

DOI: 10.12158/j.2096-3203.2020.02.020

基于最大信息系数的变压器过热故障特征选择

陈如意¹, 江军¹, 陈珉², 冯沛琪¹, 李晨², 张潮海¹

(1. 南京航空航天大学自动化学院, 江苏 南京 211106;

2. 国网浙江省电力有限公司电力科学研究院, 浙江 杭州 310014)

摘要:改良三比值法只关注若干个气体浓度比值, 信息利用不充分, 而且气体浓度的随机误差对故障诊断结果有影响, 因此文中将变压器故障特征气体扩充为单种气体增长率、多种气体比值和相对浓度等 62 个故障特征, 通过具有稳健性的最大信息系数提取与变压器故障状态相关度高的故障特征。同时为了避免筛选特征之间的冗余性, 采用相关系数筛选冗余性小的特征组合, 并采用距离相关系数和多种分类器进行检验。结果表明与油中溶解气体相比, 筛选特征集合与变压器过热故障状态联系更加紧密, 且针对过热故障类型的诊断精度, 筛选特征集合在样本类别不平衡时性能更优, 突破了单一分类器性能上限。

关键词:最大信息系数; Pearson 系数; 特征选择; 油中溶解气体; 故障诊断

中图分类号: TM407

文献标志码: A

文章编号: 2096-3203(2020)02-0140-06

0 引言

油浸式变压器是电力系统重要的组成部件, 其安全问题已成为影响电网安全运行的重要因素, 而变压器故障率是评估系统可靠性的重要指标, 能为制定调度运行、检修计划、电网规划提供重要依据^[1]。由于绝缘油在异常工况下容易分解为多种烃类气体, 利用油中溶解气体(dissolved gas analysis, DGA)分析变压器内部状态成为油浸式变压器故障诊断的一种普遍方法。

与传统的 Duval 三角法和 IEC 三比值法相比, 目前的研究主要集中在专家系统^[2-3]、模糊理论^[4-6]、支持向量机(support vector machine, SVM)^[7-11]、神经网络等方面, 提高故障诊断模型的准确性。Mileta 等人分析了 Mamdani 模型和 Sugeno 模型在模糊专家系统中的应用, 有助于提高电力变压器维护决策的正确性和及时性^[2]; 神经网络在变压器故障诊断中的应用存在瓶颈, 为了克服人工神经网络对异常值不敏感、收敛速度慢的问题, 一些学者提出了贝叶斯网络和深度信念网络^[12-13], 针对网络训练中的干扰或噪声, 其他学者提出了模糊 C 均值聚类算法和遗传算法优化反向传播(back propagation, BP)网络的初始参数^[14-15], 与 BP 网络和经典集成学习相比, 明显提高了故障诊断的准确性。

无论是传统比值法还是基于人工智能的模型构建, 其中模型输入都是 IEC 比值法推荐的气体浓

度或气体比值, 且部分直接应用监测值的模型普遍存在诊断准确性较低的问题; 多数诊断模型采用规程中的三比值法或其他编码方式进行映射, 该映射过程带有较强人为性, 在过程中容易丢失有效信息, 尚有提升空间^[16]。而且由于变压器运行环境的复杂性, 变压器负荷状况和在线监测设备的系统误差会影响到被测变压器的油色谱数据, 同时也影响与变压器状态相关性高的映射特征选择。因此, 为了避免随机误差的影响, 需采用稳健的方法选择与变压器故障相关的映射特征来构建模型^[17-20]。

文中介绍了具有稳健性的最大信息系数(maximal information coefficient, MIC)的原理和变压器特征选择的流程。按照该流程, 首先将变压器故障特征气体扩展为 62 种特征, 采用 MIC^[21]来筛选与变压器状态相关性高的映射特征。针对筛选特征之间可能存在的冗余问题, 采用线性相关系数将筛选特征划分为 4 个特征集合。为了验证筛选特征集合和故障特征气体集合对模型诊断变压器状态的影响, 文中采用多种分类器模型分别输入不同的特征集合进行训练, 观察分类器模型精度和受试者工作特征(receiver operating characteristics, ROC)曲线的变化。

