

DOI: 10.12158/j.2096-3203.2024.01.021

基于油中溶解气体特征量筛选的变压器故障诊断方法

廖才波, 杨金鑫, 胡雄, 邱志斌, 刘小天, 朱文清

(南昌大学信息工程学院, 江西 南昌 330031)

摘要:油中溶解气体分析对变压器故障预警及诊断具有重要意义。针对油中溶解气体特征量种类众多、故障关联特征分析不足等问题,文中以油浸式变压器为研究对象,提出了基于油中溶解气体特征量筛选的变压器故障诊断方法。首先,对油中溶解气体的原始特征量进行特征衍生,通过随机森林(random forest, RF)计算特征量对故障诊断的重要度,筛选得到最佳特征组合。其次,采用树结构概率密度估计(tree-structured parzen estimator, TPE)实现RF模型的参数寻优,并形成TPE-RF诊断模型。同时,结合多种评价指标,证明所提方法能够对变压器作出准确的故障诊断。最后,提出TreeSAHP模型分析特征量对各故障的重要度,优选出各故障关联的主要特征量,并根据变压器运行案例,探讨了该方法在电力行业现场应用中的适用性,验证了该方法的有效性。

关键词:油中溶解气体; 变压器; 故障诊断; 树结构概率密度估计(TPE); 随机森林(RF); 特征筛选; TreeSHAP模型
中图分类号: TM741 **文献标志码:** A **文章编号:** 2096-3203(2024)01-0192-09

0 引言

油浸式变压器是电力系统中关键的设备,承担着电能输送、电压转换等重要任务。但变压器发生过热或放电故障时,可能会停运,影响电网的正常运行^[1-3]。因此,提前预知变压器内部存在的故障,对保证电网的稳定运行具有重要意义^[4-6]。

针对变压器故障诊断,国内外相关学者、电力运维单位开展了大量应用研究工作。依据试验方式的不同,将变压器故障诊断技术分为2类。(1)在线监测:局部放电检测^[7]、油中溶解气体分析(dissolved gas analysis, DGA)^[8]、铁芯接地测量等;(2)离线试验:频率响应分析^[9]、绝缘试验^[10]、DGA、纸板老化分析^[11]等。受限于现场运行参数及诊断方法的实时性,以DGA为基础的变压器故障诊断方法成为电力行业的主流方法。在传统方法中,IEC三比值法^[12]、Rogers法^[13]等通过气体间比值诊断变压器的故障类型,原理简单、运用广泛,但存在过分依赖运维人员的现场经验和诊断精度不高等问题。

近年来,基于油中溶解气体的人工智能算法凭借其自组织性和自适应性等特点,广泛运用于数据驱动的变压器故障诊断,如信息融合^[14]、支持向量机(support vector machine, SVM)^[15]、随机森林(random forest, RF)^[16]、极端梯度提升(extreme gradient boosting, XGBoost)^[17]等。在油中溶解气体特征处

理方面,文献[17]将9个无编码比值作为诊断模型的输入特征量,建立基于XGBoost的变压器故障诊断模型。文献[18]结合SVM优选油中溶解气体特征量,提高了变压器故障诊断的准确率。文献[19]采用深度信念网络提取油中溶解气体的特征量,并对其均值聚类,实现了变压器的故障分类。上述方法均能有效提升变压器故障诊断精度,但仍有一定的完善空间,主要包括:(1)所采用的油中溶解气体特征信息的全面性和多样性方面可进一步考虑;(2)不同故障类型和特征量的重要度分析不足,无法对诊断结果的影响因素进行事后分析。

针对上述问题,文中提出基于油中溶解气体特征量筛选的变压器故障诊断方法。该方法对油中溶解气体数据进行特征衍生处理;采用RF模型计算特征量对故障诊断的重要度,优选出最佳特征组合;引入树结构概率密度估计(tree-structured parzen estimator, TPE)优化RF模型的参数空间,实现对变压器故障的准确诊断;最后,提出针对树集成模型的SHAP加法解释器计算特征量对诊断结果的边际贡献,增强了诊断模型的实用性。

1 模型建立

1.1 RF模型

RF是继Bagging算法后提出的以分类回归树为弱评估器的树状集成算法^[20]。该算法结合有放回抽样技术(Bootstrap)随机建立多棵分类回归树,并在其内部节点以基尼系数(gini)或信息熵(entropy)为指标选取特征值进行分支,最后按照多数投票法得出最终结果。RF模型结构如图1所示。

收稿日期:2023-08-22;修回日期:2023-10-13

基金项目:国家自然科学基金资助项目(62163025);江西省自然科学基金资助项目(20212ACB212007)

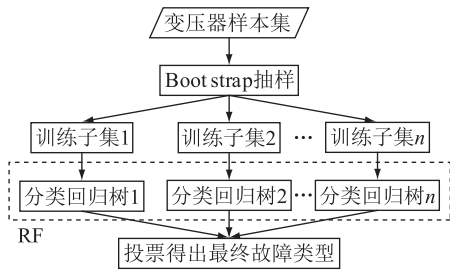


图1 RF模型结构

Fig.1 Structure of RF model

假设变压器的样本数量为 n , 初始样本集 $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 其中 $y = \{y_1, y_2, \dots, y_n\}$ 为故障目标编码, 第 i 个样本为 $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, m 为特征个数。构建 RF 分类模型的步骤如下:

(1) Bootstrap 法重采样。采取有放回抽样方法对变压器样本随机抽取 t 次, 形成 t 个训练子集, 用于构建 t 棵分类回归树。

(2) 随机特征量抽取。由于变压器训练子集存在多种故障, 为减少训练样本数据集的差异性, 选取某个特征值对树模型的根节点进行分支, 生成 2 个支路节点。变压器训练子集中的数据差异可以采用基尼系数来表示, 计算公式为:

$$G_{\text{gini}}(P) = \sum_{k=1}^K P_k(1 - P_k) \quad (1)$$

式中: K 为变压器故障类别总数; P_k 为训练子集中故障类别 k 的概率。基尼系数越小, 说明变压器故障被误判的概率也越小。

假设某训练子集为 D , 对应样本数量为 $|D|$, 其中第 k 类故障数量为 $|D_k|$, 则该训练子集的初始基尼系数为:

$$G_{\text{gini}}(D) = 1 - \sum_{k=1}^K \left(\frac{|D_k|}{|D|} \right)^2 \quad (2)$$

