

基于 SVDD 和改进 K-Means 的变压器故障诊断模型

谢旭钦[†], 刘泉辉, 赵湘文, 张清松, 林剑雄, 张帆

(中国南方电网有限责任公司 广东广州增城供电局, 广州 广东 510000)

摘 要: 变压器状态对于智能配电房的安全稳定运行具有重要意义。为实现对变压器故障的准确诊断, 在变压器油中溶解气体分析(DGA)的基础上, 提出了一种联合使用支持向量数据描述(SVDD)和改进 K-Means 聚类的变压器故障诊断方法。首先利用 SVDD 构造闭合分类曲面实现“正常”和“故障”两类判断, 然后对“故障”类样本进行 K-Means 聚类分析, 自动将其划分为低能放电、中低温过热、高能放电、高温过热和局部放电 5 种故障类型, 同时针对 K-Means 初始聚类中心选取难题, 提出局部密度概念自动确定 K-Means 初始聚类中心, 提升聚类性能。最后利用变压器故障真实数据开展实验, 结果表明, 相较于支持向量机(SVM)和 BP 神经网络模型, 所提方法的故障诊断准确率分别提升 9.8% 和 8%。

关键词: 智能配电房; 变压器故障诊断; 油中溶解气体分析; 支持向量数据描述; 多分类器联合
中图分类号: TM41 **文献标识码:** A

Transformer Fault Diagnosis Model Based on SVDD and Improved K-Means

XIE Xuqin[†], LIU Quanhui, ZHAO Xiangwen, ZHANG Qingsong, LIN Jianxiong, ZHANG Fan

(Guangzhou Zengcheng Power Supply Bureau, China Southern Power Grid Co., Ltd., Guangzhou, Guangdong 510000, China)

Abstract: The operation status of transformer is of great significance to the stability and reliability of intelligent distribution room. In order to realize the accurate diagnosis of transformer faults, based on the analysis of dissolved gases in transformer oil, a multi-classifier joint fault diagnosis method based on the combined use of support vector data description (SVDD) and improved K-Means clustering is proposed. First, SVDD is used to construct a closed classification surface to realize “normal” and “fault” judgments. Then K-Means clustering analysis is carried out on the “fault” samples, which are automatically divided into five types: low energy discharge, medium and low temperature overheat, high energy discharge, high temperature overheat and partial discharge. At the same time, the concept of local density is proposed to automatically determine the initial clustering center of K-Means to improve the clustering performance. Finally, the transformer fault data of the intelligent distribution room is used to carry out the verification experiment. The results show that compared with the traditional support vector machine (SVM) and BP neural network model, the fault diagnosis accuracy of the proposed method is improved by 9.8% and 8%, respectively.

Key words: intelligent distribution room; transformer fault diagnosis; analysis of dissolved gas in oil; support vector data description; multi-classifier association

电力变压器是智能配电房进行能量转换和传输的核心设备,其运行状态对于智能配电房至关重要,是决定智能配电房安全、可靠和稳定运行的关键^[1]。然而变压器运行工况和运行环境较为复杂,长时间运行导致的绕组变形和绝缘老化等原因都可能会引发变压器故障,给人民群众的生命财产安全带来隐患。因此,建立变压器故障诊断模型,从而实现对变压器运行状态的实时准确监控对于智能配电房具有重要意义^[2]。

