

用SVM和LS-SVM分析变压器故障诊断

张哲, 朱永利

(华北电力大学 计算机科学与技术学院, 河北 保定 071003)

摘要: 介绍并比较了支持向量机分类器和最小二乘支持向量机分类器的算法, 提出了基于支持向量机的二叉树多分类变压器故障诊断模型。将标准支持向量机(C-SVM)分类器和最小二乘支持向量机(LS-SVM)分类器分别用于变压器故障诊断, 通过网格搜索和交叉验证法取得支持向量机的参数, 准确率较高。试验结果表明, 支持向量机和最小二乘支持向量机在变压器故障诊断中具有很大的应用潜力。

关键词: 支持向量机; 分类器; 最小二乘支持向量机; 故障诊断

中图分类号: TP302.8

文献标识码: B

SVM and LS-SVM analyses transformer fault diagnosis

ZHANG Zhe, ZHU Yong Li

(College of Computer Science and Technology, North China Electric Power University, Baoding 071003, China)

Abstract: Nonlinear classifiers algorithms of support vector machine(SVM) and least squares support vector machine(LS-SVM) are discussed and compared. Transformer fault diagnosis which is based on support vector machines multi-classification binary tree model is proposed. SVM classifier and the LS-SVM classifier are used for transformer fault diagnosis, through the grid-search and cross-validation method of choice to determine parameters of support vector machines, which achieved high accuracy. Experimental verification shows that the model has high application potential in transformer fault diagnosis.

Key words: support vector machine; classifier; least squares support vector machine; fault diagnosis

电力变压器在电力系统中的地位举足轻重, 是电力系统中最重要和最关键的电气设备之一, 其运行的安全可靠直接关系到电力系统的安全与稳定。

油中溶解气体分析(DGA)技术是目前对油浸变压器进行故障诊断最方便、有效的手段之一, 它能较准确、可靠地发现逐步发展的潜伏性故障。

SVM^[1-2]是在统计学理论上发展起来的一种新的学习方法, 它能较好地解决小样本、非线性、高维数和局部极小等实际问题, 已成为神经网络研究之后机器学习领域的研究热点, 为了区别, 将Vapnik提出的支持向量机称为标准支持向量机(C-SVM)。最小二乘支持向量机LS-SVM(Least Square Support Vector Machines)是支持向量机的变形算法, LS-SVM采用二次损失函数, 将SVM中的二次规划问题转化为线性方程组求解, 在保证精度的同时大大降低了计算复杂性, 加快了求解速度。

本文提出了基于支持向量机的二叉树多分类变压器故障诊断模型, 通过网格搜索方法选择确定支持向量机参数。将C-SVM分类器和LS-SVM分类器分别用于变压器故

障诊断模型, 实验表明, 该模型具有良好的分类效果和推广能力, C-SVM分类器和LS-SVM分类器都能达到较高的故障诊断率。

1 支持向量机与最小二乘支持向量机算法

1.1 C-SVM与LS-SVM分类器算法

支持向量机的主要思想如下: 首先选择非线性变换 $\varphi(\cdot)$ 将 n 维样本 $(x_1, y_1), \dots, (x_l, y_l) \in R^n \times \{+1, -1\}$ 从原空间 R^n 映射到特征空间, 在此高维空间中构造最优线性决策函数 $y(x) = \text{sgn}[\omega \cdot \varphi(x) + b]$ 。在构造最优决策函数时, 利用结构风险最小化准则, 利用原空间的核函数取代高维特征空间的点积运算, 避免了复杂运算。

支持向量机和最小二乘支持向量机^[3-4]在利用结构风险原则的同时, 在优化目标中选取了不同的损失函数。

对于支持向量机优化目标与约束条件为:

$$\begin{aligned} \min & \left(\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \varepsilon_i \right) \\ \text{s.t. } & y_i [(\varphi(x_i) \cdot \omega + b)] \geq 1 - \varepsilon_i \\ & (\varepsilon_i \geq 0, \quad i = 1, 2, \dots, l) \end{aligned} \quad (1)$$

技术与方法 Technique and Method

式中, ε_i 是松弛变量; C 为惩罚参数。对于最小二乘支持向量机, 优化目标采用 ε_i 的平方项, 故优化问题为:

$$\begin{aligned} \min & \left(\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \varepsilon_i^2 \right) \\ \text{s.t. } & y_i [(\varphi(x_i) \cdot \omega + b)] = 1 - \varepsilon_i \\ & (i = 1, 2, \dots, l) \end{aligned} \quad (2)$$

