

基于DGA和稀疏化支持向量机的设备异常诊断

潘连荣, 张福泉, 何井龙, 杨加意

引用本文

潘连荣, 张福泉, 何井龙, 杨加意. 基于DGA和稀疏化支持向量机的设备异常诊断[J]. 计算机科学, 2024, 51(11): 292-297.

PAN Lianrong, ZHANG Fuquan, HE Jinglong, YANG Jiayi. Equipment Anomaly Diagnosis Based on DGA and Sparse Support Vector Machine [J]. Computer Science, 2024, 51(11): 292-297.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

传统机器学习模型的超参数优化技术评估

Evaluation of Hyperparameter Optimization Techniques for Traditional Machine Learning Models
计算机科学, 2024, 51(8): 242-255. <https://doi.org/10.11896/jsjcx.230600164>

SVM样本约简算法研究综述

Overview of Sample Reduction Algorithms for Support Vector Machine
计算机科学, 2024, 51(7): 59-70. <https://doi.org/10.11896/jsjcx.230400143>

基于机器学习的异常流量检测模型优化研究

Study on Optimization of Abnormal Traffic Detection Model Based on Machine Learning
计算机科学, 2024, 51(6A): 230700051-5. <https://doi.org/10.11896/jsjcx.230700051>

融合多源图特征的Kcore-GCN反欺诈算法研究

Study on Kcore-GCN Anti-fraud Algorithm Fusing Multi-source Graph Features
计算机科学, 2024, 51(6A): 230600040-7. <https://doi.org/10.11896/jsjcx.230600040>

基于推荐列表的缺陷文件识别

Buggy File Identification Based on Recommendation Lists
计算机科学, 2024, 51(6A): 230600088-8. <https://doi.org/10.11896/jsjcx.230600088>

基于 DGA 和稀疏化支持向量机的设备异常诊断

潘连荣¹ 张福泉² 何井龙¹ 杨加意¹

1 广西电网有限责任公司电力调度控制中心 南宁 530023

2 北京理工大学计算机学院 北京 100081

摘要 为了有效提高基于机器学习的设备异常诊断的精度和效率,提出了一种基于稀疏化支持向量机的故障诊断模型。首先,对异常诊断的原理和特征气体进行了分析,给出了故障类型与特征气体的关系;其次,从4个方面对数据进行预处理,包括清洗、归一化、平衡和划分;然后,针对最小二乘支持向量机普遍存在的稀疏性缺乏问题,提出将数据样本映射到高维的核空间,并通过谱聚类算法对映射后的数据进行核空间距离聚类,以实现最小二乘支持向量机的数据预处理,从而实现其稀疏化;最后,在小样本数据集上进行了具体实验分析。结果表明,对于9种类型的故障,与其他基于不同类型支持向量机的诊断模型相比,所提诊断模型仅需11次迭代就可以获得最大适应度值,平均诊断准确率为96.67%,准确率和效率均更高。

关键词: 异常诊断;机器学习;最小二乘支持向量机;油中溶解气体分析;稀疏化

中图分类号 TP391

Equipment Anomaly Diagnosis Based on DGA and Sparse Support Vector Machine

PAN Lianrong¹, ZHANG Fuquan², HE Jinglong¹ and YANG Jiayi¹

1 Power Dispatching Control Center, Guangxi Power Grid Co. LTD., Nanning 530023, China

2 School of Computer Science, Beijing Institute of Technology, Beijing 100081, China

Abstract In order to effectively improve the accuracy and efficiency of equipment anomaly diagnosis based on machine learning, a fault diagnosis model based on sparse support vector machine is proposed. Firstly, the principle of abnormal diagnosis and characteristic gas are analyzed, and the relationship between fault types and characteristic gas is given. Secondly, the data is preprocessed from 4 aspects, including cleaning, normalization, balance and division. Then, in order to solve the problem of sparsity of least squares support vector machine, a method is proposed to map data samples to a high-dimensional kernel space, and cluster the mapped data in kernel space distance by spectral clustering algorithm, to realize the data preprocessing of least squares support vector machine, so as to realize its sparseness. Finally, the specific experimental analysis is carried out on a small sample dataset. The results show that, for 9 types of faults, compared with other diagnosis models based on different types of support vector machines, the proposed diagnosis model only needs 11 iterations to obtain the maximum fitness value, and the average diagnosis accuracy rate is 96.67%, with higher accuracy and efficiency.