目前绝大多数变压器都采用油纸组成的绝缘结构,当变压器发生故障时,绝缘材料会产生各种气体,且不同故障产生的气体成分和浓度不同,因此通过油中溶解气体分析(dissolved gas analysis, DGA)可以实现对变压器不同故障类型的诊断^[3]。传统基于 DGA 的变压器故障诊断方法包括 IEC 三比值法、无编码比值法、罗杰斯比值法等^[4,5],该类方法都是通过 C_2H_2 、 C_2H_4 、 CH_4 、 H_2 、 C_2H_6 五种气体之间的比值大小进行变压器故障判断,具有原理简单、容易实现等优点,但存在信息提取维度单一、需要先验知识等问题,导致该类方法的故障诊断精度较低^[6]。近年来,随着我国智能配电房建设的推进以及人工智能技术的发展,利用人工智能算法对变压器进行故障诊断得到了众多学者的关注^[7,8]。文献[9]建立了小波变换联合反向传播(back propagation, BP)神经网络的变压器故障诊断模型,利用小波变换进行时频变换和故障特征提取,利用 BP 神经网络实现信息融合和故障诊断分类;文献[10]在 DGA 数据的基础上,将 XGBoost 模型引入变压器故障诊断领域,同时利用遗传算法对 XGBoost 模型参数进行寻优;文献[11]提出一种灰狼算法优化的支持向量机(support vector machine, SVM)模型,并对提出的 5 种 DGA 特征进行分类判决,最终实现故障诊断;文献[12]提出一种多源信息融合的特征提取方法,并利用卷积神经网络实现变压器不同故障的分类诊断。

上述已有研究都是采用单一分类器模型进行故障诊断,而 DGA 数据又表现出明显的波动性和非平稳特征,单一分类器模型由于无法全面描述不同故障类型之间的差异性,同时实际中“故障”类样本较难获取,导致变压器故障诊断往往面对不平衡样本集难题,上述有监督故障分类诊断方法的性能难以满足实际应用需求。针对该问题,提出一种联合使用支持向量数据描述(support vector data description, SVDD)和 K-Means 聚类的多分类器联合故障分类诊断方法,并利用实际变压器故障数据对其性能进行验证评估。

1 算法概述

图 1 给出了所提方法的流程图。在模型训练阶段,输入“正常”类样本,首先对其进行 DGA 分析并提取特征构成特征向量,然后利用特征向量对 SVDD 进行模型训练,得到最优闭合分类面。SVDD 作为一种一类分类器,只需要“正常”类样本即可构造分类面,从而大大降低训练阶段“故障”类样本的获取难度。

在测试阶段,输入“正常”或“故障”类样本,同样对其进行 DGA 分析并提取特征构成特征向量,然后利用训练阶段获得的最优 SVDD 分类面对特征向量进行分类判决,自动实现“正常”和“故障”两种状态的判决。接下来利用所提 ImK-Means 对“故障”数据进行进一步分析,首先定义局部密度概念,并选取局部密度最大的 K 个样本作为 K-Means 的初始聚类中心,然后利用 ImK-Means 对“故障”样本进行聚类分析,自动将其划分为“低能放电”“中低温过热”“高能放电”“高温过热”“局部放电”5 种故障类型。

接下来对上述内容分别进行介绍。

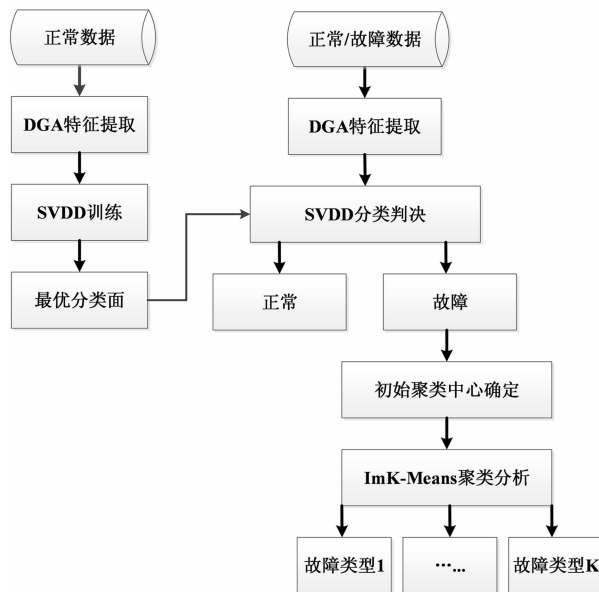


图 1 所提故障诊断方法流程图

2 SVDD

变压器绝大部分时间都处于正常工作状态,即“正常”是一种长期且稳定的状态,而“故障”是一种随机且短暂的状态,因此在实际使用过程中,通常难以获得足够多的“故障”类样本用于检测模型的训练。因此,本文首先选用 SVDD 一类分类器进

