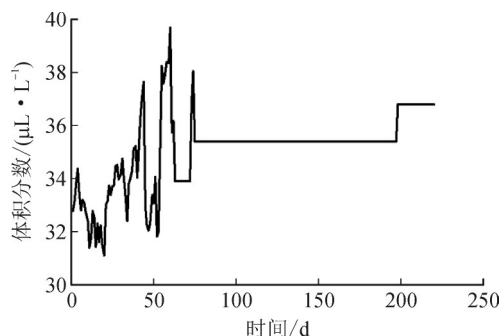


图3 数据固化现象

Fig. 3 Data curing

对于变压器油中溶解气体,其中氢气、甲烷、乙烷、乙烯、一氧化碳、二氧化碳等前6种气体的错误

模式比较相似,存在上述3种模式。一般情况下,乙炔体积分数在较低水平,甚至长期保持零值。其错误数据情况主要为突变,从零值突变到超出注意值又回到零值,类似图1(a)。因此,乙炔数据的错误模式主要是数据突变。

2 数据错误模式的识别策略

2.1 数据错误模式识别流程

在现场实际中,在线监测装置不但承担监视电力变压器绝缘状态的作用,而且肩负故障预警的作用。前者要求监测装置输出的数据尽量准确;后者要求时效性。因此,数据错误现象识别流程和结果既不能影响对突发故障的预警时效,又不能影响对设备状态的准确表征。

针对上述要求,文中设计了数据错误模式识别流程见图4。在线监测装置每隔 Δt 时间监测一次状态量 $x(t)$ 并上传数据,形成以 x_i 为元素的离散数据流 X ;数据流通过数据序列提取窗口时,被选中相对于时刻 t 而言最新的 N 个数据,形成具有 N 个元素的数据序列片段 $X(\cdot)$,其元素是 $x_i = x(t - i\Delta t)$, $i=0, 1, 2, \dots, N-1$ 。该片段进入数据错误模式诊断程序,确定该片段所属的错误模式,并确定其中错误数据的具体位置。最后,对错误数据进行标记,标记符号中包含错误模式(图4中用不同颜色深浅表示不同的错误模式)。

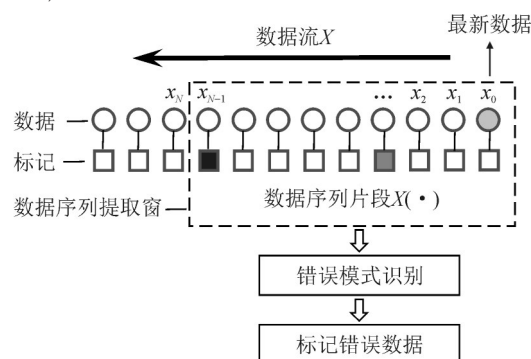


图4 数据错误模式识别流程

Fig. 4 Data error pattern recognition process

每隔 Δt 就有一个新数据进入数据流,用于错误该模式识别的数据序列片段也将从 $X(\cdot)$ 更新到 $X_{t+\Delta t}(\cdot)$ 。该新序列中的元素数据 x'_i 按照下列公式进行更新:

$$x'_0 = x(t + \Delta t) \quad (1)$$

$$x'_i = x_{i-1} = x(t + \Delta t - i\Delta t) \quad i = 1, 2, \dots, N-1 \quad (2)$$

数据序列片段被更新之后,重新运行错误模式识别程序。当装置上传了一个错误数据时,这个数据将在数据序列片段 $X(\cdot)$ 中存在的时间长度为

$N\Delta t$, 会被识别 N 次, 相应地有 N 个识别结果, 而且这 N 个结果可能会因 $X(\cdot)$ 的更新而有所差异。完成每次错误模式识别后, 需要根据最新的识别结果重新标记数据的错误模式。因此, 文中为各错误模式确定了优先等级, 从高到低依次是抖动、突变、固化模式, 最后是正常数据。在重新标记错误数据时, 优先等级高的模式取代优先等级低的模式, 反之不能。