1 特征选择

1.1 MIC

MIC 又称最大互信息系数, 常用于衡量 2 个变量之间的线性或非线性关系。MIC 在拥有足够的样本时, 可以捕获广泛的关系, 而不限定于特定的函数类型, 符合普适性和公正性的要求。MIC 的计算

收稿日期: 2019-09-13; 修回日期: 2019-10-11

基金项目: 国家自然科学基金资助项目(51807088); 江苏省自然科学基金资助项目(BK20170786)

基于信息论的互信息概念,而互信息是度量 2 个事件集合之间的相关性,定义了已知两事件集合中的一个时,对另外一个事件集合不确定度的减少量。2 个随机离散变量的互信息可以定义为:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \lg \frac{p(x,y)}{p(x)p(y)} \quad (1)$$

式中: $p(x,y)$ 为 X 和 Y 的联合概率密度函数; $p(x), p(y)$ 分别为 X 和 Y 的边缘概率密度函数。在连续随机变量的情形下,互信息公式中求和被替换成二重积分。

计算基于互信息概念的 MIC 时,让具有 2 个属性的数据点的集合分布在两维的空间中,使用 $m \times n$ 的网格划分数据空间,使落在第 (x,y) 格子中的数据点的频率作为 $P(x,y)$ 的估计,即:

$$P(x,y) = \frac{\text{第}(x,y)\text{个网络中的数据点数}}{\text{总的的数据点数}} \quad (2)$$

$P(x)$ 的估计为落在第 x 行的数据点的频率,同理获得 $P(y)$ 的估计,然后计算随机变量 X, Y 的互信息。因为 $m \times n$ 的网格划分数据点的方式不止一种,所以需获得使互信息最大的网格划分。使用归一化因子,将互信息的数值转化为 $(0,1)$ 区间内。最后,找到能使归一化互信息最大的网格分辨率,作为 MIC 的度量值。在计算过程中,网格的分辨率限制为 $m \times n < B$, $B = f(\text{data_size}) = n^{0.6}$, 因此, MIC 的计算公式为:

$$c_{\text{MIC}} = \frac{\max \left(\frac{\sum_{x \in X, y \in Y} P(x,y) \lg \frac{P(x,y)}{\sum_{x \in X} P(x,y) \sum_{y \in Y} P(x,y)}}{\lg \min(x,y)} \right)}{\lg \min(x,y)} \quad (3)$$

与其他相关性分析方法相比, MIC 能够捕获各种线性和非线性关系,因此,采用 MIC 能有效捕获 62 组故障特征与变压器故障之间的相关性。

1.2 线性相关分析

MIC 量化了 62 种故障特征包含变压器故障的信息量,而 MIC 大的多个特征可能线性相关,因此采用相关系数来评估 62 种故障特征之间的线性相关性。当选择原始数据集中信息系数最高的一种故障特征后,根据阈值将 62 种故障特征中与之线性相关性高的故障特征归为一组特征集合,直到原始数据集中 62 种故障特征完全被排除或选择,确保选择的特征集合之间冗余性较小。若直接删除相关性高的指标,可能会导致原始数据集的信息损失,因此保留相关性高的特征为一组特征集合,极大地保留了不同特征的信息,特征选择流程见图 1。

