

一种小样本民机产品的可靠性预测方法

王 焱 左洪福 蔡 景 吕德峰

(南京航空航天大学民航学院,南京,210016)

摘要:将最小二乘支持向量机(Least square support vector machine, LS-SVM)应用于小样本民机产品的可靠性预测分析。通过重构相空间的饱和嵌入维数,确定最小二乘支持向量机的最佳输入变量;然后,使用最小二乘向量机建立可靠度回归预测模型,运用自动网格搜索法,优化了最小二乘支持向量机的建模参数,实现了比现有方法精度高、泛化性好的模型。训练和测试的可靠性样本取自某机型襟翼液压锁寿命可靠性数据。与神经网络模型比较实例表明,提出的方法合理有效。

关键词:小样本;民机产品;可靠性预测;支持向量机;参数优化

中图分类号:TP18 **文献标志码:**A **文章编号:**1005-2615(2014)01-0170-05

Reliability Prediction Algorithm for Civil Aircraft Products Based on Small Sample

Wang Ye, Zuo Hongfu, Cai Jing, Lü Defeng

(College of Civil Aviation, Nanjing University of Aeronautics & Astronautics, Nanjing, 210016, China)

Abstract: The least square support vector machine (LS-SVM) is used to predict the reliability of civil aircraft products based on small sample. The optimized input variable number of LS-SVM is determined through computing the saturated embedding dimension of reconstruct phase space. Then, a reliability prediction model is established by using LS-SVM and their parameters are also optimized by an automatic grid search method. The training and validation use reliability data from the hydraulic lock of a certain type aircraft. Finally, the one step and N -step prediction results of LS-SVM and radical basis function (RBF) neural network are compared, and show that the algorithm is feasible and valid for reliability prediction based on small sample.

Key words: small sample; civil aircraft product; reliability prediction; support vector machine; parameter optimization

经典的可靠性预测分析方法是建立在渐进理论基础之上的,依据中心极限定理,认为对象的分布和参数是一个与观测过程无关的确定量,当测量数据样本达到一定数目以后,估计值接近于被估参数的真值,运用经典法进行可靠性预测分析评估时,需要较大数目的测试样本量,才能得到置信程度较高的评估结论。而在目前的支线飞机和大型

客机的研制过程中,相当多的重要零部件都是小批量、极小批量或单个产品生产,不可能进行满足一定量级的可靠性试验,而且由于民机产品的可靠性高,在整个寿命过程中很少出现故障(譬如:民机的襟翼结构),即使部分产品存在可靠性不足的问题,也通过预防性维修进行了弥补,所以无论从可靠性试验还是从使用维修中,都难以收集大量的故障数

基金项目: 国家自然科学基金(61079003, 61079013)资助项目; 国家高技术研究发展计划(“八六三”计划)(2006AA04Z427)资助项目。

收稿日期: 2013-10-28; **修订日期:** 2014-01-06

通信作者: 左洪福,男,教授,博士生导师,1959年生, E-mail: rms@nuaa.edu.cn。

据,从而给民机产品可靠性预测增加了难度。目前针对民机产品的可靠性预测大都是基于大数据量的,譬如基于发动机排气温度(Exhaust gas temperature, EGT)的发动机性能预测等。因此,迫切需要探索民机产品小子样可靠性预测分析评估的方法。

在机器智能研究领域, Vapnik 等提出了统计学习理论(Statistical learning theory, SLT)^[1], SLT 具有完备的理论基础和严格的理论体系,首次强调了小样本统计学的问题,被认为是目前针对小样本统计估计和学习预测的最佳理论。Vapnik 在这一理论基础上,提出一种新的通用学习方法——支持向量机(Support vector machine, SVM)^[2],在样本数有限的情况下, SVM 出色的推广能力使其在小子样可靠性预测分析方面具有独特的优势^[3-6]。

本文在研究了最小二乘向量机(Least square support vector machine, LS-SVM)理论^[7]的基础上,优化其参数,建立了优化 LS-SVM 的可靠度预测模型,并用民机液压锁的小样本可靠性寿命信息对 LS-SVM 预测模型与 RBF 预测模型进行检测,由此验证了所提出方法的先进性和有效性。

