

支持向量机在电力变压器故障诊断中的应用

吴晓辉¹, 刘炯¹, 梁永春¹, 汪晓明², 李彦明¹

(1. 西安交通大学电气工程学院, 710049, 西安; 2. 江西省电力公司超高压分公司, 330006, 南昌)

摘要: 利用支持向量机的学习方法, 构建了电力变压器故障诊断模型. 该模型将变压器故障分为放电性和过热性两大类, 通过统计分析寻求特征量区分故障类型, 采用支持向量机识别类内的故障类型, 利用基于交叉验证的网格搜索法来确定支持向量机的参数. 考虑到变压器油中溶解气体特征空间的紧致性原理, 利用模糊 C 均值聚类算法对所获取的样本进行预选取, 有效地解决了确定模型参数时耗时巨大的问题, 并一定程度提高了模型的推广能力. 实例验证表明, 该模型在有限样本情况下, 能达到较高的变压器故障判断率, 放电性故障样本正确判断率为 90.5%, 过热性故障样本正确判断率为 85.9%, 说明该模型具有很好的分类效果和推广能力.

关键词: 变压器; 故障诊断; 溶解气体分析; 支持向量机

中图分类号: TM855 **文献标识码:** A **文章编号:** 0253-987X(2007)06-0722-05

Application of Support Vector Machine in Transformer Fault Diagnosis

Wu Xiaohui¹, Liu Jiong¹, Liang Yongchun¹, Wang Xiaoming², Li Yanming¹

(1. School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 2. Ultra High Voltage Transmission Subcompany of Jiangxi Electric Power Corporation, Nanchang 330066, China)

Abstract: A model of transformer diagnosis based on support vector machine is proposed, where transformer faults are divided into two types, discharge fault and thermal fault, and the two fault types are recognized with their statistical features. The support vector classifier is adopted to identify fault in the two types, and the grid search method based on cross-validation is chosen to determine model parameters. Considering the compactness characteristics of dissolved gas analysis data, the achieved samples are pre-selected with the fuzzy C-means clustering method to solve the problem of long time consuming in parameter determination, thus a certain model extension ability is enhanced. The experiment shows that the model enables to detect transformer faults with a higher diagnosis rate, under condition of small samples, the diagnosis rate for discharge fault samples gets 90.5%, and 85.9% for thermal fault samples.

Keywords: transformer; fault diagnosis; dissolved gas analysis; support vector machine

电力变压器是电力系统中的重要设备, 其运行状态直接影响系统的安全性. 油中溶解气体分析(DGA)技术是目前对油浸式电力变压器进行故障诊断最方便、有效的手段, 它能有效发现变压器内部的潜伏性故障及其发展程度, 防止由此引起重大事故. 目前, 很多人工智能的方法如模糊推理、人工神经网络、聚类分析等, 被应用于电力变压器绝缘诊断

中, 并取得了较好的诊断效果^[1].

由分析变压器故障时油中溶解气体的机理可知, 单一类型的变压器故障与油中气体含量之间并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. 传统方法都是在统计学理论上发展起来的, 即当样本数量趋向于无穷大时的极限特征. 在工

程实际中,这样的前提条件往往难以满足,因此基于经验风险的最小化原理在实际应用中有很大的局限性。

本文将支持向量机(SVM)引入到变压器 DGA 故障诊断中,利用结构风险最小化原则以控制诊断中风险问题,并且用模糊均值聚类算法(FCM)对故障样本进行预选取,有效地解决了计算中的耗时问题。实例验证表明,本文采用的方法有效地解决了支持向量机应用中参数难以确定的问题,所提出的模型在变压器 DGA 中得到良好的应用。

1 支持向量机理论基础

1.1 概述

支持向量机是不同于神经网络的另外一种模式识别方法,它是在小样本学习条件下,通过选择适当的模型(最优分类面)以保持置信范围的固定,从而使经验风险最小化。它避免了学习过程中易陷入局部极小的缺点,并在提高学习一致性的同时,通过事先选择的非线性映射 Φ 将输入向量映射到一个高维特征空间 Z ,用来构造最优分类超平面。这时并不需要以显示形式来考虑特征空间,只需要能够计算特征空间中向量的点积 $Z(x) \cdot Z(y) = K(x, y)$,因此可以在对于特征空间的实际特征不清楚的条件下得出最优解,从而为问题的解决提供了便利条件^[2]。