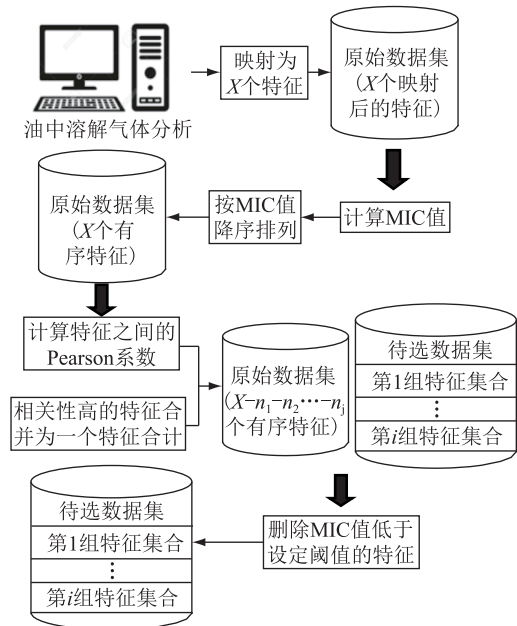


图 1 特征选择流程

Fig.1 The flow chart of features selection

2 基于 MIC 的特征选择

本次研究共使用了 2 部分的数据,第一部分数据取自浙江电网提供的中高温过热故障变压器的带电检测数据,共 14 台 899 条数据;第二部分数据取自浙江电网提供的正常变压器带电检测数据,选取了 24 台共 531 条数据。首先将变压器故障特征气体映射为单种气体浓度、单种气体浓度的月标准增长率、2 种气体比值和多种气体相对含量共 62 个指标,如表 1 所示。

为避免特征选择时选择冗余的气体指标,需考虑 62 组气体指标之间的线性关系,避免选择线性关系强的 2 组气体指标。根据检验,62 组气体指标的相关系数如图 2 所示

为了保证 MIC 结果反映气体特征与变压器故障之间的线性和非线性关系,在计算 62 组气体特征的 MIC 时,为了与 MIC 形成对比,同时计算气体特征与变压器故障的距离相关系数, MIC 最大的前 15 组气体指标的距离相关系数和 MIC 值如表 2 所示。

距离相关系数是针对 Pearson 系数只反映变量之间的简单线性关系而提出的方法,被广泛应用于衡量变量之间线性和非线性关系。MIC 数值较大的气体指标其距离相关系数数值也较大,说明 MIC 在一定程度上反映了线性和非线性关系。

采用图 1 所示的特征选择方法,最终得到特征集合,如表 3 所示。

从特征选择方法筛选的特征集合中可以得到

表 1 特征映射扩展的 62 个指标

Table 1 62 indexes of feature mapping extension

项目	指标
单种气体浓度	H ₂ , C1, C26, C24, C22, CO, CO ₂ , 总烃
单种气体浓度月标准增长率	H ₂ /150, C1/120, C26/65, C24/50, C22/1, 总烃/150, CO/350, CO ₂ /2500
2 种气体比值	C22/C24, C1/H ₂ , C24/C26, C22/C1, C22/C26, C1/C24, C1/C26, C22/H ₂ , C24/H ₂ , C26/H ₂ , CO/CO ₂ , C1/(C1+C22+C24), C24/(C1+C22+C24), C22/(C1+C22+C24), C26/(C26+C24+C22), C24/(C26+C24+C22), C22/(C26+C24+C22), C1/(C1+C26+C22), C26/(C1+C26+C22), C1/(C1+C26+C24), C26/(C1+C26+C24), C24/(C1+C26+C24), H ₂ /(H ₂ +C1+C26), C1/(H ₂ +C1+C26), C26/(H ₂ +C1+C26), H ₂ /(H ₂ +C26+C24), C26/(H ₂ +C26+C24), C24/(H ₂ +C26+C24), H ₂ /(H ₂ +C1+C24), C1/(H ₂ +C1+C24), C24/(H ₂ +C1+C24), H ₂ /(H ₂ +C1+C22), C1/(H ₂ +C1+C22), C22/(H ₂ +C1+C22), H ₂ /(H ₂ +C26+C22), C26/(H ₂ +C26+C22), C22/(H ₂ +C26+C22), H ₂ /(H ₂ +C24+C22), C24/(H ₂ +C24+C22), C22/(H ₂ +C24+C22), H ₂ /R, C1/R, C26/R, C24/R, C22/R

