

DOI: 10.19783/j.cnki.pspc.211401

基于FCBF特征选择和XGBoost原则的油纸绝缘介电响应特征量优选研究

刘庆珍, 黄昌硕

(福州大学电气工程与自动化学院, 福建省新能源发电与电能变换重点实验室, 福建 福州 350108)

摘要: 针对高维特征空间中存在的相关特征、冗余特征等导致变压器油纸绝缘综合诊断的特征空间提取存在平均误差大、分类正确率低等问题, 提出一种基于快速过滤相关算法和极限梯度上升相结合的特征量优选策略。首先, 根据变压器的介电响应实测数据, 提取多种类别的时域介电特征量形成初始高维特征空间。其次, 提出一种两级式时域特征选择方法, 第一级采用快速相关过滤算法剔除低相关、高冗余的特征量, 第二级依照极限梯度提升评估特征的重要度, 从而确定最优特征空间。最后设置不同对照组对最优特征空间进行对比论证, 有效验证了采取所提优选策略得到的最优特征空间的合理性及准确性。

关键词: 油纸绝缘; 综合诊断; 快速相关过滤算法; 极限梯度上升; 特征选择

Optimization of dielectric response characteristics of oil paper insulation based on FCBF feature selection and the XGBoost principle

LIU Qingzhen, HUANG Changshuo

(College of Electrical Engineering and Automation, Fujian Key Laboratory of New Energy Generation and Power Conversion, Fuzhou University, Fuzhou 350108, China)

Abstract: There are problems of large average error and low classification accuracy in feature space extraction of transformer oil-paper insulation comprehensive diagnosis. These problems are due to the existence of correlation and redundant features in high-dimensional feature space. Thus a feature quantity optimization strategy based on a fast filtering correlation algorithm and limit gradient rise is proposed. First, from the measured data of transformer dielectric response, various kinds of time-domain dielectric characteristics are extracted to form the initial high-dimensional feature space. Secondly, a two-stage time-domain feature selection method is proposed. In the first stage, a fast correlation filtering algorithm is used to eliminate the features with low correlation and high redundancy, and in the second stage the importance of features is evaluated according to the limit gradient, so as to determine the optimal feature space. Finally, different control groups are set for comparative demonstration of the optimal feature space. This effectively verifies the rationality and accuracy of the optimal feature space obtained by adopting the optimal strategy proposed above.

This work is supported by the National Natural Science Foundation of China (No. 61174117).

Key words: oil-paper insulation; comprehensive diagnosis; fast correlation filtering algorithm; limit gradient rise; feature selection

0 引言

电力变压器作为电力系统中电能转换和分配的重要枢纽设备, 其运行状况的优劣直接影响电力系统的安全稳定性^[1-3]。所以, 及时有效地对变压器进

行绝缘老化状态检测具有重要的应用价值。在绝缘老化状态诊断研究领域, 介质响应法包括回复电压法(RVM)和极化/去极化电流法(PDC), 相较于传统非电气量测量方法操作简单、测量无损^[4-6], 已成为该领域的研究热点。

目前, 在基于介电响应技术的变压器油纸绝缘老化研究领域已积累了许多成果。文献[7]最早提

基金项目: 国家自然科学基金项目资助(61174117)

出拓展 Debye 模型, 为后续研究奠定了基础; 文献[8]研究了 RVM 针对变压器绝缘诊断的适用性及可行性; 文献[9]通过变压器老化实验, 探究特征量的不同老化特性。

近年来学者们从微分介电谱的参数求解入手, 挖掘越来越多的特征量。文献[10]从去极化能量谱中提取峰值能量 w_{\max} 和主时间常数 t_{cdom} 两个特征量; 文献[11]根据不同老化阶段的去极化电流并结合去极化电流特性提出稳态油老化电量 Q_0 和稳态油纸界面极化电量 Q_{op} 来表征油纸绝缘状态; 文献[12]在去极化电流谱线末端采用双点解析法来获得极化最大斜率 a 和最大截距 b 。

随着特征量数目的增多, 在诊断过程中一些影响甚小和不相关的特征通常被忽视, 使得诊断结果易受某一特征的影响。在传统的绝缘老化诊断中往往采用单一特征量对老化状态进行综合诊断, 由于不同的绝缘老化条件对特征性能的影响也不尽相同, 若只想依靠单一特征量来反映绝缘老化特性, 很难对变压器的绝缘老化状态做到正确评估。

针对目前单一的时域特征量无法全面反映变压器绝缘老化的劣性问题, 许多学者开始采用多个特征量进行综合诊断研究。在现有绝缘老化的综合诊断研究中, 老化特征量包含传统特征量及新提取的特征量, 使得特征量过于冗余以致于特征量集的选取存在模糊性和随机性^[13], 往往较易忽略个别重要特征量的作用。因此, 对复杂特征量集进行优化, 找寻最优的特征量集来探究变压器绝缘状态的规律性, 是后续研究的前提。

本文基于当前研究领域下变压器油纸绝缘综合诊断的现状进行分析, 针对特征量选择存在的弊端, 提出一种基于 FCBF 特征选择和 XGBoost 重要度评估的两级式优选策略。为提高特征空间的分类准确性, 该优选策略从特征量的冗余度与重要度两个角度对特征量进行优化处理, 旨在考虑尽可能多的有效特征量的情况下保留其内在联系。特征量的性能在很大程度上会影响绝缘诊断的准确性, 因此该优选策略在优化出一组最优特征量集的同时保证各特征量的有效性。

1 时域介电响应特征量分析

在当下的研究中, 介质响应等效电路能够合理描述弛豫老化结构及其内部机理, 通常采用拓展 Debye 等效电路以及 RVM 和 PDC 两种时域介电响应技术对变压器进行油纸绝缘诊断, 其中极化/去极化电流曲线中蕴含丰富的特征量信息, 学者们通过

研究去极化电流曲线并建立回复电压极化谱来提取时域特征量。综上, 根据特征量的不同提取方式将特征量分为拓展 Debye 等效电路特征量、RVM 特征量以及 PDC 特征量。

1.1 拓展 Debye 等效电路特征量

拓展 Debye 等效电路作为广泛应用于反映复合绝缘介质弛豫响应过程的电路模型^[14-15], 其等效电路参数也被深入挖掘来建立与绝缘老化系统之间的深层关系。拓展 Debye 模型的等效电路拓扑如图 1 所示。