当特征量 Q 的某个值 α 把 D 分成 2 个样本子集 S_1 和 S_2 时, D 在分支后对应的基尼系数为:

$$G_{\text{gini}}(D, Q, \alpha) = \frac{|S_1|}{|D|} G_{\text{gini}}(S_1) + \frac{|S_2|}{|D|} G_{\text{gini}}(S_2) \quad (3)$$

式中: $|S_1|$ 、 $|S_2|$ 分别为 S_1 、 S_2 的样本数量。

(3) 最优分支节点。以最大基尼系数为原则实现树节点分支, 并重新评估分支后样本数据集的差异性。重复步骤(1)–(3), 直到基尼系数为 0。

(4) 决策分类。RF 模型按照多数投票法对最终的预测结果进行决策, 可表示为:

$$H = \arg \max_Y \sum_{j=1}^l I(h_j(x_j) = Y) \quad (4)$$

式中: H 为 RF 模型的诊断结果; h_j 为第 j 个弱评估器的诊断结果; x_j 为构建第 j 个弱评估器的样本; Y 为目标预测结果; I 为示性函数。

1.2 TPE 算法

贝叶斯优化是一种实现模型超参数自动寻优的算法^[21]。为快速找到最优解, 贝叶斯通过目标函数 f 的前序采样结果建立概率代理模型。

$$p(s | M) = \frac{p(M | s)p(s)}{p(M)} \quad (5)$$

$$M = \{(s_1, f(s_1)), (s_2, f(s_2)), \dots, (s_i, f(s_i))\} \quad (6)$$

式中: $p(s | M)$ 为参数集合 s 的后验概率分布; $p(s)$ 、 $p(M | s)$ 分别为 s 的先验概率分布和似然分布; $p(M)$ 为平均似然值。

TPE 算法通过建立目标函数的概率代理模型, 用非参数密度替换初始参数分布, 完成代理模型的更新^[22]。TPE 算法对 RF 模型的参数空间进行采样, 并在最优解区域实现快速收敛。假设某参数采样点为 w , TPE 算法将式(5)中 $p(M | s)$ 替代为:

$$p(w | z) = \begin{cases} l(w) & z < z' \\ g(w) & z \geq z' \end{cases} \quad (7)$$

式中: z 为 w 的采样阈值; z' 为最优采样阈值; $l(w)$ 、 $g(w)$ 为目标函数中的 2 种密度估计。

为了防止参数寻优过程中陷入局部最优困境, TPE 采用预期改进 (expected improvement, EI) 作为采样函数, 定义为:

$$E_z(w) = \int_{-\infty}^{z'} (z' - z) \frac{p(w | z)p(z)}{p(w)} dz \quad (8)$$

令 $\gamma = p(z < z')$, 构造 $p(w)$ 为:

$$p(w) = \int_{\mathbf{R}} p(w | z)p(z) dz = \gamma l(w) - (1 - \gamma)g(w) \quad (9)$$

$$\int_{-\infty}^{z'} (z' - z)p(w | z)p(z) dz =$$

$$\gamma z' l(w) - l(w) \int_{-\infty}^{z'} p(z) dz \quad (10)$$

将式(10)代入式(8)可得:

$$E_z(w) = \frac{\gamma z' l(w) - l(w) \int_{-\infty}^{z'} p(z) dz}{\gamma l(w) - (1 - \gamma)g(w)} \quad (11)$$

当 $l(w)$ 取最大值且 $g(w)$ 取最小值时, EI 值最大, 此时返回最优参数采样点 w 。

1.3 TreeSHAP 模型

针对树集成模型, Lundberg^[23]等人提出了基于 SHAP 的改进方法 TreeSHAP, 其运算速度快, 可以量化特征量对结果的边际贡献。文中采用

TreeSHAP 模型对 RF 模型作出一定的解释性分析。

假定特征量 λ 的数值为 b , 权重为 β , 则特征量 λ 对模型最终预测结果的贡献度为:

$$\varphi = \beta b - E(\beta b) \quad (12)$$

式中: βb 为 λ 的 SHAP 值; $E(\beta b)$ 为 λ 的期望值。

2 基于 TPE-RF 的故障诊断

2.1 故障样本和特征量筛选

变压器发生过热或放电时, 会产生不同组分或含量的气体, 考虑到气体间含量分散性较大, 常用的 IEC 三比值法诊断精度仅为 80% 左右^[24]。文献[25]总结了不同工况下的变压器油中溶解气体产生规律, 提出无编码比值用于故障诊断。文献[26]验证了无编码比值作为故障诊断模型的输入, 可以有效提升变压器故障诊断的准确率。因此, 文中选择 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 作为原始特征量, 结合无编码比值, 得出 32 种特征量, 如表 1 所示。

表 1 油中溶解气体特征量
Table 1 Features of dissolved gas in oil

| 编号 | 特征量 | 编号 | 特征量 |
|----|-----------------|----|----------------------|
| 1 | H_2 | 17 | C_2H_4/C_2H_6 |
| 2 | CH_4 | 18 | CH_4/TH |
| 3 | C_2H_6 | 19 | C_2H_4/TH |
| 4 | C_2H_4 | 20 | C_2H_6/TH |
| 5 | C_2H_2 | 21 | C_2H_2/TH |
| 6 | TH | 22 | H_2/TG |
| 7 | H_2/CH_4 | 23 | $(CH_4+C_2H_4)/TH$ |
| 8 | H_2/C_2H_2 | 24 | $(CH_4+C_2H_2)/TH$ |
| 9 | H_2/C_2H_4 | 25 | $(CH_4+C_2H_6)/TH$ |
| 10 | H_2/C_2H_6 | 26 | $(C_2H_4+C_2H_2)/TH$ |
| 11 | H_2/TH | 27 | $(C_2H_6+C_2H_2)/TH$ |
| 12 | CH_4/C_2H_2 | 28 | $(C_2H_4+C_2H_6)/TH$ |
| 13 | CH_4/C_2H_4 | 29 | $(H_2+CH_4)/TG$ |
| 14 | CH_4/C_2H_6 | 30 | $(H_2+C_2H_4)/TG$ |
| 15 | C_2H_2/C_2H_4 | 31 | $(H_2+C_2H_6)/TG$ |
| 16 | C_2H_2/C_2H_6 | 32 | $(H_2+C_2H_2)/TG$ |

