

doi: 10.3969/j.issn.1006-1576.2011.09.007

基于遗传算法优化支持向量机的航空装备维修费用预测

高鹏^{1,2}, 孙德翔³, 邢国平³, 黄勇¹(1. 空军航空大学研究生队, 长春 130022; 2. 中国人民解放军 93115 部队, 沈阳 110031;
3. 空军航空大学训练部, 长春 130022)

摘要: 针对传统航空装备维修费用预测方法难以计算得到满意结果的问题, 建立遗传算法优化支持向量机的航空装备维修费用预测模型。将遗传算法与支持向量机相结合, 利用遗传算法对支持向量机的参数进行优化, 通过实例对 GA-SVM 模型的应用进行分析对比。结果表明: 在航空装备维修费用预测中, 该模型比 SVR、BP 神经网络、偏最小二乘回归以及传统普通多元线性回归方法, 具有更高预测精度和泛化能力。

关键词: 遗传算法; 支持向量机; 维修费用

中图分类号: TJ85 **文献标志码:** A

Maintenance Cost Forecast of Aviation Equipment Based on Genetic Algorithm Optimization SVM

Gao Kun^{1,2}, Sun Dexiang³, Xing Guoping³, Huang Yong¹(1. Brigade of Postgraduate, Air Force Aviation University, Changchun 130022, China;
2. No. 93115 Unit of PLA, Shenyang 110031, China;
3. Training Department, Air Force Aviation University, Changchun 130022, China)

Abstract: Aiming at the problem that traditional forecast models can't get satisfactory results, support vector machine (SVM) optimized by genetic algorithm (GA) was built as a forecast model for maintenance cost of aviation equipment. The model combined GA and SVM, which used GA to optimize the parameters of SVM. Through the example analysis and comparison, results show that the model has more accurate results and extensibility than SVR, BP network, partial least squares regression and multivariate linear regression in the forecast of maintenance cost of aviation equipment.

Keywords: genetic algorithm; support vector machine; maintenance cost

0 引言

航空装备的维修保障费用是航空装备寿命周期费用的一部分。通过对航空装备寿命周期费用的全过程分析可知, 维修保障费用在其寿命周期费用中居主要地位, 军用飞机的使用维修费通常占寿命周期费用 65%~80%。因此, 准确地估算装备维修保障费用对寿命周期费用控制、新型武器装备研制方案决策和现役装备的科学管理有十分重要的意义^[1-2]。

航空装备维修费用预测是根据装备设计概要的情况, 初步估算航空装备(主要指各型飞机)的维修保障成本。由于维修保障费用与其影响因素之间存在着复杂的非线性关系, 还没有很好地解决非线性关系的模拟和识别及其全局优化问题, 而支持向量机这种机器学习方法具有非线性、自学习的功能, 可以用于解决该类问题。因此, 笔者针对军用飞机维修保障费用建模所面临的“小样本多元数据”问题, 通过遗传算法确定支持向量机的参数, 把遗传算法优化支持向量机引入到航空装备维修费用预测

的研究中, 建立了优化模型。

1 支持向量机

将训练样本集表示为: $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \in R^n \times R$, 其中, x_i 为输入变量; y_i 为对应的输出变量。对于非线性回归, SVM 将数据 x 映射到高维特征空间 F , 并在 F 中用估计函数 $f(x) = [\omega\phi(x)] + b$ 进行线性回归^[3], 其中, $\omega \in F$; ϕ 为从 R^m 空间到 F 空间的非线性映射; b 为偏置量。

函数拟合问题可以表示为

$$R_{reg}(f) = R_{emp}[f] + \lambda \|\omega\|^2 = \sum_{i=1}^l C(e_i) + \lambda \|\omega\|^2 \quad (1)$$

其中: $R_{reg}(f)$ 为目标函数; R_{emp} 为经验风险; λ 为常数; l 为样本数量; C 为错误惩罚因子; $\|\omega\|^2$ 反映 f 在高维空间平坦的复杂性。引入不敏感损失函数 ϵ , 由于其具有较好的稀疏性, 可得到损失函数和经验风险函数分别为