应用拉格朗日方法求解上述优化问题, 支持向量机优化问题转化为下面的二次规划:

$$\begin{aligned} \max W(\alpha) &= \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t. } & \sum_{i=1}^l \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \end{aligned} \quad (3)$$

最小二乘支持向量机优化问题转化为求解线性方程

$$\begin{bmatrix} 0 & y_1 & \cdots & y_l \\ y_1 & y_1 y_1 K(x_1, x_1) + 1/c & \cdots & y_1 y_l K(x_1, x_l) \\ \vdots & \vdots & \ddots & \vdots \\ y_l & y_l y_1 K(x_l, x_1) & \cdots & y_l y_l K(x_l, x_l) + 1/c \end{bmatrix} \times \begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_l \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (4)$$

1.2 核函数选择

由上面分析可知, C-SVM与LS-SVM分类器是由训练集和核函数完全刻画的, 构造、选择核函数是一个重要问题, 选择不同的核函数会产生不同的支持向量机算法。径向基核函数只有一个待定的参数 σ , 与其他核函数相比具有参数少的优点, 本文采用径向基函数作为核函数, 即:

$$K(x_i, x_j) = \exp\left[-\frac{\|x_i - x_j\|^2}{\sigma^2}\right] \quad (5)$$

2 基于支持向量机的变压器故障诊断

本文将支持向量机分类器和最小二乘支持向量机分类器用于变压器故障诊断中。

2.1 特征提取

通过油色谱分析法得到油中溶解气体类型和变压器内部故障性质的对应关系, 采用多种以油中溶解气体为依据的判别故障方法。目前多采用 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 5 种气体进行诊断, 本文选取列举出的 5 种气体的含量作为输入特征, 利用 SVM 对这 5 种特征量进行训练和诊断。

诊断出变压器常见的 5 种状态: 变压器的正常状态、中低温过热、高温过热、高能放电、低能放电。

2.2 数据预处理

为了避免输入向量中各变量数量级相差过大影响训练效果, 支持向量机对数据样本(训练集和测试集)进行归一化处理。

x_{\max} 和 x_{\min} 分别代表样本集合的最大值与最小值, x_i 为实际数据, \bar{x} 为归一化后的值。

$$\bar{x} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

这样使数据转换到 [0,1] 区间。

2.3 变压器故障诊断模型

基于 SVM 的二叉树多类算法是将所有类别分成两个子类, 再将两个子类分别划分成两个次级子类, 依次类推, 直到所有的节点只含有一个类别为止, 这样, 多分类问题就转化为二分类问题, 每个节点处采用 SVM 二值分类器作为分类函数。

二叉树的分类结构主要有两种: 一种是在每个内节点处, 由一个类与剩下的类构造分割面; 另一种是在内节点处, 可以是多个类与多个类的分割。本文只考虑前一种情况, 即每次分割只分割出一个类。基于支持向量机的二叉树多分类变压器故障诊断模型如图 1 所示。

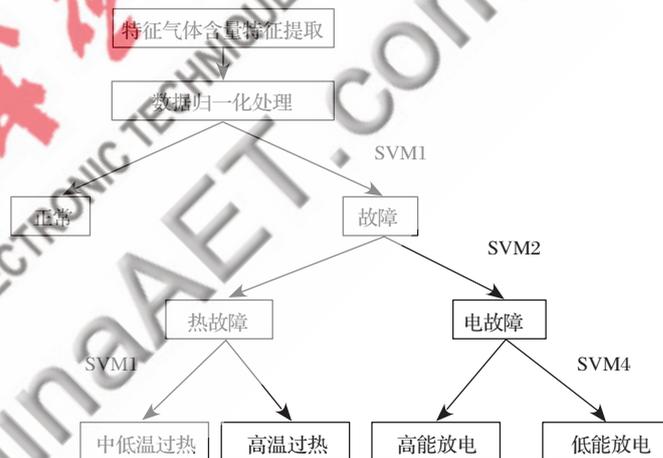


图1 基于SVM的二叉树多分类变压器故障诊断模型

在训练小样本的情况下, 该模型可以由粗到细逐步将故障进行分类, 使诊断不断深入, 逐步靠近故障的真实情况。根据变压器故障的特点, 采用 4 级 SVM 分类器对数据样本进行训练测试, 最终判断变压器所处的状态。

