

# 基于核 Fisher 判别分析技术的电力 变压器 DGA 故障诊断模型研究

吴晓辉<sup>1</sup>, 王 颂<sup>1</sup>, 方晓明<sup>1</sup>, 李延沐<sup>2</sup>, 李彦明<sup>1</sup>

(1. 西安交通大学电气工程学院, 陕西 西安 710049; 2. 西安建筑科技大学机电学院, 陕西 西安 710055)

Fault Diagnosis Model of DGA for Power Transformer Based on  
Kernel Fisher Discriminant Analysis Technology

WU Xiao-hui<sup>1</sup>, WANG Song<sup>1</sup>, FANG Xiao-ming<sup>1</sup>, LI Yan-mu<sup>2</sup>, LI Yan-ming<sup>1</sup>

(1. School of Electrical Engineering, Xi 'an Jiaotong University, Xi 'an 710049, China;

2. School of Mechanical and Electrical, Xi 'an University of Architecture and Technology, Xi 'an 710055, China)

**摘要:** 鉴于核 Fisher 判别分析技术(KFDA)在模式识别问题中表现出的良好性能,提出了基于KFDA的变压器故障诊断模型,该模型首先提出了区分放电及过热两大类故障的特征量,并用KFDA分类器来识别类内故障的具体类别。采用基于网格搜索的交叉验证法来选择模型参数,避免了参数选择的盲目性和随意性。实例分析表明,该模型具有训练时间短、不存在局部极小等优点,与IEC三比值及改良电协研法相比,具有更好的故障识别效果。

**关键词:** 变压器; 故障诊断; 油中溶解气体分析;  
核 Fisher 判别分析; 交叉验证

中图分类号: TM855

文献标志码: A

**Abstract:** In view of the good performance in pattern recognition using Kernel Fisher Discriminant Analysis (KFDA) method, this paper presented a fault diagnosis model of power transformer based on KFDA. The model extracted a feature to distinguish between discharge faults and hot faults first, and then used the KFDA classifier to distinguish in discharge faults or hot faults. To choose parameters of the model, this paper adopted the method of cross validation based on grid search, avoiding the arbitrary and capricious behavior. Practical analysis indicates that this model has advantages of short training time and no part teeny problem. Compared with three rations method of IEC and the improved three rations method, this model has considerable effectiveness of faults identification.

**Key words:** transformer; fault diagnosis;  
dissolved gases analysis(DGA); kernel fisher  
discriminant analysis; cross validation

## 0 引言

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<sup>[1-5]</sup>。其中人工神经网络技术是引入到变压器故障诊断中的智能技术中最重要的方法之一,但是,通常人工神经网络有几个比较大的缺点,如训练时间比较长、存在局部最优等问题,识别正确率和可靠性有待于进一步提高。

核机器学习算法是近几年发展起来的一类新的非线性技术,它有效地解决了小样本、高维数、非线性等学习问题,并克服了人工神经网络合理结构难以确定和存在局部最优等缺点,在模式识别、回归估计等方面都有着成功的应用<sup>[6-8]</sup>。笔者将核机器学习算法中的核 Fisher 判别分析(KFDA)引入到变压器故障诊断中,提出了基于核 Fisher 分类器的变压器故障诊断方法,并引入交叉验证法来选取核机器学习算法中所需要确定的核参数。实际验证表明,该方法可有效地求取变压器油中气体特征空间中的最优解,提高了故障诊断的准确性。

## 1 核 Fisher 判别分析

### 1.1 概述

文[9]提出了线性 Fisher 判别分析算法(FDA),它是一种旨在降低特征维数且有监督学习算法的方法,其目标是建立一个子空间(由所有投影轴构成),所有样本在这个子空间内满足类内散度最小,类间散度最大。所有样本在这些投影轴上的投影系数作