收稿日期: 2011-05-27; 修回日期: 2011-06-20

作者简介: 高鹏(1978—), 男, 辽宁人, 硕士, 工程师, 从事航空装备管理研究。

$$|y - f(x)| = \max\{0, |y - f(x) - \varepsilon|\} \quad (2)$$

$$R_{emp}^\varepsilon[f] = \frac{1}{l} \sum_{i=1}^l |y_i - f(x_i)|_\varepsilon \quad (3)$$

根据统计学理论,引入松弛因子 ξ_i^* 和 ξ_i , 支持向量机回归问题转化为对以下目标函数的最小化问题。

$$\begin{aligned} & \min \left\{ \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i^* - \xi_i) \right\} \\ & s.t. \begin{cases} y_i - \langle \omega \varphi(x_i) \rangle - b \leq \varepsilon + \xi_i^* \\ \omega \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i \quad (i=1, 2, \dots, l) \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (4)$$

利用拉格朗日算子 a_i 和 a_i^* 将其转化为对偶形式:

$$\begin{aligned} & \max \left[-\frac{1}{2} \sum_{i,j=1}^l (a_i^* - a_i)(a_j^* - a_j) K(X_i, X_j) + \right. \\ & \quad \left. \sum_{i=1}^l a_i^*(y_i - \varepsilon) - \sum_{i=1}^l a_i(y_i - \varepsilon) \right] \\ & \quad \sum_{i=1}^l a_i = \sum_{i=1}^l a_i^* \quad (0 \leq a_i, a_i^* \leq C) \end{aligned} \quad (5)$$

其中, $K(X_i, X_j)$ 为支持向量机的核函数, $K(X_i, X_j) = \langle \varphi(x_i) \varphi(x_j) \rangle$ 。对上述问题进行求解, 即得到支持向量机回归函数 $f(x) = \sum_{i=1}^l (a_i - a_i^*) K(X_i, X_j) + b$ 。其中, 当 $(a_i - a_i^*)$ 非零时对应的训练样本就是支持向量, 核函数可以是任何一个满足 Mercer 条件的正定函数常用的核函数主要有 4 类: 线性核函数、多项式核函数、gauss 径向基核函数 (RBF) 和多层感知器 sigmoid 核函数。径向基函数比线性函数好, 在选用了径向基函数后没有必要再考虑线性函数^[4]; Sigmoid 函数精确度不比径向基函数好, 而且只有满足一定条件后它才能成为有效的核函数^[5]; 多项式核函数当其阶次较高时会使参数优化更复杂, 耗费大量资源和时间, 而得到的 SVM 性能并不比径向基核函数得到的 SVM 好^[6]。因此, 笔者选取径向基函数 $K(X_i, X_j) = e^{-\gamma \|x_i - x_j\|^2}$ 作为支持向量机的核函数来进行研究。

2 遗传算法优化支持向量机

1) 遗传算法 (genetic algorithm, GA) 基本原理。遗传算法^[7]是一种基于自然选择和基因遗传学原理的优化搜索方法。它将“优胜劣汰, 适者生存”的生物进化原理引入待优化参数形成的编码串群体中, 按照一定的适配值函数及一系列遗传操作对各

个体进行筛选, 从而使适配值高的个体被保留下来, 组成新的群体。新群体中各个体适应度不断提高, 直至满足一定的极限条件。此时, 群体中适配值最高的个体即为待优化参数的最优解。正是由于 GA 独有的特性, 使之能在复杂空间进行全局优化搜索, 且具有较强的鲁棒性。GA 应用于 SVM 优化, 一方面是用来优化支持向量机的结构, 另一方面是用来优化 SVM 核函数的权值。

