

基于支持向量机的航空发动机滑油监控分析

刘洋^{1,2}, 程礼², 范家栋²

(1. 长沙航空职业技术学院, 湖南 长沙 410124; 2. 空军工程大学工程学院, 陕西 西安 710038)

摘要: 分析支持向量机用于时间序列预测的理论基础, 针对某型发动机滑油金属含量的预测分别采用传统的 AR 模型和基于现代统计理论下的支持向量回归模型对滑油时间序列进行预测建模

关键词: 支持向量机; 支持向量回归; AR 模型

中图分类号: V233.4 **文献标识码:** A **文章编号:** 1671-9654(2007)04-033-04

Analysis of Aero Engine Slide Oil Inspection Based on Support Vector Machine

LIU Yang^{1,2}, CHENG Li², FAN Jia-dong²

(1. Changsha Aeronautical Vocational and Technical college, Changsha Hunan 410124;

2. Engineering Institute, Air force Engineering University, Xi'an Shaanxi 710038)

Abstract: This paper analyses theoretical foundation of support vector machine's using in time sequence estimation. Aiming at estimation of metals content of slide oil in certain aero engine, the authors use different methods including traditional AR model and modern statistics theories of support vector regression model to estimate the slide oil time sequence.

Key words: support vector machine; support vector regression; AR model

对于发动机的传动系统, 接触滑油的零组件更趋向于高载荷、高工作温度及轻质量。在发动机内部有摩擦件的地方就有润滑油, 如转子轴承、齿轮、封严装置。这些部件中有许多是发动机重点部件, 又处在发动机的内部, 故障征兆一般很难发现。而滑油中携带着发动机运动机件状态的大量信息, 如机件磨损的数量、形状、粒度成分等, 这些信息提供了发动机有关机件的磨损程度、疲劳剥落以及磨损件位置等情况。本文依据某型飞机发动机滑油金属含量历史数据, 建立时间序列预测模型来预测分析金属含量的变化趋势。常规的时间序列预测方法主要以自回归模型为主, 这种模型在理论上十分成熟, 但其精度不高, 且容错性差, 仅仅适合做短区间预报; 而支持向量机由于采用了新型的学习机制, 使得预测模型具有很高的推广能力, 即便在较

长区间预测中仍可保证具有相当的精度。

1 支持向量机的一般理论

支持向量机是由 Vapnik 在上世纪 90 年代中期提出的一种机器学习算法。支持向量机以其良好的理论背景, 从结构风险最小化原则为机器学习提供了一个崭新的方向。支持向量机最初用于解决模式识别问题, 后又被推广到函数回归。在模式识别中, 为了发现具有推广能力的决策规则, 所选择的训练数据的一些子集称为支持向量。最佳的支持向量分离等效于所有数据的分离, 支持向量机是从线性可分情况下的最优分类面发展而来的。

2 支持向量回归与序列预测分析

2.1 支持向量回归的一般理论

支持向量回归的基本思想是通过一个非线性映射 Φ 将数据 \vec{X} 映射到高维特征空间 F , 并在这个

收稿日期: 2007-09-27

作者简介: 刘洋(1978-), 上海人, 助教, 在读硕士研究生, 研究方向为航空发动机可靠性与寿命研究。

空间进行线性回归。典型的回归型支持向量机结构如图1所示。

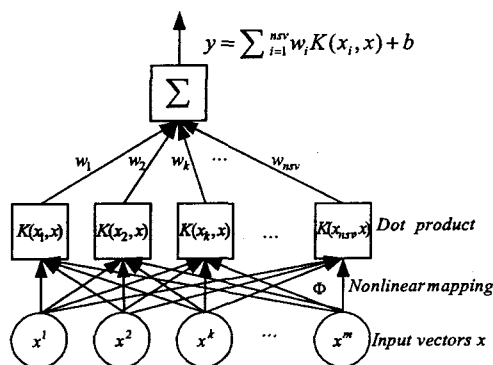


图1 典型的回归型支持向量机

$$f(\vec{X}) = (\omega \cdot \Phi(\vec{X})) + b, \Phi: R^n \rightarrow F, \omega \in F \quad (1)$$

其中 b 是域值。这样,在高维特征空间的线性回归便对应于低维输入空间的非线性回归,免去在高维空间 ω 和 $\Phi(\vec{X})$ 点积的计算。由于 Φ 是固定不变的,因此影响 ω 的有经验风险的总和 $R_{emp}[f]$,以及使其在高维空间平坦的复杂性项 $\|\omega\|^2$,则有