收稿日期: 2007-01-04; 修回日期: 2007-04-18

作者简介: 吴晓辉(1983-),男,硕士研究生,从事变压器故障诊断方面的研究。

为样本的特征向量，利用这些特征向量对样本进行分类识别。KFDA 是将核学习思想引入到 FDA 中所产生的一种新的非线性特征提取方法，它利用核学习思想，首先通过一个非线性映射，将空间  $R^d$  中的数据映射到一个高维的核空间中，通过引入核函数  $k(x, y)$ ，就可以在这个高维核空间中进行线性 Fisher 判决分析。这样，在高维核空间中所获得的样本特征相对于原空间来说就是非线性的，其更有利于分类。

### 1.2 具体算法

假定  $x$  是输入训练向量集  $X$  中的一个向量， $X$  中共有  $n$  个样本，这些样本共有  $C$  类，即

$$X = \bigcup_{k=1}^C X_k, n = \sum_{k=1}^C n_k \quad (1)$$

定义类间散度矩阵

$$S_B = \sum_{k=1}^C n_k (\mu_k - \mu) (\mu_k - \mu)^T \quad (2)$$

类内散度矩阵

$$S_W = \sum_{k=1}^C \sum_{x_j \in X_k} (x_j - \mu_k) (x_j - \mu_k)^T \quad (3)$$

其中  $\mu$  是总体均值向量； $\mu_k$  是第  $k$  类均值。则最优投影方向为

$$W_{opt} = \underset{w}{\operatorname{argmax}} \frac{|w^T S_B w|}{|w^T S_W w|} = [w_1, w_2, \dots, w_m] \quad (4)$$

这里  $[w_1, w_2, \dots, w_m]$  是对应于矩阵  $S_W^{-1} S_B$  的前  $m$  个最大特征值的特征向量，样本在这些特征向量上的投影系数就是所提取样本的特征向量值。

KFDA 算法首先通过一个非线性映射将输入数据映射到一个高维的特征空间中，即

$$: R^d \rightarrow H, x \rightarrow (x).$$

这时输入的训练样本由原来的  $x$  变为  $(x)$ ，然后在这特征空间  $H$  中进行线性 FDA。KFDA 是在  $H$  空间中求解以下问题。

$$W_{opt} = \underset{w}{\operatorname{argmax}} \frac{|(w)^T S_B w|}{|(w)^T S_W w|} = [w_1, w_2, \dots, w_m] \quad (5)$$

假定在  $H$  空间中，所有的样本都是去均值的，则类内散度矩阵为

$$S_B = \sum_{k=1}^C n_k (\mu_k) (\mu_k)^T \quad (6)$$

$$\text{其中每类的均值 } \mu_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_{ki}) \quad (7)$$

类间散度矩阵为

$$S_W = \frac{1}{n} \sum_{k=1}^C \sum_{i=1}^{n_k} (x_{ki})^T (x_{ki}) \quad (8)$$

$[w_1, w_2, \dots, w_m]$  就是特征值问题的解。

$H$  空间的维数很高，不可能进行通常的运算，为了能够在  $H$  空间中实现线性 FDA，利用内积核函数

来隐含地进行运算。引进以下的内积核函数

$$k(x_i, x_j) = \langle (x_i), (x_j) \rangle \quad (9)$$

在  $H$  空间中，任何未知模式  $(x)$  在特征向量  $[w_1, w_2, \dots, w_m]$  的投影系数(这些系数可以作为该模式的特征向量值)通过下式计算出

$$[w \cdot (x)] = \sum_{k=1}^C \sum_{i=1}^{n_k} a_{ki} k(x_{ki}, x) \quad (10)$$

### 1.3 交叉验证法

对于中等规模的问题来说，交叉验证法可以说是最可靠的模型选择方法<sup>[10]</sup>。其基本思想是<sup>[11]</sup>：首先把  $N$  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，共进行  $k$  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  $S_1, \dots, S_{i-1}, \dots, S_{i+1}, \dots, S_k$  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，便得到了  $l_1, \dots, l_k$ ，可以推想所有  $k$  次迭代中的

错误分类数  $\sum_{i=1}^k l_i$  和总样本数  $l$  之比  $\sum_{i=1}^k l_i / l$  可以作为

该算法错误率的一个估计。

### 1.4 网格搜索

网格搜索法以其简单、高效在支持向量机算法参数确定中得到广泛应用。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。笔者采用交叉验证算法，比较哪一组参数给出的交叉验证精度最高，精度最高的就是所要选取的参数。