Keywords Anomaly diagnosis, Machine learning, Least square support vector machine, Analysis of dissolved gas in oil, Rarefaction

目前,随着智能化技术的提高,基于机器学习的设备故障诊断方法受到了广泛关注^[1-2]。其中,油中溶解气体分析(Dissolved Gas-in-oil Analysis, DGA)作为一种常用的电力变压器故障诊断技术,通过分析变压器油中溶解的气体的含量和组成,可以提供有关设备运行状态和潜在故障的信息^[3-4]。另一方面,支持向量机(Support Vector Machine, SVM)作为一种机器学习方法,在处理高维数据和大规模数据时具有

较好的性能。因此,许多研究人员将DGA与SVM相结合来实现设备故障诊断。

例如, Sahri等^[5]提出了一种结合DGA和SVM的电力变压器故障诊断方法。首先从DGA数据中提取特征,包括气体浓度和不同气体的比例;然后,使用SVM作为分类器,根据提取的特征对故障类型进行分类。结果表明,该方法在识别各种类型的变压器故障(如过热、局部放电和绝缘劣化)

到稿日期:2023-05-15 返修日期:2023-08-28

基金项目:国家自然科学基金面上项目(61871204);福建省科技厅引导性项目(2018H0028);广西电网公司2023年科技项目(046000KK52222021)

This work was supported by the General Project of the National Natural Science Foundation of China(61871204), Department of Science and Technology Fujian Province Guided Project(2018H0028) and Science and Technology Project of Guangxi Power Grid in 2023(046000KK52222021).

通信作者:潘连荣(276574678@qq.com)

方面具有较高的准确性。Feng 等^[6]则采用主成分分析法(Principal Component Analysis, PCA)提取 DGA 数据变量的特征,并用海鸥优化算法(Seagull Optimization Algorithm, SOA)来提高 SVM 分类器的准确度,结果表明该 SOA-SVM 分类器在识别不同故障类型时具有较高的准确率。上述文献均表明了 DGA 与 SVM 相结合的方法的有效性。

最小二乘支持向量机(Least Squares Support Vector Machine, LS-SVM)是一种基于 SVM 的变种方法,具有非线性问题处理能力、计算效率高、可解释性强、泛化性能高等优势,这意味着它可以更好地理解模型对故障的判定依据,提高模型的可解释性,并且可以更好地处理噪声和过拟合问题,提高了模型的泛化性能,从而在实际故障诊断应用中表现良好。因此,本文提出将 DGA 与 LS-SVM 相结合来实现设备故障诊断,从而提高电力变压器故障诊断的准确性和效率。

然而,LS-SVM 普遍存在稀疏性缺乏的问题,导致在处理小样本的情况时容易出现诊断准确性降低的现象。小样本在大多数现实世界的的数据源中是十分常见的,DGA 数据集也不例外。近期,Meng 等^[7]提出了在 LS-SVM 分类器前先进行局部均值分解(Local Mean Decomposition, LMD)形态滤波降噪,以便解决稀疏性差的问题。然而,该 LMD-LS-SVM 分类器对信号做相关分析需要数据分布的先验知识,以便产生尽可能精确的能量特征。针对这一难以实现的前提条件,本文提出了一种基于核空间聚类的稀硫化 LS-SVM 分类器,适用于小样本问题。实验结果表明,在小样本条件下,提出的异常诊断模型分类精度和效率均有提升,验证了其先进性。

1 异常诊断的原理和特征气体

油浸式变压器的设计是为了高效、安全地传输电能,使用矿物油作为绝缘介质和冷却剂。然而,在不正常的运行条件下或存在内部故障的情况下,油可能会降解,导致各种气体的产生。这些气体可以提供有关故障类型和故障严重程度的有价值的信息。

1.1 设备的产气原理

油浸式变压器中气体的产生主要是由故障条件下施加在绝缘材料上的热应力和电应力所导致的,包括油和固体绝缘(如纸、纸板)。热应力可能是由过载、短路或接触不良引起的局部热点造成的。绝缘材料由于局部放电、起弧或击穿,可能会产生电应力。

当这些应力超过绝缘材料的阈值时,会引起各种化学反应和物理反应,导致气体的形成^[8]。例如,石油分子的分解可能导致氢(H₂)、甲烷(CH₄)、乙烷(C₂H₆)和其他碳氢化合物气体的产生。纤维素绝缘材料的降解会产生一氧化碳(CO)、二氧化碳(CO₂)和其他可燃气体,所产生的气体的类型和数量取决于故障的性质、温度、持续时间和绝缘材料的状况。特征气体的类型如表 1 所列。