注: C1 表示 CH₄, C26 表示 C₂H₆, C24 表示 C₂H₄, C22 表示 C₂H₂, R 表示 H₂+C1+C26+C24+C22。下同。

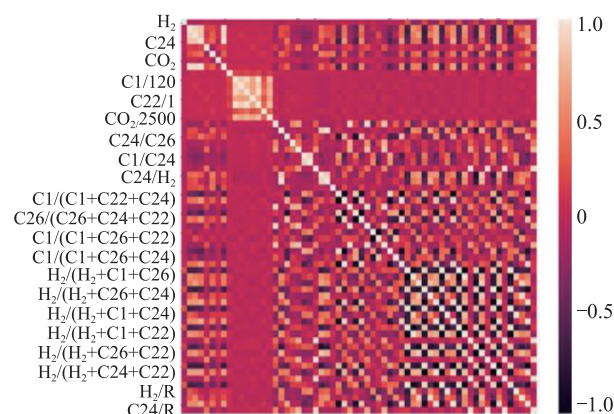


图 2 62 组气体指标之间的相关系数热力图

Fig.2 Heatmap of Pearson coefficient between 62 features

4 × 6 × 5 × 5 = 600 种特征组合, 由于特征集合是一个多维特征, MIC 无法衡量多维数据与变压器故障状态之间的关系, 因此采用距离相关系数来衡量不同特征组合与变压器故障之间的相关性。同时加入故障特征气体组成的特征集合的距离相关系数来进行比较, 采用前 4 组距离相关系数最大的特征集合和故障特征气体的距离相关系数, 如表 4 所示。

由表 4 可知, 故障特征气体(H₂, C1, C26, C24, C22)所构成的特征集合的距离相关系数结果小于前 4 组特征集合, 说明筛选特征组成的特征集合与变压器状态的关联更加紧密, 按照最优的距离相关

表 2 前 12 组气体指标的距离相关系数和 MIC

Table 2 Distance correlation coefficient and maximum information coefficient of the first 12 features

气体指标	距离相关系数	c _{MIC}
总烃	0.851	0.914
C1	0.867	0.904
C24	0.817	0.891
C26	0.792	0.799
C24/R	0.675	0.618
C24/(H ₂ +C26+C24)	0.669	0.608
C24/(H ₂ +C1+C24)	0.645	0.608
H ₂ /(H ₂ +C1+C24)	0.662	0.561
H ₂ /R	0.653	0.561
H ₂ /(H ₂ +C24+C22)	0.675	0.560
C24/H ₂	0.616	0.554
C24/(H ₂ +C24+C22)	0.675	0.551

表 3 特征选择方法筛选的特征集合

Table 3 Features sets selected by maximum information coefficient and linear correlation analysis

特征集合	气体指标
第一组	C1, C26, C24, 总烃
第二组	C24/R, C24/(C1+C26+C24), C24/(H ₂ +C26+C24), C24/(H ₂ +C1+C24), C24/(H ₂ +C24+C22), C24/(C1+C22+C24)
第三组	H ₂ /(H ₂ +C1+C24), H ₂ /(H ₂ +C1+C26), H ₂ /(H ₂ +C26+C24), H ₂ /R, H ₂ /(H ₂ +C1+C22)
第四组	C24/H ₂ , C1/H ₂ , C1/(H ₂ +C1+C22), C24/50, C26/(H ₂ +C26+C22)

表 4 前 4 组特征集合与故障特征

气体组合的距离相关系数对比

Table 4 The comparison result of first four feature combinations with the largest distance correlation coefficient and key gases set

特征集合	距离相关系数
C1, C24/R, H ₂ /(H ₂ +C1+C22), C24/H ₂	0.795 275
C1, C24/(H ₂ +C1+C24), H ₂ /(H ₂ +C1+C22), C24/H ₂	0.795 274
C1, C24/R, H ₂ /(H ₂ +C1+C26), C24/H ₂	0.795 273
C1, C24/(H ₂ +C1+C24), H ₂ /(H ₂ +C1+C26), C24/H ₂	0.795 273
H ₂ , C1, C26, C24, C22	0.692 024