## 2 基于 KFDA 的变压器故障诊断模型

笔者采用 IEC 60599 的故障分类方法对充油电力变压器故障类型进行分类，IEC 60599 粗略地将故障分为 5 类<sup>[12]</sup>：局部放电、中低温过热、高温过热、低能放电、高能放电。从文[13]可知，电力变压器中局部放电和受潮两种异常现象有时同时存在，且特征气体基本相同，因此目前从油中气体分析结果还很难加以区分，必要时应根据外部检查和其他实验结果加以综合判断，如局部放电测量和油中微水分析等。因此在判断故障类型时暂不考虑局部放电，而仅考虑中低温过热、高温过热、低能放电、高能放电。基于 KFDA 的变压器故障诊断模型见图 1。

### 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《电力设备预防性试验规程》<sup>[14]</sup>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， $C_2H_2/C_2H_4$  的比值能较好地识别放电性和过热

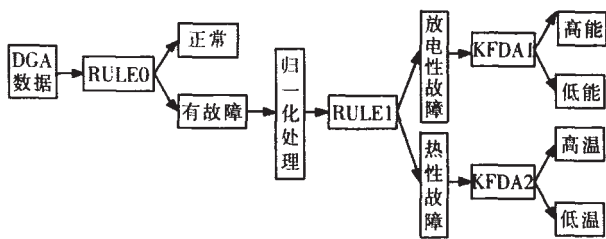


图 1 基于 KFDA 的电力变压器故障诊断模型

性故障。文[15]通过对具体 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  $C_2H_2/C_2H_4$  作为特征值已经足够, 一般不需其它特征信息即能达到很好的识别效果。笔者通过对文献收集到的 170 例放电性故障样本和 164 例热性故障样本作统计分析, 结果表明  $C_2H_2/C_2H_4$  能较好区分两类故障, 且取改良电协研法所推荐的阈值 0.1 作为区分阈值较为合理。故 RULE1 规则为: 计算样本的  $C_2H_2/C_2H_4$  值, 大于 0.1 为放电性故障, 进入 KFDA1 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 KFDA2 分类器, 进行高温、低温过热识别。

### 2.2 数据归一化处理

由于用于故障诊断的 DGA 数据来自不同形式、容量、电压等级的变压器, 样本中同一故障类的气体的体积不同, 甚至相差很大。为了保持输入向量的同一性, 需对输入气体分量归一化处理, 笔者选用的归一化处理方法为: 将  $H_2$ 、 $CH_4$ 、 $C_2H_6$ 、 $C_2H_4$  和  $C_2H_2$  的值分别进行如下处理:  $H_2/(H_2+总烃)$ 、 $CH_4/总烃$ 、 $C_2H_6/总烃$ 、 $CH_4/总烃$ 、 $C_2H_2/总烃$ 。其中总烃指所测  $CH_4$ 、 $C_2H_6$ 、 $C_2H_4$  和  $C_2H_2$  值之和。

样本的归一化处理为下一步 KFDA 分类器的训练及测试做了良好的准备。

### 2.3 KFDA 分类器设计

核机器学习算法中核函数形式及其参数的选取一直是一个比较棘手的问题, 理论上及文献上并没有形成一个统一的模式, 不同形式的核函数及核函数的不同参数对分类的结果影响很大。在常用的三种核函数中, 笔者选取高斯核函数为 KFDA 分类器所用的核函数形式, 因为高斯核可以将样本映射到一个更高维的空间, 在一定参数条件下, 高斯核函数同线性核函数具有相同的性能。与多项式核函数相比, 高斯核函数具有参数少的优点。

高斯核函数参数的确定, 并借鉴支持向量机参数确定中常用的网格搜索法, 将高斯核参数的搜索空间定为指数形式增长, 为  $\{2^{-15}, 2^{-14}, 2^{-13}, \dots, 2^1, 2^2, 2^3, \dots, 2^{15}\}$ 。在每个参数下, 用交叉验证法确定其正判率, 并选取正判率最高的参数作为其最终参数。