1 最小二乘支持向量机模型及优化

1.1 寿命时间序列预测模型框架

由 Kolmogorov 定理可知,任何一个时间序列都可以看成是由一个非线性机制 F 确定的输入输出系统。因此,时序预测本质就是依据历史数据序列寻求映射 $f: \mathbf{R}^m \rightarrow \mathbf{R}$,逼近数据中的非线性机制 F ,采用 f 作为理想预测器。故提出的基本框架是:对时间序列进行相空间重构,构建训练样本数据对,选择合适的函数逼近工具,进行参数估计和拓扑结构确定。最后,根据确定的预测器构建预测模型进行序列分析。框架结构如图 1 所示。

1.2 LS-SVM 回归算法

LS-SVM 使用最小二乘系统作为损失函数,代替了 SVM 中的不等式约束,通过解线性方程组,来解决标准 SVM 中的二次型规划问题。在应用中表现出运算快和精度高的优点^[7-8]。时序预测问题建立在函数估计基础之上,根据训练数据建立拟合函数,使用测试数据验证拟合效果,得到所需要的模型,分析实际需要预测的数据。在 LS-SVM 中,函数估计问题表述如下。给定 n 对数据样本 $(x_i, y_i), i=1, \dots, N, x \in \mathbf{R}^n, y \in \mathbf{R}$ 。对偶空间最优化问题为

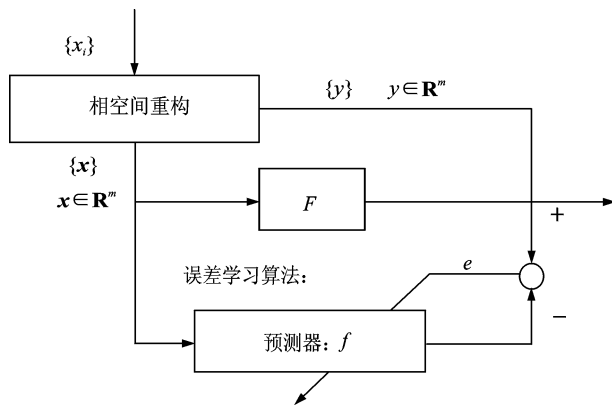


图 1 寿命时间序列预测建模的框架

Fig. 1 Framework of life time series forecast modeling

$$\min J(\mathbf{w}, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad (1)$$

$$\text{s. t. } y_k = \mathbf{w}^T \Phi(\mathbf{x}_k) + b + e_k$$

式中: $\Phi(\mathbf{x})$ 为一个从输入空间到高维特征空间的映射函数; 权向量 $\mathbf{w} \in \mathbf{R}^n$; γ 为可调整正则化参数; 误差变量 $e_k \in \mathbf{R}$; b 为常值偏差。为求解上述优化问题,把约束优化问题变成无约束优化问题,建立 Lagrange 函数

$$L(\mathbf{w}, b, e_k; \mathbf{a}) = J(\mathbf{w}, e) - \sum_{k=1}^N a_k \times \{ \mathbf{w}^T \varphi(\mathbf{x}_k) + b + e_k - y_k \} \quad (2)$$

式中: Lagrange 乘子 $a_k \in \mathbf{R}$ 。根据 Karush-Kuhn-Tucker(KKT)条件,求 L 对 \mathbf{w}, b, e_k, a_k 偏导为 0,消除变量 \mathbf{w}, e , 可得如下方程

$$\begin{bmatrix} 0 & \mathbf{I}^T \\ \mathbf{I} & \Omega + \frac{1}{\gamma} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (3)$$

式中: \mathbf{y} 为样本输出; \mathbf{I} 为单位矩阵; $\mathbf{a} = [a_1, a_2, \dots, a_n]$; $\Omega_{kl} = K(\mathbf{x}_k, \mathbf{x}_l) = (\mathbf{x}_k)^T (\mathbf{x}_l), k, l=1, \dots, N$ 。因为上述线性系统非奇异,所以可以得到 b, \mathbf{a} 的解。从而得到 LS-SVM 函数估计表达式为

$$y(\mathbf{x}) = \sum_{k=1}^N a_k K(\mathbf{x}_k, \mathbf{x}_l) + b \quad (4)$$

式中: $K(\mathbf{x}_k, \mathbf{x}_l)$ 为满足 Mercer 条件的核函数,常用的有多项式核函数、径向基核函数和双曲正切核函数。径向基核函数能够实现非线性映射且参数较少,故本文使用高斯径向基核函数,在处理非线性样本中具有优越性