系数结果筛选的特征集合为 { C1, C24/R, C24/H₂, H₂/(H₂+C1+C22) }。

3 特征对模型性能影响对比

为了验证筛选特征集合与故障特征气体组成的特征集合对模型诊断变压器状态的影响, 采用浙

江电网提供的 14 台中高温过热故障变压器和 24 台健康变压器共 1 430 条带电检测数据。由于原始数据集的故障类型均为过热故障,因此通过筛选特征和故障特征气体特征对模型诊断变压器健康和故障 2 个状态的影响,来验证筛选特征集合与变压器状态的紧密关联。

文中采用随机梯度下降(stochastic gradient descent,SGD)分类器,线性判别分析(linear discriminant analysis,LDA),二次判别分析(quadratic discriminant analysis,QDA),SVM 和随机森林(RandomForest)来检验筛选特征集合和故障特征气体集合对分类器诊断变压器状态的精度影响,同时为了观察训练集中样本类别比例对分类器性能影响,将训练集和测试集以 7:3 和 3:1 这 2 种比例按照随机抽样划分和分层抽样方法进行划分,观察不同特征集合下的数据类别比例,训练集和测试集的比例对分类器性能的影响。

文中同时从精度和 ROC 曲线 2 个指标衡量筛选特征集合和故障特征气体集合对分类器性能的影响。在不同的训练集和测试集数据比例和抽样方法下,每个分类器模型在不同特征集合下的精度如图 3 所示。

由图 3 可知,当训练集和测试集比例为 3:1 时,输入为筛选特征的 RandomForest 分类器精度最高,达到了 0.97 以上,反映筛选特征与故障特征气体相

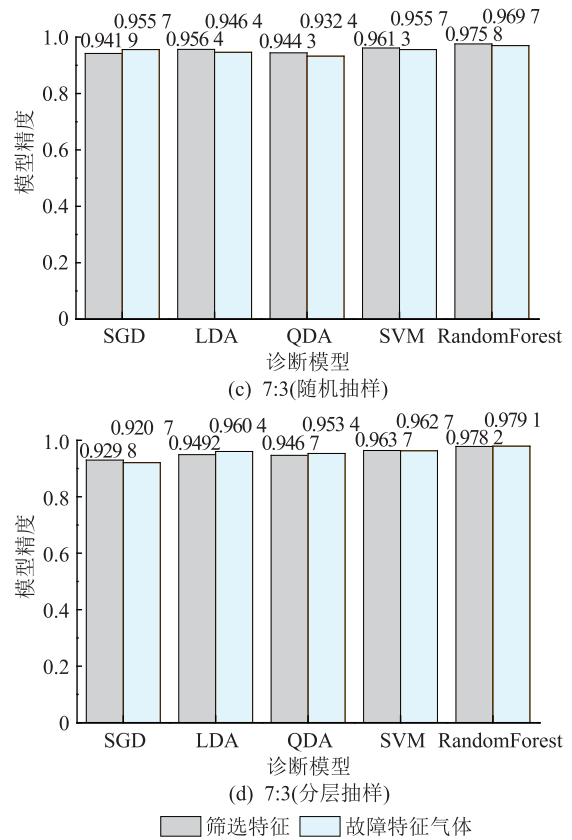
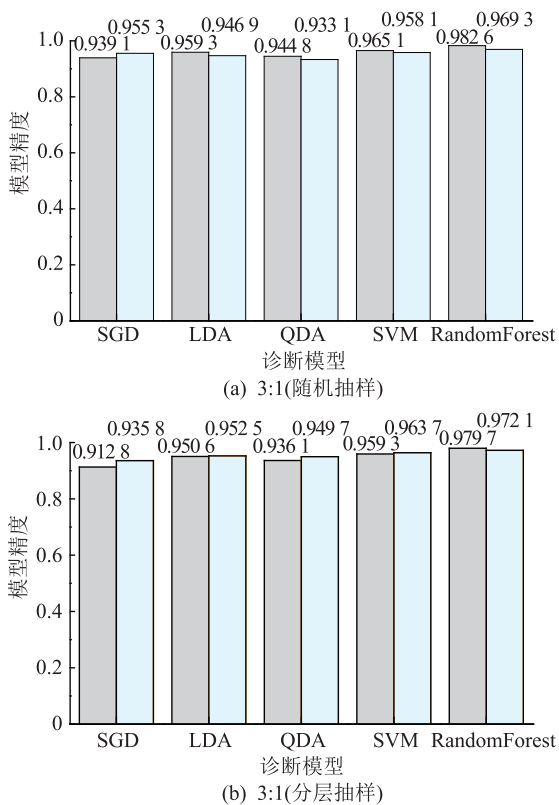