2) GA-SVM 算法。鉴于 GA 具有全局寻优能力, 而支持向量机根据有限的样本信息, 在模型的复杂性和学习能力之间寻求最佳折衷, 以期获得最好的泛化能力。与传统的神经网络相比, 支持向量机算法可转化为一个二次型寻优问题。从理论上说, 得到的将是全局最优解, 解决了在神经网络方法中无法避免的局部极值问题; 支持向量机拓扑结构由支持向量决定, 避免了传统神经网络拓扑结构需要经验试凑的方法^[8-9]。支持向量机也能以任意的精度逼近任意函数。笔者将二者结合, 采用了遗传算法优化支持向量机 (GA-SVM)。其基本思想是: 在 SVM 算法之前, 先用 GA 对 SVM 参数进行优化, 快速确定全局最优解, 再用支持向量机回归估计 (SVR) 对样本进行训练。

3 基于 GA-SVM 的航空装备维修费用预测建模

3.1 建立训练样本集

利用训练样本建立 SVM 回归模型, 其描述为 $y_i = f(a_i, a_{i-1}, \dots, a_{i-k})$, 其中 $x_i = a_i, a_{i-1}, \dots, a_{i-k}$ 为输入矢量; $y_i = a_{i+1}$ 为输出值; k 为输入矢量的维数。

3.2 基于 GA 的 SVM 参数优化

惩罚因子 C , RBF 的核参数 γ 对 SVM 回归模型的预测效果影响很大, 为了获取较高预测性能的 SVM 回归模型, 需对 C , γ 进行组合优化。优化过程如图 1。由于遗传算法具有隐含的并行性和强大全局搜索能力, 可以在很短的时间内搜索到全局最优解。因此选用遗传算法对这些参数进行优化, 具体步骤为:

1) SVM 参数编码与 GA 参数设置。将 SVM 训练参数 γ , C 以进行染色体基因编码。进行 GA 参数设置, 如交叉率, 变异率等。随机产生一组表示 SVM 参数值的染色体。

2) 计算种群中各个体的适应度。以适应度函数来衡量选取参数的优劣。

3) 根据个体适应度，通过轮盘赌法，将具有较高适应度值的染色体被复制；并通过单点交叉法将 2 条染色体中的基因进行随机的交换，交叉率为 0.8；变异是指将染色体中的基因编码“0”与“1”对换，变异率设定为 0.01。

4) 重复步骤 2)~4)，使初始确定的一组支持向量机参数不断进化，直到满足结束条件。

5) 对得到的最优解译码，得到优化的参数 γ, C 。

4 应用实例

4.1 样本的准备与处理

航空装备维修费用原始数据如表 1^[10]。其中各参数含义如下：任务可靠度 DR, %；平均故障间隔时间 MTBF, h；平均修复时间 MTTR, h；每飞行小时维修工时 DMMH/FH, h；更换发动机时间 ERT,

min；再次出动准备时间 SGR, min；故障检测率 FDR, %；故障隔离率 FIR, %；备件利用率 SPR, %。为了对建立的模型进行误差分析和预测检验，选取表 1 中前 9 个样本(即 A~I 机型)进行训练，并对第 10 个样本(即 K 机型)进行预测。

