

基于支持向量机的机械零件剩余寿命区间估计

王健,孙志礼,于震梁,柴小冬

(东北大学 机械工程与自动化学院,辽宁 沈阳 110819)

摘 要:为提高机械零件剩余寿命估计精度,提出了一种基于支持向量机(support vector machine, SVM)的剩余寿命区间估计模型.简要介绍SVM的线性及非线性理论,分析SVM输入变量与输出变量间的统计关系,将机械零件性能退化指标和剩余寿命分别作为SVM输入变量及输出变量.假设输入变量与残差相互独立且残差分布类型已知,采用极大似然法估计残差的分布参数,在此基础上推导一定置信水平下SVM输出变量置信区间.将均方误差作为SVM预测误差的衡量指标,应用变步长网格搜索法确定SVM参数.通过实例说明所提模型能够准确对机械零件剩余寿命进行区间估计,具有较强的工程应用价值及通用性.

关键词:剩余寿命;支持向量机;区间估计;机械零件;置信区间;均方误差

中图分类号:TB 114.3 文献标志码:A 文章编号:1005-3026(2016)07-0974-05

Remaining Useful Life Interval Estimation for Machine Parts Based on SVM

WANG Jian, SUN Zhi-li, YU Zhen-liang, CHAI Xiao-dong

(School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China. Corresponding author: WANG Jian, E-mail: jianwang8805@163.com)

Abstract: To improve the accuracy of remaining useful life estimation for machine parts, an interval estimation model was proposed based on the SVM (support vector machine). The linear theory and nonlinearity theory of SVM were briefly introduced, and the correlation between input variable and output variable was analyzed. Degraded index and remaining useful life of machine parts were treated as input variable and output variable, correspondingly. It was assumed that input variable and residual error were independent and the residual error's distribution pattern was known. Distribution parameters of residual error were estimated by means of the MLE (maximum likelihood estimation). Then the confidence interval of SVM output variable was obtained under a certain confidence level. The MSE (mean squared error) was used to measure the prediction of SVM. The SVM parameters were gotten by the means of variable step size grid search. A numerical example was presented to show that the proposed model can estimate the remaining useful life confidence interval precisely with the engineering application values and generality.

Key words: remaining useful life; SVM (support vector machine); interval estimation; machine parts; confidence interval; MSE (mean squared error)

重要机械零件的剩余寿命预测,特别是对航空航天、大型生产线、船舶、铁路公路运输等重要设备尤为重要.精准的剩余寿命预测为制定适当的预防性维修和更换策略提供保证,既能保证机械设备在高可靠性条件下运行,又能避免由于过频维护造成的浪费.因此机械零件(如齿轮、轴

承)的剩余寿命预测越来越得到学者关注.

随着材料制造工艺的进步,重要机械零件往往寿命较长且昂贵,大批量做全寿命试验收集大量状态数据并不现实.在少量试验或使用数据条件下建立适当模型准确预测在役零件剩余寿命是当前研究重点.针对各类设备的剩余寿命预测模

型^[1-2]大致可分为：专家系统^[3]、物理模型、统计和随机过程模型^[4-5]、机器学习算法模型^[6-7]等。以 SVM 为代表的机器学习算法模型^[8-10]对研究对象退化轨迹适应性强，具有很好的鲁棒性，得到广泛应用。

SVM 在小样本预测等问题上表现突出，在时间序列预测、结构可靠性分析、回归分析、经济金融等领域有广泛的应用，机械零件剩余寿命预测研究中也较频繁使用。Emmanuel 等^[6]通过分析轴承工作过程中振动信号提取退化信息，建立基于 SVM 的轴承剩余寿命点预测模型；Loutas 等^[7]、Caesarendra 等^[8]以非线性退化零件或系统为研究对象，将贝叶斯方法应用到 SVM 建模过程中。