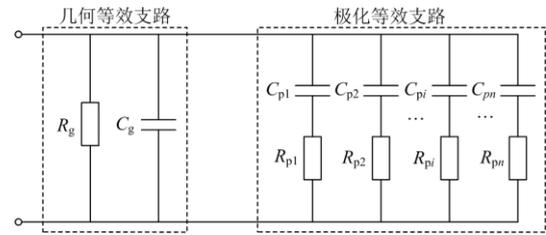


图 1 拓展 Debye 等效电路

Fig. 1 Equivalent circuit of extended Debye model

拓展 Debye 等效电路由几何等效支路和极化等效支路组成。几何等效支路由 R_g 和 C_g 并联而成, 极化等效支路是通过 N 条极化电阻 R_p 和极化电容 C_p 的串联支路相互并联来反映介质极化在不同弛豫时间下的现象^[14]。

现有研究对 Debye 模型已经提取出 R_g 、 C_g 、 K_L 、 N 、 l 、 $\bar{\tau}$ 、 τ_{\max} 、 K_τ ^[16]8 个特征量, 为方便分析绝缘介质的老化状况, 将特征量与绝缘老化的关系列于表 1 中。

表 1 Debye 特征量与绝缘老化的关系

Table 1 Relationship between Debye characteristic quantity and insulation aging

特征量名称	与绝缘老化的关系
几何电阻 R_g	老化加深, 值减小
几何电容 C_g	老化加深, 值增大
老化因子 K_L	老化加深, 值增大
弛豫机构数 N	老化加深, 值增大
极化小支路数 l	老化加深, 值增大
平均时间常数 $\bar{\tau}$	老化加深, 值增大
最大时间常数 τ_{\max}	老化加深, 值增大
弛豫时间常数比 K_τ	老化加深, 值增大

1.2 回复电压法特征量

回复电压法基于介质响应理论, 是利用绝缘材料的极化特性测得绝缘设备两端的回复电压^[17-18]。回复电压测量过程分为充电、放电、开路测量以及松弛 4 个阶段^[17]。单次回复电压测试曲线如图 2 所示。

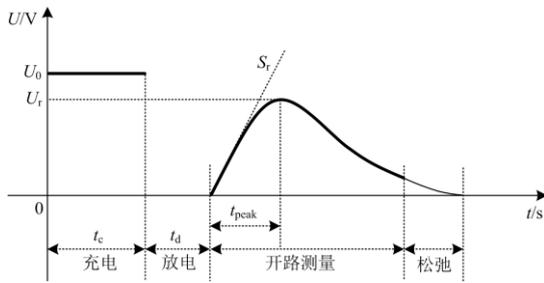


图2 回复电压测量曲线

Fig. 2 Return voltage measurement curve

在单次回复电压测量曲线中可以直接提取回复电压的初始斜率 S_r 、峰值时间 t_{peak} 两个老化特征量。通过改变充电持续时间记录每个周期的回复电压最大值 U_r ，并绘制得到回复电压极化谱^[19-20]。从极化谱中可以提取峰值电压 U_{max} 、主时间常数 t_{cdom} 和半峰周期时间 $t_{1/2}$ 等老化特征量^[21]。各回复电压法特征量与绝缘老化的关系如表2所示。

表2 回复电压法特征量与绝缘老化的关系

Table 2 Relationship between characteristic quantity of recovery voltage method and insulation aging

特征量名称	与绝缘老化的关系
最大回复电压 U_r	老化加深, 值减小
初始斜率 S_r	老化加深, 值增大
峰值时间 t_{peak}	老化加深, 值增大
主时间常数 t_{cdom}	老化加深, 值增大
半峰周期时间 $t_{1/2}$	老化加深, 值增大

1.3 极化/去极化法特征量

极化/去极化电流法的测量原理与回复电压法相类似, 在待测变压器绕组间外加直流脉冲电压 U_0 对其持续充电, 在此过程中绝缘系统开始发生极化响应, 并产生极化电流 i_p ^[22-23]; 经过 t_c 秒后断开外施电压, 将变压器短接放电, 此时内部弛豫机构发生去极化响应, 记录产生的去极化电流 i_d ^[22]。极化/去极化电流测量曲线如图3所示。

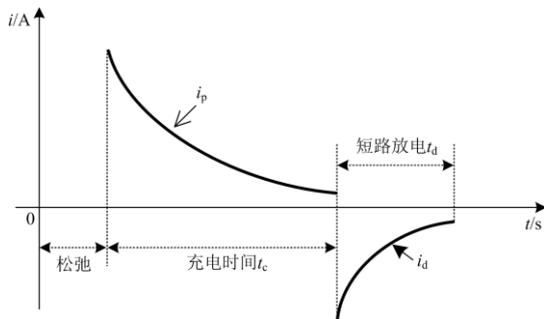


图3 极化/去极化电流测量波形图

Fig. 3 Waveform diagram of polarization/depolarization current measurement

学者们深入研究不同电压下的去极化能量谱, 从去极化能量谱中挖掘出了部分新的老化特征量, 并通过设置绝缘老化检修前后对照实验来探寻各特征量与绝缘老化之间的联系^[24], 如表3所示。

表3 极化/去极化法特征量与绝缘老化的关系

Table 3 Relationship between polarization/depolarization characteristic quantity and insulation aging

特征量名称	与绝缘老化的关系
去极化电量 Q	老化加深, 值增大
弛豫贡献度 A	老化加深, 值增大
弛豫损耗 W_g	老化加深, 值增大
弛豫能量 W_c	老化加深, 值增大
吸收比 K	老化加深, 值增大
极化指数 P_I	老化加深, 值减小
能量谱峰值 w_{max}	老化加深, 值增大
最大峰值时间 t_{max}	老化加深, 值减小
稳态极化电量 Q_{p-5000}	老化加深, 值减小
极化电量斜率 K_p	老化加深, 值减小
稳态油老化电量 Q_o	老化加深, 值减小
油纸极化电量 Q_{o-p}	老化加深, 值减小
极化最大斜率 a	老化加深, 值减小
极化最大截距 b	老化加深, 值增大
始端去极化电量 S_{oil}	老化加深, 值增大
末端去极化电量 S_{paper}	老化加深, 值增大

