

基于最大相关最小冗余准则的变压器故障 诊断特征选择

辜超¹, 杨祎¹, 张晓星², 金淼², 周思远²

(1. 国网山东省电力公司电力科学研究院, 山东 济南 250002;
2. 武汉大学电气工程学院, 湖北 武汉 430072)

摘要: 电力变压器油中溶解气体分析(DGA)技术广泛应用于变压器内典型故障诊断,其中基于DGA数据的人工智能诊断方法在变压器故障诊断领域具有较高的识别率,但该类方法在选择故障特征量时尚无统一的标准。鉴于此,本文尝试引入最大相关最小冗余算法(mRMR),以互信息理论为基础挖掘变压器故障特征量之间以及特征量与故障类型之间的关联关系,通过分析大量的DGA在线监测数据挖掘出最优的变压器故障特征量集,并采用支持向量机(SVM)分类器对比优选特征量集和传统的特征量集合在变压器故障诊断的效率。最后,通过与SVM智能分类、IEC推荐的三比值分类方法的对比测试表明该方案的故障诊断准确率优于传统的故障诊断方案,故障识别效率高于新型的人工智能诊断方案,更适合于现场的工程应用及推广。

关键词: 电力变压器; 故障诊断; 溶解气体分析; 最大相关最小冗余

DOI: 10.12067/ATEEE1709056 **文章编号:** 1003-3076(2018)07-0084-06 **中图分类号:** TM41

1 引言

电力变压器作为输变电系统中最为核心的设备之一,其故障诊断与状态评估技术的研究一直是电力行业关注的焦点^[1]。油中溶解气体分析(Dissolved Gas Analysis, DGA)是目前应用最广泛的电力变压器运行状态诊断技术^[2]。

以山东省电网为例,几乎所有110kV电压等级以上的电力变压器都配有DGA监测系统,利用检测到的DGA气体数据与IEC推荐的改良三比值法,工作人员能有效诊断变压器故障,然而该方法在现场应用中存在编码不全或者编码过界等缺陷。为此,国内外研究人员开发了以人工神经网络(Artificial Neural Network, ANN)^[3]、支持向量机(Support Vector Machine, SVM)^[4]、模糊理论(FCM)^[5]、灰色关联度^[6]、聚类算法^[7]等智能方法为基础的故障诊断方案,为变压器故障诊断提供新的技术支持。文献[8]提出的遗传支持向量机与灰色人工免疫算法相

结合的变压器故障诊断方案,关注智能分类器的训练效率,引入 k 近邻综合决策法提高故障识别率。文献[9]提出的改进人工鱼群优化粗糙集算法对故障样本输入特征维度进行约简,有效地简化了变压器的诊断工作。

然而,目前变压器智能诊断方法的输入故障特征量普遍采用 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 等主要特征气体及部分比值,在选择特征量中并没有充足的依据,因此,如何选择与故障类型更加紧密的故障特征量是目前变压器故障智能诊断中忽略的一个问题。由Peng Hanchuan^[10]提出的最大相关最小冗余准则,基于互信息理论,通过挖掘特征变量之间的关联关系,获取与目标类别相关度最大的特征参量集合,成为了备受关注的经典有效的选择特征方法,已在人工智能算法^[11]、医学疾病诊断^[12]、生物工程^[13]、电力工程^[14,15]等应用领域得到了广泛认可。

本文针对变压器内DGA数据(H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2),构建完备的变压器故障特征比值全

收稿日期: 2017-09-22

基金项目: 国家高技术研究发展计划(863计划)项目(2015AA050402)

作者简介: 辜超(1971-),男,四川籍,高级工程师,研究方向为输变电设备状态评估;

杨祎(1986-),女,山东籍,工程师,硕士,研究方向为输变电设备状态监测和状态评估。

集,引入最大相关最小冗余算法对故障特征全集进行数据挖掘分析,优选出与故障类型相关度最大的故障特征作为输入达到原始数据中的故障特征信息的最大化表达,并采用SVM分类器对优选故障特征的故障诊断效果进行测试,确定最优的故障特征,最后通过不同的故障分类方法对比验证本文提出方法的有效性。