表 1 特征气体的类型

Table 1 Types of characteristic gases

种类	名称
无机气体	氢气、氮气、一氧化碳、二氧化碳
C1 烃类	甲烷
C2 烃类	乙烷、乙烯、乙炔
C3 烃类	丙烷、丙烯、丙炔

1.2 诊断模型所需的 DGA 特征气体

油浸式变压器发生故障时产生的特征气体是故障诊断的重要指标。对这些气体进行分析可以提供有关故障类型和故障严重程度的有价值的信息。油浸式变压器的 DGA 数据分析是目前最常用的一种方法,该方法通过监测变压器油中特征气体的含量来诊断不同类型的故障。一些常见的故障类型^[9]和其在 DGA 数据中的体现如表 2 所列。

表 2 故障类型与 DGA 特征气体的关系

Table 2 Relationship between fault types and DGA characteristic gas

故障类型	特征气体
绕组过热	乙炔(C ₂ H ₂)
局部放电	甲烷(CH ₄)、乙炔(C ₂ H ₂)、乙烯(C ₂ H ₄)、乙炔(C ₂ H ₂)
绝缘老化	乙烯(C ₂ H ₄)、乙炔(C ₂ H ₂)、甲烷(CH ₄)、氢气(H ₂)
电晕放电	氢气(H ₂)、甲烷(CH ₄)
电弧放电	氢气(H ₂)、甲烷(CH ₄)、乙炔(C ₂ H ₂)、乙烯(C ₂ H ₄)
潮湿/水侵入	氢气(H ₂)、甲烷(CH ₄)
调压器故障	乙烯(C ₂ H ₄)、乙炔(C ₂ H ₂)、甲烷(CH ₄)
机械损伤	乙烯(C ₂ H ₄)、乙炔(C ₂ H ₂)、甲烷(CH ₄)
油质劣化	乙炔(C ₂ H ₂)、乙烯(C ₂ H ₄)、乙炔(C ₂ H ₂)、甲烷(CH ₄)

2 基于 DGA 和稀硫化 LS-SVM 的设备异常诊断模型

2.1 DGA 数据预处理

油浸式变压器的 DGA 数据预处理是进行故障诊断和状态评估的一个重要步骤。具体步骤为:

1)数据清洗:对 DGA 数据进行清洗,包括去除异常值、缺失值和噪声。异常值可能是由传感器误差、测量噪声或数据记录错误引起的,需要进行识别和处理。缺失值可能是由传感器故障或测量间隔引起的,本文通过删除缺失值的方式进行处理。

2)数据归一化:为了将 GGA 数据转换为统一的尺度,以便后续的数据处理和建模,本文采用了常见的标准化方法,如 Z-score 标准化^[10-11],进行归一化处理。

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

其中, x 为原始数据, x' 为标准化后的数据, μ 为原始数据的平均值, σ 为原始数据的标准差。

3)数据平衡:针对油浸式变压器 DGA 数据中的类别不平衡问题,使用人工少数类过采样法(Synthetic Minority Over-sampling Technique, SMOTE)进行数据平衡处理^[12],以保证模型在训练和评估时对不同类别的样本具有相对均衡的处理能力。SMOTE 中计算合成样本 x_{new} 的方法如下:

$$x_{\text{new}} = x_i + \text{rand}(0, 1) \times (x_{\text{neighbor}} - x_i) \quad (2)$$

其中, x_i 为每个少数类样本, $\text{rand}(0, 1)$ 为一个 0~1 之间均匀分布的随机数, x_{neighbor} 为最近邻居。

4)数据划分:将预处理后的 DGA 数据集划分为训练集

和测试集,用于模型的训练和评估。本文采用随机抽样的方法,将 DGA 数据集按照 2:1 的比例划分为训练集和测试集。

2.2 基于核空间聚类的稀疏化 LS-SVM 分类器

如上所述,小样本、不平衡样本在现实世界的大多数数据源中是十分常见的,DGA 数据集也不例外。但是,由于 LS-SVM 的稀疏性较差^[13-14],因此处理此类样本时容易出现诊断准确性降低的问题。

