最小二乘支持向量机在故障诊断中的应用*)

杨奎河[1,2] 单甘霖[1] 赵玲玲[2]
(军械工程学院光学与电子工程系 石家庄 050003)[1]
(河北科技大学信息科学与工程学院 石家庄 050054)[2]

摘 要 为了提高机械设备故障诊断的精度,将小波包分析与最小二乘支持向量机进行了有机的结合。首先对故障信号功率谱进行小波分解,简化了故障特征向量的提取。然后提出了一种基于最小二乘支持向量机的故障诊断模型,用二次损失函数取代支持向量机中的不敏感损失函数,将不等式约束条件变为等式约束,从而将二次规划问题转变为线性方程组的求解,用最小二乘法实现了支持向量机算法,并提出对核函数的 σ 参数进行动态选取,提高了诊断的准确率。仿真结果表明该模型具有较强的非线性处理和抗干扰能力。

关键词 故障诊断,最小二乘支持向量机,核函数,小波包分析

Application of Least Squares Support Vector Machine in Fault Diagnosis

YANG Kui-He[1,2] SHAN Gan-Lin[1] ZHAO Ling-Ling[2]
(Department of Optics and Electron Engineering, Ordnance Engineering College, Shijiazhuang 050003)[1]
(College of Information, Hebei University of Science and Technology, Shijiazhuang 050054)[2]

Abstract In order to enhance fault diagnosis precision, the wavelet packet analysis and least squares support vector machine (LSSVM) are combined effectively. First the power spectrum of fault signals is decomposed by wavelet analysis, which predigests choosing method of fault eigenvectors. And then a fault diagnosis model based on LSSVM is presented. In the model, the non-sensitive loss function is replaced by quadratic loss function and the inequality constraints are replaced by equality constraints. Consequently, quadratic programming problem is simplified as the problem of solving linear equation groups, and the SVM algorithm is realized by least squares method. It is presented to choose σ parameter of kernel function on dynamic, which enhances preciseness rate of diagnosis. The simulation results show the model has strong non-linear solution and anti-jamming ability.

Keywords Fault diagnosis, Least squares support vector machine, Kernel function, Wavelet packet analysis

1 引言

在多年的故障诊断实践中,涌现出了很多行之有效的方法。神经网络方法(Neural Networks, NN)在故障诊断中的应用研究取得了很多的成果,但它在网络结构选择、网络的训练及提高网络的泛化能力等方面存在的许多问题,限制了它在工程实际中的广泛应用。支持向量机(Support Vector Machine, SVM)是由 Vapnik 等人在统计学习理论的基础上建立起来的一种机器学习新方法[1],着重研究小样本情况下的统计学习规律。SVM 通过结构风险最小化原理来提高泛化能力,较好地解决了小样本、非线性、高维数、局部极小等实际问题,已在模式分类、回归预测、概率估计及控制理论等领域得到应用。

最小二乘支持向量机(Least Squares Support Vector Machine, LSSVM)是基于支持向量机方法的一种改进算法[2]。它是支持向量机在二次损失函数下的一种形式,它用二次损失函数取代 SVM 中的不敏感损失函数,通过构造损失函数将原支持向量机中算法的二次寻优变为求解线性方程,其求解速度较快,在各个领域中都得到一定的应用和进一步的研究发展。本文将小波分析技术和 LSSVM 进行了有机的结合,应用小波变换频带分析技术进行设备故障的特征向量提取,并采用 LSSVM 进行机械设备的故障诊断,取得了良好的诊断效果。

2 小波变换频带分析和特征向量提取

Fourier 分析可以作频带能量分析,其得出的不同故障的谱结构特征向量,在实际诊断中已获得了成功的运用。然而 Fourier 分析只是对信号中的正弦信号进行了统计,实际的诊断信号往往包含非平稳成分,这些信号严格讲并不能用正弦信号作为基来描述。如果用正弦信号作为基描述,则能量表示会不全面。而应用小波进行信号分析,可以描述信号的非平稳成分。尤其用小波包分析技术,可以把信号分解在任意精细的频带上。在这些频带上做能量统计,形成特征向量,更趋合理性。