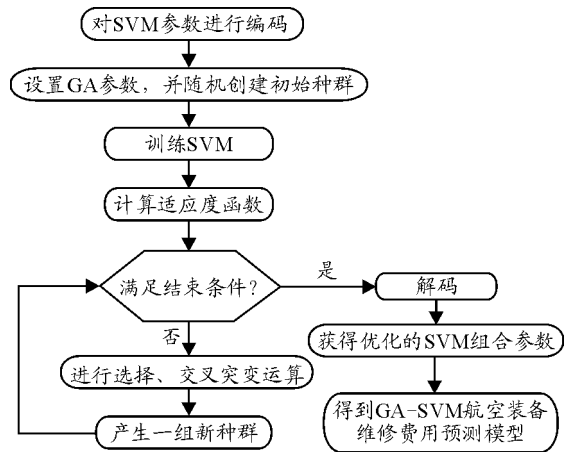


图 1 GA-SVM算法的具体流程图

表 1 航空装备维修费用原始数据

机型	A	B	C	D	E	F	G	H	I	K
DR/%	90.5	85.1	93	85	84	90	88	90.6	86	88.0
MTBF/h	1.75	2.9	5	2.6	2.1	2.5	2.9	2.7	1.5	3.0
MTTR/h	1.78	1.9	3	1.6	1.8	2	1.78	1.3	1	1.79
DMMH/h	11	12	4.5	6	7.5	8	5.6	10	11.2	5.8
ERT/min	21	30	89	149	300	150	80	640	529	128
SGR/min	15	10	13	18	42	35	15	25	34	16
FDR/%	95	96	98	96.5	97	97	95	95	94	97.0
FIR/%	95	96	98	96.5	96	98.5	95	90	92	97.5
SPR/%	95	96	98	96	95	97	97	95	94	97
COST/万元	36 652	50 515	40 150	58 220	95 511	76 251	51 567	18 131	15 427	49 865

由于各指标数值维数有所差别，在运用 GA-SVM方法预测之前，将原始数据进行预处理，消除量纲并具有可比性。根据数据的实际情况，采取最小最大化方法对数据进行泛化。根据公式 $Z_i(k) = \frac{x_i(k) - \min_k x_i(k)}{\max_k x_i(k) - \min_k x_i(k)}$, $k=1, 2, \dots, n$ 将指标各数据映射到[0,1]区间，作为GA-SVM模型的输入。

4.2 模型数据的选取

笔者选用 RBF 核函数作为 SVM 模型的核函数，模型的 ε 精度取为 0.001。选用 RBF 核函数后，惩罚因子 C 和核参数 γ 是影响模型预测性能的重要参数。最优 C 和 γ 的确定实质是组合问题，即根据经验确定 C 和 γ 的搜索范围之后，在相应范围中找出一对 C 和 γ 使得模型具有最佳的预测性能。通过试验枚举所有的组合固然能找到最优的 C 和 γ ，然而

却要耗费大量的计算时间。基于同步优化的思想，笔者采用遗传算法(GA)^[9]对这 2 个参数同时寻优，寻优区间： $C \in [1, 10\ 000]$ 和 $\gamma \in [0.01, 100]$ 。GA 的群体规模选为 20，交叉概率为 0.8，变异概率为 0.01，进化代数 100，取支持向量机 9 折交叉验证得到的 $fitness = \sum_{i=1}^9 (y_i - \hat{y}_i)^2$ 作为适应度函数， y_i 为第 i 次

验证中样本对应的实际值， \hat{y}_i 为第 i 次验证中 SVM 得到的样本预测值。根据表 1 的数据，SVM 参数寻优的结果为 $C=995.125$ ， $\gamma=1.9264$ 。

4.3 预测结果分析与对比

GA-SVM 模型训练及预测结果见表 2。为便于对比，表 2 同时给出了利用支持向量机回归理论、BP 神经网络、偏最小二乘回归及多元线性回归的建模结果^[10-11]。

表 2 GA-SVM与其他模型的预测结果及误差

方法	C/万元	误差/%	方法	C/万元	误差/%
GA-SVM	50 382	1.00	偏最小二乘回归	42 094	15.6
SVR	45 367	9.20	多元线性回归	38 620	22.6
BP神经网络	54 733	9.76			

从表 2 中可以看出,GA-SVM 模型对 K 机型维修费用的预测结果为 50 382 万元,误差仅为 1.0%,要优于 SVR 和 BP 神经网络模型的预测结果(分别为 45 367 万元,误差为 9.2%; 54 733 万元,误差为 9.76%),比偏最小二乘模型与普通多元性回归模型的预测误差(分别为 15.6%和 22.6%)低得多。因此,在航空装备维修费用预测中,运用 GA-SVM 进行建模分析是可行、有效的。

5 结束语

1) 作为最近发展起来的一种新的学习方法,支持向量机具有显著的优越性,越来越受到研究人员的关注。在利用支持向量机进行预测的过程中,参数的选取起关键性作用,若参数选取不合理,往往会造成计算的欠学习和过学习现象,从而直接影响预测精度和运行时间。2) 笔者利用遗传算法对支持向量机关键参数进行优化,以找到以 SVM 模型计算相匹配的预测精度最高的参数。实验结果证明,该方法获得了良好的改进效果,说明运用 GA-SVM

进行建模分析是可行的,并具有一定的优越性。

参考文献:

[1] 刘晓东,张恒喜.飞机可靠性与研制费用相关关系研究[J].空军工程大学学报:自然科学版,2000,1(1):63-66.