1.2 广义最优分类面

支持向量机是从线性可分情况下的最优分类面发展而成的,其基本思想是如图 1 所示的两类线性可分问题。 H_1 、 H_2 分别为各类样本中离分类线最近、且平行于分类线的直线,保证经验风险最小,该最优分类面不但能正确地将两类分开,而且要使分类间隔最大,使推广性的界中置信范围最小,从而使问题的真实风险最小。

推广至线性不可分情况,为了描述分类超平面,并考虑到存在不能被分类超平面正确分类的样本,

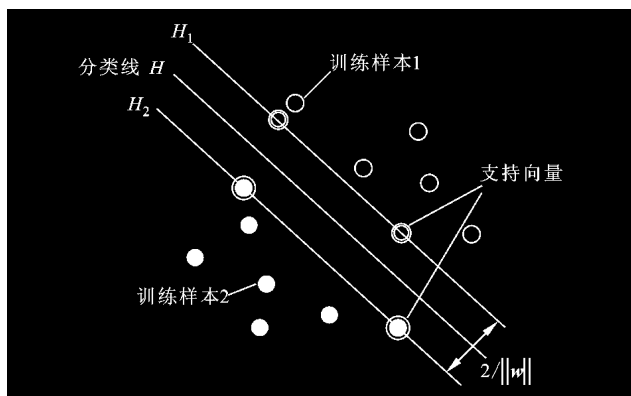


图1 最优分类超平面

引入松弛项 $\epsilon_i \geq 0$, 则超平面的约束条件为

$$y_i[(w \cdot x_i) + b] - 1 + \epsilon_i \geq 0 \quad (1)$$

式中: w 为超平面法线方向向量; b 为分类阈值。

再经过进一步的简化,在线性不可分的情况下,广义最优分类面问题可转化为

$$\Phi(w, \epsilon) = \frac{1}{2}(w \cdot w) + C \left(\sum_{i=1}^n \epsilon_i \right) \quad (2)$$

式中: C 为某个制定的常数,控制错分样本惩罚的程度,实现在错分样本比例与算法复杂度间的折衷。

1.3 最优分类面的构造

寻找最优分类面,就要求解如上所述的二次规划问题,找出惟一的极小点。在此仅讨论非线性可分的情况,该优化问题可由拉格朗日泛函的鞍点给出

$$L(w, b, \epsilon, a, \gamma) = \frac{1}{2}(w \cdot w) + C \sum_{i=1}^l \epsilon_i - \sum_{i=1}^l a_i [(w \cdot x_i - b) y_i - 1 + \epsilon_i] - \sum_{i=1}^l \gamma_i \epsilon_i \quad (3)$$

式中: a_i 为拉格朗日乘子, $a_i \geq 0$; $\gamma_i \geq 0$ 。

根据优化理论的 Kuhn-Tucker 定理^[3], 利用 $0 < a_i < C$ 所对应的样本即可求出分类阈值 b , 最后可得到基于最优超平面的决策函数

$$f(x) = \text{sgn}(w \cdot x + b) \quad (4)$$

1.4 模糊均值聚类基础

模糊 C 均值聚类算法(FCM)是一种非监督聚类算法,由 Jim Bezdek 于 1981 年提出。该算法首先随机选取若干聚类中心,所有数据点都被赋予对聚类中心一定的模糊隶属度,然后通过迭代方法不断修正聚类中心,迭代过程中以极小化所有数据点到各个聚类中心的距离与隶属度的加权和为优化目标。其基本思路^[4]为:将数据集 $X = \{x_1, x_2, \dots, x_n\} \in R^n$ 分为 c 类,任意样本 x_k 对 i 类的隶属度为 μ_{ik} , 分类结果用一个模糊隶属度矩阵 $U = \{\mu_{ik}\} \in R^{n \times c}$ 表示,FCM 通过最小化关于隶属度矩阵 U 和聚类中心 V 的目标函数 $J_m(U, V)$ 来实现

$$J_m(U, V) = \sum \sum (\mu_{ik})^m d_{ik}^2(x_k, v_i) \quad (5)$$

式中: $V = \{v_1, v_2, \dots, v_c\} \in R^c$ 为 c 个聚类中心点集; $m \in [1, \infty)$ 为加权指数。