综上所述, 本文整合现有研究的29个特征量作为特征量优选的目标, 并依此形成初始特征空间。

2 特征量优选策略

特征量性能是绝缘诊断的有效前提, 提高特征空间的分类度能够提高绝缘诊断的准确性。为确保特征量的良好性能, 优化特征空间结构, 本文采用快速相关过滤算法(FCBF)和极限梯度提升算法(XGBoost)相结合的两级式优选方法来完成特征空间优化。

2.1 基于快速相关过滤算法的特征量冗余性、相关性分析

快速相关过滤算法(FCBF)采用对称不确定性(SU)这一指标来衡量特征量之间的相关性^[25], 并通过定义各特征量与老化类别之间的相关度以及各特征量间的冗余度来筛选相关度高、冗余度低的特征量。

设初始特征集为 $X = (x_{ij})_{N \times M}$, 表征 N 个 M 维的样本向量和 M 个 N 维的特征向量。为更准确地描述对称不确定性(SU)的相关信息量, 引入熵作为对随机变量的度量手段。假定两个随机变量 $\mathbf{X} = \{x_1, x_2, \dots, x_i\} (i=1, 2, \dots, m)$ 有 m 个特征向量, $\mathbf{Y} =$

$\{y_1, y_2, \dots, y_j\} (j=1, 2, \dots, n)$ 有 n 个特征向量, 则随机变量 X 的信息熵 $H(X)$ 、随机变量 X 在变量 Y 条件下的条件熵 $H(X|Y)$ 分别定义为

$$H(X) = -\sum_i P(x_i) \log_2 P(x_i) \quad (1)$$

$$H(X|Y) = -\sum_j P(y_j) \sum_i P(x_i|y_j) \log_2 P(x_i|y_j) \quad (2)$$

式中: $P(x_i)$ 、 $P(y_j)$ 分别为随机变量 X 、 Y 取值为 x_i 、 y_j 时的概率; $P(x_i|y_j)$ 为在 $Y = y_j$ 条件下随机变量 X 取值为 x_i 的概率。

定义随机变量 X 和 Y 之间的相关信息的度量方法为互信息, 则二者之间的互信息表示为

$$I(X; Y) = H(X) - H(X|Y) \quad (3)$$

结合式(1)与式(3), 得出对称不确定性(SU)表示为

$$L_{SU}(X; Y) = 2 \frac{I(X; Y)}{H(X) + H(Y)} \quad (4)$$

由 SU 定义式可知, $SU(X; Y)$ 取值范围为(0,1), 其取值越大则表明两个随机变量的相关性越大。

为保证初步筛选的特征空间的目标特征与老化类别之间具有最大关联度、与其余特征之间的冗余度最小, 定义老化类别为变量 C , 计算原始特征空间中各特征 v_i 与老化类别 C 之间的相关度 $L_{SU}(v_i, C)$, 设置阈值 m , 若 $L_{SU}(v_i, C) < m$, 则剔除特征 v_i , 保留所有符合条件的特征量; 将剩余特征量进行二次筛选, 计算特征 v_i 与余下各特征 v_j 之间的相关度, 若 $L_{SU}(v_i, v_j) > L_{SU}(v_j, C)$, 则剔除特征 v_j 。

经过上述方法筛选后, 特征量按照从优到劣顺序排序, 得到第一级特征选择的特征空间。

2.2 基于极限梯度提升算法的特征量重要度评估

极限梯度提升算法(XGBoost)集成多棵回归树, 通过评估各特征量的重要度来挖掘数据的结构信息, 直观地反映各特征量对绝缘老化的贡献度, 合理保留重要度高的特征并淘汰影响度小的特征, 实现特征空间的进一步优化。

根据第一级选择后的特征空间 $D = \{(x_i, y_j) : i=1, \dots, p, j=1, \dots, q\}$, 其中共有 p 个样本, 每个样本有 q 个特征, 构建决策树回归模型, 如式(5)所示。该模型为融合多特征量的综合绝缘诊断的分类模型, 有效反映拟合训练数据的程度^[26]。为保证绝缘诊断的高效准确率, 防止过拟合, 引入正则化损失函数, 如式(6)所示。

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (5)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (6)$$

式中: f_k 代表一颗回归树; $\Omega(f_k)$ 代表损失函数的正则化项; γ 和 λ 为损失函数的惩罚系数; T 和 w 分别表示第 k 棵树的叶子数目及叶子权重。

损失函数越小表示回归树分类结果越优越, 对损失函数进行二阶泰勒展开可以得到一个叶子 j 的最优权重, 并相应计算得到最优目标值, 如式(7)所示。

$$Obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \lambda T \quad (7)$$

对于所有叶节点, 采用贪心算法对子树划分, 每次对一个节点进行分裂, 分裂前后的信息增益为

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{(H_L + H_R) + \lambda} \right] - \gamma \quad (8)$$

在进行节点分裂时, 计算所有特征的信息增益值, 选择信息增益最大的特征进行分裂^[27]。依次不断迭代, 将特征列向量作为树的分支节点, 从而可得到特征 j 在单颗树中的重要度为

$$\hat{J}_j^2(T) = \sum_{k=1}^{K-1} I_w^2 \quad (9)$$

式中: K 代表叶子的节点个数; 而 $K-1$ 为非叶子的节点数; I_w 表示节点 w 分裂后平方损失的减少值。

特征 j 的全局重要度表征为特征 j 在单颗树中的重要度的平均值, 如式(10)所示。

$$\hat{j}_j^2 = \frac{1}{M} \sum_{m=1}^M \hat{J}_j^2(T_m) \quad (10)$$

3 特征量优选策略的应用

3.1 建立初始特征空间

根据前文对时域介电响应特征量的论述, 基于 80 余组变压器绕组的现场实测数据进行特征计算和大量的仿真实验, 选择上述 29 个特征量作为本文的特征优选对象并构建初始特征空间。

鉴于传统的特征选择方法一般保留 30%~40% 的特征数量, 并且油纸绝缘诊断所研究的特征空间维数为 8~10 维, 因此本文案例中优选后的最优特征空间最终保留 6~8 个特征量。在两级式的优选过程中, 第一层经过 FCBF 特征筛选剔除约 1/3 的相关度低、冗余度高的特征量, 第二层采用 XGBoost 重要度评估来甄别出对目标类别重要度高的特征量, 完成特征优选。