和 Fourier 频谱分析技术一样,小波频带技术的理论依据也是 Parseval 能量积分等式。

因在 $f(x)$ 时域上的能量为:

$$\| f \|^2=\int_{-\infty}^{+\infty} |f(x)|^2 \mathrm{d}x \tag{1}$$

$f(x)$ 的小波变换为:

$$C_{j,k}=W(2^j,2^jk)=2^{-\frac{j}{2}}\int_R \bar{\psi}(2^{-j}x-k)f(x)\mathrm{d}x \tag{2}$$

二者由 Paseral 恒等式关联得:

$$\int_{-\infty}^{+\infty} |f(x)|^2 \mathrm{d}x=\sum |C_{j,k}|^2 \tag{3}$$

由上式可知,小波变换系数 $C_{j,k}$ 具有能量的量刚,因此可

*)本课题为中国博士后科学基金资助项目(编号 2005038515)。杨奎河 教授,博士后,主要研究方向:智能信息处理和故障诊断。单甘霖 教授,博士生导师,主要研究方向:装备智能故障诊断。赵玲玲 副教授,主要研究方向:支持向量机应用技术。

用于能量分析。

机械设备故障诊断技术中对旋转机械故障诊断的研究是最深入和完善的,其应用也是最成功的。把多分辨率分析应用到功率谱的特征提取上,可以方便而有效地提取出特征向量。运用小波包分析技术,可对旋转机械常见的故障进行能量分析。经过大量的实验,可以建立起旋转机械故障原因与征兆对应表[7]。用旋转机械常见的不平衡、不对中、油膜涡动等6种故障作为诊断模型的输出,利用振动信号频谱的9个频段上不同频率谱的谱峰能量值作为特征量,形成旋转机械故障的训练样本,如表1所示。

表1 旋转机械振动故障原因与征兆表

故障样本	0	1	2	3	4	5	6	7	8
频段(f为工频)	0.01～0.39f	0.40～0.49f	0.50f	0.51～0.99f	f	2f	3～5f	奇数倍f	高频>5f
0 不平衡	0.00	0.00	0.00	0.00	0.90	0.05	0.05	0.00	0.00
1 气动力偶	0.00	0.30	0.10	0.60	0.00	0.00	0.00	0.00	0.00
2 不对中	0.00	0.00	0.00	0.00	0.40	0.50	0.10	0.00	0.00
3 油膜涡动	0.10	0.80	0.00	0.10	0.00	0.00	0.00	0.00	0.00
4 转子径向碰摩	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10
5 共生松动故障	0.00	0.00	0.00	0.00	0.20	0.15	0.40	0.00	0.25

3 LSSVM诊断模型

3.1 SVM原理

SVM是从线性可分情况下的最优分类面发展而来的[4]。设有N个样本x_i及其所属类别y_i,表示为$\{(x_i,y_i)\}$,$x\in R^d$,$y_i\in\{1,-1\}$,$i=1,\cdots,N$,d为输入空间的维数。对于标准的SVM,其分类间隔为$\frac{2}{\|w\|}$,使分类间隔最大相当于是$\|w\|^2$最小,因此,使分类间隔最大的优化问题可表示为下面的二次规划问题:

$$\min_{w,b,\xi} J(w,\xi)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i \tag{4}$$

$$\text{Subject to } y_i[w^T\Phi(x_i)+b]\geqslant 1-\xi_i,i=1,\cdots,N \tag{5}$$

$$\xi_i>0,i=1,\cdots,N \tag{6}$$

其中,$\xi_i\geqslant 0(i=1,\cdots,N)$是松弛因子,用来保证在线性不可分情况下分类的正确性;惩罚因子C为某个指定的常数,起到控制对错分样本惩罚程度的作用,实现在错分样本的比例和算法复杂程度之间的折衷。Φ为非线性变换函数,可通过Φ将非线性问题转化为某个高维空间中的线性问题,在变换空间求最优分类面。