2.2 错误模式识别算法

一方面, 数据错误模式的识别基于数据序列片段的整体情况, 需要观察前后数据之间的关联性。在某一片段中, 可能同时存在多种错误模式。诊断结果与片段的长度密切相关。另一方面, 目前对于各类错误模式缺乏明确的界定标准, 主要依靠人工识别。因此, 文中拟采用多分类支持向量机算法, 建立自动分类器, 代替人工识别。自动分类器的建立过程见图5。首先, 基于前文数据错误现象, 确立潜在的系列特征参数; 然后基于人工识别结果, 建立各种数据错误模式、正常模式的训练样本集和测试样本集; 再次, 基于多分类支持向量机算法, 建立许多个分类器(其中每各分类器对应于不同的特征参数组合), 并进行模型训练; 最后, 对比各个分类器的准确度, 从中选取准确度最高的分类器。由于在多分类支持向量机模型中, 特征量的种类和数量与分类器是一一对应的, 则最佳的特征量组合对应于最佳的分类器。

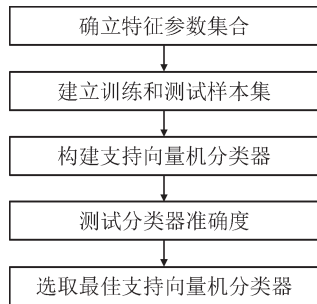


图5 数据错误模式识别算法的确立过程

Fig. 5 The establishment process of data error pattern recognition algorithm

2.3 特征向量的确立

在现场实际应用中, 要求数据错误模式识别算法具有尽量短的运行时间, 从而能够在装置的检测时间间隔 Δt 内完成错误数据的识别及后续处理。因此, 要求在保证一定准确度的前提下采用尽量少的特征参数, 以减少模型的运算时间。特征参数作为元素, 组成特征向量 T 。不同种类和数量的特征参数构成不同的特征向量。

文中首先确立了6个特征参数, 然后分别通过

主成分分析法和排列组合法来压缩特征参数的数量, 建立了多个特征向量, 最后通过对比各特征向量对应的支持向量机分类器的准确度, 得到最佳特征向量。

用于时间序列异常诊断的常用特征参数有变异系数 V 、标准差 S 、平均绝对误差 E_a 、极差 D ^[15-16]:

$$V = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \bar{x})^2} / \left(\frac{1}{N} \sum_{i=0}^{N-1} x_i \right) \quad (3)$$

$$S = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \bar{x})^2} \quad (4)$$

$$E_a = \frac{1}{N} \sum_{i=0}^{N-1} |x_i - \bar{x}| \quad (5)$$

$$D = \max X(\cdot) - \min X(\cdot) \quad (6)$$

式(3)-(6)中: $X(\cdot)$ 是 t 时刻的数据序列片段, 省去了下标 t ; N 是该序列中元素个数; x_i 是该序列的元素; $\bar{x}$ 是该序列所有元素的平均值; $\max()$ 和 $\min()$ 函数分别求取序列中的最大值和最小值。

此外, 文中还结合前文错误模式的特点, 定义了两个新的特征参数: 一阶差分数据的零值比 K_L 和超高变化率占比 K_S 。 K_L 大致反映了在一段时间内数据固定不变的数量, 能有效表征数据固化模式, 其表达式为

$$K_L = \frac{1}{N} \text{Num}[D_x(\cdot), d_i = 0] \quad (7)$$

式(7)中: $\text{Num}(\cdot)$ 函数求取其自变量序列中符合条件的元素的个数, 在这里是数值等于0的元素的个数; $D_x(\cdot)$ 是 $X(\cdot)$ 序列的一阶差分序列, 其元素 d_i 由式(8)确定

$$d_i = x_i - x_{i+1} \quad i = 0, 1, 2, \dots, N-2 \quad (8)$$

K_S 大致反映了在一段时间内数据发生剧烈增大和减小的次数, 能有效表征数据抖动模式, 其表达式为

$$K_S = \frac{1}{N} \text{Num}[D_x(\cdot), |d_i| > s_p] \quad (9)$$

式(9)中, $\text{Num}(\cdot)$ 函数求取了 $D_x(\cdot)$ 序列中绝对值大于阈值 s_p 的元素个数。

2.4 样本集的构建

数据序列片段 $X(\cdot)$ 的长度 N 的选取对于序列的识别准确度和花费时间有较大影响。文中参考现有研究成果, 将 N 设定为15, 不但具有较高的准确度, 而且运算时间较小^[17]。

在构建样本集时, 采用人工识别的方法从现场实际在线监测数据中挑选出前文所述的3种错误模式的数据序列样本, 作为训练样本和测试样本。数据错误模式识别程序的作用是发现在线装置上传数据中的错误, 而不是诊断电力变压器是否存在故