2 最大相关最小冗余(mRMR)算法模型

2.1 最大相关最小冗余基本原理

最大相关最小冗余准则以互信息^[11]为衡量变量之间相关度的基础标准,通过最大相关最小冗余的准则进行特征优选。

(1) 互信息理论

互信息是mRMR算法中相关度与冗余度的计算基础,其理论定义如下:

给定的两个随机变量 x 和 y ,它们的概率密度与联合概率密度定为 $p(x)$ 、 $p(y)$ 、 $p(x, y)$,则两个变量之间的互信息 I 的公式定义为^[11]:

$$I(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (1)$$

(2) 最大相关原则

该准则要求所选的特征集与目标类别有最大的依赖性,其衡量式为^[11]:

$$\max D(S, c), D = \frac{1}{|S|} \sum_{f_i \in S} I(f_i, c) \quad (2)$$

式中, S 为特征子集; $|S|$ 为特征量的个数; f_i 为第 i 个特征; c 为目标类别。

(3) 最小冗余准则

该准则要求所选的特征集中各特征间的相关性最小,其衡量标准用特征间的互信息最小化来定义^[11]:

$$\min R(S), R = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i, f_j) \quad (3)$$

2.2 mRMR 算法搜索规则

mRMR算法结合最大相关和最小冗余准则,分别定义了算子 Φ_1 :信息差(Mutual Information Difference, MID)和算子 Φ_2 :信息熵(Mutual Information Quotient, MIQ),如式(4)所示,用以指导最优特征子集的选择^[12]。

$$\begin{cases} \max \Phi_1(S, c), \Phi_1 = D - R \\ \max \Phi_2(S, c), \Phi_2 = \frac{D}{R} \end{cases} \quad (4)$$

该算法的搜索方法为:基于最大相关最小冗余

原则实现最优特征集 S_m 的选取,现假定已经获取了 $m-1$ 个特征所组成的特征集 S_{m-1} ,则第 m 个特征可通过式(5)中的算子来进行搜索^[12]:

$$\begin{cases} \max \nabla_{MID}, \nabla_{MID} = \max \left\{ I(f_j, c) - \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I(f_j, f_i) \right\} \\ \max \nabla_{MIQ}, \nabla_{MIQ} = \max \left\{ I(f_j, c) / \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I(f_j, f_i) \right\} \end{cases} \quad (5)$$

式中, f_j 为原始特征集中不包含 S_{m-1} 中特征量的其他特征。

3 基于mRMR的变压器故障特征量优选

3.1 变压器内故障类型

变压器内的故障类型主要为过热故障和放电故障,故障出现后变压器内的绝缘油和绝缘纸会分解出与故障类型相关的故障特征气体,如目前检测到的 N_2 、 O_2 、 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 、 C_2H_6 、 CO 、 CO_2 等气体,这些故障气体与故障类型之间存在着密切的关系,通过分析油中溶解气体可以辨识出变压器内发生的故障。

本文故障诊断研究主要针对目前变压器内存在的五种典型故障类型,故障类型及其相关说明如表1所示。

表1 油浸式变压器内典型故障类型

Tab.1 Typical fault types in oil-immersed transformer		
类型	典型故障类型	故障描述
T1	中低温过热 (低于 700℃)	变压器过负荷运行或者油流阻塞等原因引起的中低温过热
T2	高温过热 (高于 700℃)	变压器铁心内硅钢片发生短路等原因引起的高温过热
T3	低能放电	连接不良形成的不同电位或悬浮电位等原因引起的诸如变压器套管与箱壁等部件之间的低能放电
T4	高能放电	变压器套管与箱体、变压器内线圈短路等原因引起的沿面放电、电弧等高能放电
T5	局部放电	变压器内浸渍不完全、变压器纸湿度过高等原因引起的局部放电

3.2 故障特征集构造

以变压器内主要溶解气体 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 为典型的故障特征气体,参照IEC三比值法的故障特征构造规则,本文提出的各种可能故障特征比值如表2所示,其中总烃为 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 的含量之和。