3.2 特征量相关度、冗余度优选

根据 FCBF 原则, 设定阈值 m 为 0.5, 计算各个特征量与老化类别之间的相关度以及各特征量间的相关度, 如表 4 以及式(11)所示, 由于篇幅限制, 只展示部分矩阵。

表 4 各特征量与老化类别之间的相关度 SU

Table 4 Correlation between each characteristic quantity and aging category SU

排名	特征	$SU(v_i, C)$	排名	特征	$SU(v_j, C)$
1	t_{cdom}	0.8515	16	w_{max}	0.6203
2	R_g	0.8382	17	Q	0.6189
3	l	0.8267	18	$\bar{\tau}_{max}$	0.5908
4	a	0.8013	19	A	0.5734
5	$t_{1/2}$	0.7888	20	t_p	0.5486
6	U_{rmax}	0.7759	21	K_r	0.5366
7	N	0.7630	22	P_1	0.5092
8	t_{max}	0.7487	23	Q_{o-p}	0.4873
9	W_c	0.7362	24	C_g	0.4625
10	Q_o	0.7158	25	W_g	0.4331
11	K	0.6941	26	S_{paper}	0.4104
12	S_r	0.6811	27	$\bar{\tau}$	0.3890
13	K_p	0.6627	28	S_{oil}	0.3525
14	b	0.6433	29	Q_{p-5000}	0.3173
15	K_L	0.6295	—	—	—

由表 4 各特征量与老化类别间的相关度排序可知, Q_{o-p} 、 C_g 、 W_g 、 S_{paper} 、 $\bar{\tau}$ 、 S_{oil} 、 Q_{p-5000} 这些特征对于油纸绝缘老化状态的敏感度较低, 与老化类

别间的关联度较小, 故剔除并保留剩余特征量进行二次筛选。鉴于式(11)计算各特征量间的冗余度, 可以得出 Q 、 W_c 、 W_g 等特征高度线性相关, 而 A 、 Q_o 、 P_1 、 K_p 、 K_L 这些特征属于在特征空间下全局相关性低、与其余特征量间的冗余度高的特征量, 故选择剔除。

$$SU(v_i, v_j) = \begin{matrix} t_{cdom} & \begin{bmatrix} - & 0.857 & 0.808 & L & 0.368 \\ R_g & 0.857 & - & 0.824 & L & 0.392 \\ l & 0.808 & 0.824 & - & L & 0.421 \\ M & M & M & M & M & M \\ Q_{p-5000} & 0.368 & 0.392 & 0.421 & L & - \end{bmatrix} \end{matrix} \quad (11)$$

综上所述, 剔除总体特征中对于相关度、冗余度表现较为劣态的 12 个特征量, 并形成第一级优选后的特征空间 $V_1 = \{t_{cdom}$ 、 R_g 、 l 、 a 、 $t_{1/2}$ 、 U_{rmax} 、 N 、 t_{max} 、 W_c 、 K 、 S_r 、 b 、 w_{max} 、 Q 、 τ_{max} 、 t_p 、 $K_r\}$ 。

3.3 特征量重要度评估

XGBoost 集成若干回归树后每个节点不断进行特征分裂, 在这一背景下, 可以将选中特征的分裂次数作为该特征的重要度^[27]。将特征空间 V_1 放入 XGBoost 模型中进行特征训练, 并计算特征分裂次数来分析特征重要度。本文共进行 4 次特征重要度筛选, 每次剔除 1~2 个重要度最低的特征并生成新的特征空间, 直至特征空间维度缩减至最小阈值。特征重要度筛选过程如图 4 所示。

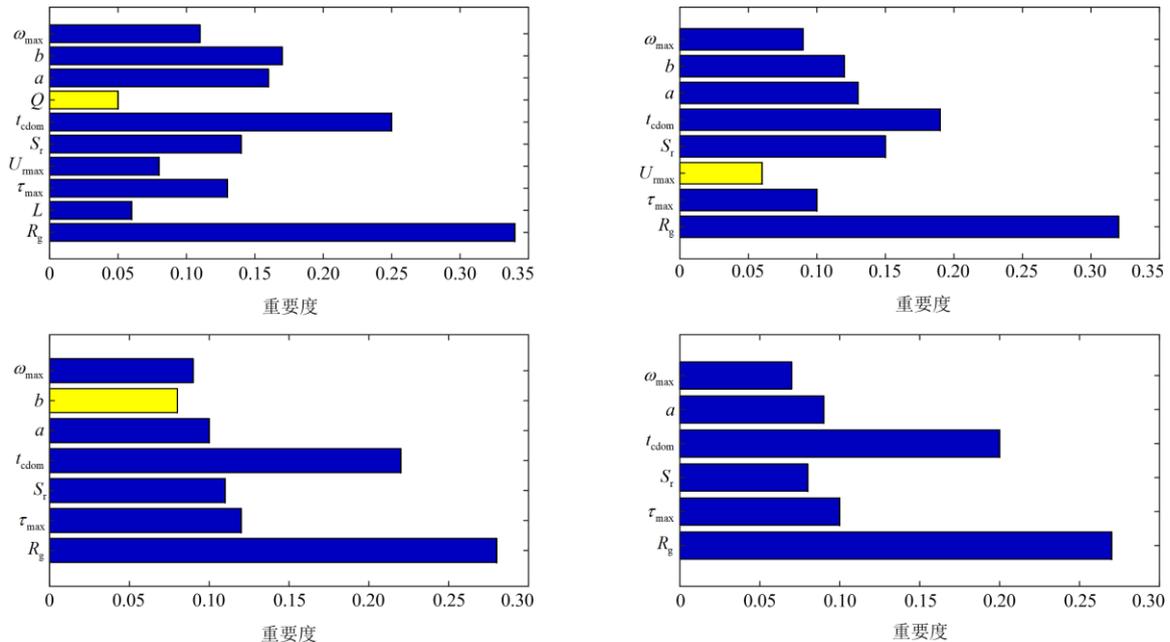


图 4 特征重要度筛选

Fig. 4 Feature importance screening

由图 4 可知, 6 维特征空间 V_2 、7 维特征空间 V_3 以及 8 维特征空间 V_4 均符合目标特征空间的维度要求。为保证选择的特征量性能最优, 对以上 3 个待选特征空间采用主成分分析法, 对 R_g 、 τ_{\max} 、 U_{\max} 、 S_r 、 t_{cdom} 、 a 、 b 、 w_{\max} 这 8 个特征量进行主