行变压器“故障”检测,只需要“正常”类样本即可完成最优分类面的训练。

对于“正常”类样本集合 $\mathbf{X} = [x_1, x_2, \dots, x_N]^T$, SVDD 分类模型可以用式(1)表示:

$$\begin{cases} \min B(c, r) = r^2 + C \sum_{i=1}^N \xi_i \\ \text{s. t. } (x_i - c)^T (x_i - c) \geq r^2 + \xi_i, \\ \xi_i \leq 0, \forall i = 1, \dots, N \end{cases} \quad (1)$$

式中: c 为 SVDD 闭合超球体分类面的球心, r 为对应的半径; ξ_i 和 C 分别为避免过拟合而设置的松弛变量和惩罚因子。由于真实环境中训练样本集通常不符合球状分布,因此 SVDD 引入核函数将训练样本向高维空间映射,从而使其在高维空间中呈现出球状分布,常用的核函数为式(2)所示径向基形式:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right) \quad (2)$$

式中: σ 为核参数。对式(1)引入拉格朗日乘子和核函数,可以将其转化为式(3)所示对偶形式:

$$\begin{cases} \max L = \sum_{i=1}^N \alpha'_i K(x_i, x_i) - \sum_{i,j} \alpha'_i \alpha'_j (x_i, x_j) \\ \text{s. t. } \sum_{i=1}^N \alpha'_i = 1, 0 \leq \alpha'_i \leq C, \forall i = 1, \dots, N \end{cases} \quad (3)$$

式中: α'_i 为支撑向量标识符,当 $\alpha'_i > 0$ 时,对应的 x_i 为超球面上的支撑向量。根据式(3)获得的所有支撑向量可以得到 SVDD 最优超球体为:

$$\begin{cases} c = \sum_{i=1}^N \alpha'_i x_i \\ r = \|K(x_i) - K(c)\|^2 \end{cases} \quad (4)$$

得到最优超球体后,对于任意测试样本 x^* ,只需要根据其到球心之间的距离是否大于半径即可实现“正常”和“故障”的判断,即 SVDD 判决函数为:

$$f(x^*) = \|K(x^*) - K(c)\|^2 - r^2 \quad (5)$$

若 $f(x^*) \leq 0$,即 x^* 处于超球体内部,表明其为“正常”类样本;若 $f(x^*) > 0$,即 x^* 处于超球体外,表明其为“故障”类样本。

3 ImK-Means

得到“故障”类样本后,需要对其故障类型进行进一步判断。K-Means 是一种经典的无监督聚类方法,根据样本之间的欧式距离大小即可实现不同子集的划分,具有原理简单、运算效率高的优势。

但是 K-Means 聚类性能受初始聚类中心的选取影响较大,而选择合适的初始聚类中心并非易事。一种好的聚类结果是同一类别内样本之间的聚集性好,而不同类别聚类中心之间的分散性高。因此合适的初始聚类中心应该选择在样本分布较为密集的区域,同时不同初始聚类中心之间的距离应尽可能大。基于此,本文提出样本邻域半径和局部密度的概念,进而建立改进的 K-Means (Improved K-Means, ImK-Means) 聚类模型。

对于 SVDD 获得的 M 个“故障”类样本 $X = \{\bar{x}_n\}_{n=1}^M$,利用所提 ImK-Means 对其进行聚类分析的步骤如下。

步骤 1:根据“故障”类别确定最优聚类个数 K 。

步骤 2:选取初始聚类中心。

① 定义式(6)所示样本空间密度概念,并计算每个样本的空间密度值 $dist(\bar{x}_n)$;

$$dist(\bar{x}_n) = \sum_{j=1}^M \frac{\|\bar{x}_n - \bar{x}_j\|_{l_2}}{\sum_{i=1, i \neq n}^N \|\bar{x}_n - \bar{x}_i\|_{l_2}} \quad (6)$$