表 2 故障特征比值

Tab.2 Ratio of characteristic gas

编号	特征量	编号	特征量
<i>F1</i>	C_2H_2/H_2	<i>F9</i>	C_2H_6/CH_4
<i>F2</i>	C_2H_2/CH_4	<i>F10</i>	CH_4/H_2
<i>F3</i>	C_2H_2/C_2H_6	<i>F11</i>	$H_2/\text{总烃}$
<i>F4</i>	C_2H_2/C_2H_4	<i>F12</i>	$CH_4/\text{总烃}$
<i>F5</i>	C_2H_4/H_2	<i>F13</i>	$C_2H_6/\text{总烃}$
<i>F6</i>	C_2H_4/CH_4	<i>F14</i>	$C_2H_4/\text{总烃}$
<i>F7</i>	C_2H_4/C_2H_6	<i>F15</i>	$C_2H_2/\text{总烃}$
<i>F8</i>	C_2H_6/H_2	<i>F16</i>	$(CH_4 + C_2H_6)/\text{总烃}$

表 3 油浸式变压器内样本数据

Tab.3 DGA sample data

数据集	<i>T1</i>	<i>T2</i>	<i>T3</i>	<i>T4</i>	<i>T5</i>	合计
mRMR 样本	93	64	63	71	61	352
训练样本	62	42	42	47	41	234
测试样本	31	22	21	24	20	118

(1) 样本数据归一化处理

为了去除计算得到的特征参数物理单位干扰,仅从量值上来分析,将所有特征比值的数据归一化到[0,1]之间,本文采取了以下的归一化预处理:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad i = 1, 2, \dots \quad (6)$$

式中, x_i 与 x_i^* 分别为原始数据与归一化数据; x_{\min} 与 x_{\max} 分别为输入特征数据的最小值与最大值。

(2) 特征数据离散化处理

本文提取的特征量为连续型变量,而基于互信息理论的 mRMR 算法建立是以离散变量的分析为基础的,因此本文通过对归一化的特征数据进行等间隔划分归类处理获取对应的离散变量类型^[15,16]。

以特征量 *F1*、*F2* 数据为例,其中特征量 *F1* 或 *F2* 的数据总个数为 *N*,将特征量 *F1*、*F2* 的归一化数据按照数值大小升序排列,将大小范围在 [0,1] 的归一化特征数据等分为 *N_a* 和 *N_b* 个子区间,分别计算 *F1* 中落入第 *i* 个区间的数据个数 *n_a*、*F2* 中落入第 *j* 个区间的个数 *n_b*。然后计算各故障类型中 *F1* 数据第 *i* 个子区间与 *F2* 数据第 *j* 个子区间的数据 *n_{ab}(i, j)*,式(1)可变为离散化形式:

$$I(F1, F2) = \sum_{i=1}^{N_a} \sum_{j=1}^{N_b} \frac{n_{ab}(i, j)}{N} \log \frac{N^2 n_{ab}(i, j)}{n_a(i) n_b(j)} \quad (7)$$

此外,每个数据对应的故障类别分别用数据标签[1,2,3,4,5]标注,为离散化数据,无需处理。

3.3 DGA 数据样本

本文选取国网山东省电力公司电力科学研究院 220kV 等级同一型号的变压器 2010 ~ 2015 年的油色谱数据,从中遴选出各故障样本数据 352 条,其中所有数据样本作为 mRMR 算法的输入进行特征优选,智能分类器的训练样本和测试样本如表 3 所示。

3.4 最优变压器故障特征集搜索步骤

基于 mRMR 算法中的基本理论,本文提出的最优特征选择步骤如下:

(1) 基于原始的 DGA 故障特征气体数据,采用表 2 构建原始的故障特征全集 *F*,并确定故障数据对应的表 1 中的故障类型 *T*。

(2) 假定最优的故障特征集合的个数为 *n*,*n* 小于全集 *F* 中故障特征量总数 *m*。

(3) 计算原始特征集 *F* 中所有特征量的互信息 *I(f, T)*,选取互信息最大的 *f_x*,令最优特征集为 *S = {f_x}*,需要计算的故障特征集变为 *F = F - S*。

(4) 遍历特征集 *F* 中的所有特征量,分别找到满足 MID、MIQ 两种规则下的特征 *f_x*,此时最优特征集更新为 *S = S ∪ {f_x}*,需要计算的原始故障特征集更新后,减为 *F = F - {f_x}*。