障。将正确数据误判成错误数据可能会导致应该报警而未能预警, 而将错误数据漏判成准确数据可能会引发虚假报警。对于两者后果权衡后, 应尽量降低误判率, 并按此原则确定人工识别标准。

在区分数据序列的正常波动与数据抖动时, 需要通过统计历史数据获取其阈值 s_p , 文中通过统计四省 200 台装置一年内的装置气体增速分布规律, 得到气体增速分布函数, 然后设置关键分位点确定 s_p 值。获取的 H_2 速率分布见图 6, 利用拟合工具发现其最符合 Logistic 分布, 计算得其 90 分位点为 1.4, 将该值作为异常变化速度的阈值, 其他气体的判断阈值也可以通过该方法得到。 K_s 与 K_L 值的设定, 需要根据历史错误数据情况进行设定。在对历史数据统计之后, 发现发生数据剧烈变化以及数据固化的情况超过数据的 40% 便可以进行人工判断, 故将人工判断抖动和固化的 K_s 和 K_L 的值设定为 0.4^[15]。

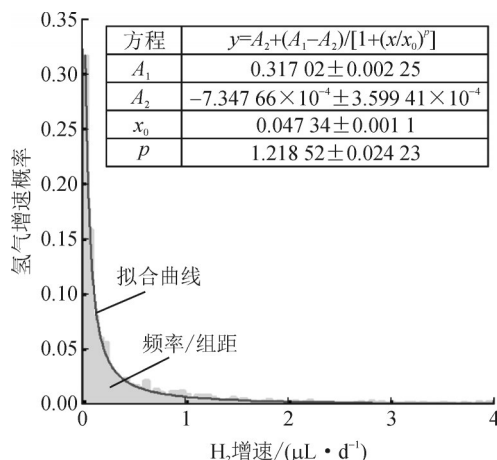


图6 氢气增速频率分布图

Fig. 6 Frequency distribution of Hydrogen increase rate

此外, 在识别是否错误数据时, 需要根据相应状态量的特点, 对于判断标准和所用特征参数进行适当调整。氢气、甲烷、乙烷、乙烯、一氧化碳、二氧化碳等 6 种气体的错误模式比较相似, 可以采用相同的特征向量和分类器; 而乙炔的错误模式主要是数据突变, 数据抖动和数据固化模式和特征量 K_L 均不适用, 需要单独的特征向量和分类器。

3 基于支持向量机的分类方法

3.1 多分类支持向量机算法简介

支持向量机是以结构风险最小化原则作为基础实现的分类器。基本的 SVM 是一种二分类器, 它通过构建最优超平面作为分类界面, 将数据分成正类和反类。这个超平面是由特征向量在高维空间形成的^[17-21]。文中要识别 4 类数据序列, 即正常数据、数据突变、数据抖动、数据固化序列(分别用编号

1、2、3、4 表示), 需要构建多分类 SVM。文中利用 MATLAB 软件平台中的 LIBSVM 函数来具体构建多分类 SVM。其中, SVM 的核函数选用常用的高斯核, 其最佳参数通过网格搜索法得到。

多分类 SVM 由多个二分类 SVM 组成。如果存在 M 类数据待划分, 则每两个类别之间需要训练一个二分类 SVM, 总共需要 $M(M-1)/2$ 个二分类 SVM。这些二分类 SVM 的集合加上投票机制就形成了多分类 SVM。进行模式识别时, 利用多分类 SVM 中的每一个二分类 SVM 对需要识别的数据进行分类, 统计各分类器的输出结果, 以得票最多的类别作为待识别数据的最终类别^[9]。多分类 SVM 的工作流程见图 7。

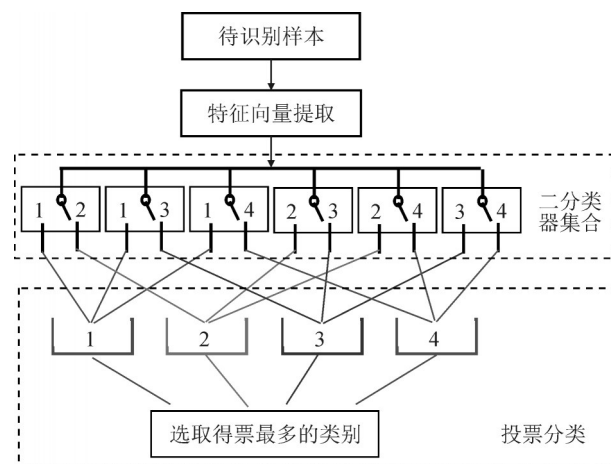


图7 多分类支持向量机工作流程示意

Fig. 7 Multi-classification support vector machine workflow diagram

对于每一个二分类支持向量机, 其分类超平面函数均由包含这两类的训练样本来决定

$$\hat{y} = \text{sgn} \left(\sum_{i=1}^n \alpha_i y_i k(\mathbf{t}_i, \mathbf{t}) + b \right) \quad (10)$$