图 3 不同数据比例和抽样方法下各模型的精度
Fig.3 The accuracy of models under different data ratios and sampling methods

比突破了分类器性能的上限。此外,虽然分层抽样划分训练集和测试集时,筛选特征在部分分类器上性能表现不如故障特征气体,但是抽样方法为随机抽样时,训练集和测试集中数据类别的比例可能出现明显差别,即变压器健康样本数量可能远远高于故障样本,输入筛选特征的分类器其性能表现更加优异。这反映了筛选特征在原始数据集样本类别不平衡时能包含更多的变压器状态信息,从而让训练后的模型性能提升。

为了直观对比不同特征集合下分类器性能表现最优的情况,采用 ROC 曲线只管衡量分类器性能。由图 3 可知,筛选特征集合在 3:1(随机抽样)条件下 RandomForest 分类器精度在所有输入筛选特征的分类器中最优,而故障特征气体集合在 7:3(分层抽样)的条件下 RandomForest 分类器精度在所有输入故障特征气体的分类器中最优。不同特征集合在模型精度均达到最优时,ROC 曲线如图 4 所示。

图 4 中筛选特征的 ROC 曲线包围了故障特征气体的 ROC 曲线,说明输入筛选特征的模型性能表现优于输入故障特征气体的模型,突破了其性能上限。

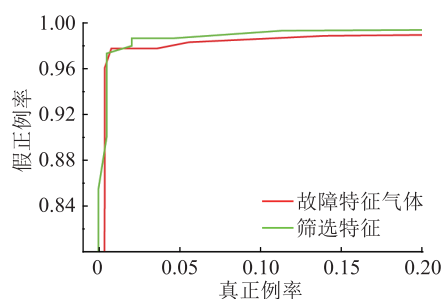


图4 不同特征集合的 ROC 曲线

Fig.4 The ROC curve of different feature sets

4 结论

针对现有模型只利用 IEC 比值法、Duval 三角法等提供的故障特征气体和固定类型的气体比值,文中提出了基于 MIC 和 Pearson 系数的特征选择方法,构建了从特征扩展、特征相关性分析、冗余性分析到特征提取的完整方案。实例分析表明,相比于故障特征气体构成的特征集合,文中筛选的特征集合与变压器过热故障状态之间的关联更加紧密。具体结论如下:

(1) 采用具有稳健性的 MIC 来衡量 62 种特征中与变压器状态相关性高的特征,有助于减小气体浓度的误差对特征选择结果的影响。

(2) 筛选特征在原始数据集样本类别不平衡时能包含更多的变压器过热故障和健康状态信息,从而让训练后的模型性能提升。

(3) 在确保采用合适训练集和测试集比例及抽样方法时,模型性能达到最优,筛选特征突破了模型对过热故障类型分类性能的上限。

本文得到国网浙江省电力有限公司科技项目(5211DS17001C)资助,谨此致谢!

参考文献:

[1] 陆云才,胡汉巧,蔚超,等. 基于超声波法的变压器重症监护系统研制及应用[J]. 电力工程技术,2017,36(2):94-98.
 LU Yuncai, HU Hanqiao, WEI Chao, et al. Development and application of transformer intensive care system based on ultrasonic method[J]. Electric Power Engineering Technology, 2017, 36(2):94-98.

[2] ŽARKOVI M, STOJKOVI Z. Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics[J]. Electric Power Systems Research, 2017, 149: 125-136.