以上文献主要针对剩余寿命点估计取得了很好的效果。在基于可靠性的维护策略制定中，研究对象寿命概率分布情况或其置信区间对维护策略制定也至关重要。

本文在假定机械零件退化指标及对应阈值已知情况下，分别将零件退化指标和与其对应的剩余寿命作为 SVM 输入变量和输出变量，使用零件全寿命退化数据建立基于 SVM 机械零件剩余寿命预测模型，并根据 SVM 概率估计理论推导剩余寿命一定置信水平下的置信区间。

1 剩余寿命

机械零件从开始工作到失效大致要经历 3 个过程：稳定运行、性能持续退化、最终失效。图 1 为某双列滚动轴承整个寿命期间振动均方根值（root mean square, RMS）变化情况，在稳定运行阶段其 RMS 没有明显变化，初始损伤后其退化现象随着运行时间较明显加剧，当退化量达到某一数值时最终失效。机械零件在初始损伤前零件退化不明显，较难建立数学模型进行分析，因此主要研究初始损伤之后其剩余寿命分布情况。

设某机械零件退化程度的指标为 $x(t)$ ， t 为累计运行时间。不失一般性假设，随着工作时间 t 的增长， $x(t)$ 呈现上升“趋势”，这是由于退化特征提取准确性、工作环境、传感器精度、其他零件等因素影响， $x(t)$ 不一定表现严格单调。当 $x(t)$ 等于或大于预设阈值 x_0 时认为该零件已无法满足正常工作要求，应立即更换或进行必要维修。则 t 时刻的剩余寿命可定义为

$$y(t) = \inf\{t_1 | x(t_1) \geq x_0\} - t.$$

其中 $\inf\{\}$ 表示下确界。

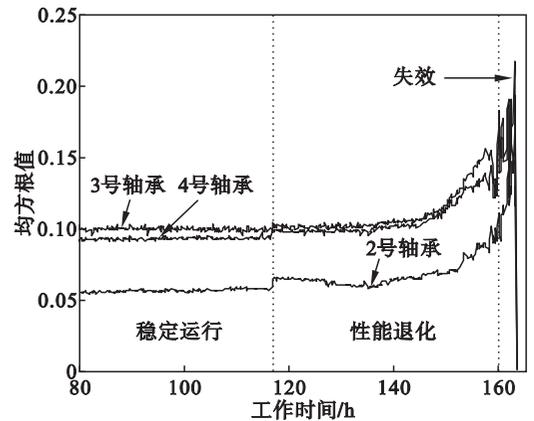


图 1 ZA-2115 双列滚动轴承 RMS 曲线
Fig. 1 RMS curves of a ZA-2115 double row bearings

2 支持向量机模型

2.1 支持向量机

支持向量机是建立在统计学习理论上以结构风险最小化为原则的机器学习算法，在处理小样本问题中表现出良好性质。样本集为

$$S = \{(x_n, y_n) | x_n \in \mathbf{R}^l, y_n \in \mathbf{R}, n = 1, 2, \dots, N\}.$$

式中 x_n 为输入变量，可以为标量或 $(l > 1)$ 维向量， y_n 为与 x_n 对应的预期输出值，一般为标量，回归函数为

$$f(x) = w \cdot x + b. \quad (1)$$

式中 w 为与 x 维数相同的权值向量， $b \in \mathbf{R}$ ， $w \cdot x$ 表示 w 与 x 的点积。 w 和 b 需通过求解如下优化问题得到：

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{n=1}^N (\xi_n + \xi_n^*) \\ \text{s. t.} \quad & \begin{cases} y_n - w^T x - b \leq \varepsilon + \xi_n, \\ w^T x + b - y_n \leq \varepsilon + \xi_n^*, \\ n = 1, 2, \dots, N. \end{cases} \end{aligned}$$

C 和 ε 分别为惩罚因子和不敏感因子， ξ_n 和 ξ_n^* 为松弛因子。根据 Karush - Kuhn - Tucker 条件，以上问题可转化为凸二次优化问题。

$$\begin{aligned} \max \quad & \sum_{n=1}^N y_n (\alpha_n - \alpha_n^*) - \varepsilon \sum_{n=1}^N (\alpha_n + \alpha_n^*) - \\ & \frac{1}{2} \sum_{m,n} (\alpha_m - \alpha_m^*) (\alpha_n - \alpha_n^*) (x_m \cdot x_n). \\ \text{s. t.} \quad & \begin{cases} \sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0, \\ 0 \leq \alpha_n, \alpha_n^* \leq C. \end{cases} \end{aligned}$$

$$\hat{f}(x) = \sum_{n=1}^N (\hat{\alpha}_n - \hat{\alpha}_n^*) \chi(x_n; x) + \hat{b}. \quad (2)$$