2 变压器故障的诊断模型

本文采用 IEC60599 的故障分类方法对油浸式电力变压器故障类型进行分类,IEC60599 粗略地将故障分为局部放电、中低温过热、高温过热、低能放电、高能放电^[5]。从文献^[6]可知,目前从油中气体分析结果很难加以区分局部放电和受潮故障,应根据

外部检查和其他实验结果加以综合判断,故本文将变压器的故障类型分为中低温过热、高温过热、低能放电、高能放电.基于 SVM 和 FCM 的变压器故障诊断模型如图 2 所示.

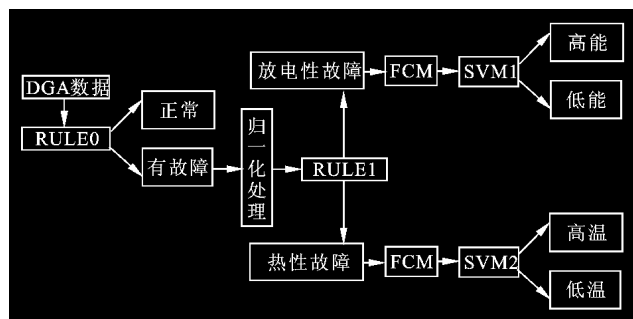


图2 基于 FCM 和 SVM 的变压器故障诊断模型

应用支持向量机处理具体问题时,对存储和计算量两方面都有一定的要求^[7].在算法实施时要存储与训练集相应的核矩阵,而存储核矩阵所需的内存是随着训练集中样本点个数的平方增长的.支持向量机算法包含大量的矩阵运算,所需的运算时间往往过长,加之用网格搜索法来确定参数,所需运算时间更是难以接受.但是,支持向量机的最优分类超平面主要由支持向量来确定,那些距最优分类超平面比较远的样本对 SVM 训练的作用很小,如果能根据实际问题预先取出这些样本,会大大减小计算量及存储空间.在以 DGA 电气设备故障的模式识别研究中,DGA 数据中表征同一故障的特征数据占有故障特征空间的某一子空间,其子空间所表达的故障类都是有界的,而且该类中任一由上述特征数据构成的特征向量与本类中所有特征向量之间的距离小于与其他各类所有或大多数特征向量之间的距离^[8].DGA 数据的这些特性使得用模糊 C 均值聚类对所获取样本进行预选取成为可能.本文的处理思路为:①利用 FCM 算法对获取的 DGA 样本进行预处理,提取所有可能是支持向量的样本即有效约束;②所提取的样本进入支持向量机进行训练计算,采用基于交叉验证的网格搜索法确定支持向量机的参数.

本文针对收集到的 482 台故障明确的变压器油中气体含量的数据,去除一些数据不全及测量结果明显有误的数据后,得到 348 组故障变压器的 DGA 数据,从各类型故障中选取 50 个做训练样本,余下的做检验样本,数据分布如表 1 所示.

2.1 诊断规则 RULE0 和 RULE1

RULE0: 根据电力系统多年的运行经验,采用

表 1 所收集变压器故障样本的分布情况

故障类型	样本总数	训练样本数	检验样本数
高能放电	98	50	48
低能放电	65	50	15
高温过热	108	50	58
中低温过热	77	50	27

《电力设备预防性试验规程》^[9]中规定的色谱注意值及产气率作为区分变压器是否存在异常的界限.

RULE1: 对 IEC 三比值及改良电协研法的分析发现, $\varphi(\text{C}_2\text{H}_2)/\varphi(\text{C}_2\text{H}_4)$ 能较好地识别放电性和过热性故障.文献^[10]通过对 717 个故障样本(414 个过热样本和 303 个放电样本)的统计得出结论:仅需将故障分为过热和放电两类,采用 $\varphi(\text{C}_2\text{H}_2)/\varphi(\text{C}_2\text{H}_4)$ 作为特征值已经足够,一般不需其他特征信息即能达到很好的识别效果.本文对收集到的放电性故障样本和热性故障样本做统计分析,结果表明 $\varphi(\text{C}_2\text{H}_2)/\varphi(\text{C}_2\text{H}_4)$ 能较好区分两类故障,且取改良电协研法所推荐的 0.1 作为区分阈值较为合理,故 RULE1 规则为:计算样本的 $\varphi(\text{C}_2\text{H}_2)/\varphi(\text{C}_2\text{H}_4)$,大于 0.1 为放电性故障,进入 SVM1 分类器,进行高能、低能放电识别,小于 0.1 为过热性故障,进入 SVM2 分类器,进行高温、低温过热识别.