注: TH 为烃类气体体积分数之和; TG 为 5 种特征气体体积分数之和。

依据 DL/T 722—2014《油中溶解气体分析和判断导则》^[27], 变压器故障诊断结果可分为中低温过热(T1)、高温过热(T2)、局部放电(D1)、低能放电(D2)、高能放电(D3)和正常(N)。结合现场案例及文献[28], 选取 1 057 条已知故障类型的在线或离线油中溶解气体样本数据, 变压器各故障样本分布如表 2 所示。

在变压器故障诊断过程中, 面临数量众多的油中溶解气体特征量, 分类器应以故障诊断的准确率

表 2 故障样本分布

Table 2 Distribution of fault samples

| 类别 | 数量 | 类别 | 数量 |
|----|-----|----|-----|
| N | 228 | D1 | 82 |
| T1 | 180 | D2 | 125 |
| T2 | 245 | D3 | 197 |

为目标筛选有效特征量。因此, 文中对训练集进行 5 折交叉验证, 以 RF 模型的最大平均测试准确率 (mean accuracy, MA) 为目标值, 调用 scikit-learn 库中 RF 内部的 feature_importances_ 模块计算表 1 中 32 个特征量的重要度。为筛选变压器有效特征量, 根据特征量的影响度降序逐次抽取, 不同特征数量对应故障诊断效果如图 2 所示。

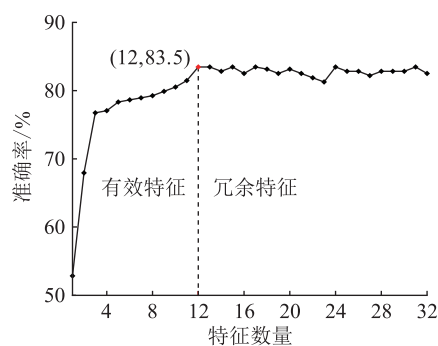


图 2 不同特征数量的分类效果

Fig.2 The effect of classification for different number of features

可以看出, 当油中溶解气体特征数量为 12 时, 基于 RF 的变压器故障诊断模型准确率为 83.5%, 此时分类效果已达到最佳状态, 当继续增加特征量时, 故障诊断结果在一定程度上已处于收敛状态。为提高变压器故障诊断的准确率和运行效率, 文中共筛选出 12 个有效特征量, 用于表征油中溶解气体特征信息, 具体见表 3, 并对后续变压器故障诊断展开分析。

表 3 最佳特征组合

Table 3 Optimal combination of features

| 序号 | 特征量 | 序号 | 特征量 |
|----|-------------------|----|--------------------|
| 1 | TH | 7 | C_2H_2/C_2H_6 |
| 2 | C_2H_2/TH | 8 | H_2 |
| 3 | C_2H_4/C_2H_6 | 9 | C_2H_4/TH |
| 4 | C_2H_2/C_2H_4 | 10 | H_2/CH_4 |
| 5 | C_2H_4 | 11 | CH_4/C_2H_2 |
| 6 | $(H_2+C_2H_2)/TG$ | 12 | $(CH_4+C_2H_4)/TH$ |

2.2 TPE 优化 RF 模型

传统的网格搜索依赖经验和参数遍历, 而 TPE 算法搜索的参数空间相对稀疏, 因此寻找最优参数组合更高效。由于 TPE 算法只支持寻找目标函数

的最小值,因此选择错误率 ξ 作为损失函数^[29]。

$$\xi = \frac{1}{n} \sum_{i=1}^n R(y_i \neq y'_i) \quad (14)$$

式中: y_i 为第 i 个样本的实际值; y'_i 为第 i 个样本的模型预测值; R 为计数函数。

TPE 优化 RF 模型步骤如下,流程如图 3 所示。

(1) 构建 RF 模型参数空间。

(2) 在训练集下,引入 TPE 对 RF 模型参数空间进行概率密度估计,利用式(11)计算其 EI 值。

(3) 对多组参数采样结果,判断 EI 值是否为当前最大值,若满足迭代要求,则输出对应的参数组合,否则重复步骤(1)–(3)。

(4) 在最佳组合参数下形成基于 TPE-RF 的变压器故障诊断模型,并输出对应的故障类型。

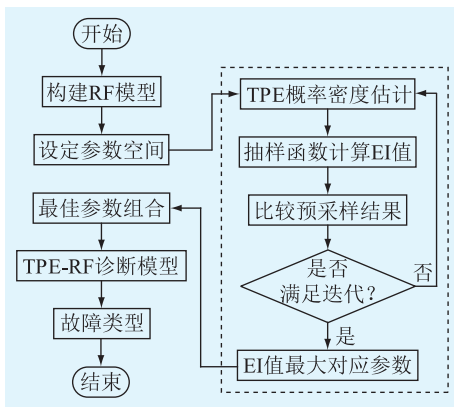


图 3 TPE 优化 RF 模型

Fig.3 TPE optimized RF model

选取 RF 模型中较为重要的 6 个参数变量,其中连续变量设定参数区间步长为 1,如表 4 所示。

表 4 RF 参数空间

Table 4 Spaces of RF parameters

| 参数名称 | 参数意义 | 参数空间 |
|------------------|---------|--------------------|
| n_estimators | 树模型数量 | [20,200] |
| criterion | 不纯度指标 | ['gini','entropy'] |
| max_depth | 最大树深度 | [5,50] |
| min_sample_split | 最小分支数量 | [2,12] |
| max_sample_leaf | 最大叶节点数量 | [1,11] |
| max_features | 最大分支特征数 | [3,6,12] |

