

# 最小二乘支持向量机方法在农用柴油机故障诊断中的应用研究

谭玉玲 (罗定职业技术学院, 广东罗定 527200)

**摘要** 采用最小二乘支持向量机法对农用柴油机故障进行计算机仿真诊断。结果表明:该方法能够提高故障诊断的准确性,减少误诊。

**关键词** 农用柴油机;最小二乘支持向量机;故障诊断

**中图分类号** S218.5 **文献标识码** A **文章编号** 0517-6611(2009)22-10698-02

**Study on Least Square Support Vector Machine Method and Its Application in Agricultural Diesel Engine Fault Diagnosis**

TAN Yu-ling (Luoding Polytechnic College, Luoding, Guangdong 527200)

**Abstract** Adopting the method of least square support vector machine, the computer simulation diagnosis of agricultural diesel engine fault was performed. The results showed that the accuracy of fault diagnosis was increased and the misdiagnosis was reduced.

**Key words** Agricultural diesel engine; LS-SVM; Fault diagnosis

柴油发动机是农用大型机械的核心动力,从结构上,柴油机由点火系、供油系、冷却系、润滑系、控制系和机械系等几个子系统组成,各子系统又由多种部件构成。由于各个子系统协同工作,相互间的耦合度较大,导致许多故障现象对应的原因可能是同一个系统故障,或者不同子系统出现故障却导致同一参数发生异常变化<sup>[1]</sup>,这使得农用柴油机故障难以检测。

柴油机故障诊断主要采用的方法包括非智能方法和智能方法两种。非智能方法有热力参数分析法和油液分析法。智能方法主要有专家系统和神经网络,专家系统存在专家知识难以获得,规则提取困难,且容易出现匹配冲突、组合爆炸以及无穷递归等问题,因此,专家系统更多的应用是和其他方法相结合。神经网络则存在下列缺点:网络结构的选取基于经验;网络权值学习算法采用经验风险最小化原则,容易产生过学习和欠学习问题,泛化能力差;网络权值求取的非凸问题具有多个最优值,学习过程中容易陷入局部最优解。

目前,农用柴油机的检测设备不完全,诊断系统的征兆数据属于小样本集。支持向量机(SVM)方法<sup>[2]</sup>是建立在统计学习理论和结构风险最小原理的基础上,该方法较好地解决了小样本、非线性、局部极小点等实际问题。最小二乘支持向量机(LS-SVM)是标准SVM的一种扩展,优化指标采用平方项,并用等式约束代替标准SVM的不等式约束,即将二次规划问题转化为线性方程组求解,降低了计算复杂性,加快了求解速度。因此,LS-SVM可用于柴油机故障诊断。

## 1 LS-SVM 算法

支持向量机方法是 Vapnik 等在 20 世纪 70 年代末提出的,该方法是在有限样本统计学习理论上发展起来的一种新的机器学习方法,本质上是一种线性二值分类器。对于非线性分类问题,可以通过核函数将非线性输入空间映射到一个高维线性空间,然后,在这个新空间中求取最优分类面。支持向量机法的特点是:将最优分类面求取问题转化为二次规划求解问题,根据求解得到最接近分类面的支持向量,分类判别函数是支持向量的线性组合。由于支持向量机法获得的是最优分类面,因此,有较好的分类能力和容错性。

与其他方法相比,支持向量机方法有以下优点:①它是专门针对有限样本情况的,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;②算法最终将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;③算法将实际问题通过非线性变换转换到高维的特征空间,在高维空间中构造线性判别函数来实现原空间中的非线性判别函数,保证机器有较好的推广能力,同时,它巧妙地解决了维数问题,其算法复杂度与样本维数无关。

LS-SVM 方法是 Suykens 等人在 SVM 方法的基础上提出的一种改进方法,LS-SVM 方法用最小二乘法进行参数估计,替代了原 SVM 通过二次规划求解分类面参数,减小了计算代价,能够进一步提高农用柴油机故障诊断的计算效率,其原理如下。