2.2 数据归一化处理

考虑到各种气体组分含量之间的巨大差异性及其分散性,为降低其相互之间由于量值差异造成的影响,对 DGA 原始数据进行归一化处理,以降低气体之间的互斥性.本文选用的归一化处理方法为: $\varphi(\text{H}_2)/(\varphi(\text{H}_2)+\text{总烃})$, CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 的体积分数分别与总烃相比,其中总烃指所测 CH_4 、 C_2H_6 、 C_2H_4 和 C_2H_2 体积分数之和.

样本的归一化处理为下一步 SVM 分类器的训练及测试做了良好的准备.

3 支持向量机参数确定及实现

3.1 支持向量机的参数

支持向量机算法中需确定的参数有以下几个:①核函数的参数 γ ;②支持向量的数目;③非零拉格朗日因子 α_i 及其对应的支持向量;④阈值;⑤惩罚因子 C 值.其中②、③、④项参数在训练中解决二次规划问题时得到,而①、⑤项需要事先选定.选择合适的参数对问题的结果起着重要作用,基于交叉验

证的网格搜索法是一种简单有效的方法,但其缺点是计算代价大.针对此问题,本文采用模糊均值聚类算法预选取样本,使问题得以有效解决.

3.2 网格搜索法

在支持向量机算法参数确定中,网格搜索法以其简单、高效得到了广泛的应用.网格搜索法首先确定各个参数的取值范围,然后对每个参数取值范围按照一定规律插值,得出若干组参数组合,对每个网格点采用一定算法进行计算.本文采用交叉验证算法,基本思想是^[8]:把 N 个数据样本随机地分成 k 个互不相交的子集,即 k -折 $S_1, S_2, \dots, S_k$,每个折的大小大致相等;进行 k 次训练与测试,即对 $i=1, 2, \dots, k$ 进行 k 次迭代;第 i 次迭代的做法是选择 S_i 为测试集,其余 $S_1, \dots, S_{i-1}, S_{i+1}, \dots, S_k$ 的和集为训练集,然后根据训练集求出决策函数后,即可对测试集 S_i 进行测试;记其中错误分类的样本点个数为 l_i , k 次迭代完成后,得到 $l_1, \dots, l_k$;比较哪一组参数给出的交叉验证精度最高,精度最高的参数就是所要选取的参数.

3.3 支持向量机的训练

训练样本进行模糊均值聚类划分后,样本会按照其在数据空间固有的分布情况进行聚类划分,把聚类结果与事先已知故障类型吻合的样本称为正确归类样本,不吻合的称为错误归类样本.由于变压器 DGA 数据分布的复杂性,会有一些样本被错误归类,选择被错误归类的样本以及被正确归类但隶属度 μ_{ij} 小于一定值的样本做训练样本,去除以较大的隶属度被正确归类的样本.之所以这样选取是因为:正确归类但隶属度低的样本相对来说处于聚类的边缘,错误归类的样本本身就说明其处于数据空间的边缘,这两类样本成为支持向量的可能性最大,去除对支持向量机的影响很小的样本.

3.4 支持向量机的实现

经模糊均值聚类算法预选取的样本作为支持向量机的训练样本.选取高斯型核函数作为支持向量机算法中的核函数形式.用基于交叉验证的网格搜索法对支持向量机的参数进行确定.