## 3 实例验证及结果

通过对所搜集到的 170 例放电性故障样本和

164 例过热性故障样本进行筛选, 选取放电性故障样本、过热性故障样本各 100 例, 其中高能放电、低能放电、高温过热、低温过热各 50 例。确定分类器参数时所用的交叉验证为 10-折交叉验证, 即将放电性故障样本和过热性故障样本分别随机分为 10 组, 迭代中每次取所分 10 组中的 9 组作为训练样本, 而余下的 1 组为检验样。10 次迭代完成后将每次迭代过程中正确分类的样本数累加, 累加值与总样本数的比值为正判率。图 2 给出了分类器 KFDA1 和 KFDA2 的正判率随核函数参数的变化情况。

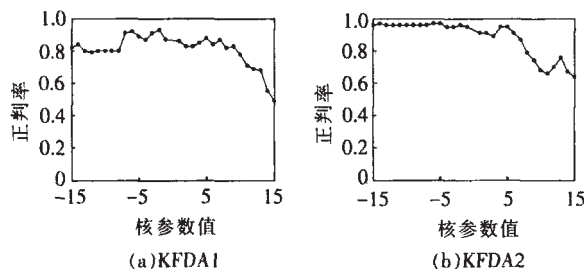


图 2 分类器 KFDA1 和 KFDA2 训练及检验结果

从图 2 可知, 核参数值对分类器 KFDA1 的结果(即对放电性故障样本)影响比较大, 而对分类器 KFDA2 的结果(即对过热性故障样本)影响不大, 在一定参数值范围都能保持较高的正判率。综合考虑, 选取  $=2^{-2}$  作为分类器 KFDA1 的核参数值, 此时正判率为 92%;  $=2^{-6}$  作为分类器 KFDA2 的核参数值, 此时正判率为 97%。

为了说明所提出的基于核 Fisher 判别分析的变压器故障诊断模型的有效性, 另从文献中选取具有一定的典型性的数据对构造的诊断模型进行检验。其中, 电弧放电故障样本 7 例、火花放电故障样本 4 例、高温过热故障样本 7 例和低温过热故障样本 7 例。分别用 IEC 三比值法、改良电协研法和本文所提出的基于 KFDA 诊断模型进行判别, 表 1 列出了应用 3 种方法的结果。

方法名称	正判率/%
IEC 三比值法	60
改良电协研法	72
基于 KFDA 诊断模型	92

由表 1 可以看出, 基于 KFDA 的变压器诊断模型与规程中所推荐的 IEC 三比值及改良电协研法相比有较大优势。

## 4 结论

(1) 核 Fisher 判别分析通过核函数的非线性映射, 将低维空间的数据映射到高维空间, 从而可以在高维空间中应用线性鉴别技术得到相应于低维空间的非线性区分, 适于充油电力变压器油中溶解

气体分析中气体组分与故障类型之间的复杂对应关系。笔者建立了基于 KFDA 的变压器故障诊断模型,通过分析具体选取了高斯型核函数,并采用基于交叉验证的网格搜索法来确定核函数的参数,取得了良好的效果。

(2)笔者将核 Fisher 判别分析技术初步应用于变压器故障诊断,还有一些问题需要深入研究。如何将核函数的具体形式同变压器油中溶解气体分析的先验知识相结合,并从理论上去分析选取最优参数,都有待于进一步的探讨和研究。

### 参考文献:

[1] DUVAL M, LANGDEAU F, GERVAIS P, et al. Acceptable Gas-in-oil Levels in Generation and Transmission Power Transformers [C]// Electrical Insulation and Dielectric Phenomena, Annual Report, Conference on 1990: 325-330.

[2] 徐文,王大忠,周泽存,等. 神经网络在变压器先生气体法故障诊断中的应用[J]. 高电压技术, 1996, 22(2): 27-30.

[3] 颜湘莲,文远芳. 模糊神经网络在变压器故障诊断中的应用研究[J]. 变压器, 2002, 39(7): 41-43.