成分分析, 利用方差贡献率来选取主成分。主成分的方差越大, 表明其弛豫贡献度越高, 特征涵盖弛豫信息越丰富。计算各主成分的方差及其方差贡献率, 如表 5 所示。

表 5 各主成分的方差及其方差贡献率

Table 5 Variance of each principal component and its variance contribution rate

主成分	1	2	3	4	5	6	7	8
方差	1.298	0.863	0.285	0.131	0.097	0.052	0.036	0.008
方差贡献率/%	38.74	24.81	11.67	4.29	2.02	1.73	0.87	0.34
累计贡献率/%	38.74	63.55	75.22	79.51	79.53	81.26	84.13	84.47

由表 5 可以看出, 前 6 个主成分的方差累计贡献率已经达到 81.26%, 一般规定超过 80% 即可代表所有方差^[28]。同时, 这 6 个特征完全符合 XGBoost 第 4 轮的选择结果, 且包含 3 种类型的介电特征, 涵盖弛豫信息广。因此选择特征空间 V_2 作为最终优化的特征空间 Ω_Z , 如表 6 所示。

表 6 特征空间 Ω_Z 的特征量分布Table 6 Distribution of characteristic quantity in feature space Ω_Z

特征空间	维度	类别/个数	特征量	
Ω_Z	6	RVM/2	S_r	t_{cdom}
		PDC/2	w_{\max}	a
		等效电路/2	R_g	\bar{r}_{\max}

4 验证与比较分析

4.1 设置对比组

为验证本文所提的变压器时域特征优选策略的有效性及其可靠性, 根据近些年在变压器绝缘老化诊断方面的研究成果, 搜集学者们提取的几组特征空间以及本文在优选过程中生成的特征空间来对比论证最优特征空间的优越性。

基于收集到的变压器现场实例数据, 本文选用如下 4 个特征空间作为对照组进行对比分析: 使用 FCBF 优选后的优先级前 6 位的 6 维特征空间 Ω_1 ; 文献[29]基于多指标回归选择的特征空间 Ω_2 ; 文献[30]中选取作为灰靶理论老化诊断的特征空间 Ω_3 ; 文献[31]中采用 AHP-TOPSIS 法评估的特征空间 Ω_4 。各特征空间包含特征情况如表 7 所示。

4.2 诊断结果分析

本文拟采用支持向量机(SVM)^[32]、K 阶近邻法(KNN)^[32]、梯度上升树(GBDT)^[32]和极限梯度上升(XGBoost) 4 种不同的诊断方法以及 SVM+XGBoost、KNN+GBDT 两种的结合诊断方法来检验最优空间 Ω_Z 的性能。

表 7 待验证特征空间的特征量分布

Table 7 Feature quantity distribution of feature space to be verified

特征空间	维度	类别/个数	特征量		
Ω_{20}	29	3 类/29	初始特征空间全部特征量		
Ω_1	6	RVM/3	U_{\max}	t_{cdom}	$t_{1/2}$
		PDC/1 等效电路/2	a R_g	—	—
Ω_2	5	RVM/3	U_{\max}	S_r	t_{cdom}
		PDC/2	t_{\max}	P_I	—
Ω_3	8	RVM/3	U_{\max}	S_r	t_{cdom}
		PDC/2	Q	w_{\max}	—
		等效电路/3	R_g	\bar{r}_{\max}	N
Ω_4	6	RVM/3	U_{\max}	S_r	t_{cdom}
		PDC/1 等效电路/2	t_{\max} R_g	—	—

为减小算法模型的差异性, 对不同的特征空间分别进行三折交叉验证^[33], 计算各分类算法对于特征空间 Ω_1 — Ω_4 的准确率, 结果如表 8 所示。表 9 根据表 8 的诊断结果对平均准确率、区间宽度、最大误差以及基于诊断准确率平均值的方差做了统计。表 10 是对各特征空间诊断时间的统计。

根据表 8、表 9 的诊断结果, 进行如下分析。

1) 诊断准确率分析: 由表 8 可知, 不管采用哪种诊断方法, 最优特征空间 Ω_Z 的特征量的诊断准确率都超过了 90%, 采用组合算法诊断的准确率更是超过了 95%, 表 9 中的平均诊断准确率为 94.88%, 明显高于其他对照组的诊断结果, 这说明本文提出的特征优选策略能够明显提高绝缘老化诊断的准确率。

2) 对诊断方法的适应性分析: 由表 8 还可以看出, 各组特征量应用于不同的诊断方法, 优选的特征空间 Ω_Z 的特征量相对于不同诊断方法的诊断结果差异不大, 而其他各对照组相对于不同的诊断方

表 8 各诊断算法下特征空间的准确率

Table 8 Accuracy of feature space under each diagnostic algorithm

特征空间	各诊断算法准确率/%					
	SVM	KNN	GBDT	XGBoost	SVM+XGBoost	KNN+GBDT
Ω_Z	92.32	90.11	94.83	95.19	98.58	98.26
Ω_{Z0}	79.72	87.54	89.62	93.07	94.29	94.66
Ω_1	82.35	83.78	90.39	92.74	94.18	93.92
Ω_2	80.91	81.63	87.32	90.72	92.41	92.58
Ω_3	84.82	86.51	92.48	93.37	94.74	94.69
Ω_4	84.25	87.18	92.53	92.05	92.07	92.36

表 9 各特征空间评价指标结果

Table 9 Evaluation index results of each feature space

特征空间	平均准确率/%	区间宽度/%	最大误差/%	方差
Ω_Z	94.88	8.47	4.77	3.0144
Ω_{Z0}	89.82	14.92	10.10	5.1836
Ω_1	89.56	11.83	7.21	4.7705
Ω_2	87.60	11.67	6.69	4.7984
Ω_3	91.10	9.92	6.28	4.1211
Ω_4	90.07	4.28	5.82	3.2000

法存在较大差异。表 9 中的最大误差与方差指标也表明优化的特征空间对于不同的诊断方法依然具有较强的适应性，且波动性小，这对于变压器的绝缘老化分析具有特别重要的意义。