核空间距离聚类(Kernel Space Distance Clustering, KS-DC)是一种数据预处理方法^[15],用于将数据样本映射到高维的核空间,从而实现对数据的降维和特征提取。

因此,本文采用 KSDC 方法对 LS-SVM 的数据进行预处理,以实现 LS-SVM 模型的稀疏化,从而提高模型的效率和泛化性能。此外,本文采用基于图论的谱聚类算法^[16],对映射后的数据进行聚类,实现具体步骤如下。

首先,在 LS-SVM 中,使用的核函数与传统 SVM 类似,其通过将输入样本映射到高维特征空间来实现非线性建模能力。本文选择多项式核函数来实现核空间距离聚类的预处理。

$$K(x, x') = (\gamma x^T x' + r)^d \quad (3)$$

其中, γ 为尺度参数, r 是常数项, d 是多项式的次数。

使用多项式核函数将输入数据样本映射到高维核空间,具体采用了隐式映射方法和核技巧(Kernel Trick)来避免显式地计算高维空间中的特征向量,从而降低了计算复杂度。隐式映射的计算式如下:

$$K'(x, y) = \phi(x) \cdot \phi(y) \quad (4)$$

其中, $K'(x, y)$ 表示在低维空间中计算得到的核函数值, $\phi(x)$ 和 $\phi(y)$ 分别表示输入数据样本 x 和 y 在高维核空间中的隐式映射。

使用欧氏距离度量方法在高维核空间中计算样本之间的距离,其在核空间中的计算式如式(5)所示:

$$D(\phi(x_i), \phi(x_j)) = \sqrt{\|\phi(x_i)\|^2 + \|\phi(x_j)\|^2 - 2\langle \phi(x_i), \phi(x_j) \rangle} \quad (5)$$

其中, x_i 和 x_j 分别为两个样本在输入空间中的特征, $\phi(x_i)$ 和 $\phi(x_j)$ 分别为 x_i 和 x_j 在高维核空间中的表示, $\phi(\cdot)$ 表示核映射函数, $\|\cdot\|$ 表示向量的范数, $\langle \cdot, \cdot \rangle$ 表示向量的内积。

在计算时,可以使用核函数 $K(x_i, x_j)$ 来替代内积 $\langle \phi(x_i), \phi(x_j) \rangle$,从而避免显式地计算高维核空间中的向量表示,即:

$$D(\phi(x_i), \phi(x_j)) = \sqrt{K(x_i, x_i) + K(x_j, x_j) - 2K(x_i, x_j)} \quad (6)$$

这样,我们可以通过在输入空间中计算核函数来得到高维核空间中样本之间的欧氏距离,从而在基于核空间的稀疏化 LS-SVM 分类器中使用欧氏距离度量方法进行样本聚类。

其次,对映射后的样本在核空间中应用谱聚类。谱聚类通过图分割方法来实现聚类类别划分,其分割子集的数量和聚类类别相等,谱聚类需要求解顶点相似度及分割子集的特征值。

设图 $G=(V, E)$ 共包含 $V=\{v_1, v_2, \dots, v_n\}$ 个顶点, n 个顶点形成的边关系集合为 $E=\{e_{ij}=\langle v_i, v_j \rangle \mid v_i, v_j \in V\}$ 。

其中顶点 v_i 和 v_j 之间的相似程度为 w_{ij} , $w_{ij} \geq 0$,计算式为:

$$w_{ij} = \begin{cases} \exp\left(\frac{-d(v_i, v_j)^2}{\sigma^2}\right), & e_{ij} \in E \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

其中, σ 为常量, $d(v_i, v_j)$ 表示两个顶点之间的距离。

设归一化割集将图 G 划分为 k 个类别 (A_1, A_2, \dots, A_k) ,计算时 k 个类别用集合 $h_j = \{h_{1j}, h_{2j}, \dots, h_{nj}\}$ 表示, h_{ij} 表示为:

$$h_{ij} = \begin{cases} \frac{1}{\sqrt{\text{vol}(A_j)}}, & v_i \in A_j \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$\text{vol}(A_i) = \sum_{v_i \in A_i} d_i \quad (9)$$

$$d_i = \sum_{j=1}^n w_{ij} \quad (10)$$

其中, $i=1, 2, \dots, n, j=1, 2, \dots, k$,将类别集构建成子集合 $H=\{h_1, h_2, \dots, h_k\}$ 。