其中 $\hat{y}$ 是被辨识样本的分类结果(正类取 1, 负类取 -1); a_i 是拉格朗日乘数; y_i 是第 i 个样本的类别值(正类取 1, 负类取 -1); $\mathbf{t}_i$ 是第 i 个样本的特征向量; $\mathbf{t}$ 是被辨识样本的特征向量; b 是超平面曲线的截距; $k(\mathbf{t}_i, \mathbf{t})$ 是内积函数文中取高斯内积函数。

3.2 特征向量的优化

为了提高计算速度和分类准确性, 在对数据序列片段进行识别时, 需要首先提取其中的多个特征参数, 形成特征向量 $\mathbf{t}$; 再将 $\mathbf{t}$ 代入分类器进行模式识别。文中初定的 6 个特征参数, 可能存在相关性, 需要进一步优化压缩。文中分别通过核主成分分析法(kernel principal component analysis, KCPA)和特征参数排列组合遍历寻优法来选取最优的特征向量。

3.2.1 核主成分分析法降维

KPCA是一种常用的数据融合和维数约简的方法,原理为:构造高维特征空间,把原始变量转化为空间内积矩阵,并利用原始变量空间的核函数去代替内积函数。KCPA进行特征量降维算法流程如下^[21-25]。

1)利用 n 个样本的 n 个特征向量构建特征向量矩阵 $T_{n \times m}$, m 为特征向量的维数;并进行z-score标准化处理。

2)选取高斯核函数,计算得到核矩阵 K 。核矩阵 K 中元素的计算式为

$$K(t_i, t_j) = \exp\left(-\frac{\|t_i - t_j\|^2}{2\sigma^2}\right) \quad (11)$$

式(11)中: t_i, t_j 为特征向量矩阵行元素; d 为核参数。

3)将核矩阵 K 去中心化后得到矩阵 K_p ,求取 K_p 的特征值 λ_i 及其特征向量 X_i 。

4)将 K_p 的特征值 λ_i 降序排列,对特征向量 X_i 的排序随之变化,形成 K_p 的特征向量矩阵 Q 。再利用施密特正交法对 Q 进行单位正交处理。

5)根据贡献率大小选取前 s 个特征值 λ_i 对应特征向量,即 Q 的前 s 列,构成矩阵 Q_s 。文中选取总贡献率达到90%的前 s 个特征值 λ_i 。

6)计算降维后的特征向量矩阵 Z

$$Z = K_p \cdot Q_s \quad (12)$$

Z 的列数 s 一般小于原来的列数 m ,起到了降维作用。但是保留了原特征向量矩阵 T 的主要特征。

3.2.2 排列组合遍历寻优

文中初定的特征参数为6个。在这6个特征参数中任选两种及以上,可以得到56种不同的组合,即56种特征向量。相应地,构建56种多分类SVM,对比各个多分类SVM的准确度,选择具有最高准确度的特征向量及其对应的多分类SVM。

4 多分类支持向量机测试结果

4.1 样本集

从4个网省公司多达200余台监测装置的2020年度全年在线监测数据中,选取正常数据、数据突变、数据抖动、数据固化等4类数据序列片段,并且分别用编号1、2、3、4来代表其类别。每个数据序列片段的长度 N 等于15。每个类别选30个样本,共120个样本,作为训练样本用于构造SVM。

为了同时检验多分类SVM的泛化能力,用2021年度在线监测数据构建测试样本集。共构建了20个测试样本集。每个测试样本集中的样本数量为20个,对应于4个数据类型、每个数据类型5个样本。