基于硬件条件为 Intel(R) Core(TM) i7、1.8 GHz、内存 16 GB 的计算机上所搭建的 Python 3.7 测试环境,采用 TPE 实现 RF 模型参数寻优。TPE 迭代次数与准确率的关系如图 4 所示。

当迭代次数为 66 次时,准确率为 88.3%,且连续 50 次没有出现下降,TPE 参数寻优结束,形成 TPE-RF 诊断模型。TPE-RF 模型对应最佳参数组合如表 5 所示。

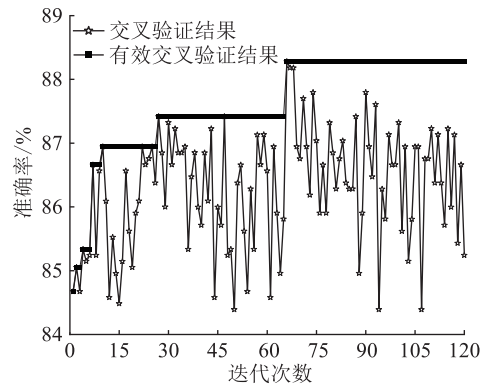


图 4 TPE 迭代过程

Fig.4 Iteration process of TPE

表 5 最佳参数组合

Table 5 Optimal combination of parameters

| 参数名称 | 参数值 | 参数名称 | 参数值 |
|--------------|--------|------------------|-----|
| n_estimators | 63 | min_sample_split | 6 |
| criterion | 'gini' | min_sample_leaf | 1 |
| max_depth | 32 | max_features | 3 |

2.3 故障诊断结果分析

结合油中溶解气体最佳特征组合,TPE-RF 诊断模型在 317 组测试集样本下准确率为 88.5%,而在 120 组验证集样本下准确率为 90%。图 5 为验证集下各故障的诊断效果。可以看出,TPE-RF 诊断模型对正常状态和高温过热的诊断精度均为 100%。同时,各故障类型沿混淆矩阵的对角线分布,说明该诊断算法具有较高的准确性。

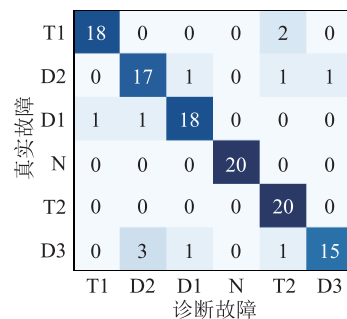


图 5 验证集下的分类结果

Fig.5 Classification results under validation set

文中结合接收者操作特征(receiver operating characteristic, ROC)曲线对逻辑回归(logistic regression, LR)、高斯朴素贝叶斯(gaussian naive bayes, GNB)、线性判别分析(linear discriminant analysis, LDA)、RF 和 TPE-RF 模型进行对比分析。ROC 曲线纳入分类器对每个故障的概率实现模型性能的评估。在二分类中,ROC 曲线纵坐标定义为正阳率,表示正类中正确判定的比例,横坐标定义为假阳率,表示负类中误判的比例。ROC 曲线描述的是

正阳率和假阳率从概率最小阈值到最大的变化情况,同时 ROC 曲线下方面积 (area under curve, AUC) 和模型诊断性能成正比^[30]。在变压器故障诊断中,对所有的故障类型进行“独热编码”^[31]处理,并和每种故障概率对应,从而将多故障转化成二分类问题处理。每种诊断模型的 ROC 曲线如图 6 所示,其中 S 为 AUC 对应的数值。

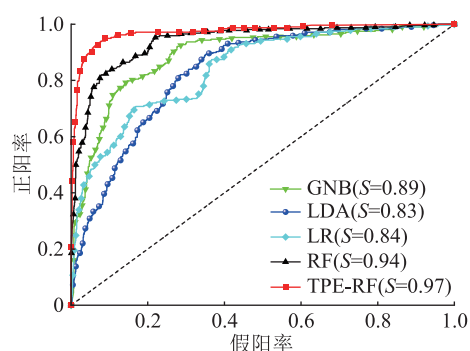


图 6 不同诊断模型对应的 ROC 曲线

Fig.6 The ROC curves of different diagnostic models

可以看出,线性分类模型 LDA、LR 的 S 值要低于 GNB 和 RF 模型,说明变压器故障诊断表现出一定的非线性。另外,TPE-RF 模型的 S 值大于其他 4 种模型。采用机器学习中精确度、召回率、 F_1 值和准确率这 4 种评价指标对 TPE-RF 模型的诊断性能进行进一步分析,如表 6 所示。

表 6 不同评价指标下的诊断效果

Table 6 Diagnostic effects under different indicators of evaluation

| 诊断模型 | 精确度 | 召回率 | F_1 值 | 准确率 |
|--------|------|------|---------|------|
| LDA | 0.69 | 0.65 | 0.66 | 0.70 |
| LR | 0.74 | 0.72 | 0.73 | 0.76 |
| GNB | 0.81 | 0.81 | 0.81 | 0.82 |
| RF | 0.85 | 0.83 | 0.84 | 0.85 |
| TPE-RF | 0.87 | 0.86 | 0.86 | 0.90 |

表 6 中 4 种评价指标越接近 1 则模型效果越好。TPE-RF 诊断模型在 4 种评价指标下的诊断效果均在 85% 以上,其中准确率为 90%。不同评价指标的 TPE-RF 模型的诊断效果均优于其他 4 种模型,表明 TPE-RF 模型具有较好的故障识别能力。

3 故障关联特征量优选

在 2.1 节中根据特征量对故障诊断的影响度筛选出 12 个特征量,但针对不同故障类型的差异化特征量研究相对较少。对此,引入 TreeSHAP 模型,通过计算特征量的 SHAP 值,得到不同特征量对 6 种故障的重要度关系,如图 7 所示。