对于 k 个类别,令 $i=1, 2, \dots, k$,可以将归一化割集优化为 L_s 的 k 个特征值所对应的特征,从而获得对应 L_r 的特征。谱聚类算法将样本划分到不同的类别中,形成聚类结果。

最后,根据聚类结果,筛选出每个类别的代表性样本,例如选择每个类别的聚类中心作为代表性样本。使用筛选出的代表性样本作为训练集,训练 LS-SVM 模型。由于经过核空间距离聚类的预处理,所需训练集中的样本数量较少,从而实现了 LS-SVM 分类器的稀疏化。

举例来说,假设 N 为训练样本数据集,则对于训练集中 A 类样本 M_1 和 B 类样本 M_2 来说,需要各自运用谱聚类算法计算样本簇中心作为代表性样本。

$$d_1 = \sum_{j=1}^k d(x_i, y_j) \quad (11)$$

$$d_2 = \sum_{i=1}^k d(y_j, x_i) \quad (12)$$

其中, d_1 为 M_1 中各类到 M_2 的核距离, d_2 为 M_2 中各类到 M_1 的核距离。

将筛选出的代表性样本记作新训练集 N' ,使用 N' 进行建模,从而获得基于核空间聚类的稀疏化 LS-SVM 分类器。核空间聚类的稀疏化原理如图 1 所示。

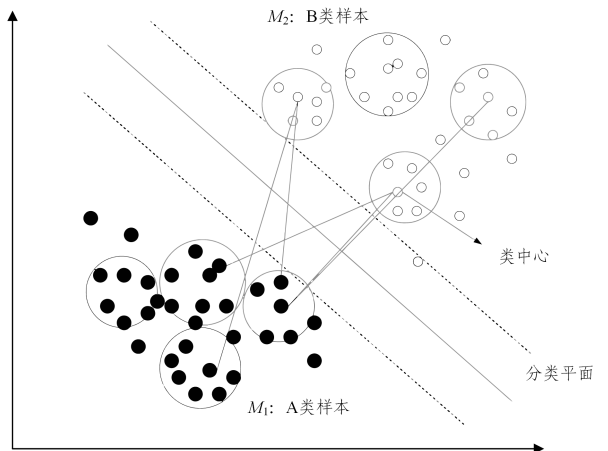


图 1 核空间聚类的稀疏化原理

Fig. 1 Sparse principle of kernel space clustering

基于核空间聚类的稀疏化 LS-SVM 的流程如图 2 所示。

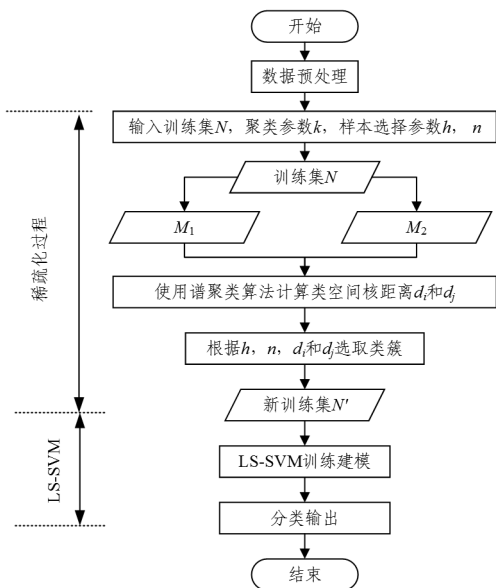


图2 所提稀疏化 LS-SVM 的流程

Fig.2 Flow of the proposed sparsing LS-SVM

2.3 设备异常诊断的流程

基于 DGA 和稀疏化 LS-SVM 的设备异常诊断流程如下:

步骤 1 对 DGA 数据进行编号分类和预处理,包括数据清洗、数据归一化、数据平衡和数据划分,训练集和测试集的比例为为 2:1;

步骤 2 从 DGA 数据中提取每种气体的浓度值作为特征,并使用最大值、最小值和平均值统计指标来描述气体浓度的分布情况;

步骤 3 使用所提稀疏化 LS-SVM 分类器对训练集中的数据进行模型训练,通过最小化目标函数来求解分类器的参数;

步骤 4 通过交叉验证方法来寻找最优的稀疏化 LS-SVM 分类器参数组合,如尺度参数 γ 和多项式的次数 d ,以提高分类器的性能;