[3] 师瑞峰,史永锋,牟军,等. 油中溶解气体电力变压器故障诊断专家系统[J]. 电力系统及其自动化学报,2014,26(12): 49-54.
 SHI Ruifeng, SHI Yongfeng, MU Jun, et al. Power transformer fault diagnosis expert system with dissolved gas analysis in oil

[J]. Proceedings of the CSU-EPSS, 2014, 26(12):49-54.

[4] 谷凯凯,郭江. 紧致融合模糊集和故障树的变压器故障诊断[J]. 高电压技术,2014,40(5):1507-1513.
 GU Kaikai, GUO Jiang. Transformer fault diagnosis method based on compact fusion of fuzzy set and fault tree[J]. High Voltage Engineering, 2014, 40(5):1507-1513.

[5] 司马莉萍,舒乃秋,左婧,等. 基于灰关联和模糊支持向量机的变压器油中溶解气体浓度的预测[J]. 电力系统保护与控制,2012,40(19):41-46.
 SIMA Lipin, SHU Naiqiu, ZUO Jing, et al. Concentration prediction of dissolved gases in transformer oil based on grey relational analysis and fuzzy support vector machines[J]. Power System Protection and Control, 2012, 40(19):41-46.

[6] 李恩文,王力农,宋斌,等. 基于改进模糊聚类算法的变压器油色谱分析[J]. 电工技术学报,2018,33(19):4594-4602.
 LI Enwen, WANG Linong, SONG Bin, et al. Analysis of transformer oil chromatography based on improved fuzzy clustering algorithm[J]. Transactions of China Electrotechnical Society, 2018, 33(19):4594-4602.

[7] 李春茂,周妹末,刘亚婕,等. 基于邻域粗糙集与多核支持向量机的变压器多级故障诊断[J]. 高电压技术,2018,44(11):3474-3482.
 LI Chunmao, ZHOU Momo, LIU Yajie, et al. Multi-level fault diagnosis of transformer based on neighborhood rough set and multiple kernel support vector machine[J]. High Voltage Engineering, 2018, 44(11):3474-3482.

[8] 陈欢,彭辉,舒乃秋,等. 基于蝙蝠算法优化最小二乘双支持向量机的变压器故障诊断[J]. 高电压技术,2018,44(11): 3664-3671.
 CHEN Huan, PENG Hui, SHU Naiqiu, et al. Fault diagnosis of transformer based on LS-TSVM optimized by bat algorithm[J]. High Voltage Engineering, 2018, 44(11):3664-3671.

[9] 吴广宁,袁海满,宋臻杰,等. 基于粗糙集与多类支持向量机的电力变压器故障诊断[J]. 高电压技术,2017,43(11): 3668-3674.
 WU Guangning, YUAN Haiman, SONG Zhenjie, et al. Fault diagnosis for power transformer based on rough set and multi-class support vector machine[J]. High Voltage Engineering, 2017, 43(11):3668-3674.

[10] 肖怀硕,李清泉,施亚林,等. 灰色理论-变分模态分解和 NSGA-II 优化的支持向量机在变压器油中气体预测中的应用[J]. 中国电机工程学报,2017,37(12):3643-3653.
 XIAO Huaihuo, LI Qingquan, SHI Yalin, et al. Prediction of dissolved gases in oil for transformer based on grey theory-variational mode decomposition and support vector machine improved by NSGA-II [J]. Proceedings of the CSEE, 2017, 37(12):3643-3653.

[11] QI B, WANG Y, ZHANG P, et al. A novel deep recurrent belief network model for trend prediction of transformer DGA data[J]. IEEE Access, 2019(7):80069-80078.