设有非线性映射$\Phi:R^d\rightarrow H$将输入空间的样本映射到高维的特征空间H中。当在特征空间H中构造最优超平面时,训练算法仅使用空间中的内积,即$\Phi(x_i)\cdot\Phi(x_j)$,而没有单独$\Phi(x_i)$的出现。因此,如果能够找到一个函数K使得$K(x_i,x_j)=\Phi(x_i)\cdot\Phi(x_j)$,这样,高维空间的内积运算就可以用原空间中的函数实现,我们甚至没有必要知道变换Φ的形式。统计学习理论指出,根据Hilbert-Schmidt原理,只要一个函数$K(x_i,x_j)$满足Mercer条件[5],它就对应某一变换空间中的内积。因此,在最优分类面中用适当的内积核函数$K(x_i,x_j)$就可以实现从低维空间向高维空间的映射,从而实现某一非线性变换后的线性分类,而计算复杂度却没有增加。

SVM算法通过求解上述二次规划问题来实现对样本的正确分类。

3.2 LSSVM分类算法

LSSVM则对SVM进行了改进。对于LSSVM,它用二次损失函数取代SVM中的不敏感损失函数,将不等式约束条件变为等式约束,寻优目标函数式应变为

$$\min_{w,b,\xi} J_{LS}(w,\xi)=\frac{1}{2}\|w\|^2+\frac{1}{2}\gamma\sum_{i=1}^{N}\xi_i^2 \tag{7}$$

$$\text{Subject to } y_i[w^T\Phi(x_i)+b]=1-\xi_i,i=1,\cdots,N \tag{8}$$

其中参数γ是类似SVM中的参数C,用于对$J_{LS}(w,\xi)$进行控制。为求解这个问题优化问题,引入拉格朗日函数:

$$L_{LS}(w,b,\xi,\alpha)=J_{LS}(w,\xi)-\sum_{i=1}^{N}\alpha_i\{y_i[w^T\Phi(x_i)+b]-1+\xi_i\} \tag{9}$$

其中α_i是Lagrangian乘子。在鞍点(极值)处把上式分别对w,b,ξ和α_i求导并令它们等于零,从而得

$$\frac{dL_{LS}}{dw}=0,w=\sum_{i=1}^{N}\alpha_i y_i\Phi(x_i) \tag{10}$$

$$\frac{dL_{LS}}{db}=0,\sum_{i=1}^{N}\alpha_i y_i=0 \tag{11}$$

$$\frac{dL_{LS}}{d\xi}=0,\alpha_{i_i}=\gamma\xi_i \tag{12}$$

$$\frac{dL_{LS}}{d\alpha_i}=0,y_i[w^T\Phi(x_i)+b]-1+\xi_i=0 \tag{13}$$

上述条件可以写为一个线性系统:

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix}\begin{bmatrix} w \\ b \\ \xi \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{1} \end{bmatrix} \tag{14}$$

其中$Z=[\Phi(x_1)^Ty_1,\cdots,\Phi(x_N)^Ty_N]^T$,$Y=[y_1,\cdots,y_N]^T$,$\vec{1}=[1,\cdots,1]^T$,$\xi=[\xi_1,\cdots,\xi_N]^T$,$\alpha=[\alpha_1,\cdots,\alpha_N]^T$,$I\in R^{(N\times N)}$为单位矩阵。消除$w$和$\xi$后式(14)简化为:

$$\begin{bmatrix} 0 & Y^T \\ Y & ZZ^T+\gamma^{-1}I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \tag{15}$$

定义$\Omega=ZZ^T=[q_{ij}]_{N\times N}$,并且对$\Omega$矩阵应用Mercer条件,该矩阵的元素可以表示为如下形式:

$$q_{ij}=y_iy_j\Phi(x_i)^T\Phi(x_j)=y_iy_jK(x_i,x_j) \tag{16}$$

其中$K(x_i,x_j)$为核函数。如果能够找到一个函数K使得$K(x_i,x_j)=\Phi(x_i)\cdot\Phi(x_j)$,那么,在高维空间实际上只需进行内积运算。常用的核函数有:

(1) 多项式核函数 $K(x_i,x_j)=((x_i\cdot x_j)+\theta)^d \quad d=1,2,\cdots$

(2) RBF(Radial Basis Function) 核函数 $K(x_i,x_j)=\exp\left(\frac{-\|x_i-x_j\|^2}{\sigma^2}\right)$

(3) Sigmoid 核函数 $K(x_i,x_j)=\tanh(\beta(x_i\cdot x_j)+\theta)$

式(15)用最小二乘法即可解。在LSSVM算法中,将二次规划问题转变为线性方程组的求解,简化了计算的复杂性。求解上述问题后得到的最优分类函数是:

$$y(x)=sign[\sum_{i=1}^{N}\alpha_i y_j \Phi(x,x_i)+b] \tag{17}$$

3.3 LSSVM 中参数选取算法的改进

由于 RBF 函数计算简单，只有一个标准化参数 σ，本文采用 RBF 函数作为核函数。在 LSSVM 算法中，规则化参数 γ 和 RBF 核函数的标准化参数 σ 一般都是根据经验选取一个固定的值，但是针对不同的样本集，最优的参数值是变化的，因此在一定程度上影响了故障的诊断正确率。经试验发现，规则化参数 γ 的取值对计算的结果影响不明显，取 $\gamma=10$，而随着 σ 的取值不同，计算的结果波动较大。

设诊断出错率 $y=f(x)$ 是区间 $[a, b]$ 上的下单峰函数，在此区间上 t 处函数有唯一的极小点，若在此区间内取任两点 a_1 和 b_1，$a_1<b_1$，并计算函数值 $f(a_1)$ 和 $f(b_1)$，可能出现以下两种情况：

(1) $f(a_1)<f(b_1)$，这时 t 必在区间 $[a, b_1]$ 内。

(2) $f(a_1)\geqslant f(b_1)$，这时 t 必在区间 $[a_1, b]$ 内。

这说明，只要在区间 $[a, b]$ 内取两个不同点，并算出它们的函数值加以比较，就可以把搜索区间 $[a, b]$ 缩小成 $[a, b_1]$ 或 $[a_1, b]$，缩小后的区间仍包含极小点。如果要继续缩小搜索区间 $[a, b_1]$ 或 $[a_1, b]$，就需要在上述区间内再取出一点算出其函数值，并与 $f(a_1)$ 或 $f(b_1)$ 加以比较即可。区间的缩小率与函数的计算次数有关。本文在具体计算时取 $a=0.1$，$b=1$，a_1 和 b_1 取在 $[a, b]$ 的三等分处。

4 仿真结果

为了验证 LSSVM 故障诊断模型对故障的诊断准确率，本文分别用两种诊断模型进行仿真计算。模型一采用概率神经网络(Probabilistic neural networks, PNN)模型[8,9]。PNN 的理论依据是贝叶斯决策理论中的贝叶斯最小风险准则，PNN 可以用来解决分类和故障诊断问题，当训练样本足够多时，PNN 收敛于一个贝叶斯分类器。

模型二采用 LSSVM 故障诊断模型，用 one-against-one 方法实现多类分类。首先用训练样本对两种诊断模型分别进行训练，然后利用训练后的模型对仿真故障进行诊断。

设 D_1 为加入噪声前的仿真样本数据矩阵，D_2 为加入噪声后的仿真样本数据矩阵，仿真需要的含有噪声的样本数据矩阵元素通过下式获得：

$$D_2(i,j)=D_1(i,j)\times(1+\alpha\times rands(1)) \tag{18}$$

其中 α 为噪声控制系数，分别取 $\alpha=0, 0.2, 0.5, 0.8$；$rands(1)$ 为随机函数，可生成一个 -1 到 1 之间的随机函数。

利用上式对每种故障情况分别产生 80 组带噪声的测量参数，共 480 组样本，用 300 组作为训练集，用 180 组作为测试集，在未经任何预处理的情况下直接供 PNN 及 LSSVM 进行诊断，诊断结果如表 2 所示。