当输入变量 x 与 y 之间是非线性关系时,通过引入核函数将 x 映入高维空间,将低维空间非线性问题转化为高维空间的线性问题.此时二次优化问题只需将式(2)中的 $x_n \cdot x$ 换成 $K(x_n, x)$,回归函数为

$$\hat{f}(x) = \sum_{n=1}^N (\hat{\alpha}_n - \hat{\alpha}_n^*) K(x_n, x) + \hat{b}. \quad (3)$$

式中, $K(\cdot, \cdot)$ 是满足 Mercer 条件的核函数,核函数有多种形式,常见的核函数有线性核函数、多项式核函数、Gauss 径向基核函数、sigmoid 核函数,本文采用 Gauss 径向基核函数,其形式为

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2).$$

式中 γ 是核函数参数.

2.2 输出变量区间估计

根据 SVM 理论,输入 x 与输出 y 之间可以表示为

$$y = \hat{f}(x) + \zeta.$$

其中 ζ 为随机变量,为了得到输出 y 的置信区间,假设 ζ 与输入 x 相互独立.理论上这种假设与实际情况不相符,对于固定的 $\hat{f}(x)$ 输出 y 的残差 ζ 必然与 x 存在很大程度的相关性,当假设 ζ 与 x 相互独立时,所计算得到的 y 的置信区间在实际中并不会产生很大误差^[10],能够满足工程需要.基于独立性假设的理论推导复杂程度及计算复杂度有极大简化.

设

$$\zeta_n = y_n - \hat{f}(x_n),$$

通常情况 ζ 服从零均值的高斯分布或拉普拉斯分布^[10],更多情况下 ζ 更好地服从拉普拉斯分布.参数为 σ 的拉普拉斯分布密度函数为

$$p(z) = \frac{1}{2\sigma} \exp\left(-\frac{|z|}{\sigma}\right).$$

以 $\zeta_n (n=1 \dots N)$ 为样本采用极大似然法估计 σ ,极大似然函数为

$$L(\sigma, \zeta) = \left(\frac{1}{2\sigma}\right)^N \exp\left(-\frac{|\zeta_1| + \dots + |\zeta_N|}{\sigma}\right).$$

$$\text{令 } \frac{\partial \ln L}{\partial \sigma} = 0, \text{ 得}$$

$$\hat{\sigma} = \frac{1}{N} \sum_{n=1}^N |\zeta_n|.$$

$$\text{令 } P(\hat{y} - \Delta \leq y \leq \hat{y} + \Delta) = p_0, \text{ 则}$$

$$p_0 = \int_{-\Delta}^{\Delta} p(\zeta) d\zeta = 1 - \exp\left(-\frac{\Delta}{\hat{\sigma}}\right),$$

$$\Delta = -\hat{\sigma} \ln(1 - p_0).$$

所以 SVM 输出变量 y 的置信度为 p_0 的置信区间为

$$[\hat{y} + \hat{\sigma} \ln(1 - p_0), \hat{y} - \hat{\sigma} \ln(1 - p_0)].$$

若 $t_n (n=1, 2, \dots, N)$ 为传感器采集数据时间点,各时间点对应的机械零件的退化指标为 $x(t_n)$,剩余寿命为 $y(t_n)$,若将 $x(t_n)$ 作为输入变量,将 $y(t_n)$ 作为输出变量,即可训练 SVM 得到 x 与 y 的对应关系模型,进而使用该模型实现机械零件剩余寿命区间估计.