[2] 张恒喜,刘晓东,郭基联.飞机全寿命费用预测模型(ALCCE-2)研究报告[R].西安:空军工程大学工程学院,1994.

[3] 吴少雄,黄恩洲.基于支持向量机的控制图模式识别[J].计算机应用,2007,27(1):61-64.

[4] Keerth I, Lin C J. Asymptotic behaviors of support vector machines with Gaussian kernel[J]. Neural Computation, 2003, 15(7):1667-1689.

[5] Lin H T, Lin C J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods[R]. Taipei: National Taiwan University, 2003.

[6] 李盼池,许少华.支持向量机在模式识别中的核函数特性分析[J].计算机工程与设计,2005,26(2):302-304.

[7] 阎树田.基于遗传神经网络的航空电机表面温升的预测[J].电机与控制应用,2007,34(6):17-19.

[8] 谭志杨,林友红,张小海,等.基于粗糙集的支持向量机回归建模及应用[J].四川兵工学报,2010,31(8):53.

[9] 王世卿,曹彦.基于遗传算法和支持向量机的特征选择研究[J].计算机工程与设计,2010,31(18):4088-4092.

[10] 郭风,张恒喜,李寿安,等.基于偏最小二乘回归的飞机维修保障费用预测[J].空军工程大学学报:自然科学版,2005,6(3):10-15.

[11] 刘晖,穆志韬,牟丽君.军用飞机维修保障费用机器学习预测方法[J].航空计算技术,2009,39(3):38-41.

(上接第 17 页)

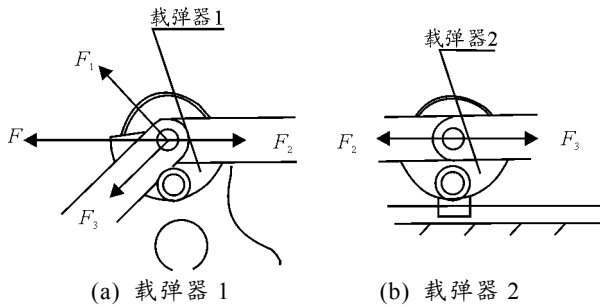


图 7 新设计的载弹器 1 和载弹器 2 的受力分析

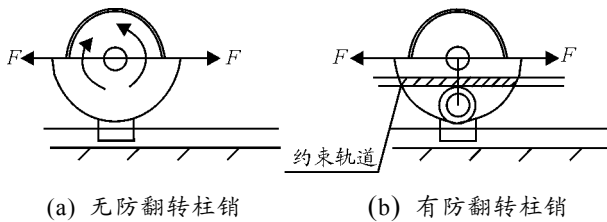


图 8 有无防翻转柱销情况对比

如图 8(a),在载弹器没有防翻转柱销时,在圆周方向无约束,尤其在沿链条轴向缺少约束明显,在外力作用下可造成载弹器绕轴心做小幅的摆动。

在设计了防翻转柱销后,柱销在约束轨道内运动直接限制了载弹器的翻转运动,使载弹器在工作过程中更加稳定(如图 8(b))。结合图 6 的结构特点,新设计的闭合弹带装置有 2 个优点:

- 1) 闭合弹带工作时运行稳定可靠;
- 2) 在驱动轮区域,由于各载弹器之间相互影响较小,极大减弱降低了闭合弹带的多边形效应,使载弹器的运动轨迹更加光滑,为弹丸的运输及交接创造了有利条件。

4 结论

经过对该结构进行设计和运动学仿真,结果表明:采用闭合弹带技术供弹,在输弹过程中运行平稳、冲击小、弹丸交接可靠,且布置灵活,可进行单路和双路供弹。

参考文献:

[1] 成大先.机械设计手册[M].北京:机械工业出版社,2008.

[2] 《现代机械传动手册》编辑委员会.现代机械传动手册[M].北京:机械工业出版社,2002.