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2) \quad (5)$$

LS-SVM 算法优化问题转化为以最小二乘法求解式(3)表示的线性方程组,避免了标准支持向

量机中求解凸二次规划问题,解线性方程组比求解二次规划更为简单快速。

1.3 建模参数的优化

LS-SVM 建模中,正则化参数 γ 及核参数 σ 的选择属于二维最优化问题,如何优化建模参数 γ, σ 具有重要意义,已有学者进行了探索^[9],但仍缺乏明确的理论指导。因为无法得到建模参数之间的显示表达式,而且不能保证变化的连续性单调性,所以不能使用最陡剃度法、牛顿法等最优化方法。有学者^[10-13]利用遗传算法(Genetic algorithm, GA)和模拟退火算法(Simulated annealing, SA)优化模型参数,并在理论上获得了较好的效果,但实际应用中,GA 容易陷入局部最优,稳定性不高^[10],SA 优化的实现算法比较复杂^[12]。网格搜索法(Grid search, GS)易于实现,广泛应用于工程实际,但存在搜索速度问题。本文运用自动网格搜索法,根据模型误差自动调整搜索步长,从而在保证寻优精度的基础上提高搜索速度。

均方根误差(Root mean square error, RMSE)可用于评价模型泛化性能^[2],本文在固定步长网格搜索法的基础上,以 RMSE 为目标函数,自适应调整搜索步长,实现建模参数的最优化。给定数据样本后,目标函数 RMSE 由 LS-SVM 模型计算而得,其自变量为 γ, σ 。从而在 RMSE 取极小值时, γ, σ 为 LS-SVM 建模参数。RMSE 函数记为 $R = f(\mathbf{p})$, \mathbf{p} 为二维参数组合向量, $\mathbf{p} = (\gamma, \sigma)$ 。优化算法如下:

(1) 初始化搜索区域 p_0, p_u , 步长 Λ_0 , 极小值 R_{\min} ;

(2) $p_i = p_{i-1} + \Lambda_{i-1}, i = 1$;

(3) 计算 $\Delta R_i = f(p_i) - f(p_{i-1})$;

(4) 若 $\Delta R_i < 0$, 则 $R_{\min} = f(p_i)$;

(5) 更新步长 $\Lambda_i = \Lambda_{i-1} + h(\Delta R_i)$;

(6) 判断 $p_i + \Lambda_i < p_u$, 如是, 则置 $i = i + 1$, 转步骤(3), 如否, 则转步骤(7);

(7) R_{\min} 为搜索所得目标函数极小值, 对应的 σ 即为 LS-SVM 建模参数。

其中

$$h(\Delta R_i) = \frac{\Delta R_i}{(4|\Delta R_i|)} \Lambda_{i-1} \quad (6)$$

自动优化过程中, 根据目标函数 RMSE 变化, 自动地调整网格搜索步长 Λ 。在目标函数下降较快的区域, 减小搜索步长, 下降较慢的区域, 增加搜索步长。与普通网格搜索法相比较, 自动网格搜索法可明显减少搜索次数。

后面将验证步长的自适应调整, 它既提高了网格搜索速度, 又保证了最优化效果; 同时, 由于网格搜索法是全局范围内的直接寻优, 无需目标函数具有单调、连续可导等性质。

2 基于优化 LS-SVM 的可靠度预测模型

LS-SVM 机应用小样本预测的效果很好, 能够避免学习不足的问题, 且民机产品本身寿命数据样本量不大, 所以此方法适合于民机产品的可靠性预测分析。根据上文提出的优化参数后的 LS-SVM 算法, 民机产品的可靠度模型建立如下:

(1) 根据需要预测的可靠度问题, 确定 LS-SVM 系统相关的输入输出变量;

(2) 对原始数据进行预处理, 并构建 LS-SVM 学习样本;