(5) 此时,如果最优特征集 *S* 中的故障特征个数为 *n*,则停止搜索,依照搜索出的顺序,依次输入 MID 规则下和 MIQ 规则下的最优特征集。如果 *S* 中的故障特征个数小于 *n*,则继续进行步骤(3),直到最优特征集 *S* 中的故障特征个数为 *n*。

故障类型识别方法如下:选用合适的分类器,将得到的最优特征集作为输入,分别选定合适的训练集和测试集,获取故障准确率;选定故障准确率最高的特征集的个数为最终选定的最优故障特征集。选用不同的诊断方法验证本方案的有效性。

支持向量机是目前模式识别领域应用最为成熟的算法之一,其基本原理是在数据组成的特征空间中,考虑寻找一个超平面,使得不同类别的训练样本正好位于超平面的两侧^[17,18]。本文直接选定 SVM 作为变压器故障诊断分类器。

4 算例分析

4.1 基于 mRMR 的故障特征优选

将原始油色谱故障数据处理为 16 维特征比值量,构建故障特征集的原始数据为 352 × 16 的矩阵数据,以其作为 mRMR 输入,并且故障数据对应的故障类型(*T1*, *T2*, *T3*, *T4*, *T5*)构成了标签向量矩阵。

假定最优特征集的个数定为 16,利用 MID 和 MIQ 搜索规则获取的原始特征集中所有故障特征量优劣排序如表 4 所示。其中 NO 代表表 2 中 $F1 \sim F16$ 的原始故障特征量排列顺序。由表 4 可知,MIQ 与 MID 两种搜索规则下的特征量优劣排序差别不大,几乎一致;并且,最优故障特征量排序中的前几个特征量中都包括 4、7、10,即 IEC 导则推荐的三个特征比值量 C_2H_2/C_2H_4 、 C_2H_4/C_2H_6 、 CH_4/H_2 ,该结果说明数据分析的结果与传统的变压器故障诊断理论中故障特征比值的选取保持一致。

表 4 MID 和 MIQ 搜索规则获取的故障特征结果

Tab.4 Feature selection results based on MID and MIQ search rule

搜索规则	特征量优劣排序
NO	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16
MIQ	15, 7, 11, 4, 10, 16, 6, 14, 13, 1, 3, 12, 9, 5, 2, 8
MID	15, 7, 11, 4, 10, 6, 16, 14, 1, 13, 3, 12, 9, 8, 5, 2

4.2 基于 SVM 的故障分类测试

将表 4 中计算得到的 MID、MIQ 和 NO 特征排序逐一增加特征,并以表 3 选定的训练集与测试集样本作为 SVM 分类器的输入进行故障分类器的测试。特征集个数 5~16 时所得到的变压器的故障分类准确率如图 1 所示。

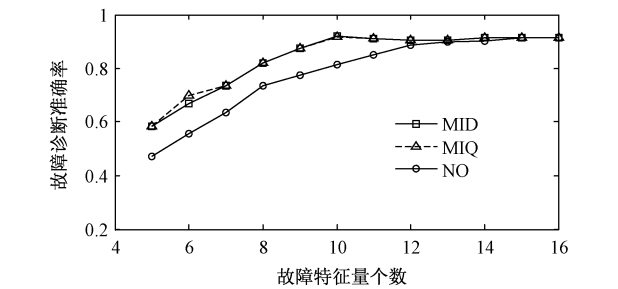


图 1 不同特征集下的故障分类准确率

Fig.1 Fault classification accuracy under different feature set

由图 1 可知,依照 MID、MIQ 与 NO 三种特征排序逐次增加故障特征量,变压器的故障诊断率提高,并且在故障特征量为 9 个时,故障诊断率达到最高。之后,随着特征量个数增加,故障诊断率略微下降,最终维持在一个恒定的诊断率,这说明该特征全集里存在着冗余的特征,甚至某些特征对于故障诊断存在副作用,该结果表明了特征优选的必要性。

此外,优选特征的选择输入使输入数据量小,降

低了 SVM 分类器所需的训练时间和分类时间,从而提高了故障诊断的效率。

4.3 不同故障分类方法效果对比

本文选取了三比值法、BPNN 算法、SVM 算法三种方法进行变压器故障诊断对比测试。其中三比值法选取 C_2H_2/C_2H_4 、 C_2H_4/C_2H_6 、 CH_4/H_2 IEC 导则推荐的三个特征比值量,BPNN 算法与 SVM 算法则用 mRMR 特征优选的故障特征集作为输入。选用同样的数据进行测试,测试结果如表 5 所示。