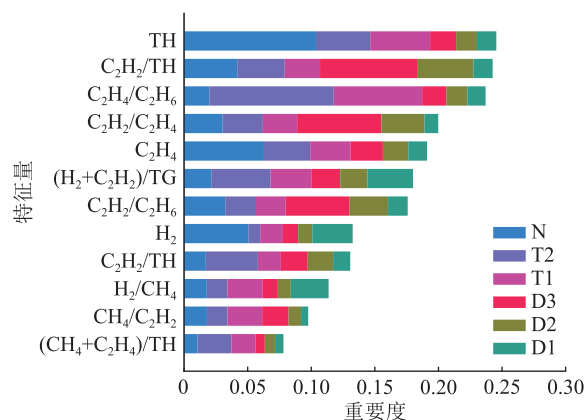


图 7 特征量重要度分析

Fig.7 Analysis of characteristic importance

结果表明, C_2H_4/C_2H_6 为影响过热故障最重要的特征量,而 C_2H_2/TH 为影响放电故障最重要的特征量。在评估变压器是否处于正常状态时,TH、 H_2 、 C_2H_4 和 C_2H_2/TH 对其有较大的影响。结合图 7,可以针对每种故障类型关联的主要特征量进一步优选,结果如表 7 所示。

表 7 故障关联特征量的优选

Table 7 Optimization of fault correlation features

| 故障类型 | 特征量 | 故障类型 | 特征量 |
|-------|-------------------|------|-------------------|
| 正常 | TH | 局部放电 | $(H_2+C_2H_2)/TG$ |
| | C_2H_4 | | H_2 |
| | H_2 | | H_2/CH_4 |
| | C_2H_2/TH | | C_2H_2/C_2H_6 |
| 中低温过热 | C_2H_4/C_2H_6 | 低能放电 | C_2H_2/TH |
| | TH | | C_2H_2/C_2H_4 |
| | $(H_2+C_2H_2)/TG$ | | C_2H_2/C_2H_6 |
| 高温过热 | C_2H_4 | 高能放电 | C_2H_4/TH |
| | C_2H_4/C_2H_6 | | C_2H_2/TH |
| | $(H_2+C_2H_2)/TG$ | | C_2H_2/C_2H_4 |
| | TH | | C_2H_2/C_2H_6 |
| | C_2H_4/TH | | $(H_2+C_2H_2)/TG$ |

4 模型应用流程

基于上述油中溶解气体特征量筛选、故障诊断和特征量分析,文中构建了基于油中溶解气体特征量筛选的变压器故障诊断方法,其流程如图 8 所示。

基于油中溶解气体特征量筛选进行变压器故障诊断,步骤如下:

(1) 收集已知故障的油中溶解气体作为变压器样本集,并提取其特征气体作为原始特征量。

(2) 对原始特征量进行特征衍生处理,形成 32 维变压器特征集合,利用 RF 模型剔除冗余特征量,得到最佳特征组合,并将变压器样本集按一定比例

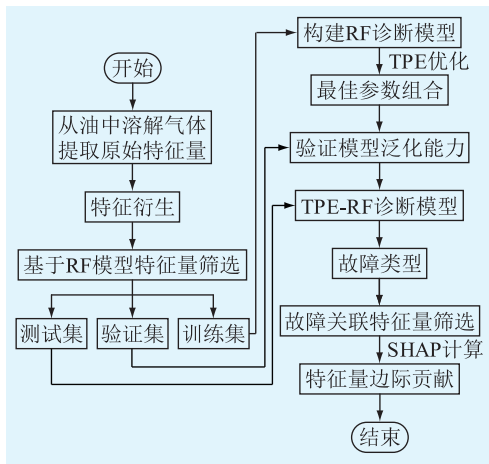


图8 模型应用流程

Fig.8 Flow chart of model application

划分为训练集、测试集和验证集。

(3) 构建 RF 诊断模型,在训练集下利用 TPE 算法实现 RF 模型参数寻优,并在验证集下验证模型的泛化能力,形成 TPE-RF 诊断模型。最终,在测试集下测试模型的诊断效果。

(4) 引入 TreeSHAP 模型分析特征量对各故障类型的重要度,选出每种故障类型关联的主要特征量,得出特征量对诊断结果的边际贡献,并给出一定的解释性分析。

5 算例分析

为验证文中方法的有效性,从文献[17]和现场案例中选择表8所示12组油中溶解气体样本数据。根据表8,对TPE-RF模型和IEC三比值法进行对比分析,结果如表9所示。

表8 油中溶解气体数据

Table 8 Data of dissolved gas in oil

| 序号 | 油中溶解气体体积分数/ $(\mu\text{L}\cdot\text{L}^{-1})$ | | | | |
|----|---|-----------------|-------------------------------|-------------------------------|-------------------------------|
| | H ₂ | CH ₄ | C ₂ H ₆ | C ₂ H ₄ | C ₂ H ₂ |
| 1 | 30.25 | 18.33 | 0.56 | 9.40 | 0 |
| 2 | 40.35 | 25.29 | 1.26 | 15.36 | 0 |
| 3 | 35.50 | 51.70 | 8.01 | 5.47 | 0 |
| 4 | 27 | 90 | 42 | 63 | 0.2 |
| 5 | 1 100 | 1 600 | 221 | 2 010 | 26 |
| 6 | 20.6 | 19.9 | 7.4 | 61.3 | 1.5 |
| 7 | 157.8 | 3.2 | 0.6 | 0.7 | 0 |
| 8 | 29.53 | 15.00 | 0.08 | 0.53 | 0.44 |
| 9 | 147.10 | 5.62 | 1.07 | 3.04 | 0.88 |
| 10 | 19.24 | 1.15 | 6.18 | 0.41 | 6.54 |
| 11 | 6.30 | 7.87 | 6.97 | 3.03 | 11.1 |
| 12 | 39.90 | 7.29 | 1.58 | 4.15 | 6.87 |