农用柴油机故障诊断问题是一类非线性不可分问题,因此,选择一个非线性变换  $\theta(\cdot)$  将  $n$  维输入、输出样本向量  $(x_i, y_i)$  从原空间映射到高维特征空间  $F$  (其中  $x_i \in R^n$  为输入向量,  $y_i \in [-1, 1]$  为输出类别,  $i = 1, 2, \dots, N, N$  为训练样本个数),在此高维特征空间构造最优线性分类函数:

$$f(x) = \text{sign}[w \cdot \theta(x) + b] \quad (1)$$

式中,  $w$  为权值向量;  $b$  为偏差。

根据结构风险最小化原则,寻找  $w$  和  $b$  就是最小化

$$R = \|w\|^2/2 + C \cdot R_p \quad (2)$$

式中,  $\|w\|^2$  为控制模型的复杂度;  $C$  为控制对输出误差的样本惩罚程度;  $R_p$  为误差控制函数。选取不同的损失函数,可构造不同形式的 SVM。LS-SVM 在优化目标中选择的损失函数为误差  $\xi_i$  的二范数。因此优化问题为

$$\begin{cases} \min & Q(w, b, \xi) = \frac{1}{2} w^T w + \frac{1}{2} C \sum_{i=1}^N \xi_i^2 \\ \text{st} & y_i [w^T \theta(x_i) + b] = 1 - \xi_i \quad i = 1, 2, \dots, N \end{cases} \quad (3)$$

一般地,由于  $w$  可能为无限维,直接求解式(3)的优化问题极其困难。因此将求解式(3)的优化问题转化为求解如下线性方程组

$$\begin{bmatrix} 0 & 1^T \\ 1 & K + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (4)$$

式中,  $1 = [1, \dots, 1]^T$ ;  $K = \theta(x_i)^T \theta(x_j)$  是核函数,选用径向基核函数(RBF),其中  $K = \frac{-\|x - x_i\|^2}{2(\delta)^2}$ ,  $\delta$  为核宽度;  $Y = [y_1,$

$\dots, y_N]^T; \alpha = [\alpha_1, \dots, \alpha_N]^T; I$  为单位矩阵。采用最小二乘法得到  $b$  和  $\alpha$  值, 进一步得到农用柴油机故障分类函数:

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

### 2 仿真研究

福田谷神 4YD-2B 玉米收割机的柴油机燃油系统故障总会引起高压油管中燃油压力波形的变化, 同一类故障基本上具有相似波形。笔者从柴油机燃油压力的时域波形上提取波形的 8 个最能体现柴油机运行状况的结构特征点。即, 最大压力( $P1$ )、次最大压力( $P2$ )、波形幅度( $P3$ )、上升沿宽度( $P4$ )、波形宽度( $P5$ )、最大余波的宽度( $P6$ )、波形的面积( $P7$ )和起喷压力( $P8$ )。笔者将燃油系统常见的故障分为 8 种, 主要有 100% 供油量(正常情况  $T1$ ), 75% 供油量( $T2$ ), 25% 供油量( $T3$ ), 怠速油量( $T4$ ), 针阀卡死(小油量  $T5$ ), 针阀卡死(标定油量  $T6$ ), 针阀泄漏( $T7$ )和出油阀失效( $T8$ )。根据现场专家判断和分析共选得到 1 305 组故障数据, 其中  $T1$  数据 326 组,  $T2$  数据 141 组,  $T3$  数据 71 组,  $T4$  数据 167 组,  $T5$  数据 121 组,  $T6$  数据 179 组,  $T7$  数据 108 组和  $T8$  数据 182 组。

根据柴油发动机工作原理和故障征兆描述分析确定诊断系统输入参数为: 最大压力( $P1$ )、次最大压力( $P2$ )、波形幅度( $P3$ )、上升沿宽度( $P4$ )、波形宽度( $P5$ )、最大余波的宽度( $P6$ )、波形的面积( $P7$ )和起喷压力( $P8$ )。为了求取最优