步骤 5 使用测试集对训练好的稀疏化 LS-SVM 模型进行评估。

3 实验结果与分析

3.1 数据集与实验配置

实验所用的 DGA 数据来自国家电网公司所属市级变电站,类型为三相油浸式变压器,型号为 S9-M-800/10,共存在 9 种故障类型。运行环境参数:温度 28℃,湿度 60%,海拔高度 32m,空气中硫化氢含 2.3 mg/m³。S9-M-800/10 的额定容量 800kVA,额定电压 10kV,空载损耗 2.4kW,短路阻抗电压 4.5%。训练集和测试集的比例为 2:1,共计 270 组数据。通常将单类样本数量不超过 30 的情况称为小样本,而该数据集符合此要求。DGA 数据集的配置如表 3 所列。

实验环境软硬件参数为: Intel Core i7970 Processor 12 M Cache 3.20GHz CPU,8GB 内存,固态硬盘传输速率 200 MB/s。软件环境为 Window10 64 位, MATLAB r2017a。

表 3 DGA 数据集的配置

Table 3 Configuration of DGA dataset

故障类型	编号	训练集	测试集	总数据集
绕组过热	1	20	10	30
局部放电	2	20	10	30
绝缘老化	3	20	10	30
电晕放电	4	20	10	30
电弧放电	5	20	10	30
潮湿/水侵入	6	20	10	30
调压器故障	7	20	10	30
机械损伤	8	20	10	30
油质劣化	9	20	10	30

交叉验证使用的 Matlab 函数为 crossval。稀疏化 LS-SVM 分类器中尺度参数 γ 和多项式的次数 d 的寻优示意图如图 3 所示。

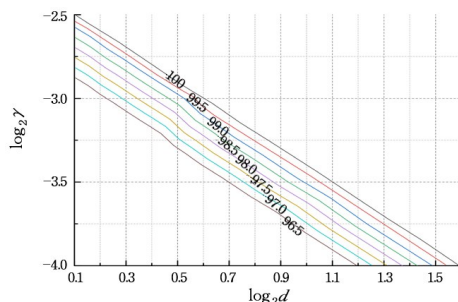


图 3 LS-SVM 分类器参数的寻优示意图

Fig.3 Schematic diagram of parameter optimization of LS-SVM classifier

可以看出,尺度参数 γ 和多项式的次数 d 的最佳取值分别为 0.8409 和 3(1.58496),后续实验中 LS-SVM 分类器参数均采用这个最佳配置。

3.2 诊断性能分析

首先,使用基于核空间聚类的稀疏化 LS-SVM 分类器对预处理后的 DGA 数据进行故障诊断。在 90 个样本的测试集中,验证了稀疏化 LS-SVM 分类器的性能,结果如图 4 所示。

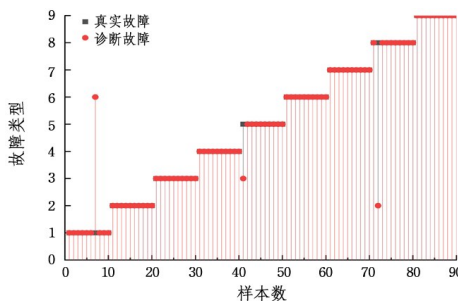


图 4 所提模型的故障诊断结果

Fig.4 Fault diagnosis results of the proposed model

可以看出,所提稀疏化 LS-SVM 分类器在 90 个样本的分类判断中仅出现了 3 次错误,分别为样本 7、样本 41 和样本 72,也就是说,稀疏化 LS-SVM 分类器的诊断准确率为 96.67%。

3.3 性能对比

为了充分验证所提模型的故障诊断能力,将 SOA-SVM 和 LMD-LS-SVM 作为对照组,使用 9 组共 180 个样本对 3 个

诊断模型进行训练,利用混淆矩阵来显示 3 个诊断模型的诊断结果,如图 5 所示。

表 4 3 个模型的诊断正确率对比

Table 4 Comparison of diagnostic accuracy of three models

故障类型	诊断正确率 (%)		
	稀疏化 LS-SVM	SOA-SVM	LMD-LS-SVM
绕组过热	90.00	80.00	80.00
局部放电	100.00	90.00	90.00
绝缘老化	100.00	90.00	90.00
电晕放电	100.00	80.00	80.00
电弧放电	90.00	80.00	100.00
潮湿/水侵入	100.00	80.00	100.00
调压器故障	100.00	80.00	90.00
机械损伤	90.00	80.00	100.00
油质劣化	100.00	80.00	90.00
平均正确率	96.67	82.22	91.11