3) 差异性分析：对比计算中的各组特征空间，

表 10 各特征空间的诊断时间

Table 10 Diagnostic time of each feature space

特征空间	诊断时间/s					
	SVM	KNN	GBDT	XGBoost	SVM+XGBoost	KNN+GBDT
Ω_Z	3.03	2.83	2.24	2.16	1.97	1.62
Ω_{Z0}	9.06	8.91	9.12	8.68	7.82	8.53
Ω_1	4.11	4.06	3.98	4.37	4.52	3.66
Ω_2	3.05	3.47	3.41	2.98	2.88	2.52
Ω_3	6.14	6.05	5.96	5.95	5.71	5.88
Ω_4	5.62	4.48	5.27	4.96	5.16	4.91

综上所述，本文采取两级式优选策略所提时域介电最优特征空间 Ω_Z 不仅能提高绝缘诊断的准确率和诊断方法的适应性，还能够缩短诊断时间，保证特征量类别的多样性，这对绝缘老化诊断尤为重要。

另外，优化的特征空间的特征量具有普遍适应性，意味着进行油纸绝缘老化诊断时，可以只考虑优化空间的特征量，而无需提取所有的相关特征量，避免了特征量提取过程大量复杂的计算，从而提高了绝缘老化诊断的效率。

4.3 影响因素分析

1) 样本数据容量

前文对老化特征量的提取及计算工作是基于

Ω_Z 的特征量包含了 3 种类型，且不含类区分度较低的特征，其表现出的诊断结果差异性很小； Ω_2 仅有两种特征种类，其诊断准确率最低，诊断结果差异性最大。这说明，特征空间特征量类型的多样性是影响诊断准确率和诊断方法适应性的重要因素。因此，对绝缘老化状态的诊断尽可能保证特征空间中特征量类型的多样化。

根据表 10 的诊断时间结果，在所有特征空间中，采用最优特征空间 Ω_Z 进行诊断的诊断时间最短，诊断效率大幅提高。这是由于最优特征空间 Ω_Z 对特征维度的有效把控，故特征样本的训练时间大大提高；反观 Ω_{Z0} — Ω_4 ，特征量数目的增多进一步导致样本复杂性增加，同时某些样本数据中存在离群点，也会影响诊断计算时间。

80 余组的变压器数据完成的。但随着研究的进一步深入，变压器样本数据在不断地更新扩充，老化特征量的数目也在持续扩张，对特征量的提取将建立在更庞大的数据规模上。数据量的增加会减小随机因素带来的影响，能够进一步提升诊断的准确率，但同时会导致样本的训练速率减缓。

2) 变压器类型

由于各类别特征量蕴含的弛豫信息不同，对变压器类型的反馈程度也不同。对于 RVM 类型的特征量来说是通过实测曲线得到，适用于大部分类型的电力变压器；而某些 PDC 类型的特征量是根据对调压器等绝缘设备进行油纸绝缘老化实验得到，故

对部分特定型号的变压器的反应更为灵敏。因此, 变压器类型的差别也会影响对绝缘老化诊断的效果。

4.4 实例验证

为了验证本文所提的两级式优选方法在实际变压器绝缘老化诊断的有效性, 先采用一台型号为 SZG-31500/110 的实际油浸式变压器, 标号为 T1, 于 2004 年生产, 目前已运行 17 年。经过介电响应试验并计算出其最优特征空间特征量。

采用 T1 的特征空间应用于不同的诊断方法以实现 T1 的绝缘老化诊断。鉴于篇幅所限, 本文只列出 XGBoost 和 SVM+XGBoost 两种诊断方法的诊断结果, 如表 11 所示。

表 11 特征空间 Ω_{T1} 的诊断准确率及诊断时间

Table 11 Diagnostic accuracy and time of characteristic space Ω_{T1}

特征空间	诊断结果			
	XGBoost		SVM+XGBoost	
	准确率/%	诊断时间/s	准确率/%	诊断时间/s
Ω_{T1}	95.32	2.56	98.18	1.78

从表 11 中可以看出, T1 的最优特征空间的诊断准确率均在 95% 以上, 且诊断时间较短。这说明采用本文所提优化算法所确定的最优特征空间在实际变压器绝缘状态诊断应用中也具有很高的诊断准确率和诊断效率, 在工程上具有实际应用价值。

本文所提的两级式时域特征优选策略能够有效提升老化诊断的准确率。但由于对特征量冗余性、相关性以及重要性的考察是两级独立的, 这会导致有时候会重复考虑或者遗漏某些弛豫信息。部分特征量的老化联系被切断, 可能会对特征量的择优选择造成影响。因此, 在后续考察老化特征量的性能时尽可能多地保留其内在联系, 在全面评估的同时涵盖更丰富的弛豫信息, 优化特征空间结构。

5 结论

1) 本文提出一种基于快速过滤相关算法和极限梯度上升相结合的两级式的时域特征优选方法, 对低冗余、强相关的特征量进行定向筛选, 通过主成分分析对特征空间阈值维度进行选取, 实现高维特征空间的降维。

2) 依据统计指标对所提最优特征空间进行验证, 结果表明最优特征空间的平均准确率明显高于其余对照特征空间, 区间宽度、最大误差以及方差几项适应性指标也几乎是所有结果中最小的, 有效论证了根据所提优选方法得到的空间集适用于各种诊断算法。

本文所提的特征量优选策略可为评估变压器油纸绝缘状态过程中的特征空间的选择依据提供新思路。但由于本文变压器样本数量有限, 在后续的研究中仍可继续扩充数据库以达到更理想的效果。

