

混沌支持向量机在变压器局部放电检测中的应用

李继胜^{1,2}, 陈立彬³, 赵学风¹, 孙振权¹, 罗勇芬¹, 李彦明¹

(1.西安交通大学电气工程学院, 陕西 西安 710049; 2.陕西师范大学物理学及信息技术学院, 陕西 西安 710062;
3.西安陆军学院军训教研室, 陕西 西安 710108)

摘要: 针对变压器故障诊断中缺少实际典型故障样本的问题,提出了混沌支持向量机(CSVMs)变压器局部放电检测方法。该方法采用 K 均值聚类(KMC)对变压器油中 5 种特征气体样本进行预选取作为特征向量,输入到混沌优化多分类支持向量机中进行训练,建立 CSVMs 诊断模型,实现对故障样本的诊断分类。实例分析表明,KMC 算法浓缩了故障信息,有效地解决了确定模型参数时耗时巨大的问题;混沌优化较好地提高了模型的推广能力。该方法在有限样本情况下,能够达到较高的故障正判率,满足变压器故障自动诊断的目的。

关键词: 变压器; 故障诊断; K 均值聚类; 混沌优化; 支持向量机

中图分类号: TM855

文献标志码: A

文章编号: 1001-1609(2009)05-0104-03

The Application of Chaos Support Vector Machines in Transformer Partial Discharge Measurements

LI Ji-sheng^{1,2}, CHEN Li-bin³, ZHAO Xue-feng¹, SUN Zhen-quan¹, LUO Yong-fen¹, LI Yan-ming¹

(1. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 2. College of Physics & Information Technology, Shaanxi Normal University, Xi'an 710062, China;
3. Military Training Department, Xi'an Army Institutes, Xi'an 710108, China)

Abstract: Due to the lack of typical damage samples in the transformer fault diagnosis, a new method based on chaos support vector machines (CSVMs) was proposed. According to the method, the five characteristic gases dissolved in transformer oil were extracted by the K -means clustering (KMC) method as feature vectors, which were input to chaotic optimal multi-classified SVMs for training. Then the CSVMs diagnosis model was established to implement fault samples classification. The experiment shows that by adopting feature extraction with KMC, the diagnosis information is concentrated and the time-consuming in parameter determination is solved effectively. On the other hand, chaos optimization better enhanced model extension ability. Moreover, the presented method enables to detect transformer faults with a higher correct judgment rate, and can be used as an automation approach for diagnosis under condition of small samples.

Key words: transformer; fault diagnosis; K -means clustering; chaos optimization; support vector machines

0 引言

变压器是电力系统中最重要设备之一,其运行状态直接影响着整个系统的安全性。油中溶解气体分析(DGA)被认为是一种变压器故障诊断和进行绝缘寿命评估最方便、有效的手段之一。然而在变压器故障诊断的实践中,利用 DGA 数据进行故障诊断的三比值法、改良三比值法存在“缺编码”、“编码边

界过于绝对”等不足,近年来,很多人工智能方法如专家系统、神经网络、模糊理论等方法被应用于变压器绝缘诊断中,取得了一定的效果^[1-4]。

由分析变压器故障时油中溶解气体的机理可知,油中气体含量与变压器故障类型之间没有明确的函数映射关系,气体含量的分布特性也很难推测,而实际现场数据的采集精度及数量也很有限,因此基于经验风险最小化原理的传统诊断分类器得不到足够的知识学习,导致分类精度不高,诊断误差较大。

收稿日期:2009-04-25; 修回日期:2009-06-02

基金项目:国家自然科学基金项目(50877064)。

作者简介:李继胜(1977),男,博士研究生,讲师,从事局部放电检测技术研究。

支持向量机(SVMs)通过寻求结构风险最小化达到在样本量较少的情况下获得很好的学习效果。克服了过学习、欠学习、局部极小值等缺陷。由于出色的学习性能,已在很多领域得到了成功的应用。笔者将SVMs引入到基于DGA的变压器故障诊断中,采用K均值聚类算法(KMC)对故障样本进行预选,有效地提取了支持向量个数;混沌优化减小了SVMs参数选择的盲目性,提高了诊断模型的推广能力。实例分析表明,该方法在基于DGA的变压器故障诊断中具有很好的适应性。

1 相关基本理论

1.1 SVMs 二分类算法

设有样本 x_i 及其所属类别 y_i 表示为 (x_i, y_i) , $x \in R^d, y \in \{+1, -1\}, i=1 \dots n$, 超平面方程 $w \cdot x + b = 0$ 能将两类样本正确区分,并使分类间隔最大的优化问题可表示为^[5,6]

$$\min \varphi(w) = \frac{1}{2} (w \cdot w) + C \sum_{i=1}^n \xi_i \quad (1)$$

约束条件: $y_i [(w \cdot x_i) + b] \geq 1 - \xi_i$, w 为权重矩阵。式(1)中 ξ_i 为松弛因子,常数 C 起到控制对错样本惩罚程度的作用,实现在错分样本的比例核算法复杂程度之间的折中。