样本所属的类型,按照前文所述的样本集构造策略,由人工识别确定。通过对比SVM的识别结果与人工识别结果,来确定SVM的准确度。

4.2 最佳SVM分类器的确定

按照前文所述方法,构建57个多分类SVM,分别对应于来自核主成分分析的特征向量和排列组合的特征向量。将训练集依次代入每个多分类SVM,构建各二分类器的超平面。然后代入测试样本集进行测试,计算各个多分类SVM的平均准确度。

以 H_2 为例,准确度较高的几个多分类SVM排序见表1。用KCPA方法,选用90%的贡献率能将特征向量降到四维,准确度为89%。从排列组合选取的特征向量优选结果可以看出,高准确率的结果总是包含特征参数 K_s 和 K_L ;而且只需 V 、 K_s 和 K_L 3个特征参数即可达到97.5%的准确度。

表1 各多分类支持向量机对 H_2 数据分类准确度对比

Table 1 Comparison of classification accuracy of each multi-classification support vector machines for H_2 data

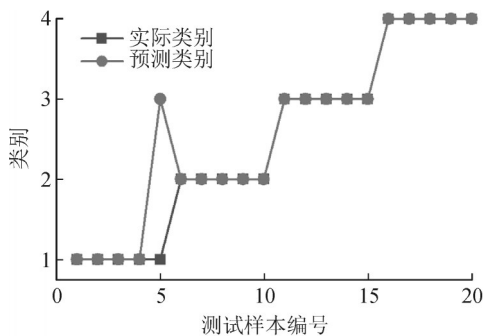
特征向量	准确度/%
KPCA输出特征向量	89.0
$\{V, S, E_a, D, K_s, K_L\}$	97.5
$\{V, S, K_s, K_L\}$	97.5
$\{V, K_s, K_L\}$	97.5
$\{K_s, K_L\}$	85.0

部分测试结果见图8。在以 $\{V, K_s, K_L\}$ 为特征向量的多分类SVM中,误判全是将正常数据误判为错误数据,而将错误数据误判为正常数据的情况为0。这意味着通过错误数据识别程序对监测数据进行处理可以有效避免错误数据带来的误报警等不良后果;而且没有出现错误数据被分错或者漏掉的现象,非常有利于后续查找装置故障原因。

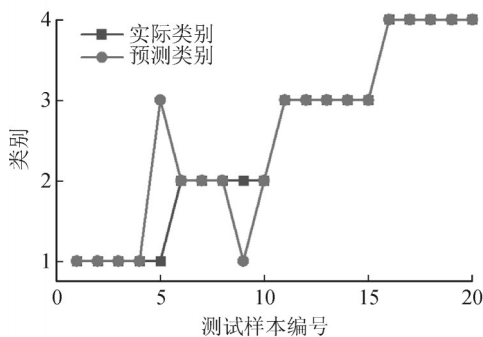
对于如图8(a)中的将正常数据判断成错误数据的情况,对错判案例集中分析后,发现错误主要来自增速为设置值边界位置的数据。这种情况较少,缺乏足够多的样本进行准确分类,通过增设训练样本可以减少此类错判。

结合上述分析 $\{V, K_s, K_L\}$ 是最佳特征向量,对应的多分类支持向量机是最佳的分类器,对错误类型的识别准确率达到97.5%。

利用该方法得到除乙炔外的5种气体具有相类似的结果,其中不同气体的超高变化率占比 K_s 中判断阈值 s_p 的设置需要随气体组分特点而改变^[15]。利用以 $\{V, K_s, K_L\}$ 为特征向量的SVM对各气体的错误数据分类准确度测试结果见表2,都达到90%以上。表明该分类器对其他气体也同样适用。



(a) 三特征参数



(b) KPCA降维特征向量

图8 典型多分类SVM分类器测试结果

Fig. 8 Test results of typical multi-classification

SVM classifier

表2 SVM对各气体错误数据识别准确率

Table 2 SVM identification accuracy of each gas

error data	
气体类型	准确度/%
H ₂	97.5
CH ₄	90.5
C ₂ H ₆	94.0
C ₂ H ₄	91.5
CO	97.5
CO ₂	98.5
C ₂ H ₂	100.0

由于乙炔气体的特殊性,错误数据类别主要有数据突变模式这一个类别,仅需一个二分类器。我们只用了特征参数 K_s ,就实现了100%的识别准确率。因此对于所有7类气体,文中的方法都能实现较高准确度的识别。