可以看出,IEC三比值法用于第1、2、8组样本

表9 故障诊断结果

Table 9 Results of fault diagnosis

| 序号 | 真实故障 | IEC三比值法 | | TPE-RF模型 | |
|----|------|---------|-----|----------|------|
| | | 编码 | 结果 | 原始特征量 | 新特征量 |
| 1 | N | — | — | N | N |
| 2 | N | — | — | N | N |
| 3 | T1 | 020 | T1 | T1 | T1 |
| 4 | T1 | 021 | T1 | T1 | T1 |
| 5 | T2 | 022 | T2 | T2 | T2 |
| 6 | T2 | 002 | T2 | T1* | T2 |
| 7 | D1 | 011 | 无* | D1 | D1 |
| 8 | D1 | — | — | D1 | D1 |
| 9 | D2 | 111 | D3* | D2 | D2 |
| 10 | D2 | 210 | D2 | D3* | D3* |
| 11 | D3 | 220 | D2* | D2* | D3 |
| 12 | D3 | 100 | D3 | D3 | D3 |

注:“*”表示诊断错误。

时,H₂、C₂H₂、总烃浓度均低于DL/T 722—2014《油中溶解气体分析和判断导则》规定的注意值,不满足IEC三比值使用前的判定条件^[27]。当诊断第7组样本时,三比值编码为011,存在编码对应缺失问题,无法诊断出对应的故障类型。同时,在处理第9、11组样本时,由于三比值编码界限过于绝对,在低能放电和高能放电中发生了误判。结果表明,TPE-RF诊断模型能够较好地解决IEC三比值法的不足,同时在选取新特征量组合后,故障诊断效果相比原始特征量得到了提升。

以某地区500 kV主变为例,进一步探究TPE-RF模型的适用性。2019年2月16日该主变C₂H₂出现异常增长情况,缩短色谱跟踪频次至2月27日,在线色谱中C₂H₂显示为0.44 μL/L,对其进行多次离线取样检测,发现27日C₂H₂达到6.74 μL/L,28日C₂H₂达到6.87 μL/L。2月27日在线监测数据和2月28日离线数据分别对应表8中第8、12组样本。TPE-RF模型对在线数据的诊断结果为局部放电,离线色谱数据的诊断结果为高能放电。考虑到现场中在线色谱数据易受到外界干扰,而离线色谱相对可靠,因此文中方法以在线数据诊断结果为辅助参考,以近期多次离线数据的诊断结果为准。

为分析特征量对诊断结果的边际贡献,以高能放电为例,通过计算SHAP值引入特征贡献图,结果如图9所示,其中p为SHAP值。

图9中红色表示正向贡献度,蓝色表示负向贡献度。可以看出,C₂H₂/TH对高能放电结果影响最大,其SHAP值为0.15。C₂H₂/C₂H₄、C₂H₂/C₂H₆、(H₂+C₂H₂)/TG对应SHAP值分别为0.12、0.09和0.09,均促进高能放电的发生。C₂H₄浓度为4.15

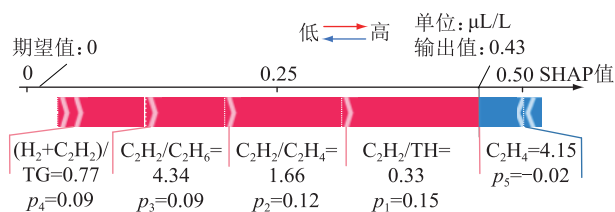


图9 特征贡献图

Fig.9 Diagram of characteristic contribution

μL/L, 对应 SHAP 值为-0.02, 呈现负向贡献度。

结合变压器返厂解体检查结果, 该主变绝缘筒在纵绝缘方向存在明显的电弧放电, 并在高压侧线圈纵绝缘方向发展, 导致线圈内撑条出现灼伤痕迹, 如图 10 所示。故障原因为绝缘筒内部存在缺陷, 绝缘纸发生老化, 并在长期的电场作用下发生局部放电直至产生电弧。

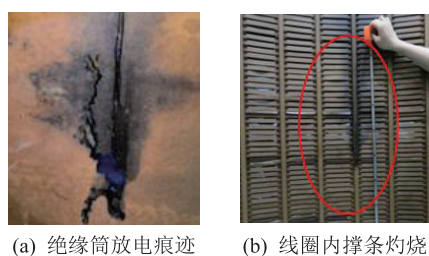


图 10 绕组故障检查结果

Fig.10 Inspection results of winding fault

6 结论

针对传统变压器故障诊断的特征量选取过于依赖现场经验和故障关联特征量分析不足等问题, 文中提出了基于油中溶解气体特征量筛选的变压器故障诊断方法, 得到结论如下:

(1) 利用 RF 模型计算特征量对故障诊断的重要度, 结合故障诊断效果筛选出表征油中溶解气体信息的最优特征量组合。

(2) 采用 TPE 算法优化 RF 模型参数, 有效解决了集成算法的参数众多、调参复杂等问题, 同时结合多种评估指标对 TPE-RF 模型诊断性能进行评估, 并与 IEC 三比值法进行对比分析, 结果表明 TPE-RF 诊断模型有较高的准确性和可靠性。

(3) 融合 TPE-RF 模型, 引入 TreeSHAP 分析特征量对 6 种故障的贡献度, 优选出每种故障关联的特征量。结合特征贡献图, 对诊断结果的主要影响因素进行分析, 增强了诊断模型的可解释性和实用性。

参考文献:

[1] 徐鹏, 薛帅, 杨斌, 等. 一起油浸式并联电抗器总烃异常问题分析[J]. 电力电容器与无功补偿, 2022, 43(1): 63-67.

XU Peng, XUE Shuai, YANG Bin, et al. Analysis of abnormal total hydrocarbon in oil immersed shunt reactor[J]. Power Capacitor & Reactive Power Compensation, 2022, 43(1): 63-67.

[2] 许广虎, 金铭, 张陵, 等. 500 kV 变压器高压套管闪络事故模拟试验分析与预防[J]. 电瓷避雷器, 2022(5): 178-185.

XU Guanghu, JIN Ming, ZHANG Ling, et al. Simulation test analysis and prevention of high voltage bushing flashover accident of 500 kV transformer[J]. Insulators and Surge Arresters, 2022(5): 178-185.