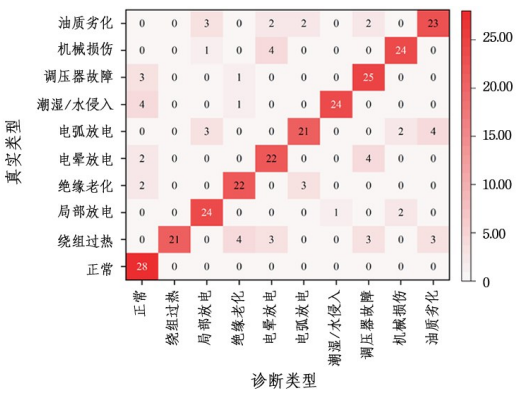
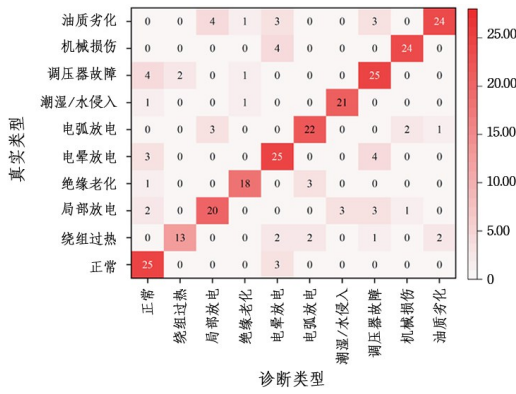
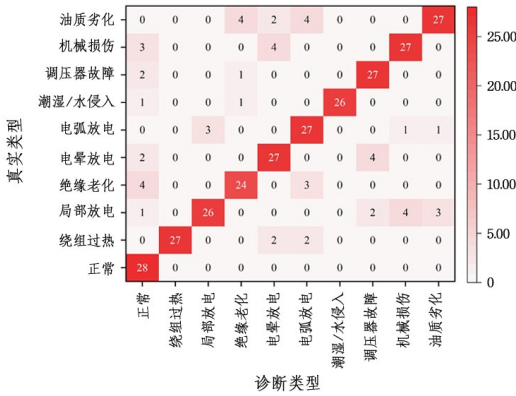


图 5 3 个模型的诊断结果

Fig. 5 Diagnostic results of three models

可以看出,相比 SOA-SVM 和 LMD-LS-SVM,稀疏化 LS-SVM 分类器的大部分结果均集中分布在混淆矩阵的对角线上,说明诊断的准确率更高。以绕组过热为例,SOA-SVM 分类器仅正确诊断了 13 例,LMD-LS-SVM 分类器正确诊断了 21 例,而稀疏化 LS-SVM 分类器正确诊断了 27 例,明显高于另外 2 个分类器,说明 SOA-SVM 分类器在绕组过热类型故障诊断中的适用性较差。综合对比诊断结果可以看出,稀疏化 LS-SVM 分类器适用于所有 9 种故障类型,而 SOA-SVM 分类器和 LMD-LS-SVM 分类器则对绕组过热类型存在适用性较差的问题。3 个模型的诊断正确率对比结果如表 4 所列。

对于局部放电、绝缘老化、电晕放电、潮湿/水侵入、调压器故障和油质劣化这 6 种故障类型,稀疏化 LS-SVM 的判断正确率均达到 100%,其他故障类型也达到了 90%,平均正确率为 96.67%。SOA-SVM 和 LMD-LS-SVM 的平均正确率分别为 82.22% 和 91.11%,明显低于稀疏化 LS-SVM。

此外,相比其他两个模型,稀疏化 LS-SVM 的适应度曲线收敛速度更快,仅需 11 次迭代就可以获得最大的适应度值 96.67,3 个模型的适应度曲线收敛结果对比如表 5 所列。

表 5 适应度曲线收敛结果对比

Table 5 Comparison of convergence results of fitness curve

模型	最优适应度值	收敛代数
SOA-SVM	82.22	22
LMD-LS-SVM	91.11	17
稀疏化 LS-SVM	96.67	11

综上所述,对于 9 种类型的 DGA 数据,相比 SOA-SVM 和 LMD-LS-SVM,稀疏化 LS-SVM 分类器的诊断准确率更高,诊断时间更短,验证了其先进性。