参考文献

- [1] HAN J, KONG X, LI P, et al. A novel low voltage ride through strategy for cascaded power electronic transformer[J]. Protection and Control of Modern Power Systems, 2019, 4(3): 227-238.
- [2] 党建, 魏慧, 贾嵘, 等. 基于定量修正层次分析的变压器性能评估方法研究[J]. 高压电器, 2020, 56(7): 133-141, 151.
DANG Jian, WEI Hui, JIA Rong, et al. Research on transformer performance evaluation based on quantitative correction analytic hierarchy process[J]. High Voltage Apparatus, 2020, 56(7): 133-141, 151.
- [3] 熊一, 廖晓红, 柯方超, 等. 基于多体系数据融合的主变压器全寿命成本分析[J]. 电力科学与技术学报, 2020, 35(6): 3-11.
XIONG Yi, LIAO Xiaohong, KE Fangchao, et al. Life cycle cost analysis of main transformer based on the multi-system data fusion[J]. Journal of Electric Power Science and Technology, 2020, 35(6): 3-11.
- [4] 董明, 王丽, 吴雪舟, 等. 油纸绝缘介电响应检测技术研究现状与发展[J]. 高电压技术, 2016, 42(4): 1179-1189.
DONG Ming, WANG Li, WU Xuezhou, et al. Status and progress in study of dielectric response technology for oil-paper insulation[J]. High Voltage Engineering, 2016, 42(4): 1179-1189.
- [5] 陈明星, 高波, 胡广才, 等. 基于时域介电法的变压器油纸绝缘老化特征参量分析[J]. 高压电器, 2020, 56(2): 108-113, 121.
CHEN Mingxing, GAO Bo, HU Guangcai, et al. Aging characteristic parameters analysis of transformer oil-paper insulation based on time domain dielectric method[J]. High Voltage Apparatus, 2020, 56(2): 108-113, 121.
- [6] 李良书, 万康鸿, 王琼苑, 等. 二硫化钼传感器对于绝缘介质特征分解气体的敏感特性研究[J]. 高压电器, 2021, 57(10): 36-43.
LI Liangshu, WAN Kanghong, WANG Qiongyuan, et al. Study on sensitivity characteristic of decomposed gas of MoS₂ sensor with insulating medium feature[J]. High Voltage Apparatus, 2021, 57(10): 36-43.
- [7] BOGNAR A, CSEPES G, KALOCSAI L, et al. Spectrum of polarization phenomena of long time-constant as a diagnostic method of oil-paper insulating systems[C] // Proceedings of the 3rd International Conference on

- Properties and Applications of Dielectric Materials, July 8-12, 1991, Tokyo, Japan.
- [8] 邹阳, 蔡金锭. 变压器极化谱特征量与绝缘状态关系研究[J]. 仪器仪表学报, 2015, 36(3): 608-614.
ZOU Yang, CAI Jinding. Study on the relationship between polarization spectrum characteristic quantity and insulation condition of oil-paper transformer[J]. Chinese Journal of Scientific Instrument, 2015, 36(3): 608-614.
- [9] 杨丽君, 廖瑞金, 孙会刚, 等. 变压器油纸绝缘热老化特性及特征量研究[J]. 电工技术学报, 2009, 24(8): 27-33.
YANG Lijun, LIAO Ruijin, SUN Huigang, et al. Investigation on properties and characteristics of oil-paper insulation in transformer during thermal degradation process[J]. Transactions of China Electrotechnical Society, 2009, 24(8): 27-33.
- [10] 黄云程, 蔡金锭. 应用去极化能量谱评估变压器绝缘老化受潮状态[J]. 电工电能新技术, 2016, 35(1): 53-59.
HUANG Yuncheng, CAI Jinding. Diagnosis of insulation condition of transformers by using depolarization energy spectrum[J]. Advanced Technology of Electrical Engineering and Energy, 2016, 35(1): 53-59.
- [11] 吴广宁, 宋臻杰, 杨飞豹, 等. 基于时域介电谱和去极化电量的变压器油纸绝缘老化特征量研究[J]. 高电压技术, 2017, 43(1): 195-202.
WU Guangning, SONG Zhenjie, YANG Feibao, et al. Study on aging characteristics of transformer oil-paper insulation based on the time domain dielectric spectroscopy and depolarization charge quantity[J]. High Voltage Engineering, 2017, 43(1): 195-202.
- [12] 刘庆珍, 张晓燕, 蔡金锭. 油纸绝缘弛豫法谱线特征量提取及老化诊断[J]. 电机与控制学报, 2020, 24(5): 124-134.
LIU Qingzhen, ZHANG Xiaoyan, CAI Jinding. Relaxation spectrum characteristic parameters extraction and aging diagnosis for oil-paper insulation[J]. Electric Machines and Control, 2020, 24(5): 124-134.
- [13] 高浩, 刘庆珍, 蔡金锭. 基于去极化电流 Prony 拟合的油纸绝缘 Debye 参数辨识方法[J]. 高压电器, 2020, 56(11): 210-218.
GAO Hao, LIU Qingzhen, CAI Jinding. Debye parameter identification method for oil-paper insulation based on depolarization current Prony algorithm fitting[J]. High Voltage Apparatus, 2020, 56(11): 210-218.
- [14] TANG Mingjie, LEI Min, XU Hao, et al. Frequency domain characteristics and insulation condition evaluation of power transformer oil-paper insulation[C] // 2012 IEEE International Conference on High Voltage Engineering and Application, September 17-20, 2012, Shanghai, China: 443-446.
- [15] 刘庆珍, 张晓燕, 蔡金锭. 基于多元模糊联系度模型的变压器油纸绝缘老化评价[J]. 高压电器, 2020, 56(5): 47-54, 61.
LIU Qingzhen, ZHANG Xiaoyan, CAI Jinding. Aging state evaluation of oil-paper insulation transformer based on multivariate fuzzy relation degree model[J]. High Voltage Apparatus, 2020, 56(5): 47-54, 61.
- [16] 林智勇, 张达敏, 郑运鸿, 等. 混联等效电路 p 参数诊断变压器油纸绝缘老化研究[J]. 电力系统保护与控制, 2020, 48(15): 164-169.
LIN Zhiyong, ZHANG Damin, ZHENG Yunhong, et al. Investigation on the p parameter of hybrid equivalent circuit to diagnose transformer oil-paper insulation aging[J]. Power System Protection and Control, 2020, 48(15): 164-169.
- [17] 黄云程, 蔡金锭. 油纸绝缘系统回复电压函数建模及拓扑结构辨识[J]. 电力自动化设备, 2016, 36(3): 149-153.
HUANG Yuncheng, CAI Jinding. Return voltage function modeling and topology recognition of oil-paper insulation system[J]. Electric Power Automation Equipment, 2016, 36(3): 149-153.
- [18] 李春茂, 陈子宣, 杨雁, 等. 基于修正 Cole-Cole 模型的氧化锌压敏电阻老化状态评估[J]. 高压电器, 2020, 56(2): 85-92.
LI Chunmao, CHEN Zixuan, YANG Yan, et al. Aging state evaluation of ZnO varistor based on Cole-Cole modified model[J]. High Voltage Apparatus, 2020, 56(2): 85-92.
- [19] 陈汉城, 蔡金锭. 基于多时域特征参量的变压器油纸绝缘状态综合评估[J]. 电力自动化设备, 2017, 37(7): 184-190.
CHEN Hancheng, CAI Jinding. Synthetic insulation state evaluation based on multiple time-domain characteristic parameters for transformer oil-paper[J]. Electric Power Automation Equipment, 2017, 37(7): 184-190.
- [20] 张毅涛, 齐波, 林元棣, 等. 变压器油纸绝缘套管受潮缺陷频域介电谱特征[J]. 电力工程技术, 2021, 40(2): 135-140.
ZHANG Yitao, QI Bo, LIN Yuandi, et al. FDS characteristic of damp-affected oil-paper insulation transformer bushings[J]. Electric Power Engineering Technology, 2021, 40(2): 135-140.
- [21] 张晓燕, 刘庆珍, 蔡金锭. 基于降维技术与 K-means 聚类的油纸绝缘状态综合灰评估[J]. 电力系统保护与控制, 2019, 47(8): 62-70.