利用 Lagrange 优化方法把上述问题转化为其对偶问题,即

$$\max L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (2)$$

约束条件

$$\sum_{i=1}^n y_i \alpha_i = 0, 0 \leq \alpha_i < C, i=1, 2, \dots, n$$

解式(2)得到最优分类函数 $f(x)$ 为

$$f(x) = \text{sgn} \{ (w \cdot x) + b \} = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right\} \quad (3)$$

式(3)中, α^* 和 b^* 分别是 Lagrange 系数最优解和分类阈值。对于非常复杂的非线性分类,将样本空间映射到高维特征空间,使其线性可分,根据泛函的有关理论,只要核函数 $K(x_i \cdot x_j)$ 满足 Mercer 条件,它就对应某一变换空间的内积,即 $k(x_i \cdot x_j) = \phi(x_i) \cdot \phi(x_j)$, ϕ 为非线性映射函数。这样最优分类函数变为

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i K(x_i, x) + b^* \right\} \quad (4)$$

1.2 SVMs 多分类算法

变压器故障诊断属于多值分类问题。多值分类

(k 分类, $k > 2$) 是以二值分类为基础的。设有 k 类训练数据,这 k 类数据两两组合,共可构建 $M = C_m^2 = k(k-1)/2$ 个训练集,分别使用 SVM 二值分类算法对这 M 个训练集进行学习,产生 M 个分类器。当决定样本 $x \in R^d$ 所属类别时,采用投票决策法^[7],其主要思想是用所有的 $k(k-1)/2$ 个分类器对 x 进行分类,在第 m 类和第 n 类之间分类时,若该分类器判断 x 属于 m 类,则 m 类的票数加 1,否则 n 类的票数加 1。最后将归为得票数最多的那一类。分类函数为

$$f^{mn}(x) = \text{sgn} \left[\sum_{i=1}^n \alpha_i^{mn} y_i^{mn} K(x_i, x) + b^{mn} \right] \quad (5)$$

采用高斯径向基核函数,表达式为

$$K(x_i, x_j) = \exp \left(-\frac{|x_i - x_j|^2}{\sigma^2} \right) \quad (6)$$

式(6)中, σ 是核宽度参数。

1.3 基于混沌优化算法选取 SVMs 参数

惩罚因子 C 和核宽度 σ 对 SVMs 分类器诊断和泛化能力有很大的影响。只有合理选择参数,才能得到满意的诊断精度。一般最佳参数可通过交叉验证或网格搜索方法实现。前者非常耗时,后者需要人为设定参数的取值范围,识别效果随机性较大。笔者采用鲁棒性较好的混沌优化算法来实现 SVMs 最佳参数选择^[8]。

混沌运动具有遍历性、随机性、规律性等特点。由于混沌运动可在一定范围内不重复地经历所有状态,使其成为一种十分有效的搜索工具^[9]。

以故障误判率 E 来估计 SVMs 分类器的性能,作为混沌优化算法的目标函数。设惩罚因子 C 和核宽度 σ 为变量 z_1, z_2 , 则

$$\min E(z_1, z_2) \quad (7)$$

约束条件: $a_1 \leq z_1 \leq b_1, a_2 \leq z_2 \leq b_2$

由 Logistic 映射产生混沌序列

$$x(n+1) = kx(n)[1-x(n)] \quad (8)$$

约束条件: $0 \leq x(n) \leq 1, 0 \leq k \leq 4$

当 $k=4$ 时,系统处于完全混沌状态。基于混沌优化算法寻找 $((z_1, z_2))$ 最优组合的步骤如下。

(1) 初始化 $n=0, z_1^* = a_1, z_2^* = a_2, x^* = x(0) = 0.4$, 当前最优目标函数值 E^* 初始化为一个较大正数;

(2) 将 $x(n)$ 映射到优化变量 z_1 和 z_2 的取值区间:

$$z_i(n) = a_i + (b_i - a_i)x(n), i=1, 2 \quad (9)$$

(3) 优化搜索,若 $E[z_i(n)] \leq E^*$, 则 $E^* = E[z_i(n)]$, $x_i^* = x_i(n), z_i^* = z_i(n)$, 否则继续;

(4) $n=n+1, x_i(n+1) = kx_i(n)[1-x_i(n)]$;

(5) 重复步骤(2)~(4), 直到达到最大迭代次

数,则 $z_i^*=a_i+(b_i-a_i)x_i^*$ 为当前最优解。

1.4 K-均值聚类(KMC)

KMC 是 MacQueen 提出的一种非监督实时聚类算法。其基本思路^[10]为:把 n 个向量 $x_j(j=1,2,\dots,n)$ 分成 c 类 $G_i(i=1,2,\dots,c)$,并求每类的聚类中心,使得目标函数