结束语 本文提出了一种基于 DGA 和稀疏化 LS-SVM 的设备故障诊断模型。对 DGA 数据进行编号分类和预处理,且训练集和测试集的比例为 2:1。得出如下结论:

1) 通过交叉验证方法获得最优的稀疏化 LS-SVM 分类器的参数组合取值(尺度参数和多项式的次数)分别为 0.840 9 和 3(1.58496);

2) 稀疏化 LS-SVM 分类器适用于所有 9 种故障类型,而 SOA-SVM 分类器和 LMD-LS-SVM 分类器则对绕组过热类型存在适用性较差的问题;

3) 相比 SOA-SVM 和 LMD-LS-SVM,稀疏化 LS-SVM 分类器能够在小样本数据集上获得较好的准确率和执行效率,准确率达到 96.67%。

后续将对所提模型在非平衡样本集中的性能或者样本泛化性开展进一步研究。

参考文献

[1] SONG L, WANG H, CHEN P. Step-by-step fuzzy diagnosis method for equipment based on symptom extraction and trivalent logic fuzzy diagnosis theory [J]. IEEE Transactions on Fuzzy Systems, 2018, 26(6): 3467-3478.

[2] HSU J Y, WANG Y F, LIN K C, et al. Wind turbine fault diag-

- nosis and predictive maintenance through statistical process control and machine learning [J]. *IEEE Access*, 2020, 8: 23427-23439.
- [3] GAO J H, ZHANG Y. Fault diagnosis of ship regional distribution power system based on FWA-PSO-MSVM [J]. *Computer Science*, 2022, 49(S2): 956-960.
- [4] LI X S, ZHANG Z Y. Fault diagnosis algorithm of transformer windings based on bayesian classification [J]. *Journal of University of Jinan(Science and Technology)*, 2021, 35(4): 412-416.
- [5] SAHRI Z B, MALAYSIA U T. Support vector machine-based fault diagnosis of power transformer using k nearest-neighbor imputed DGA dataset[J]. *Journal of Computer and Communications*, 2014, 2(9): 22-31.
- [6] FENG Z L, XIAO H Q, REN W F, et al. Transformer fault diagnosis based on principal component analysis and seagull optimization support vector machine [J]. *China Measurement & Testing Technology*, 2023, 49(2): 99-105.
- [7] MENG L, XU T L, MA J Y, et al. There Search of Ls-SVM Based on LMD Morphology filter [J]. *Journal of Harbin University of Science and Technology*, 2022, 27(1): 92-99.
- [8] BUSTAMANTE S, MANANA M, ARROYO A, et al. Dissolved gas analysis equipment for online monitoring of transformer oil: A review[J]. *Sensors*, 2019, 19(19): 4057-4068.
- [9] QIAN G, HU J, WANG S, et al. Adsorption and Sensing Properties of Dissolved Gas in Oil on Cr-Doped InN Monolayer: A Density Functional Theory Study [J]. *Chemosensors*, 2022, 10(1): 30-42.
- [10] ALTMAN E I, IWANICZ-DROZDOWSKA M, LAITINEN E K, et al. Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model[J]. *Journal of International Financial Management & Accounting*, 2017, 28(2): 131-171.
- [11] DEVORE G R. Computing the Z score and centiles for cross-sectional analysis; a practical approach [J]. *Journal of Ultrasound in Medicine*, 2017, 36(3): 459-473.
- [12] SOLTANZADEH P, HASHEMZADEH M. RCSMOTE: Range-Controlled synthetic minority over-sampling technique for handling the class imbalance problem [J]. *Information Sciences*, 2021, 542: 92-111.
- [13] DENG W, YAO R, ZHAO H, et al. A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm [J]. *Soft Computing*, 2019, 23: 2445-2462.
- [14] XIAO C, XIA W, JIANG J. Stock price forecast based on combined model of ARI-MA-LS-SVM [J]. *Neural Computing and Applications*, 2020, 32: 5379-5388.
- [15] CHENG Y, ZHU H, WU J, et al. Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks [J]. *IEEE Transactions on Industrial Informatics*, 2018, 15(2): 987-997.
- [16] MENG Q J, YAO H C. Information assets behavior anomaly detection based on spectral clustering algorithm [J]. *Journal of Nanjing University of Science and Technology*, 2021, 45(2): 205-213.



PAN Lianrong, born in 1985. His main research interests include deep learning, data processing and device anomaly analysis.

(责任编辑:何杨)