基于网格搜索的参数确定方法实现如下:选定 1 组 C, γ 的范围,在 C, γ 坐标系上构成一个二维网络.对应网格上每一组 C, γ 的值,按照交叉验证法计算出预测准确率,最后比较在各个网格上的准确率,确定最佳 C, γ 值.通过比较,对于电性故障分类器 SVM1, γ 和 C 的选择范围均为 $[2^{-15}, 2^{-14}, \dots, 2^{14}, 2^{15}]$,对于热性故障分类器 SVM2, γ 和 C 的选

择范围均为 $[2^{-10}, 2^{-9}, \dots, 2^9, 2^{10}]$.

经反复比较后确定 FCM 算法中选取正确归类样本的隶属度阈值为:放电性故障为 0.8,过热性故障为 0.93.确定分类器参数时所用的交叉验证为 10-折交叉验证.图 3 和图 4 给出了分类器 SVM1 和 SVM2 的正判率随核函数参数的变化情况.为了比较,图 5 和图 6 给出了未经 FCM 算法预选取样本的分类器 SVM1 和 SVM2 的正判率随核函数参数的变化情况.

图 3~图 6 中, n_2 轴表示参数 γ 的取值,且 $\gamma=2^{n_2}$, n_1 轴表示参数 C 的取值,且 $C=2^{n_1}$.由图并结合网格搜索法综合比较,选取 SVM1 参数 $\gamma=2^3$, $C=2$,此时交叉验证正判率为 95.3%.选取 SVM1 参数 $\gamma=2^2$, $C=2^8$,此时交叉验证正判率为 90.5%.

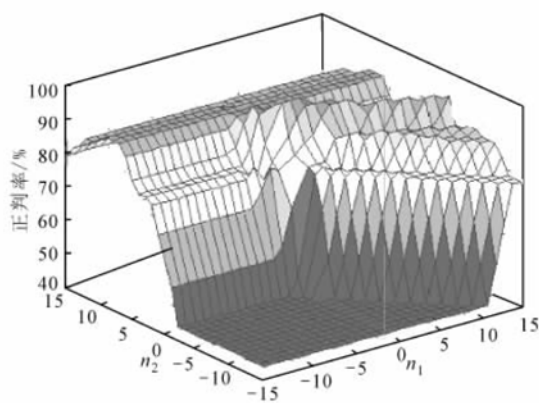


图3 电性故障 FCM+SVM1 分类器交叉验证结果

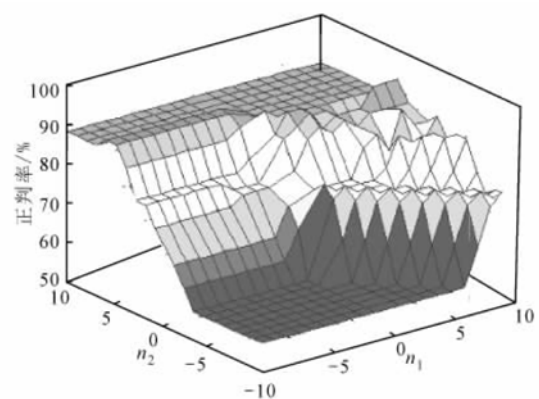


图4 热性故障 FCM+SVM2 分类器交叉验证结果

4 实例验证

放电性故障 100 例样本经 FCM 算法预选取后,错误归类的样本为 32 个,正确归类但隶属度小于 0.8 的为 32 个.选取这 64 个样本作为 SVM1 分类器的训练样本,训练及检验结果如表 2 所示,为了比较将未经 FCM 算法预选取样本的分类器 SVM1 的训练及检验结果一并列出.