3 参数选择

核函数参数 γ 和惩罚因子 C 是影响 SVM 训练效率和推广能力的核心参数,因此 γ 和 C 的选择直接影响剩余寿命预测精度. SVM 模型推广能力和预测精度是通过其在验证集上的预测表现来衡量的.

3.1 均方误差及交叉验证法

设验证集 T 为

$$T = \{(x_m, y_m) | x_m \in \mathbf{R}^l, y_m \in \mathbf{R}, m=1, \dots, M\}.$$

SVM 模型 $\hat{f}(x)$ 推广能力及预测精度可由其在验证集 T 上的预测均方误差(mean squared error, MSE)度量,

$$\text{MSE}_T = \frac{1}{M} \sum_{m=1}^M (y_m - \hat{y}_m)^2 = \frac{1}{M} \sum_{m=1}^M (y_m - \hat{f}(x_m))^2.$$

交叉验证法(cross validation, CV)用来验证训练所得 SVM 模型推广能力及预测精度. K-fold CV 的基本思想是将训练集 S 随机分成 K 组(一般采用近似均分方式),对于同一组 γ 和 C 值, K 个子集分别做一次验证集,其余的 $K-1$ 个子集的并集做训练集,得到 K 个模型及对应的 MSE_k ,取 K 个 MSE 值的均值作为与 γ 和 C 对应 SVM 模型的推广能力及预测精度度量.

$$\text{MSE} = \frac{1}{K} \sum_{k=1}^K \text{MSE}_k.$$

3.2 变步长网格搜索算法

目前还没有系统准确地确定最佳 γ 和 C 的方法,被比较广泛使用的算法有蚁群算法、粒子群算法、网格搜索算法等智能或近似算法.由于蚁群和粒子群算法伴随往往以较大的计算量为代价才能确定较好的 (γ, C) ,且不能保证算法每次搜索都收敛于同一点,网格搜索算法在处理二维优化问题时稳定性较高,计算量处于可接受范围,本文

采用变步长网格搜索算法确定最佳 (γ, C) , 保证在不明显增加计算量情况下搜索到较精确 $(\hat{\gamma}, \hat{C})$.

变步长网格搜索步骤：

1) 初步确定 γ 和 C 的取值范围. 为保证最后搜索到最优 (γ, C) , 初始范围应当适当大些, 如 γ 和 C 范围都是 $[10^{-15}, 10^{15}]$.

2) 设定搜索步长. 在双对数直角坐标系中设定搜索步长, 如 0.1. 这样就在双对数直角坐标系中生成了一个二维网格, 每个交叉点都对应一组 (γ, C) 取值, 对每一组 (γ, C) 采用交叉验证法计算其对应的 MSE 值. 在双对数直角坐标系中绘制等高线.

3) 缩小 γ 和 C 范围. 由步骤 2) 中等高线缩小搜索范围, 在双对数直角坐标系中缩小步长, 重复步骤 2) 直至 γ 和 C 的范围满足预先设定的精度要求.

4 实 例

本文采用 IMS 数据库中滚动轴承全寿命试验数据验证所提出基于 SVM 剩余寿命区间估计模型. 轴承的全寿命试验分 3 组进行, 每组 4 个轴承. 试验转速为 2 000 r/min, 加载在轴承径向载荷为 27 kN, 采用强制润滑. 轴承型号为美国莱克斯诺工业集团 ZA-2115 双列滚动轴承, 加速度传感器安装在轴承座上. 每 10 min 采集一次轴承振动信号, 每次采集时长为 1 s, 收集 20 480 个加速度值.

针对轴承振动信号退化特征提取方法主要有时域分析、频域分析和时频域分析, 结合 PAC 及 ISOMAP 等线性或非线性降维方法得到最终的能够反映轴承工作状态的表征量. 本文借鉴文献 [9] 选用相对均方根值 (RRMS) 作为轴承退化特征. 经过平滑处理^[9] 后第二组试验 2、3、4 号轴承 RRMS 如图 2 所示. 定义 RRMS 值 1.05 和 1.5 分别为轴承衰退期起始阈值和最终失效阈值, 当 RRMS 处于 1.05 和 1.5 之间时认为轴承处于衰退期. 2、3、4 号轴承的起始衰退时间分别为 100, 103 和 118 h. 3 个轴承起始衰退后 RRMS 与剩余寿命对应集分别为 S_2, S_3, S_4 .