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|x_k - c_i\|^2 \right) \quad (10)$$

达到最小。 c_i 为当第 i 类 G_i 的聚类中心。

2 基于 CSVMs 的变压器故障诊断

2.1 故障分类及特征向量选择

IEC 60599 规定,故障类型可分为 5 种^[11]:局部放电(PD)、低能放电(D_1)、高能放电(D_2)、中低温过热(T_1)、高温过热(T_2)。另外再加正常状态(NC),共有 6 种诊断类型。

用于变压器故障诊断的油中溶解特征气体主要有: H_2 、 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 。考虑到各种气体组分含量之间的巨大差异性 & 分散性,为降低其相互之间由于量值差异造成的影响,对样本数据进行归一化处理,以降低气体之间的互斥性。归一化方法为

$$x_i' = \frac{x_i}{\sum_{i=1}^5 x_i} \quad (11)$$

式(11)中, $x_i(i=1,2,\dots,5)$ 为原始的气体浓度数据。

将归一化处理后获取的样本进行 KMC 运算,提取可能是支持向量的样本作为 SVMs 多分类器输入特征向量。

2.2 诊断方法

首先,对原始样本的 5 种气体浓度进行归一化处理,减小数据的分散性;

其次,通过 KMC 提取训练样本集中可能成为支持向量的样本作为 SVMs 分类器的输入特征向量,混沌优化 SVMs 参数,建立 CSVMs 分类模型;

最后,利用已建好的 CSVMs 分类器对测试样本进行诊断分类。诊断流程见图 1。

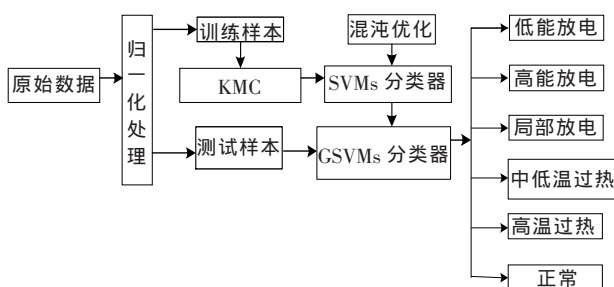


图 1 CSVMs 变压器故障诊断模型

2.3 实例分析

针对收集到的 122 组有明确结论的变压器故障数据,选择其中 84 组样本作为训练集,剩余的 38 组作为检验样本,样本分布见表 1。

表 1 变压器故障样本分布情况

故障类型	样本总数	训练样本	检验样本
NC	10	7	3
PD	15	10	5
D_1	20	12	8
D_2	22	15	7
T_1	25	20	5
T_2	30	20	10

训练样本经 KMC 算法预选取后,对应六种故障类型,错误归类的样本和类内样本点到聚类中心距离大于 0.5 的样本总和为 58 个,将这 58 个样本作为 SVMs 分类器的训练样本。在本算例中,和的取值范围均为 $[0.5, 500]$ 。以故障误判率为目标函数,混沌优化得到的 SVMs 参数值为 $C=24, \sigma=3$ 。为了比较,分别用 BP 神经网络诊断方法和 IEC 三比值法对相同的故障样本进行诊断,结果见表 2。

表 2 诊断结果比较

诊断方法	KMC+GSVMs	GSVMs	BP	三比值
检验样本正判台次	35/38	30/38	27/38	25/38
正判率/%	92	79	71	66

由表 2 诊断结果可知:

(1) 经过 KMC 算法预选取支持向量的诊断结果明显优于原始 GSVMs 分类效果,正判率由 79% 提高到 92%。

(2) 在小样本的情况下,GSVMs 较之 BP 神经网络,诊断效果有显著提高,且速度快。

以上结果表明,经过 KMC 算法的 GSVMs 分类器效果比故障样本直接进行 GSVMs 分类的效果有了较大的提高,KMC 算法预选取支持向量使其故障信息有效浓缩,解决了确定模型参数时耗时巨大的问题;混沌优化 SVMs 参数一定程度上提高了模型的推广能力。

3 结论

针对变压器故障诊断中,现场数据采集精度及数量有限,典型故障样本缺少的问题,结合 KMC 算法有效浓缩故障信息、解决确定模型参数时耗时巨大的问题和 SVMs 解决小样本分类问题的优势,提

(下转第 111 页)