5 多分类支持向量机用于错误数据模式统计与实例标记

将文中构建的多分类SVM应用到2020年度200台装置的氢气体积分数数据分析中,识别并统计错误数据,结果见图9。目前的在线检测数据只有72.86%是正常的,错误数据占27.14%。错误数据中,数据陡动占比较大,为14.74%;而数据突变与数

据固化占比差不多,大约为6%。在2020年乙炔体积分数据中,错误占比仅为1.75%。

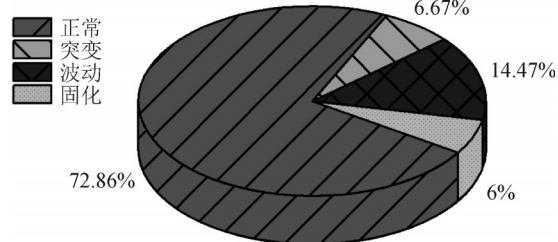


图9 氢气监测数据中各类错误数据占比

Fig. 9 Proportion of all kinds of wrong data in Hydrogen monitoring data

某省某变电站二号主变在6月到12月阶段的实际氢气测量数据见图10。通过文中流程的数据识别后,对其数据进行了标记,获取了其正常数据与各类错误数据,有助于数据的利用与错误原因的定位。结合PMS的检修记录,数据异常点未发现变压器异常运行状态,因此可以判断数据异常来自于装置本身的监测错误,需要对监测装置进行检修维护。

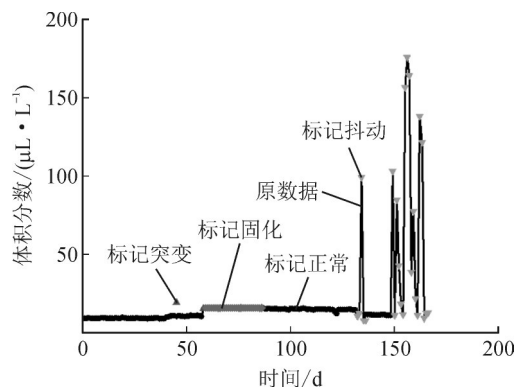


图10 异常数据识别实例

Fig. 10 Example of abnormal data identification

6 结论与展望

文中通过分析4个网省公司200余台油中溶解气体在线监测装置的历史数据,研究了基于多分类支持向量机的数据错误模式识别策略、方法和准确度。主要结论如下:

1)数据错误模式可分为数据突变、数据抖动和数据固化3类。

2)最佳特征向量由变异系数 V 、超高变化率占比 K_s 和一阶差分数据的零值比 K_L 组成,相应的所分类支持向量机的准确率对H₂达到97.5%,对CH₄、C₂H₆、C₂H₄、CO、CO₂、C₂H₂达到90%以上。

3)2020年度全年H₂的错误数据占全部数据的27.14%,C₂H₂的错误数据占全部数据的1.75%。

上述结论基于对错误数据模式的人工识别策略和判据,具有较大的主观性。后续还需要统计更

多年份的数据、继续细分数据错误模式、深入研究判据和特征参数,进一步提高错误数据模式识别的客观性和准确度;并进行相应的数据修复。

参考文献:

- [1] 万 达,伍志荣,孟玉婵,等. 简析涉及变压器类设备纸绝缘放电的油气相色谱特征[J]. 变压器,2023,60(2):10-21.
WAN Da, WU Zhirong, MENG Yuchan, et al. Analysis of oil gas chromatographic characteristics of paper insulation discharge in transformer equipment[J]. Transformer, 2023, 60(2): 10-21.
- [2] 鲁丽萍,刘 娟,邹丹平,等. 基于混沌特征的电力变压器油色谱监测装置有效性评估[J]. 变压器,2022,59(6):36-41.
LU Liping, LIU Juan, ZOU Danping, et al. Effectiveness evaluation of DGA monitoring devices in power transformer based on chaotic characteristics[J]. Transformer, 2022, 59(6): 36-41.
- [3] 鲁 永,赵映宇,陈 婕,等. 基于油色谱分析技术的变压器故障分析[J]. 变压器,2023,60(6):64-67.
LU Yong, ZHAO Yingyu, CHEN Jie, et al. Transformer fault analysis based on oil chromatography analysis technology[J]. Transformer, 2023, 60(6): 64-67.
- [4] 方 静,彭小圣,刘泰蔚,等. 电力设备状态监测大数据发展综述[J]. 电力系统保护与控制,2020,48(23):176-186.
FANG Jing, PENG Xiaosheng, LIU Taiwei, et al. Development trend and application prospects of big data-based condition monitoring of power apparatus[J]. Power System Protection and Control, 2020, 48(23): 176-186.
- [5] 潘育龙. 变压器油中溶解气体在线监测技术的研究[D]. 郑州:郑州大学,2019.
PAN Yulong. Research on on-line monitoring technology of dissolved gas in transformer oil[D]. Zhengzhou: Zhengzhou University, 2019.
- [6] 许易经,韩学山,杨 明,等. 基于设备在线监测的电网状态检修决策模型[J]. 电力系统自动化,2020,44(23):72-81.
XU Yijing, HAN Xueshan, YANG Ming, et al. Decision-making model of condition-based maintenance for power grid with equipment on-line monitoring[J]. Automation of Electric Power Systems, 2020, 44(23): 72-81.
- [7] 王 艳,李 伟,赵洪山,等. 基于油中溶解气体分析的DBN-SSAELM变压器故障诊断方法[J]. 电力系统保护与控制,2023,51(4):32-42.
WANG Yan, LI Wei, ZHAO Hongshan, et al. Transformer DGA fault diagnosis method based on DBN-SSAELM[J]. Power System Protection and Control, 2023, 51(4): 32-42.
- [8] 曹 旺,万 元,唐 伟,等. 变压器油中溶解气体在线监测系统[J]. 变压器,2023,60(5):31-37.
CAO Wang, WAN Yuan, TANG Wei, et al. Design of on-line monitoring system for dissolved gas in transformer oil[J]. Transformer, 2023, 60(5): 31-37.
- [9] 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.
- [10] 何 尧,梁宏池,连鸿松,等. 基于滑动窗口和多元高斯分布的变压器油色谱异常值检测[J]. 高压电器,2020,56(1):203-209.
HE Yao, LIANG Hongchi, LIAN Hongsong, et al. Outlier detection of power transformer oil chromatographic data based on algorithm sliding windows and multivariate gaussian distribution[J]. High Voltage Apparatus, 2020, 56(1): 203-209.
- [11] 赵天辉,张 耀,王建学. 基于空间密度聚类 and 错误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.
ZHAO Tianhui, ZHANG Yao, WANG Jianxue. Load outlier identification method based on density-based spatial clustering and outlier boundaries[J]. Automation of Electric Power Systems, 2021, 45(10): 97-105.
- [12] 严英杰,盛戈皞,陈玉峰,等. 基于时间序列分析的输变电设备状态大数据清洗方法[J]. 电力系统自动化,2015,39(7):138-144.
YAN Yingjie, SHENG Gehao, CHEN Yufeng, et al. Cleaning method for big data of power transmission and transformation equipment state based on time sequence analysis[J]. Automation of Electric Power Systems, 2015, 39(7): 138-144.
- [13] 毛李帆. 电网规划中长期负荷预测技术的研究[D]. 长沙:湖南大学,2011.
MAO Lifan. Research on medium and long term load forecasting technology for power grid planning[D]. Changsha: Hunan University, 2011.
- [14] 汤 琳. 基于马尔科夫预测的多传感器故障检测与诊断机制[J]. 传感器与微系统,2014,33(11):52-55.
TANG Lin. Multi-sensor fault detection and diagnosis mechanism based on markov prediction[J]. Transducer and Microsystem Technologies, 2014, 33(11): 52-55.
- [15] 齐 波,张 鹏,荣智海,等. 基于数据驱动和多判据融合的油色谱监测传感器有效性评估方法[J]. 电网技术,2017,41(11):3662-3669.
QI Bo, ZHANG Peng, RONG Zhihai, et al. Validity assessment method of DGA sensors based on data driven and multiple criterion integration[J]. Power System