3 参数优化

在支持向量机算法参数的确定中, 基于网格搜索和交叉验证^[5]的参数寻优方法简单实用, 有效地解决了 SVM 参数设计的难题, 增加了诊断模型的推广性。网格搜索方法首先确定各个参数的范围, 然后对每个参数范围按照一定规律插值, 得出若干组参数组合。交叉验证法的思想是: 将训练样本 M 随机分为大致相等的 k 个互不相交的子集, 即 $S = S_1 \cup S_2 \cup \dots \cup S_k$, 共进行 k 次训练和测试, 第 i 次迭代做法是: 选择 S_i 作为测试集, 其余子集的和为训练集。算法根据训练集合求出决策函数后, 对测试集合 S_i 进行测试, 记错误分类的样本点个数为 l_i 。 k 次迭代完成后, 得到了 $l_1 \cdots l_k$, 比较哪一组参数得到的交叉验证精度高, 精度高的参数就是所要选取的参数。

表1 基于LS-SVM的二叉树多分类故障诊断结果

LS-SVM	变压器状态	检验样本数	正判台次	正判率%	运行时间	最优参数选择(gam;sig2)	
LS-SVM1	正常	30	30	100	23.5	58.489	0.1615
	故障	45	43	95.5			
LS-SVM2	放电故障	21	19	90.4	12.3	37.176 6	0.632
	过热故障	24	24	100			
LS-SVM3	中低温过热	4	4	100	6.9	212.77	1.078
	高温过热	20	20	100			
LS-SVM4	高能放电	17	16	94.1	9.5	219.35	0.166
	低能放电	4	3	75			

表2 基于SVM的二叉树多分类故障诊断结果

SVM	变压器状态	检验样本数	正判台次	正判率%	运行时间	最优参数选择(C; g)	
SVM1	正常	30	30	100	30.5	32 768	0.5
	故障	45	43	95.5			
SVM2	放电故障	21	21	100	15.8	2 048	0.5
	过热故障	24	23	95.8			
SVM3	中低温过热	4	4	100	9.6	2 048	8
	高温过热	20	20	100			
SVM4	高能放电	17	16	94.1	10.6	2 048	0.031 5
	低能放电	4	4	100			

4 变压器故障诊断实例分析

本文针对采集到的240组故障电力变压器油中气体含量的历史数据信息,先剔除一些数据不全以及测量结果明显有误的数据。最后得到228组故障明确的电力变压器油中气体数据,其中2/3样本用于训练集合,1/3用于测试集合。

将此数据对基于支持向量机的二叉树多分类变压器故障诊断效果进行了实例验证。

分别利用支持向量机分类器与最小二乘支持向量机分类器,选用相同的训练样本集和测试样本集,通过网格搜索和交叉验证的参数寻优方法确定参数,训练和测试结果分别如表1和表2所示。

为了更好地评价基于支持向量机的二叉树多分类方法。本文也采用神经网络对同样的数据进行试验对比,得到测试结果如表3所列。

通过对基于C-SVM与LS-SVM的二叉树多分类故障诊断模型的试验验证,并与其他模型比较,可得出如下结论:

(1)基于支持向量机的二叉树多级分类故障诊断模型具

表3 基于ANN的二叉树多分类故障诊断结果

ANN	变压器状态	检验样本	正判台次	正判率/%	运行时间/s
ANN1	正常	30	28	93.3	39.6
	故障	45	40	88.8	
ANN2	放电	21	18	85.7	20.5
	过热	24	23	95.8	
ANN3	中低温过热	4	4	100	12.3
	高温过热	20	17	85	
ANN4	高能放电	17	15	88.2	11.6
	低能放电	4	3	75	

有很高的正确率,大大降低了参数选择的时间,使模型构造简单。

(2)参数寻优能提高SVM分类器模型的分正确率,基于网格搜索和交叉验证的参数寻优方法为解决SVM分类器的实际应用提供了有效的途径。从表1与表2可知,标准SVM具有较高的准确度,然而LS-SVM具有较快的收敛速度。

(3)基于支持向量机的二叉树多级分类故障诊断模型训练时间比神经网络的训练时间短,且精确度高。

本文比较了标准支持向量机分类器和最小二乘支持向量机分类器,并将它们应用于二叉树的多分类变压器故障诊断中,使用网格搜索和交叉验证的参数寻优方法,取得了良好的分类效果,应用于故障诊断领域应当具有良好的前景。

参考文献

- [1] VAPNIK V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.
- [2] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000,26(1):32-42.
- [3] BURGESS C J C. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 2 (2) :1-47.
- [4] 阎威武, 邵惠鹤. 支持向量机和最小二乘支持向量机的比较及其应用研究[J]. 控制与决策, 2003, 18(3):358-360.
- [5] 董明, 孟源源, 徐长响, 等. 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 [J]. 中国电机工程学报, 2003, 23(7):88-92.

(收稿日期:2009-01-13)