(3) 选择径向基核函数, 初始化正则化参数以及核参数 σ , 利用上述优化方法得到 σ ;

(4) 建立优化 LS-SVM 的可靠度预测模型, 并对民机产品的可靠度问题进行预测。

3 民机襟翼某液压锁的可靠性预测

3.1 可靠性数据的预处理

航空公司某型号民机的襟翼液压锁, 襟翼液压锁是襟翼的关键部件, 直接影响襟翼是否能正常放下和收起, 真正因为其关键性作用, 液压锁失效数据很少, 属于小样本数据, 现收集了 36 个平均无故障间隔时间数据, 1.4, 1.7, 2.0, 2.4, 2.6, 2.8, 3.1, 3.2, 3.4, 3.7, 3.9, 4.0, 4.2, 4.6, 4.8, 5.1, 5.3, 5.6, 5.7, 5.9, 6.3, 6.8, 7.1, 7.4, 7.6, 7.9, 8.1, 8.2, 8.4, 8.5, 8.8, 9.0, 9.3, 9.5, 9.6, 9.9 (1 000 h)。上述数据反应了襟翼液压锁的平均故障时间, 即可靠性水平, 根据秩评定法^[14]对其进行分析整理, 分别得到其对应的可靠度。

3.2 LS-SVM 的可靠度预测模型参数的优化

将数据分成两组, 前 30 个数据作为训练数据, 后 6 个数据作为预测数据测试样本集。采用 FPE 准则优化选取支持向量回归机的嵌入维数, 得到 FPE 值与嵌入维数 m 之间的关系。得出当 $m = 6$ 时, FPE 具有最小值, 由确定维数 $m = 6$ 来构造建模数据集, $x_i = f(x_{i-1}, x_{i-2}, \dots, x_{i-6})$, 进行 LS-SVM 建模, 核函数为高斯径向基函数核。对正则化参数 C 与核宽度参数进行自适应搜索寻优。搜索初始值设置如下: $C \in [0, 600]$, $\sigma \in [0, 60]$,

搜索步长 $\Delta_c = 1, \Delta_s = 0.1$, 搜索过程中, Δ 随 RMSE 变化而自适应调整。优化过程中, 当 $C = 407, \sigma = 0.7$ 时, 目标函数 RMSE 有最小值 1.58。根据自适应搜索算法可知, LS-SVM 建模参数的优化过程是非单调的, 该自动网格搜索法在目标函数下降, 有可能出现极小值的区域降低了搜索步长, 以提高精度, 而在目标函数稳定和上升的区域, 逐渐增加搜索步长, 减少搜索计算量。

3.3 可靠性预测模型的性能分析

为分析支持向量回归机与神经网络算法的预测效果, 取 RBF 神经网络模型和 LS-SVM 模型比较。取均方根相对误差(Root mean square relative error, RMSRE)来评价两种模型的预测效果

$$RMSRE_p = \sqrt{\frac{1}{N - n_{tr}} \sum_{i=1}^N \left(\frac{x_i - y_{i,p}}{x_i} \right)^2} \quad (7)$$

式中: x_i 为测试样本; $y_{i,p}$ 为 x_i 的 P 步逼近值, 即 LS-SVM 与 RBF 神经网络模型的输出值。

根据本文建立的 LS-SVM 可靠度回归模型训练相空间重构后的样本集, 并同时使用 Matlab 所带的 RBF 程序包训练样本集, 训练集同样为 30 个可靠性寿命数据, 测试集为 6 个可靠性寿命数据。两种方法得出的预测数据如表 1 所示。

表 1 LS-SVM 模型与 RBF 模型 P 步预测结果

Tab. 1 P -step forecasting results of LS-SVM and RBF models

$R(T_i)$ 实际值	$R(T_i)$ 单步测试		$R(T_i)$ 6 步预测	
	LS-SVM	RBF	LS-SVM	RBF
0.710 1	0.710 2	0.692 7	0.699 3	0.710 1
0.711 2	0.712 8	0.689 6	0.704 1	0.711 2
0.691 9	0.687 3	0.682 5	0.696 7	0.691 9
0.687 7	0.673 6	0.678 1	0.671 4	0.687 7
0.672 1	0.665 7	0.691 4	0.659 3	0.672 1
0.652 6	0.649 5	0.663 1	0.647 4	0.652 6