- Technology, 2017, 41(11): 3662-3669.
- [16] 孙滢涛, 张锋明, 陈水标, 等. 基于多域特征提取的电力数据异常检测方法[J]. 电力系统及其自动化学报, 2022, 34(6): 105-113.
- SUN Yingtao, ZHANG Fengming, CHEN Shuibiao, et al. Power data anomaly detection algorithm based on multi-domain feature extraction[J]. Proceedings of the CSU - EPSA, 2022, 34(6): 105-113.
- [17] 严英杰, 盛戈皞, 刘亚东, 等. 基于滑动窗口和聚类算法的变压器状态异常检测[J]. 高电压技术, 2016, 42(12): 4020-4025.
- YAN Yingjie, SHENG Gehao, LIU Yadong, et al. Anomalous state detection of power transformer based on algorithm sliding windows and clustering[J]. High Voltage Engineering, 2016, 42(12): 4020-4025.
- [18] 钱宇骋, 甄超, 季坤, 等. 变压器在线监测数据异常值检测与清洗[J]. 哈尔滨理工大学学报, 2020, 25(5): 15-22.
- QIAN Yucheng, ZHEN Chao, JI Kun, et al. Transformer online monitoring data abnormal value detection and cleaning[J]. Journal of Harbin University of Science and Technology, 2020, 25(5): 15-22.
- [19] 刘伟, 韩彦华, 王荆, 等. 基于粒子群算法优化支持向量机的变压器绕组变形分类方法[J]. 高压电器, 2020, 56(3): 72-78.
- LIU Wei, HAN Yanhua, WANG Jing, et al. Transformer winding deformation classification method based on particle swarm algorithm optimizing support vector machine[J]. High Voltage Apparatus, 2020, 56(3): 72-78.
- [20] 海涛, 范恒, 王楷杰, 等. 基于 PSO-SVM 算法的风电机组结冰故障诊断[J]. 智慧电力, 2021, 49(4): 1-6.
- HAI Tao, FAN Heng, WANG Kaijie, et al. Icing fault diagnosis of wind turbines based on PSO-SVM algorithm[J]. Smart Power, 2021, 49(4): 1-6.
- [21] 蒋伟, 王挺, 盛文, 等. 基于多分类支持向量机的相控阵雷达 T/R 组件寿命分布仿真识别[J]. 兵器装备工程学报, 2018, 39(11): 89-93.
- JIANG Wei, WANG Ting, SHENG Wen, et al. Identification of T/R module's life distribution of phased array radar with multi-class support vector machine[J]. Journal of Sichuan Ordnance, 2018, 39(11): 89-93.
- [23] 胡启国, 白熊, 杜春超. 基于 KPCA-BLSTM 的航空发动机多信息融合剩余寿命预测[J]. 航空工程进展, 2022, 13(3): 157-163.
- HU Qiguo, BAI Xiong, DU Chunchao. Remaining useful life prediction of aero-engine multi-information fusion based on KPCA-BLSTM[J]. Advances in Aeronautical Science and Engineering, 2022, 13(3): 157-163.
- [23] 刘云鹏, 郭春影, 秦明宽, 等. 基于 PCA-SVM 算法对稀土元素与稀土判别指标耦合数据集的铀矿床分类[J]. 吉林大学学报(地球科学版), 2021, 51(3): 723-733.
- LIU Yunpeng, GUO Chunying, QIN Mingkuan, et al. Classification of Uranium deposits based on PCA-SVM algorithm for coupling data set of rare earth elements and rare discrimination indexes[J]. Journal of Jilin University (Earth Science Edition), 2021, 51(3): 723-733.
- [24] 刘冬梅, 霍龙龙, 王浩然, 等. 基于 PSO-SVM 的电流放大器故障诊断研究[J]. 传感器与微系统, 2021, 40(8): 50-52.
- LIU Dongmei, HUO Longlong, WANG Haoran, et al. Study on fault diagnosis for current amplifier based on PSO-SVM[J]. Transducer and Microsystem Technologies, 2021, 40(8): 50-52.
- [25] 樊浩, 李兴文, 苏海博, 等. 基于主成分分析-支持向量机优化模型的断路器故障诊断方法研究[J]. 高压电器, 2020, 56(6): 143-151.
- FAN Hao, LI Xingwen, SU Haibo, et al. Research on circuit breaker fault diagnosis method based on principal component analysis-support vector machine optimization model[J]. High Voltage Apparatus, 2020, 56(6): 143-151.

何宁辉(1986—), 男, 博士, 高级工程师, 研究方向为电网设备状态监测技术(E-mail: 232464433@qq.com)。

杨擎柱(1997—), 男, 硕士, 研究方向为变压器状态评估(E-mail: 1216349079@qq.com)。

程养春(1974—), 男, 博士, 教授, 研究方向为高压电气设备在线监测、故障诊断与状态评估、新型传感器及其应用、高电压测量技术(E-mail: chych@ncepu.edu.cn)。