[3] 徐鹏, 吴鹏, 甘强, 等. 500 kV 变压器纯瓷套管漏油事故调查及防范措施[J]. 电瓷避雷器, 2022(3): 186-190.

XU Peng, WU Peng, GAN Qiang, et al. Investigation on oil leakage accident of pure porcelain bushing of 500 kV transformer and preventive measures[J]. Insulators and Surge Arresters, 2022(3): 186-190.

[4] 武天府, 刘征, 王志强, 等. 基于 Focal 损失 SSDAE 的变压器故障诊断方法[J]. 电力工程技术, 2021, 40(6): 18-24.

WU Tianfu, LIU Zheng, WANG Zhiqiang, et al. Transformer fault diagnosis method based on Focal loss SSDAE[J]. Electric Power Engineering Technology, 2021, 40(6): 18-24.

[5] 谢佳, 杨海涛, 丁国成, 等. 变压器高压套管末屏在线监测系统的设计[J]. 电瓷避雷器, 2023(4): 213-220.

XIE Jia, YANG Haitao, DING Guocheng, et al. On-line monitoring system for end shield of high voltage bushing of transformer[J]. Insulators and Surge Arresters, 2023(4): 213-220.

[6] 沙伟燕, 李秀广, 何宁辉, 等. 基于油色谱时频域信息和残差注意网络的变压器故障诊断方法[J]. 电网与清洁能源, 2022, 38(1): 66-75.

SHA Weiyang, LI Xiuguang, HE Ninghui, et al. Transformer fault diagnosis method based on oil chromatogram time-frequency domain information and residual attention network[J]. Power System and Clean Energy, 2022, 38(1): 66-75.

[7] 蔡黎, 袁文泽, 张轩瑞, 等. 基于特高频自感知的变压器局部放电检测方法[J]. 高电压技术, 2021, 47(6): 2041-2050.

CAI Jun, YUAN Wenzhe, ZHANG Xuanrui, et al. Ultra-high frequency self-sensing detection method for transformer partial discharge[J]. High Voltage Engineering, 2021, 47(6): 2041-2050.

[8] 张鹏, 齐波, 李成榕, 等. 电力变压器油中溶解气体特性影响因素的量化分析[J]. 中国电机工程学报, 2021, 41(10): 3620-3631, 3686.

ZHANG Peng, QI Bo, LI Chengrong, et al. Quantitative analysis of influence factors of dissolved gas characteristics in power transformer oil[J]. Proceedings of the CSEE, 2021, 41(10): 3620-3631, 3686.

[9] ARSHAD M, ISLAM S M. Significance of cellulose power transformer condition assessment[J]. IEEE Transactions on Dielectrics and Electrical Insulation, 2011, 18(5): 1591-1598.

[10] 刘恒, 陈文鸿, 沈剑韬, 等. 一种电力变压器绝缘电阻的高压高阻电桥测量方法[J]. 高压电器, 2020, 56(7): 224-228.

LIU Heng, CHEN Wenhong, SHEN Jiantao, et al. Method for measuring power transformer insulation resistance based on high-voltage and high-resistance bridge[J]. High Voltage Ap-