所谓单步预测, 是指对 T 时刻以前的所有样本学习后, 只用于预测 T 时刻的可靠度 $R(T)$; 而多步预测, 则要用于紧随其后的多个时刻 $T, T+1, \dots, T+k$ 的可靠度预测。图 2 给出了单步预测的结果, 图 3 为多步(6 步)预测的结果。

一步预测中, LS-SVM 和 RBF 网络的 RMSRE 值分别为 2.9%, 5.1%。LS-SVM 回归模型一步预测误差稳定在较小范围内, RBF 神经网络在某些点上的预测误差相对较大。进一步考察两种模型长期预测能力, 对两种方法 P 步预测的 RMSRE 指标进行比较, 得 RMSRE 的值如表 2 所示。

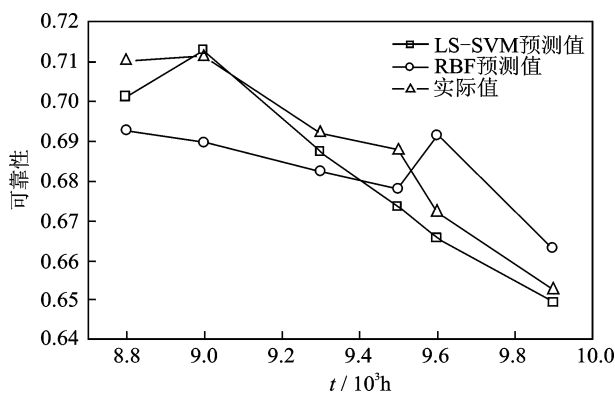


图 2 单步预测结果比较

Fig. 2 Comparison of single-step forecasting results

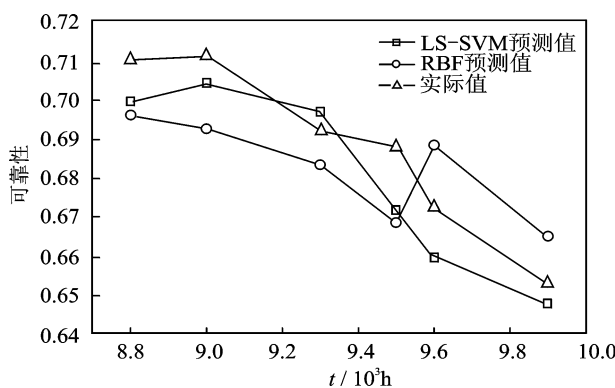


图 3 6 步预测结果比较

Fig. 3 Comparison of six-step forecasting results

表 2 LS-SVM 与 RBF 预测 RMSRE 效果表

Tab. 2 RMSRE forecasting results of LS-SVM and RBF models

P 步	1	2	3	4	5	6
SVM	2.9	3.5	4.3	5.1	5.7	6.6
RBF	5.1	6.2	7.9	9.3	11.8	13.9

RMSRE 分布如图 4 所示, 方块为 LS-SVM 模型的 RMSRE, 圆圈为 RBF 模型的 RMSRE, 分别比较两种模型从 1 步预测到 6 步预测的泛化能力。

可见, 多步预测中, LS-SVM 模型的 RMSRE 值稳定在较小的范围内。可知, LS-SVM 对于可靠性的短期预测与长期预测, 都具有较好的预测精度, 预测误差较为稳定。RBF 模型的 RMSRE 值逐渐变大, 特别是在长期预测中, 其误差明显要大于 LS-SVM 模型, 这是由于神经网络模型可能出现的过学习或学习不足的现象引起的, 网络结构的选择同样也无确定的方法, 而 LS-SVM 在“结构风险最小化”概念下, 建立了一种具有稀疏性质的网络模型, 具有较好的可靠度预测性能。因此, LS-SVM 算法更加适合于此类数据的回归建模。