[4] 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.

[5] 王由华,刘振娟,李宏光. 混合型集成神经网络故障诊断方

法研究[J]. 北京化工大学学报: 自然科学版, 2003, 30(1): 65-68.

[6] BAUDAT G, ANOUAR F. Generalized Discriminant Analysis Using a Kernel Approach [J]. Neural Computation, 2000, 12(10): 2385-2404.

[7] MULLER K R, MIKA S, GUNNAR R, et al. An Introduction to Kernel Based Learning Algorithms [J]. IEEE Trans Neural Networks, 2001, 12(2): 181-201.

[8] MA J H, PERKINS S, THEILER, et al. Modified Kernel Based Nonlinear Feature Extraction [C]// International Conference on Machine Learning and Applications, 2002: 127-132.

[9] 边肇琪,张学工. 模式识别[M]. 北京: 清华大学出版社, 2000.

[10] BROWNE M W. Cross-validation Methods [J]. Journal of Mathematical Psychology, 2000, 44(1): 108-132.

[11] 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.

[12] IEC 60599. Mineral Oil-impregnated Electrical Equipment in Service-guide to the Interpretation of Dissolved and Free Gases Analysis, Second Edition[S]. 1999.

[13] 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.

[14] 中华人民共和国电力工业部. 电力设备预防性试验规程[Z]. DL/T 596-1996. 北京: 中国电力出版社, 1996.

[15] 张蕊,郭瑞君. 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.

(上接第 213 页)

$$5\sin(0.5 t)e^{0.01t}, t \in [0, 500].$$

对上述两个函数进行离散采样,采样间隔为 1,采样 501 个数据点作为初始采样数据进行计算验证。笔者根据上述方法编制了相关的软件,对初始采样数据进行的参数计算,计算结果见表 1。f(t) 具有解析函数表达式,通过表达式根据参数的定义可计算较精确的参数值作为参考值来检验该方法的精确性。

表 1 参数计算结果

	拟合法计算参数值				标准参考值
	f(t)	相对误差/%	f <sub>2</sub> (t)	相对误差/%	
U <sub>max</sub>	199.980	0.010	199.980	0.010	200.000
t <sub>3</sub>	276.421	0.149	276.549	0.195	276.010
t <sub>d</sub>	38.934	2.058	38.705	1.457	38.149

对 TRV 波形进行峰值计算的抛物线拟合时,相关指数 R<sup>2</sup>=0.999 0;对电压波形的起始点至峰值点进行曲线拟合时,相关指数 R<sup>2</sup>=0.999 9。可见采用最小二乘法对 TRV 波形进行拟合的结果非常逼近采样数据,具有良好的拟合效果,说明所采用的基于最小二乘法和牛顿迭代法计算 TRV 波形参数是一种非常有效且精确度很高的方法。

### 3 结论

(1) 基于最小二乘法原理,采用多项式函数对 TRV 波形进行曲线拟合,可以得到很好的逼近,多项式最高项的次数可以根据波形的运动趋势进行合理选择。

(2) 采用牛顿迭代法计算 TRV 波形的参数,可以得到精确的结果。

(3) 笔者采用的方法用于计算只需 2 个参数表征的 TRV 波形的参数,具有一定的稳定性和较高的精度,能满足实际计算的要求。

### 参考文献:

[1] 王锡凡. 电力工程基础[M]. 西安: 西安交通大学出版社, 1998.

[2] 林莘. 现代高压电器技术[M]. 北京: 机械工业出版社, 2002.

[3] STL Technical Report. Harmonization of Data Processing Methods for High Power Laboratories[R]. 2003.

[4] 王晶,牟磊,李彦明,等. 大容量实验室电流波形参数计算机提取算法的研究[J]. 高压电器, 2006, 42(1): 44-46.

[5] 邓建中,刘之行. 计算方法[M]. 西安: 西安交通大学出版社, 2001.

[6] 解西伦,张陆仁. 指数曲线拟合传染病发病率探讨[J]. 济宁医学院学报, 1997, 20(1): 55-56.