表 5 各故障诊断方法诊断率对比

Tab.5 Fault recognition accuracy comparison of different methods

故障类型	故障诊断率(%)		
	三比值法	BPNN	SVM
T1	76.53	87.66	89.66
T2	80.26	89.71	94.12
T3	77.17	86.66	92.27
T4	74.65	85.20	90.27
T5	81.24	84.11	89.21
平均	77.97	86.68	91.11

对比三种诊断方法的诊断率可知,本文提出的特征优选与智能诊断方法具有最佳的诊断效率,该方法克服了传统三比值法的故障诊断失效;并且利用 BPNN 与 SVM 智能分类器对本文优选的 9 个特征量构成的特征集分别进行训练和识别测试,两个分类器的识别时间分别为全特征量测试集识别时间的 56% 和 64%,这证明了本文方案的高诊断效率。

此外,同一优选特征集下,不同智能分类方法都有较高的故障诊断率,因此相较于选择更加高效的智能分类器,本文提出的变压器故障特征量优选的思路,从原始的故障样本数据中挖掘更多的故障特征信息,显著提升了故障诊断效果。

5 结论

本文引入最大相关最小冗余准则开展变压器故障诊断特征量的优选,通过对目前积累的大量变压器油中溶解气体数据的分析,获取了最能表征故障类型的特征优选集;并通过 SVM 智能分类方法进行训练与测试,对比了不同故障特征集的诊断效果,最后用现场监测的故障数据进行特征优选与故障诊断效率测试,得出如下结论:

- (1)最大相关最小冗余算法计算到的最优故障特征量集中较少的优选特征量集就能达到较好的故障诊断效果,该方案降低了分类器的计算成本。

(2)相较于传统的三比值法和人工智能诊断方法,本文提出的特征优选与智能诊断有更稳定和更高效的故障诊断率。在DGA数据广泛积累的今天,该方案具有更好的实用性。

参考文献 (References):