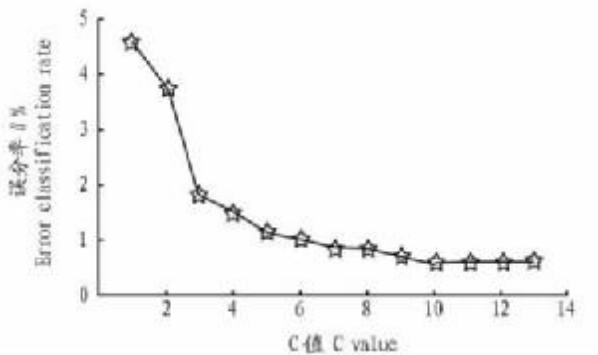


图 1  $C=11$  时 LS-SVM 误分率

Fig. 1 LS-SVM error classification rate when  $C$  is equal to 11

表 2 故障诊断估计结果误差对比

Table 2 Error comparison of fault diagnosis estimation results

误差 Error	LS-SVM		神经网络 Neural network		专家系统 Expert system		
	拟合数据	预测数据	拟合数据	预测数据	拟合数据	预测数据	
	Fitting data	Forecasting data	Fitting data	Forecasting data	Fitting data	Forecasting data	
绝对误差 Absolute error	max	0.640	0.233	0.754	0.504	0.911	0.838
	min	0.001	0.003	0.016	0.012	0.004	0.014
	ave	0.105	0.109	0.177	0.220	0.191	0.275
相对误差 Relative error	max	1.69	0.49	1.6	0.8	1.93	1.75
	min	0.00	0.01	0.0	0.0	0.01	0.03
	ave	0.57	0.23	0.41	0.4	0.4	0.57

### 3 结论

通过采样数据进行系统仿真, 结果表明: 人为确定的参数(如 RBF 核函数的  $C$  值和  $\delta$  值)对分类效果有很大的影响; 优化的参数下, 正常状况诊断准确率高于 98%, 各异常状况诊断准确率高于 90%; 同时, LS-SVM 与传统的神经网络和专家系统相比, 估计误差指标更好。该方法不仅可以用于柴

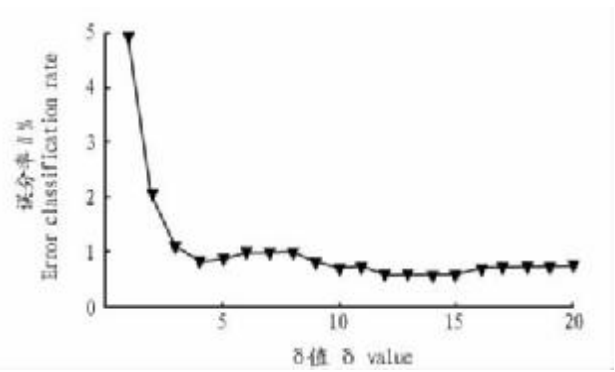


图 2  $\delta=13$  时 LS-SVM 误分率

Fig. 2 LS-SVM error classification rate when  $\delta$  is equal to 13  
分类器参数, 采用  $k$ -折交叉验证( $k=5$ ), 选取误判、错判最少的作为最终的分类器。LS-SVM 方法的参数  $C$  和径向基函数的  $\delta$  由人工给定, 因此, 它们的选取对分类结果有一定的影响。从图 1、图 2 可以看出, 通过分析, 在  $C$  值和  $\delta$  值较大或较小时, 分类错误率都很高, 最终选取了合适的  $C$  值和  $\delta$  值, 使各分类器具有很好的分类效果, 其分类结果如表 1 所示。

表 1 各分类器在 5-折交叉验证下的分类效果

Table 1 The classification effects of each classifier under the 5-fold cross verification

序号 Serial No.	故障诊断误分率 // % Fault diagnosis error classification rate								
	T0	T1	T2	T3	T4	T5	T6	T7	T8
1	1.39	5.7	6.3	6.2	3.9	4.5	2.9	2.1	2.6
2	0	6.3	6.3	5.0	3.7	4.6	3.3	2.1	2.7
3	0	8.3	5.6	4.8	4.2	5.8	2.8	2.8	2.5
4	0.69	4.2	6.3	6.4	3.7	6.1	3.4	3.5	3.0
5	0.69	4.2	5.6	5.6	3.8	6.0	3.2	4.9	3.1
平均 Average	0.6	5.7	6.0	5.6	3.86	3.6	3.2	3.1	2.8

从仿真结果可见, 柴油机正常情况的诊断准确率为 98%, 即, 异常状况的诊断准确率为 98%, 只是在各种异常运行状况的判断上, 误报率略有上升, 但仍能达到 90% 以上。将同种数据代入其他方法, 可以看出采用 LS-SVM 方法的故障诊断误差率明显小于其他方法。

油机诊断, 还可以进一步推广到其他领域的诊断中。

### 参考文献

[1] 胡奕涛, 武和雷. 柴油机故障的混合智能诊断方法研究[J]. 小型内燃机与摩托车, 2003, 32(4): 11-14.  
[2] 肖燕彩, 陈秀海, 朱衡君. 基于最小二乘支持向量机的变压器油中气体浓度预测[J]. 电网技术, 2006, 30(11): 91-94.