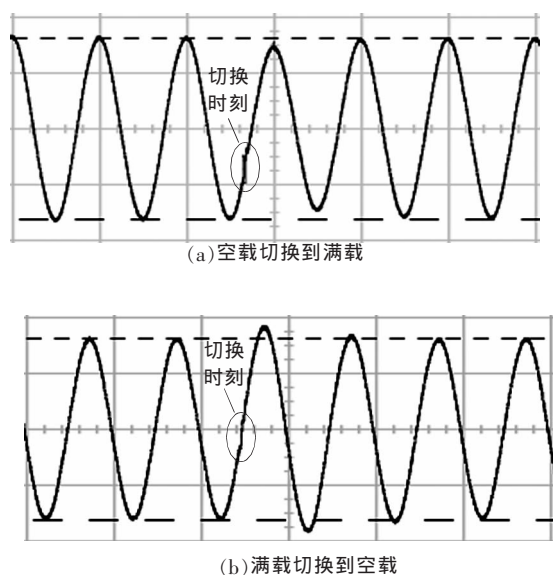


图9 20 Hz 不同条件下的输出电压波形

用于电流互感器暂态特性测试的变频测试电源工作稳定,可靠性高,波形质量好,完全能够满足工业现场对电流互感器暂态特性测试电源的具体要求。

(上接第106页)

出了混沌支持向量机 (CSVMs) 变压器故障诊断方法。该方法将变压器油中5种特征气体样本KMC聚类后作为特征向量输入到混沌优化多分类支持向量机中建立CSVMs诊断模型,实现对故障样本的诊断分类。特征气体样本经KMC聚类后使得故障信息有效浓缩,进而有效地解决了确定模型参数时耗时巨大的问题;混沌优化SVMs参数能够快速准确地在大范围找到优化值,较好地提高了模型的推广能力。通过对实际故障样本诊断,达到预期的效果,证明了该方法的有效性和实用性,为变压器故障诊断向智能化方向发展提供了一个新的途径。

参考文献:

- [1] CASTRO A R G, MIRANDA V. Knowledge Discovery in Neural Networks with Application to Transformer Failure Diagnosis[J]. IEEE Transactions on Power Systems, 2005, 20 (2): 717-724.
- [2] QIAN Su-xiang, JIAN wei-dong, HU hong-sheng, et al. Transformer Power Fault Diagnosis System Design Based on the HMM Method [C]//Proceeding of the IEEE International Conference on Automation and Logistics, 2007: 18-21.
- [3] DONG M, XU D K. Fault Diagnosis Model for Power Transformer Based on Statistical Learning Theory and Dissolved Gas Analysis [C]//IEEE International Symposium on Electrical

参考文献:

- [1] 国家技术监督局. GB 16847-1997 保护用电流互感器暂态特性技术要求[S]. 1997.
- [2] 袁季修, 盛和乐, 吴聚业. 保护用电流互感器应用指南[M]. 北京: 中国电力出版社, 2004.
- [3] 国家技术监督局. GB 1208-1997 电流互感器[S]. 1997.
- [4] 张占松, 蔡宣三. 开关电源的原理与设计[M]. 北京: 电子工业出版社, 1998.
- [5] Texas Instruments Inc. TMS320F281x Dataheet [K]. <http://www.ti.com>, 2003.
- [6] 付慧生. 复杂可编程逻辑器件与应用设计[M]. 徐州: 中国矿业大学出版社, 2003.
- [7] 王兆安, 黄俊. 电力电子技术[M]. 北京: 机械工业出版社, 2000.
- [8] 杨旭, 裴云庆, 王兆安. 开关电源技术[M]. 北京: 机械工业出版社, 2004.
- [9] 张红, 王诚梅. 电力系统常用采样方法比较[J]. 华北电力技术, 1999(4): 25-27.
- [10] 陶永华. 新型PID控制及其应用[M]. 北京: 机械工业出版社, 2002.

Insulation, 2004; 85-88.

- [4] WANG M H. Grey-extension Method for Incipient Fault Forecasting of Oil-immersed Power Transformer [J]. Electric Power Components and Systems, 2004, 32 (10): 959-975.
- [5] VAPNIK V N. The Nature of Statistical Learning Theory [M]. New York Springer-Verlag, 1995.
- [6] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 2 (1): 32-42.
- [7] HSU C W, LIN C A. Comparison of Methods for Multiclass Support Vector Machines [J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415-425.
- [8] YUAN Xiao-fang, WANG Yao-nan. Parameter Selection of Support Vector Machine for Function Approximation Based on Chaos Optimization [J]. Journal of Systems Engineering and Electronics, 2008, 19(1): 191-197.
- [9] ZHANG H D, HE Y Y. Comparative Study of Chaotic Neural Networks with Different Models of Chaotic Noise [C]//Proceedings of First International Conference on Natural Computation, 2005: 273-282.
- [10] KANUNGO T, MOUNT D M. An Efficient K-means Clustering Algorithm: Analysis and Implementation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24 (7): 812-892.
- [11] IEC 60599. Mineral oil-impregnated Electrical Equipment in Service Guide to the Interpretation of Dissolved and Free Gases Analysis[S].