- paratus, 2020, 56(7): 224-228.
- [11] 崔家齐,董海鹰,李帅兵,等. 基于图像特征识别的绝缘纸老化状态评估[J]. 高电压技术, 2022, 48(2): 636-643.
CUI Jiaqi, DONG Haiying, LI Shuaibing, et al. Aging state evaluation of insulating paper based on image feature recognition [J]. High Voltage Engineering, 2022, 48(2): 636-643.
- [12] DUVAL M, DEPABLA A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases [J]. IEEE Electrical Insulation Magazine, 2001, 17(2): 31-41.
- [13] ROGERS R R. IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis [J]. IEEE Transactions on Electrical Insulation, 1978, EI-13(5): 349-354.
- [14] 刘文泽,张俊,邓焱. 基于深度置信网络和多维信息融合的变压器故障诊断方法[J]. 电力工程技术, 2019, 38(6): 16-23.
LIU Wenze, ZHANG Jun, DENG Yan. Transformer fault diagnosis method based on deep learning and multi-dimensional information fusion [J]. Electric Power Engineering Technology, 2019, 38(6): 16-23.
- [15] 陈铁,吕长钦,张欣,等. 基于KPCA-WPA-SVM的变压器故障诊断模型[J]. 电测与仪表, 2021, 58(4): 158-164.
CHEN Tie, LYU Changqin, ZHANG Xin, et al. Transformer fault diagnosis model based on KPCA-WPA-SVM [J]. Electrical Measurement & Instrumentation, 2021, 58(4): 158-164.
- [16] 王雪,韩韬. 基于贝叶斯优化随机森林的变压器故障诊断[J]. 电测与仪表, 2021, 58(6): 167-173.
WANG Xue, HAN Tao. Transformer fault diagnosis based on Bayesian optimized random forest [J]. Electrical Measurement & Instrumentation, 2021, 58(6): 167-173.
- [17] 张又文,冯斌,陈页,等. 基于遗传算法优化XGBoost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.
ZHANG Youwen, FENG Bin, CHEN Ye, et al. Fault diagnosis method for oil-immersed transformer based on XGBoost optimized by genetic algorithm [J]. Electric Power Automation Equipment, 2021, 41(2): 200-206.
- [18] 汪可,李金忠,张书琦,等. 变压器故障诊断用油中溶解气体新特征参量[J]. 中国电机工程学报, 2016, 36(23): 6570-6578, 6625.
WANG Ke, LI Jinzhong, ZHANG Shuqi, et al. New features derived from dissolved gas analysis for fault diagnosis of power transformers [J]. Proceedings of the CSEE, 2016, 36(23): 6570-6578, 6625.
- [19] 刘仲民,翟玉晓,张鑫,等. 基于DBN-IFCM的变压器故障诊断方法[J]. 高电压技术, 2020, 46(12): 4258-4265.
LIU Zhongmin, ZHAI Yuxiao, ZHANG Xin, et al. Transformer fault diagnosis method based on deep belief network and improved fuzzy C-means clustering [J]. High Voltage Engineering, 2020, 46(12): 4258-4265.
- [20] 罗云芳,李力. 一种基于矩阵分解和随机森林算法的推荐模型[J]. 计算机应用与软件, 2021, 38(8): 315-321.
LUO Yunfang, LI Li. A recommendation model based on matrix decomposition and random forest algorithm [J]. Computer Applications and Software, 2021, 38(8): 315-321.
- [21] 龚雪娇,朱瑞金,唐波. 基于贝叶斯优化XGBoost的短期峰值负荷预测[J]. 电力工程技术, 2020, 39(6): 76-81.
GONG Xuejiao, ZHU Ruijin, TANG Bo. Short-term peak load forecasting based on Bayesian optimization XGBoost [J]. Electric Power Engineering Technology, 2020, 39(6): 76-81.
- [22] 翁克瑞,刘淼,刘钱. TPE-XGBOOST与LassoLars组合下PM2.5浓度分解集成预测模型研究[J]. 系统工程理论与实践, 2020, 40(3): 748-760.
WENG Kerui, LIU Miao, LIU Qian. An integrated prediction model of PM2.5 concentration based on TPE-XGBOOST and LassoLars [J]. Systems Engineering-Theory & Practice, 2020, 40(3): 748-760.
- [23] LUNDBERG S, LEE S I. A unified approach to interpreting model predictions [C] // Conference and Workshop on Neural Information Processing Systems. California, USA, 2017: 4765-4774.
- [24] JWG D1/A2.47. Advances in DGA interpretation [R]. CIGRE, 2019.
- [25] 孙才新. 电气设备油中气体在线监测与故障诊断技术 [M]. 北京: 科学出版社, 2003.
SUN Caixin. On-line monitoring and fault diagnosis technology of gas in oil of electrical equipment [M]. Beijing: Science Press, 2003.
- [26] 代杰杰,宋辉,杨祎,等. 基于油中气体分析的变压器故障诊断ReLU-DBN方法[J]. 电网技术, 2018, 42(2): 658-664.
DAI Jiejie, SONG Hui, YANG Yi, et al. Dissolved gas analysis of insulating oil for power transformer fault diagnosis based on ReLU-DBN [J]. Power System Technology, 2018, 42(2): 658-664.
- [27] 王健一,李金忠,凌愨,等. 新版电力行业标准《变压器油中溶解气体分析判断导则》解读[J]. 变压器, 2014, 51(12): 49-53.
WANG Jianyi, LI Jinzhong, LING Min, et al. Interpretation of the new electricity industry standard guide to the analysis and the diagnosis of gases dissolved in transformer oil [J]. Transformer, 2014, 51(12): 49-53.
- [28] 孙琛. 基于机器学习算法的变压器故障诊断 [D]. 上海: 上海交通大学, 2019.
SUN Chen. Transformer fault diagnosis based on machine learning algorithm [D]. Shanghai: Shanghai Jiao Tong University, 2019.
- [29] 翟夕阳,王晓丹,雷蕾,等. 基于多类指数损失函数逐步添加模型的改进多分类AdaBoost算法[J]. 计算机应用, 2017, 37(6): 1692-1696.
ZHAI Xiyang, WANG Xiaodan, LEI Lei, et al. Improved multi-class AdaBoost algorithm based on stagewise additive modeling using a multi-class exponential loss function [J]. Journal of Computer Applications, 2017, 37(6): 1692-1696.
- [30] 王彦兵,王聪,赵亚丽,等. 基于ROC曲线的永久散射体识

别最佳阈值定量筛选[J]. 遥感学报, 2021, 25(10): 2083-2094.

WANG Yanbing, WANG Cong, ZHAO Yali, et al. Quantitative determination of the optimal threshold of permanent scatterer based on ROC curve [J]. National Remote Sensing Bulletin, 2021, 25(10): 2083-2094.

[31] 吴天傲, 李江, 张薇, 等. 基于类别特征编码的参考作物蒸散量预报模型[J]. 应用基础与工程科学学报, 2022, 30(6): 1402-1419.

WU Tian'ao, LI Jiang, ZHANG Wei, et al. A novel reference evapotranspiration forecasting model based on categorical fea-

ture encoding methods[J]. Journal of Basic Science and Engineering, 2022, 30(6): 1402-1419.

作者简介:



廖才波

廖才波(1989),男,博士,副教授,研究方向为输变电设备多物理场仿真分析及状态评估(E-mail:lcb1752@126.com);

杨金鑫(1998),男,硕士在读,研究方向为电力变压器故障诊断技术;

胡雄(1985),男,博士,讲师,研究方向为智能电网的分布式控制与优化。

Fault diagnosis method for transformers based on feature selection of dissolved gas in oil

LIAO Caibo, YANG Jinxin, HU Xiong, QIU Zhibin, LIU Xiaotian, ZHU Wenqing
(School of Information Engineering, Nanchang University, Nanchang 330031, China)

Abstract: Dissolved gas analysis is important for the early warning and diagnosis of transformer faults. Aiming at the problems of numerous types of features for dissolved gas in oil and the insufficient analysis of fault associated features, a new fault diagnosis method for oil-immersed transformers based on feature selection of dissolved gas in oil is proposed. Firstly, the derivation of original features for dissolved gases is completed. The optimal combination of features is selected by calculating the importance of features for fault diagnosis based on random forest (RF). Then, the tree-structured parzen estimator (TPE) is used to realize the parameter optimization of the RF model, and the TPE-RF diagnostic model is obtained. Combined with the various evaluated indicators, the proposed method is proved to be able to diagnosis the transformer faults accurately. Finally, the TreeSHAP model is introduced to analyze the importance of the features corresponding to each fault, and the specialized features for each fault are selected. According to the case of transformer in operation, the applicability of the method in the power system is discussed, and the effectiveness of the method is verified.

Keywords: gas dissolved in oil; power transformer; fault diagnosis; tree-structured parzen estimator (TPE); random forest (RF); feature selection; TreeSHAP model

(编辑 方晶)