$$S_2 = \{ (x_{2n}, y_{2n}) \mid n = 1, 2, \dots, 146 \},$$

$$S_3 = \{ (x_{3n}, y_{3n}) \mid n = 1, 2, \dots, 148 \},$$

$$S_4 = \{ (x_{4n}, y_{4n}) \mid n = 1, 2, \dots, 126 \}.$$

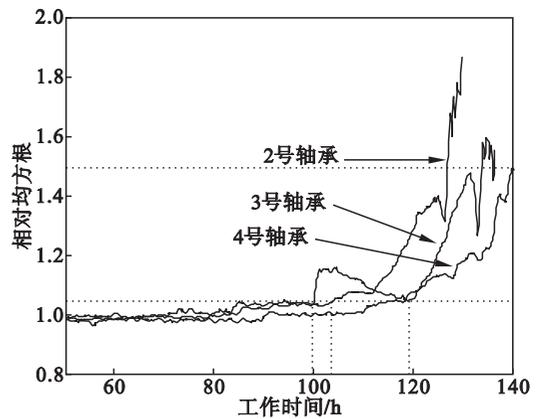


图 2 RRMS 曲线

Fig. 2 Curves of relative root mean square

设 $\varepsilon = 0.5$, 初始 γ 和 C 范围都是 $[10^{-8}, 10^8]$. 将 $S_2 \cup S_3$ 作为训练集, 采用 3.2 中改进步长网格搜索确定 (γ, C) , 得到最佳 (γ, C) 为

$$\hat{\gamma}_1 = 1.6466, \hat{C}_1 = 597.86.$$

将 $(\hat{\gamma}_1, \hat{C}_1)$ 及 $S_2 \cup S_3$ 代入 SVM 模型中进行训练得到回归模型 1, 并对 σ 做极大似然估计:

$$\hat{\sigma}_1 = 9.4073.$$

用 S_4 作为验证集预测 4 号轴承剩余寿命, 并得到其置信度 $p_0 = 0.9$ 的置信区间, 如图 3a 所示. 分别以 $S_2 \cup S_4$ 和 $S_3 \cup S_4$ 作为训练集计算 $(\hat{\gamma}_2, \hat{C}_2)$ 和 $(\hat{\gamma}_3, \hat{C}_3)$, 进而得到回归模型 2 和回归模型 3 并估计 σ_2, σ_3 , 再分别以 S_3 和 S_2 为验证集得到置信度 $p_0 = 0.9$ 的剩余寿命区间估计结果, 如图 3b 和 3c 所示. 表 1 为 3 个回归模型参数估计结果.

表 1 γ, C, σ 估计值
Table 1 Estimations of γ, C and σ

参数	模型 1	模型 2	模型 3
$\hat{\gamma}$	1.6466	1.4564	1.5486
\hat{C}	597.86	4265.1	1596.9
$\hat{\sigma}$	9.4073	7.3349	8.8307

由图 3 可知, 基于 SVM 得到的轴承剩余寿命预测均值与准确值可能存在较大偏差 (见图 3b), 导致剩余寿命真实值处在置信区间上下限附近甚至不在置信区间内. 导致预测均值误差较大的主要原因是: ① 轴承的制造加工误差. 滚动轴承特别是双列滚动轴承对制造加工误差极敏感, 导致其退化轨迹存在较大差异 (如图 2 所示); ② 由于目前还没有准确计算 SVM 模型参数方法, 本文采用应用较广泛的交叉验证法计算 SVM 模型均方误差作为确定模型精度指标. 该方法具有一定随机性, 因而不能在理论上保证所确定的参