式中: $\|\cdot\|_{l_2}$ 表示计算 l_2 范数算子。

② 定义式(7)所示邻域半径概念,并计算每个样本的邻域半径值 $r(\bar{x}_n)$;

$$r(\bar{x}_n) = \frac{1}{M} \sum_{n=1}^M \exp(-dist(\bar{x}_n)) \quad (7)$$

③ 在样本空间密度和邻域半径的基础上,进一步定义式(8)所示样本局部密度概念,并计算每个样本的局部密度值 $\rho(\bar{x}_n^*)$:

$$\rho(\bar{x}_n^*) = dist(\bar{x}_n^*) \times \log(r(\bar{x}_n^*)) \quad (8)$$

④ 选取局部密度最大的 K 个样本作为初始聚类中心。

步骤 3:将样本集合中除初始聚类中心外的样本按照欧式距离的大小划分至最近的聚类中。

步骤 4:按式(9)计算得到新的聚类中心。

$$\tilde{x}_k = \frac{1}{n_k} \sum_{n=1}^{n_k} \bar{x}_n \quad (9)$$

式中: n_k 为第 k 个子类中的样本数。

步骤 5:重复步骤 3~步骤 4,直至相邻两次聚类结果相同,得到最终聚类结果。

从样本空间密度的定义可知,空间密度描述的是样本之间的距离,空间密度越大,表明该样本与其他样本之间的距离越大;从邻域半径的定义可知,邻域半径描述的是样本之间的聚集程度,样本邻域半径越大,表明该样本处在周围样本的均值附近。局部密度同时考虑了样本空间密度和邻域半径,选择局部密度大的样本作为初始聚类中心能够保证不同聚类中心之间的距离较远,同时每个聚类中心都处在样本分布较为密集的区域。

4 实验及结果分析

4.1 实验数据及评估指标

为了验证所提方法的故障检测性能,从 IEC TC10 数据库及中国电力行业标准 DL/T 722—2014 故障案例中,分析整理出 413 条变压器故障数据及对应的 5 种故障类型:低能放电 59 条、中低温放电 86 条、高能放电 96 条、高温过热 114 条和局部放电 58 条,与 400 条变压器正常数据构成实验数据集,数据集中同时包含 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 五种 DGA 数据。为了便于分析,对正常数据和不同故障类型数据进行标号,每种数据的标号及样本数如表 1 所示。

实验首先将“正常”类样本按 6 : 4 的比例进行划分,其中 60% (240 个) 作为训练样本,完成 SVDD 模型训练,剩余 40% (160 个) 作为测试样本,同时将所有 413 个故障样本作为 1 类同正常样本一起构成测试数据集。完成 SVDD 分类后,进一步对“故障”类样本进行 ImK-Means 聚类分析。

实验用数据分析平台为联想 ThinkPad X1 Carbon 便携式计算机,采用 Intel Core i7 处理器,2.1 GHz 主频,32 GB 内存,数据分析软件平台为 MATLAB R2018b。

表 1 不同故障类型数据分布

故障类型	类别标号	样本数
正常	0	400
低能放电	1	59
中低温过热	2	86
高能放电	3	96
高温过热	4	114
局部放电	5	58

采用表 2 所示混淆矩阵对所提方法的故障检测性能进行分析,其中 TPi_P 为将实际第 i ($i=0, 1, \dots, 5$) 类数据正确判决为第 i 类标签的样本数, TPi_N 为将实际第 i 类数据错误判决为其他类标签的样本数。

表 2 故障检测及分类结果混淆矩阵

	正常	低能放电	中低温过热	高能放电	高温过热	局部放电
正常	$TP0_P$	$TP0_N$	$TP0_N$	$TP0_N$	$TP0_N$	$TP0_N$
低能放电	$TP1_N$	$TP1_P$	$TP1_N$	$TP1_N$	$TP1_N$	$TP1_N$
中低温过热	$TP2_N$	$TP2_N$	$TP2_P$	$TP2_N$	$TP2_N$	$TP2_N$
高能放电	$TP3_N$	$TP3_N$	$TP3_N$	$TP3_P$	$TP3_N$	$TP2_N$
高温过热	$TP4_N$	$TP4_N$	$TP4_N$	$TP4_N$	$TP4_P$	$TP4_N$
局部放电	$TP5_N$	$TP5_N$	$TP5_N$	$TP5_N$	$TP5_N$	$TP5_P$