[12] 王德文,雷倩. 基于贝叶斯正则化深度信念网络的电力变压器故障诊断方法[J]. 电力自动化设备,2018,38(5):

- 129-135.
WANG Dewen, LEI Qian. Fault diagnosis of power transformer based on BR-DBN[J]. Electric Power Automation Equipment, 2018, 38(5):129-135.
- [13] 代杰杰, 宋辉, 杨祎, 等. 基于深度信念网络的变压器油中溶解气体浓度预测方法[J]. 电网技术, 2017, 41(8): 2737-2742.
DAI Jiejie, SONG Hui, YANG Yi, et al. Concentration prediction of dissolved gases in transformer oil based on deep belief networks[J]. Power System Technology, 2017, 41(8): 2737-2742.
- [14] BHALLA D, BANSAL R K, GUPTA H O. Function analysis based rule extraction from artificial neural networks for transformer incipient fault diagnosis[J]. International Journal of Electrical Power & Energy Systems, 2012, 43(1):1196-1203.
- [15] ZHANG K, YUAN F, GUO J, et al. A novel neural network approach to transformer fault diagnosis based on momentum-embedded BP neural network optimized by genetic algorithm and fuzzy c-means[J]. Arabian Journal for Science and Engineering, 2016, 41(9):3451-3461.
- [16] 郑一鸣, 何文林, 孙翔, 等. 基于油色谱超立方映射的电力变压器缺陷援例诊断模型[J]. 电力工程技术, 2017, 36(4):48-53.
ZHENG Yiming, HE Wenlin, SUN Xiang, et al. Case based power transformer defects diagnose model using hypercube mapping of oil chromatography[J]. Electric Power Engineering Technology, 2017, 36(4):48-53.
- [17] 郑元兵, 孙才新, 李剑, 等. 变压器故障特征量可信度的关联规则分析[J]. 高电压技术, 2012, 38(1):82-88.
ZHENG Yuanbing, SUN Caixin, LI Jian, et al. Association rule analysis on confidence of features for transformer fault[J]. High Voltage Engineering, 2012, 38(1):82-88.
- [18] 廖瑞金, 邓小聘, 杨丽君, 等. 油纸绝缘热老化特征参量的多元统计分析[J]. 高电压技术, 2010, 36(11):2621-2628.
LIAO Ruijin, DENG Xiaopin, YANG Lijun, et al. Multivariate statistical analysis of thermal aging characteristic parameters of oil-paper insulation[J]. High Voltage Engineering, 2010, 36(11):2621-2628.
- [19] 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.
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.
- [20] 吴广宁, 袁海满, 高波, 等. 基于特征评估与核主元分析的电力变压器故障诊断[J]. 高电压技术, 2017, 43(8):2533-2540.
WU Guangning, YUAN Haiman, GAO Bo, et al. Fault diagnosis of power transformer based on feature evaluation and kernel principal component analysis[J]. High Voltage Engineering, 2017, 43(8):2533-2540.
- [21] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large data sets[J]. Science, 2011, 334(6062):1518-1524.

作者简介:



陈如意

陈如意(1996),男,硕士在读,研究方向为电气设备状态检测与故障诊断(E-mail: cry508069@163.com);

江军(1988),男,博士,副教授,研究方向为电气设备状态监测与故障诊断;

陈珉(1989),男,硕士,工程师,从事电气设备状态监测与带电检测工作。

Feature selection of dissolved gases in power transformer based on maximal information coefficient

CHEN Ruyi¹, JIANG Jun¹, CHEN Min², FENG Ruiqi¹, LI Chen², ZHANG Chaohai¹

(1. School of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

2. State Grid Zhejiang Electric Power Co., Ltd. Research Institute, Hangzhou 310014, China)

Abstract: The key gas method and IEC method for dissolved gases analysis only focus several gas index that may be affected by random error and relative percentages, and maybe it is not perfect to be used to conduct preventive maintenance and inspection of transformers. To solve the problem, key gases are mapped into 62 features including traditional combinations of key gases and other new features. The features highly correlated to transformer status are selected by maximum information coefficient, to escape random error of gas concentration. To avoid redundancy between selected features, Pearson coefficient is used to filter features to reduce the redundancy of the selected features. The result showed that correlation coefficient between the selected features and the transformer fault is relatively high. And the result of distance correlation coefficient also shows that the feature set formed by the selected features is more closely related to the transformer overheat fault state than that of the traditional fault characteristic gas.

Keywords: maximal information coefficient; Pearson correlation coefficient; feature selection; dissolved gas analysis; fault diagnosis

(编辑 方晶)