- [1] 郭创新, 朱承治, 张琳, 等 (Guo Chuangxin, Zhu Chengzhi, Zhang Lin, et al.). 应用多分类多核学习支持向量机的变压器故障诊断方法 (A fault diagnosis method for power transformer based on multiclass multiple-kernel learning support vector machine) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2010, 30 (13): 128-134.
- [2] 董明, 孟源源, 徐长响, 等 (Dong Ming, Meng Yuanyuan, Xu Changxiang, et al.). 基于支持向量机及油中溶解气体分析的大型电力变压器故障诊断模型研究 (Fault diagnosis model for power transformer based on support vector machine and dissolved gas analysis) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2003, 23 (7): 88-92.
- [3] 王晓霞, 王涛 (Wang Xiaoxia, Wang Tao). 基于粒子群优化神经网络的变压器故障诊断 (Power transformer fault diagnosis based on neural network evolved by particle swarm optimization) [J]. 高电压技术 (High Voltage Engineering), 2008, 34 (11): 2362-2367.
- [4] Bacha K, Souahlia S, Gossa M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine [J]. Electric Power Systems Research, 2012, 83 (1): 73-79.
- [5] Huang Y C, Sun H C. Dissolved gas analysis of mineral oil for power transformer fault diagnosis using fuzzy logic [J]. IEEE Transactions on Dielectrics & Electrical Insulation, 2013, 20 (3): 974-981.
- [6] 赵峰, 李硕 (Zhao Feng, Li Shuo). 基于熵权和灰关联度的变压器故障诊断 (Fault diagnosis in transformer based on entropy weighted and grey incidence) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2015, 34 (1): 57-61.
- [7] 宋人杰, 刘瑞英, 王林 (Song Renjie, Liu Ruiying, Wang Lin). 灰色定权聚类 and 变权模式在变压器状态评估中的应用研究 (Application of grey fixed weight clustering and variable weight model in transformer condition evaluation) [J]. 电工电能新技术 (Advanced Technology of Electrical Engineering and Energy), 2017, 36 (3): 75-80.
- [8] 郑蕊蕊, 赵继印, 赵婷婷, 等 (Zheng Ruirui, Zhao Jiyin, Zhao Tingting, et al.). 基于遗传支持向量机和灰色人工免疫算法的电力变压器故障诊断 (Power transformer fault diagnosis based on genetic support vector machine and gray artificial immune algorithm) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2011, 31 (7): 56-63.
- [9] 陈小青, 刘觉民, 黄英伟, 等 (Chen Xiaoqing, Liu Juemin, Huang Yingwei, et al.). 采用改进人工鱼群优化粗糙集算法的变压器故障诊断 (Transformer fault diagnosis using improved artificial fish swarm with rough set algorithm) [J]. 高电压技术 (High Voltage Engineering), 2012, 38 (6): 1403-1409.
- [10] Peng Hanchuan, Long Fuhui, Ding Chris. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27 (8): 1226-1238.
- [11] Mandal M, Mukhopadhyay A. An improved minimum redundancy maximum relevance approach for feature selection in gene expression data [J]. Procedia Technology, 2013, 10 (1): 20-27.
- [12] Ahmad F, Isa N A M, Hussain Z, et al. A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer [J]. Pattern Analysis and Applications, 2015, 18 (4): 861-870.
- [13] Xu Y, Ding Y X, Ding J, et al. Mal-Lys: Prediction of lysine malonylation sites in proteins integrated sequence-based features with mRMR feature selection [J]. Scientific Reports, 2016, 6: 38318.
- [14] 李扬, 顾雪平 (Li Yang, Gu Xueping). 基于改进最大相关最小冗余判据的暂态稳定评估特征选择 (Feature selection for transient stability assessment based on improved maximal relevance and minimal redundancy criterion) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2013, 33 (34): 179-186.
- [15] 赵永宁, 叶林 (Zhao Yongning, Ye Lin). 区域风电场短期风电功率预测的最大相关-最小冗余数值天气预报特征选取策略 (A numerical weather prediction feature selection approach based on minimal redundancy maximal relevance strategy for short-term regional wind power prediction) [J]. 中国电机工程学报 (Proceedings of the CSEE), 2015, 35 (23): 5985-5994.
- [16] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information [J]. Physical Review E, 2004, 69 (6): 066138.
- [17] Hirschm Hirschmuller. Stereo processing by semiglobal matching and mutual information [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30 (2): 328-341.

[18] 张凯, 黄华平, 杨海涛, 等 (Zhang Kai, Huang Huaping, Yang Haitao, et al.). 利用改进遗传算法与 LS-SVM 进行变压器故障诊断 (A transformer fault diagnosis method integrating improved genetic algorithm with least square support vector machine) [J]. 电网技术 (Power System Technology), 2010, 34 (2): 164-168.

Feature selection for transformer fault diagnosis based on maximal relevance and minimal redundancy criterion

GU Chao¹, YANG Yi¹, ZHANG Xiao-xing², JIN Miao², ZHOU Si-yuan²

- (1. Electric Power Research Institute of Shandong Power Supply Company of State Grid, Ji'nan 250002, China;
2. School of Electrical Engineering, Wuhan University, Wuhan 430072, China)

Abstract: The power transformer oil dissolved gas analysis (DGA) is widely applied in the transformer fault diagnosis, and the artificial intelligent diagnosis method based on DGA has high data recognition rate in the field of transformer fault diagnosis. However this intelligent method has no unified standard in transformer fault selection. This paper tries to introduce the maximal relevance and minimal redundancy (mRMR) algorithm on the basis of the principle of mutual information. Based on the DGA on-line monitoring data, the mRMR get the optimal features set through analyzing the relationship between the features and the relationship between the features and the fault types. The SVM classifier is employed to compare the fault diagnostic effect with original feature set and the optimized feature set. Finally, compared with the SVM classification and the IEC recommended three ratio classification methods, the proposed fault diagnosis accuracy is superior to the traditional fault diagnosis method and recognition speed is faster than the intelligent diagnosis method, and this method is more suitable for engineering application on site.

Key words: power transformer; fault diagnosis; dissolved gas analysis (DGA); maximal relevance and minimal redundancy (mRMR)