数为全局最佳。

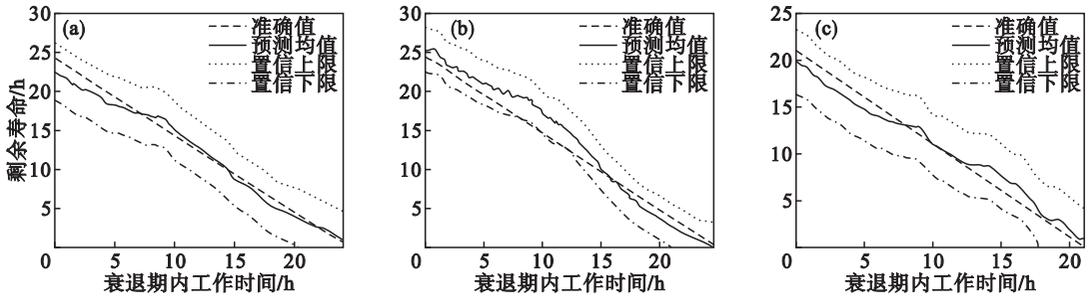


图 3 区间估计结果与真实值比较

Fig. 3 The comparisons between interval estimations and true values

(a)—4 号轴承 ;(b)—3 号轴承 ;(c)—2 号轴承。

5 结 论

基于支持向量机的剩余寿命点估计能够较准确地预测零件剩余寿命,但当全寿命状态数据较少且零件对制造加工误差敏感时预测结果并不稳定,有时预测值与真实剩余寿命有较大偏差。预测精度随真实剩余寿命变化有明显波动,剩余寿命估计曲线不光滑。本文提出的剩余寿命区间估计方法能够准确估计零件剩余寿命所在置信区间,有效弥补了单纯点估计的不稳定性的不足,能够为制定合理的维修或更换计划提供参考。

参考文献 :

- [1] Sikorska J Z ,Hodkiewicz M ,Ma L. Prognostic modelling options for remaining useful life estimation by industry[J]. *Mechanical Systems and Signal Processing* ,2011 ,25(5) : 1803 - 1836.
- [2] Si X S ,Wang W B ,Hu C H ,et al. Remaining useful life estimation—a review on the statistical data driven approaches [J]. *European Journal of Operational Research* ,2011 ,213 (1) :1 - 14.
- [3] Sanders D A ,Jasper G J ,Alexander G. Improving ability of tele-operators to complete progressively more difficult mobile robot paths using simple expert systems and ultrasonic sensors [J]. *Industrial Robot* 2010 ,37(5) :431 - 440.
- [4] Si X S ,Wang W B ,Hu C H ,et al. A Wiener-process-based degradation model with a recursive filter algorithm for

remaining useful life estimation[J]. *Mechanical Systems and Signal Processing* 2013 ,35(1/2) :219 - 237.

- [5] Su C ,Shen J Y. A novel multi-hidden semi-Markov model for degradation state identification and remaining useful life estimation [J]. *Quality and Reliability Engineering International* 2013 ,29(8) :1181 - 1192.
- [6] Emmanuel R ,Rafael G. Remaining useful life estimation by classification of predictions based on a neuro-fuzzy system and theory of belief functions[J]. *IEEE Transactions on Reliability* 2014 ,63(2) :555 - 566.
- [7] Loutas T H ,Roulias D ,Georgoulas G et al Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic E-support vectors regression [J]. *IEEE Transactions on Reliability* 2013 ,62(2) :821 - 832.
- [8] Caesarendra W ,Widodo A ,Yang B S. Application of relevance vector machine and logistic regression for machine degradation assessment[J]. *Mechanical Systems and Signal Processing* 2010 ,24(4) :1161 - 1171.
- [9] 申中杰 ,陈雪峰 ,何正嘉 ,等. 基于相对特征和多变量支持向量机的滚动轴承剩余寿命[J]. *机械工程学报* ,2013 ,49 (2) :183 - 189.
(Shen Zhong-jie , Chen Xue-feng , He Zheng-jia , et al. Remaining life predictions of rolling bearing based on relative features and multivariable support vector machine [J]. *Journal of Mechanical Engineering* ,2013 ,49(2) :183 - 189.)
- [10] Lin C J ,Weng R C. Simple probabilistic predictions for support vector regression [R]. Taipei : National Taiwan University 2004.