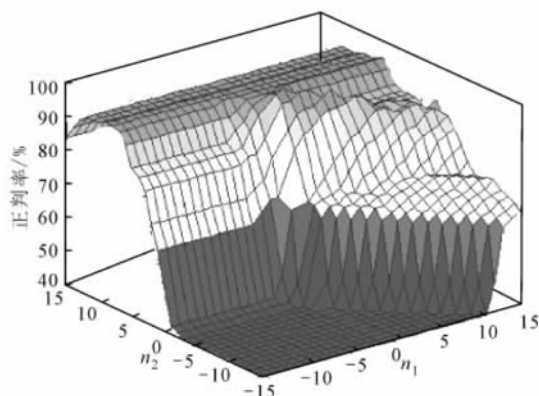


图5 电性故障 SVM1 分类器交叉验证结果

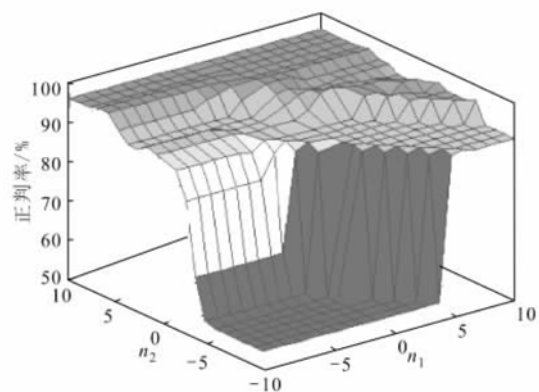


图6 热性故障 SVM2 分类器交叉验证结果

表2 放电性故障的训练时间及检验结果

方法	训练 时间/s	检验样本 正判台次	正判率/ %	去除样本 正判台次
FCM+SVM1	8 650	57/63	90.5	35/36
SVM1	19 798	53/63	84.1	无

过热性故障 100 例样本经 FCM 算法预选取后,错误归类的样本为 9 个,正确归类但隶属度小于 0.93 的有 33 个.选取这 42 个样本作为 SVM1 分类器的训练样本,训练及检验结果如表 3 所示,为了比较将未经 FCM 算法预选取样本的分类器 SVM1 的训练及检验结果一并列出.

表3 过热性故障的训练时间及检验结果

方法	训练 时间/s	检验样本 正判台次	正判率/ %	去除样本 正判台次
FCM+SVM2	1 949	73/85	85.9	58/58
SVM2	8 190	60/85	71	无

注:本文所用计算机硬件情况为 Celeron(R) CPU 1.70 GB、256 MB 内存,软件为 Matlab 6.5 版本.

5 结 论

支持向量机作为一种以有限样本统计学习理论为基础发展起来的新的学习方法,有助于深入挖掘变压器 DGA 数据中所含的故障信息,以便在有限的小训练样本的条件下也能得到比较准确的诊断结果,具有很强的推广性.基于变压器 DGA 数据特征空间中的紧致性原理,采用模糊均值聚类算法对样本进行预选取,有效地解决了支持向量机参数确定时耗时巨大的问题,同时由于对训练样本进行了筛选,使得支持向量机的泛化能力得到一定程度提高.如何充分挖掘 DGA 数据中所含的故障信息,并结合支持向量机小样本条件下的优点,使所做出的诊断结果能有效地指导现场维修是进一步的工作方向.

参考文献:

- [1] 张勇,李昕,刘君华,等.数据融合技术在识别变压器油中四种特征气体的研究[J].中国电机工程学报,2001,21(8):10-14.
Zhang Yong, Li Xin, Liu Junhua, et al. A new datum fusion technique for pattern recognition of four characteristic gases in transformer oil[J]. Proceedings of the CSEE, 2001,21(8):10-14.
- [2] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.
- [3] Bertsekas D P. Nonlinear programming [M]. Belmont: Athena Scientific, 1995.
- [4] Jan A K, Murty M N, Flynn P J. Data clustering: a review[J]. ACM Computer Survey, 1999,31(3):264-323.
- [5] IEC. IEC-60599 mineral oil-impregnated electrical equipment in service-guide to the interpretation of dissolved and free gases analysis[EB/OL]. [2006-11-02]. <http://webstore.iec.ch/webstore/webstore.nsf/artnum/023752>.
- [6] 操敦奎.变压器油中气体分析诊断与故障检查[M].北京:中国电力出版社,2005.
- [7] 邓乃扬,田英杰.数据挖掘中的新方法——支持向量机[M].北京:科学出版社,2004.
- [8] 孙才新,陈伟根,李俭,等.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.
- [9] 中华人民共和国电力工业部.电力设备预防性试验规程 DL/T 596-1996[S].北京:中国电力出版社,1996.
- [10] 张蕊,郭瑞君,李华,等.基于变压器故障分类的 DGA 特征提取[J].高电压技术,2005,31(4):32-33.
Zhang Rui, Guo Ruijun, Li Hua, et al. Feature selection of DGA data based on transformer fault classification[J]. High Voltage Engineering, 2005,31(4):32-33.

(编辑 杜秀杰)