- ZHANG Xiaoyan, LIU Qingzhen, CAI Jinding. Comprehensive grey evaluation for oil-paper insulation based on dimension reduction techniques and K-means cluster[J]. Power System Protection and Control, 2019, 47(8): 62-70.
- [22] XIE Yiming, RUAN Jiangjun. Parameters identification and application of equivalent circuit at low frequency of oil-paper insulation in transformer[J]. IEEE Access, 2020, 8: 86651-86658.
- [23] 张晨萌, 谢施君, 谭思文, 等. 基于极化去极化电流法的电力电容器绝缘状态检测方法研究[J]. 高压电器, 2020, 56(6): 296-301.
- ZHANG Chenmeng, XIE Shijun, TAN Siwen, et al. Research on power capacitor insulation state detection based on polarization depolarization current method[J]. High Voltage Apparatus, 2020, 56(6): 296-301.
- [24] 谭贵生, 曹生现, 赵波, 等. 基于关联规则与变权重系数的变压器状态综合评估方法[J]. 电力系统保护与控制, 2020, 48(1): 88-95.
- TAN Guisheng, CAO Shengxian, ZHAO Bo, et al. An assessment of power transformers based on association rules and variable weight coefficients[J]. Power System Protection and Control, 2020, 48(1): 88-95.
- [25] 董治强. 基于 DTW-FCBF-LSTM 模型的超短期风速预测[J]. 电测与仪表, 2020, 57(4): 93-98.
- DONG Zhiqiang. Ultra-short term wind speed prediction based on DTW-FCBF-LSTM model[J]. Electrical Measurement & Instrumentation, 2020, 57(4): 93-98.
- [26] 唐圳雄, 唐东林, 丁超, 等. 基于 XGBoost 特征重要度的储油罐缺陷 ANN 面积量化模型[J]. 电子测量与仪器学报, 2020, 34(8): 109-115.
- TANG Zhenxiong, TANG Donglin, DING Chao, et al. Quantitative model of ANN area of tank defects based on XGBoost feature importance[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(8): 109-115.
- [27] 陈明华, 刘群英, 张家枢, 等. 基于 XGBoost 的电力系统暂态稳定预测方法[J]. 电网技术, 2020, 44(3): 1026-1033.
- CHEN Minghua, LIU Qunying, ZHANG Jiashu, et al. XGBoost-based algorithm for post-fault transient stability status prediction[J]. Power System Technology, 2020, 44(3): 1026-1033.
- [28] PORTNOY I, MELENDEZ K, PINZON H, et al. An improved weighted recursive PCA algorithm for adaptive fault detection[J]. Control Engineering Practice, 2016, 50(5): 69-79.
- [29] 蔡金锭, 叶荣, 陈汉城. 回复电压多元参数回归分析的油纸绝缘老化诊断方法[J]. 电工技术学报, 2018, 33(21): 5080-5089.
- CAI Jinding, YE Rong, CHEN Hancheng. Aging diagnosis method of oil-paper insulation based on multiple parameter regression analysis of recovery voltage[J]. Transactions of China Electrotechnical Society, 2018, 33(21): 5080-5089.
- [30] 蔡金锭, 陈汉城. 基于样本集的变压器油纸绝缘状态区间灰靶分类及老化诊断[J]. 高电压技术, 2018, 44(8): 2486-2492.
- CAI Jinding, CHEN Hancheng. Grey target classification and aging diagnosis of transformer oil-paper insulation state interval based on sample set[J]. High Voltage Engineering, 2018, 44(8): 2486-2492.
- [31] 张宇, 蔡金锭. 基于层次分析和逼近理想解法的绝缘状态评估[J]. 仪器仪表学报, 2018, 39(11): 35-42.
- ZHANG Ning, CAI Jinding. Evaluation of insulation state based on the combination of analytical hierarchy process and TOPSIS[J]. Chinese Journal of Scientific Instrument, 2018, 39(11): 35-42.
- [32] 邓威, 郭钊秀, 李勇, 等. 基于特征选择和 Stacking 集成学习的配电网网损预测[J]. 电力系统保护与控制, 2020, 48(15): 108-115.
- DENG Wei, GUO Yixiu, LI Yong, et al. Power losses prediction based on feature selection and Stacking integrated learning[J]. Power System Protection and Control, 2020, 48(15): 108-115.
- [33] 朱文广, 李雪映, 杨为群, 等. 基于 K-折交叉验证和 Stacking 融合的短期负荷预测[J]. 电力科学与技术学报, 2021, 36(1): 87-95.
- ZHU Wenguang, LI Xueying, YANG Weiqun, et al. Short-term load forecasting based on the K-fold cross-validation and Stacking ensemble[J]. Journal of Electric Power Science and Technology, 2021, 36(1): 87-95.

收稿日期: 2021-10-18; 修回日期: 2021-12-29

作者简介:

刘庆珍(1971—), 女, 博士, 副教授, 主要从事电力设备故障诊断与系统优化方面的研究; E-mail: lqz515@126.com

黄昌硕(1997—), 男, 硕士研究生, 研究方向为电气绝缘老化设备诊断。E-mail: 605584342@qq.com

(编辑 周金梅)