在表 2 的基础上,定义式(10)所示每类样本的正确分类率(ACC)和所有样本的平均正确分类率(ACC_MEANS)2 项指标对所提方法的故障诊断性能进行定量评估。

$$\begin{cases} ACC_i = \frac{TPi_P}{TPi_P + \sum TPi_N} \\ ACC_MEANS = \frac{\sum_{i=0}^K TPi_P}{\sum_{i=0}^K TPi_P + \sum_{i=0}^K TPi_N} \end{cases} \quad (10)$$

4.2 实验结果及分析

图 2 给出了利用 SVDD 进行故障检测得到结果的二维投影,其中“—”为利用正常类样本训练得到的 SVDD 最优分类面。可以看出,利用最优分类面对测试样本进行分类判决,所有“正常”类测试样本被包围在最优分类面以内,同时所有“故障”

类测试样本均分布在最优分类面以外,表明 SVDD 能够有效地实现“正常”和“故障”的分类。

图 3 给出了利用 ImK-Means 对 SVDD 判决为“异常”类的数据进行聚类分析得到的聚类结果,可以看出,聚类后不同故障类型对应的数据都表现出了较好的类内聚集性,同时不同类别之间的样本距离较远,表现出了较好的类间分离性。

对 SVDD 和 ImK-Means 聚类结果进行综合,得到所提方法的最终故障检测结果如表 3 所示,同时为了对比,表 3 中一并给出了在相同条件下采用文献[11]所提 SVM 方法和文献[9]所提 BP 神经网络方法进行故障检测得到的结果。可以看出,所提方法对于每一类数据的分类结果均优于两种对比方法,对所有类别的平均检测结果相较于 SVM 方法提升 9.8%,相较于 BP 神经网络方法提升 8%。究其原因在于 SVM 和 BP 神经网络方法需要直接对 1 类“正常”样本和 5 类“故障”样本进行检测分类,一方面多类分类会造成分类模型较为复

杂,同时从表 1 所示每种类别样本数量可知,低能放电和局部放电 2 类“故障”类样本数量较少,在分类器学习过程中会导致该 2 类所占权重较低,导致最终检测结果中该 2 类“故障”样本的分类性能较低。而所提方法在利用 ImK-Means 聚类时,通过设置聚类个数可以有效避免样本不均衡导致的分类结果下降问题。