表 2 LSSVM 及 PNN 模型诊断结果

噪声控制系数	α=0.0		α=0.2		α=0.5		α=0.8	
诊断模型	PNN	LSSVM	PNN	LSSVM	PNN	LSSVM	PNN	LSSVM
诊断准确率	100	100%	98.2%	98.6%	92.7%	96.3%	81.6%	93.1%

表 2 为 PNN 及 LSSVM 两种网络模型的诊断结果。诊断结果表明，网络的诊断准确率受到噪声控制系数 α 的影响，当样本数据中不包含噪声($\alpha=0.0$)时为理想的输入数据，两种网络都可以达到 100% 诊断准确率；当噪声较小($\alpha=0.2$)时，LSSVM 的诊断准确率为 98.6%，PNN 亦具有很高的诊断准确率，为 98.2%；当测量参数的噪声较大($\alpha=0.5, 0.8$)时，LSSVM 则保持了 96.3% 和 93.1% 的高准确率，而 PNN 只能达到 92.7% 和 81.6% 的诊断准确率。从上面诊断准确率可以看出，虽然随着样本噪声的增加两种模型的诊断准确率都出现下降，但 LSSVM 模型的下降速度明显低于 PNN 模型，显示出较强的鲁棒性。

结论 仿真结果表明 LSSVM 模型的故障诊断能力远大于 PNN 模型，显示了 LSSVM 较强的诊断鲁棒性，LSSVM 是一种可用于模式分类的算法，与 PNN 相比较，SVM 模型是专门针对小样本情况的，其目标是得到现有信息下的最优解而不仅仅是样本数趋于无穷大时的最优值；算法最终将分类问题转化成为一个二次型寻优问题，从理论上说，得到的将是全局最优点，解决了在神经网络方法中无法避免的局部极值问题。LLSVM 则对 SVM 进行了改进，用二次损失函数取代 SVM 中的不敏感损失函数，将不等式约束条件变为等式约束，将二次规划问题转变为线性方程组的求解，用最小二乘法实现了 SVM 算法，简化了计算的复杂性。仿真结果表明，LSSVM 在用于故障诊断时在抗干扰能力及故障识别准确率方面有明显的优势。

参 考 文 献

1 Cortes C, Vapnik V. Support-Vector Network. Machine Learning, 1995, 20: 273～297

2 Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Support Vector Machines. IEEE Trans on Neural Networks, 2002, 13(2):415～425

3 Chuah T C, Sharif B S, Hinton O R. Robust Adaptive Spread-Spectrum Receiver with Neural-Net Preprocessing in Non-Gaussian Noise. IEEE Transactions on Neural Networks, 2001, 12(3): 546～558

4 Chen S, Samingan A K, Hanzo L. Support Vector Machine Multiuser Receiver for DS-CDMA Signals in Multipath Channels. IEEE Transactions on Neural Networks, 2001, 12(3): 604～611

5 Sebald D J, Bucklew J A. Support Vector Machine Techniques for Nonlinear Equalization. IEEE Transactions on Signal Processing, 2000, 48(11): 3217～3226

6 Engel Y, Mannor S, Meir R. The Kernel Recursive Least-Squares Algorithm. IEEE Trans on Signal Processing, 2004, 52(8):2275～2285

7 虞和济，陈长征，张省，等. 基于神经网络的智能诊断. 北京：冶金工业出版社，2002

8 叶志锋，孙健国. 基于概率神经网络的发动机故障诊断. 航空学报，2002, 23(2):155～157

9 李冬辉，刘浩. 基于概率神经网络的故障诊断方法及应用. 系统工程与电子技术，2004, 26(7):997～999

10 Lendasse A, Simon G, Wertz V, et al. Fast Bootstrap for Least-Square Support Vector Machines. Proceedings of European Symposium on Artificial Neural Networks. Bruges, Belgium, April 2004. 525～530

11 Vapnik V, Golowich S, Smola A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. Advances in Neural Information Processing Systems, 1996, 9:281～287

12 Vila J P, Wagner V, Neveu P. Bayesian Nonlinear Model Selection and Neural Networks: A Conjugate Prior Approach. IEEE Trans on Neural Networks, 2000, 11(2):265～278