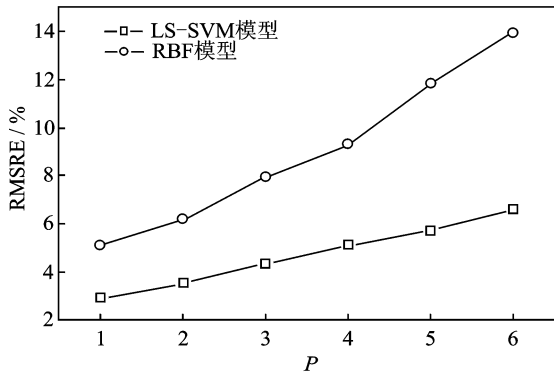


图4 SVM和RBF模型 P 步预测比较

Fig. 4 P -step forecasting comparison between LS-SVM and RBF models

4 结 论

本文建立了基于优化最小二乘支持向量机的民机产品可靠度时序预测模型,提出了一种新的可靠性预测分析方法,结论如下:

(1)实例验证表明此模型具有较高的短期预测和长期预测精度。利用襟翼液压锁的寿命数据,对液压锁的可靠度进行预测,具有很高的应用价值,预测结果可为小样本民机部件的可靠性和维修性分析提供良好的数据基础。

(2)采用相空间重构的方法对寿命时间序列重构能够获取更多有用信息。

(3)与RBF网络模型相比,本文方法具有很强的推广能力,在短期预测中相差不大,但长期预测中优化LS-SVM模型效果明显占优。

参考文献:

[1] Vapnik V. The nature of statistical learning theory [M]. New York: Springer, 1999.

[2] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation, and signal processing[C]// Proceedings of the 1996 Neural Information Processing Systems Conference NIPS. Dever: MIT Press, 1996: 281-287.

[3] Pai Pingfeng. System reliability forecasting by support vector machines with genetic algorithms[J]. Mathematical and Computer Modelling, 2006, 43(3/4): 262-274.

[4] Chen Kuanyu. Forecasting system reliability based on support vector regression with genetic algorithms [J]. Reliability Engineering & System Safety, 2007, 92(4):423-432.

[5] Hong Weichiang, Pai Pingfeng. Predicting engine reliability by support vector machines [J]. Int J Adv Manuf Technol, 2005, 17(4):2340-2348.

[6] Rocco C M, Moreno J A. Fast Monte Carlo reliability evaluation using support vector machine[J]. Reliability Engineering and System Safety, 2002, 76(3): 237-243.

[7] Suykens J, Gestel T. Least squares support vector machines[M]. Singapore: World Scientific Pub Co, 2002.

[8] Gestel T, Suykens J, Baestaens D. Financial time series prediction using least squares support vector machines with in the evidence framework [J]. IEEE Transaction on Neural Networks, 2001, 12(4):809-821.

[9] Chapelle O, Vapnik V. Choosing multiple parameters for support vector machines [J]. Machine Learning, 2002, 46(3):131-159.

[10] Liang Xuefeng, Liu Fang. Choosing multiple parameters for SVM based on genetic algorithm signal processing[C]// 2002 6th International Conference on signal Processing Proceedings. Beijing: Institute of Electrical and Electronics Engineers Inc., 2002: 117-119.

[11] 秦业,袁海斌,袁海文,等. SVM和神经网络在电能质量扰动分类应用中的对比[J]. 南京航空航天大学学报,2011,43(S):74-78.

Qin Ye, Yuan Haibin, Yuan Haiwen, et al. Comparison study on classification of power quality disturbances with SVM and neural network[J]. Journal of Nanjing University of Aeronautics and Astronautics, 2011, 43(S):74-78.

[12] Chuang Chenchia, Su Shunfeng, Jeng Jintsong, et al. Robust support vector regression networks for function approximation with outliers [J]. IEEE Transaction on Neural Networks, 2002, 13(6): 1322-1330.

[13] Michinari M, Kristin P. A pattern search method for model selection of support vector regression [C]// Proceedings of the SIAM International Conference on Data Mining. Arlington: SIAM, 2002: 247-260.

[14] 麻晓敏,张士杰,胡丽琴,等. 可靠性数据威布尔分析中秋评定算法改进研究[J]. 核科学与工程, 2007, 27(2):152-155.

Ma Xiaomin, Zhang Shijie, Hu Liqin, et al. An improved rank assessment method for Weibull analysis of reliability data[J]. Chinese Journal of Nuclear Science and Engineering, 2007, 27(2):152-155.