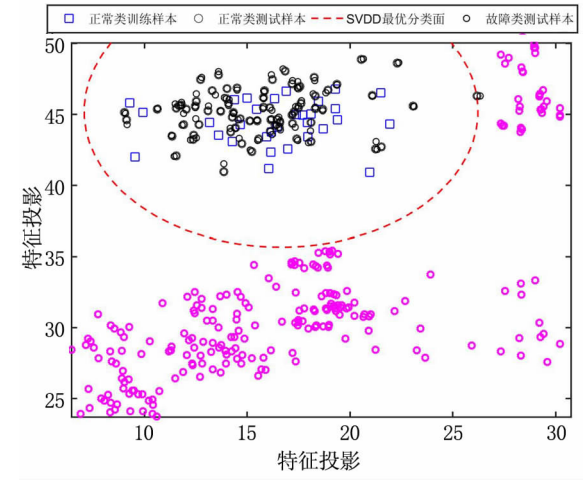


图 2 SVDD 分类结果

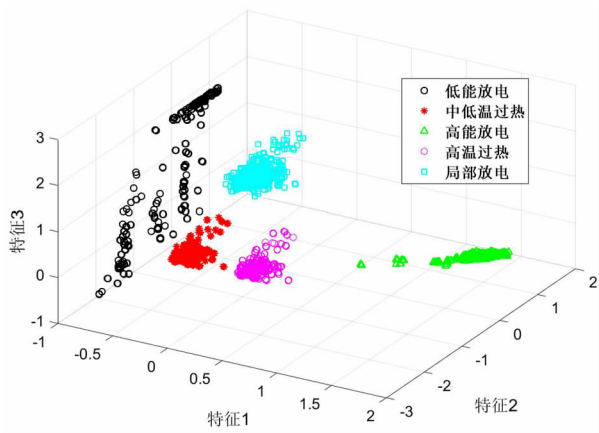


图 3 ImK-Means 聚类结果

表 3 不同方法故障诊断结果

	SVM	BP	所提方法
正常	93.1%	94.7%	100%
低能放电	82.5%	84.4%	99.8%
中低温过热	92.2%	95.5%	100%
高能放电	90.3%	93.3%	98.2%
高温过热	95.9%	97.6%	99.8%
局部放电	84.7%	87.5%	100%
ACC_MEANS	89.8%	91.6%	99.6%

5 结 论

本文提出了一种联合使用 SVDD 和 ImK-Means 聚类的变压器故障诊断模型,并利用实际数据验证了模型的有效性,得到以下结论。

(1) 针对变压器故障诊断面临的样本非平衡问题,提出利用 SVDD 一类分类器首先进行故障诊断,然后利用 ImK-Means 进行无监督聚类的方法,该方法不需要“故障”类训练样本即可实现故障诊断,大大降低了样本获取难度;

(2) 针对传统 K-Means 初始聚类中心确定难题,提出局部密度概念,进而得到 ImK-Means 聚类算法,ImK-Means 能够自动选择局部密度最大的 K 个样本作为初始聚类中心,从而提升聚类性能;

(3) 基于实际变压器故障诊断数据的实验结果表明,相对于传统 SVM 和 BP 神经网络模型,所提方法对于每种故障类型均能获得更优的分类结果。

参考文献

[1] 李婷,肖京,刘簪,等. 基于信息融合和神经网络的变压器内部故障快速诊断方法[J/OL]. 电测与仪表, 2022, DOI: <http://kns.cnki.net/kcms/detail/23.1201.TH.20211101.1014.006.html>.

[2] 张育杰,李典阳,冯健,等. 基于多模型选择性统合的变压器在线故障诊断[J]. 电力系统自动化, 2021,45(13):95—101.

[3] 吴国强,姚建锋,管敏渊,等. 基于 DBSCAN 的电力变压器故障诊断[J]. 武汉大学学报(工学版), 2021, 54(12):1172—1179.

[4] 罗伟. 基于支持向量机与化学反映优化算法的变压器故障诊断[D]. 长沙:湖南大学, 2017.

[5] 朱保军,咸日常,范慧芳,等. WRSR 与改进朴素贝叶斯融合的变压器故障诊断技术研究[J]. 电力系统保护与控制, 2021, 49(20):120—128.

[6] 王东,朱永利. 基于模糊 C 均值聚类和改进相关向量机的变压器故障诊断[J]. 电测与仪表, 2019, 56(12):14—19.

[7] 张育杰,冯健,李典阳,等. 基于油色谱数据的变压器故障征兆新优选策略[J]. 电网技术, 2021, 45(08):3324—3332.

[8] 刘可真,苟家其,骆钊,等. 基于粒子群优化-长短期记忆网络模型的变压器油中溶解气体浓度预测方法[J]. 电网技术, 2020, 44(7):2778—2884.

[9] 李腾飞,郝玉杰,袁方,等. 基于多源特征信息融合的油浸式变压器故障智能诊断模型[J]. 电工电能新技术, 2023, 42(1):48—57.

[10] 张又文,冯斌,陈页,等. 基于遗传算法优化 XG-Boost 的油浸式变压器故障诊断方法[J]. 电力自动化设备, 2021, 41(2):200—206.

[11] 李云洪,咸日常,张海强,等. 基于改进灰狼算法与最小二乘支持向量机耦合的电力变压器故障诊断方法[J/OL]. 电网技术, 2023, <https://kns.cnki.net/kcms/detail/11.2410.TM.20230222.1629.005.html>

[12] 孙叶,唐琳,苏冰,等. 基于深度学习算法的变压器故障识别研究[J]. 微型电脑应用, 2021, 37(12):74—77.