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

其中 $e_i = f(\vec{x}_i) - y_i$, S 表示样本容量, $C(\cdot)$ 是损失函数, λ 为规则化常数。

2.2 基于支持向量回归的时间序列预测模型

2.2.1 数据准备

建立自相关输入 $\vec{x}_t = \{x_{t-1}, x_{t-2}, \dots, x_{t-p}\}$ 与输出 $y_t = \{x_t\}$ 之间一一映射关系 $f: R^p \rightarrow R$, 其中 p 值称为嵌入维数,可以看出 p 值的选取实际隐含反映了转换后矩阵蕴涵的知识量。而在支持向量回归平稳时间序列预测中,其选取尚无严格理论指导。同时为了与 AR 模型比较,本文也采用 FPE 准则评价预测误差,并根据误差值大小来优化选取嵌入维数 P 。经过这样的变换之后,可以得到用于向量机学习的样本:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \\ x_2 & x_3 & \cdots & x_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-p} & x_{n-p+1} & \cdots & x_{n-1} \end{bmatrix}, Y = \begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix}$$

2.2.2 支持向量预测模型

在得到学习样本之后,就可对支持向量机进行训练,得到的回归函数表示如下:

$$Y_t = \sum_{i=1}^{n-p} (a_i - a_i^*) K(\vec{x}_i \cdot \vec{x}_t) + b, t = p+1, \dots, n \quad (3)$$

一般的可以得到预测第 l 步的支持向量机数学模型为:

$$Y_{n+l} = \sum_{i=1}^{n-p} (a_i - a_i^*) K(\vec{x}_i \cdot \vec{x}_{n-p+l}) + b \quad (4)$$

$$\text{其中 } \vec{x}_{n-p+l} = \{x_{n-p+l}, \dots, x_{n+1}, \dots, x_{n+l-1}\}$$

2.3 自回归模型

对于一个平稳的时间序列 $\{x_1, x_2, \dots, x_n\}$, 可以建立一个线性时间反演的时序模型:

$$x_t - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p} = a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} \quad (5)$$

其中 $x_t, t = 1, 2, \dots, n$ 为离散时间数列; $\varphi_i, i = 1, 2, \dots, p$ 为自回归系数; $\theta_i, i = 1, 2, \dots, q$ 为滑动平均系数, $a_t, t = 1, 2, \dots, n$ 为方差为 σ_a^2 的白噪声序列。

该时序模型称为 (p, q) 阶自回归滑动平均模型简称为: $ARMA(p, q)$ 。当 $q = 0$, 得到的即为 p 阶自回归模型, 简称为 $AR(p)$ 。

2.3.1 AR模型的系数和阶次确定

时间序列 $\{x_i\}, i = 1, 2, \dots, n$ 可写成以下矩阵形式的线性组合。

$$Y = X\varphi + a \quad (6)$$

其中

$$Y = \begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix}, \varphi = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_p \end{bmatrix}, a = \begin{bmatrix} a_{p+1} \\ a_{p+2} \\ \vdots \\ a_n \end{bmatrix}$$

$$X = \begin{bmatrix} x_p & x_{p-1} & \cdots & x_1 \\ x_{p+1} & x_{p+2} & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & \cdots & x_{n-p} \end{bmatrix}$$

根据多元回归理论,可得到参数矩阵的最小二

乘估计为

$$\hat{\varphi} = (X^T X)^{-1} X^T Y \quad (7)$$

可以证明 $\hat{\varphi}$ 为真值 φ 的无偏估计,因此采用最小二乘估计精度较高。从理论上讲,ARMA 模型成立的条件是 $\{a_t\}$ 为白噪声。因此,模型适用性的准则是检验 $\{a_t\}$ 是否为白噪声。实际中常用的准则有:FPE 准则、AIC 准则、BIC 准则等。本文采用最终预报误差(FPE) 准则。

$$FPE(k) = \frac{n+k}{n-k} \sigma_a^2 \quad (8)$$

其中 $\sigma_a^2 = E(a_n^2) = \frac{1}{n-k} \sum_{t=k+1}^n [x_t - \varphi_1 x_{t-1} - \dots - \varphi_k x_{t-k}]^2$, 当值增大时,残差 σ_a^2 将减少,因此可以找到一个最优值 p 使得 FPE 达到最小。

2.3.2 AR 模型预测机理

AR 模型在 t 时刻向前预测 l 步的结果为:

$$\hat{x}_t(l) = \sum_{i=1}^p \varphi_i x_{t+l-i} + \sum_{i=1}^{l-1} \varphi_i \hat{x}_t(l-i) + \sum_{i=1}^p \varphi_i x_{t+l-i} + \sum_{i=1}^p \varphi_i \hat{x}_t(l-i) \quad (9)$$

从中可以看出,其中预测结果只与相关 $x_t, x_{t-1}, \dots, x_{t+1-p}$, 而不需要求解 $\{a_t\}$ 。

2.4 基于支持向量机的航空发动机滑油监控分析

本文采用的数据来自某型发动机从一次换油到另一次换油的完整工作阶段的滑油光谱分析数据,并按照光谱分析的时间排序得到金属含量数据序列(取 94 个),因为在滑油监测中以 Fe 的金属含量为主要监控对象,因此本文仅以含 Fe 量的时间序列分析为例。

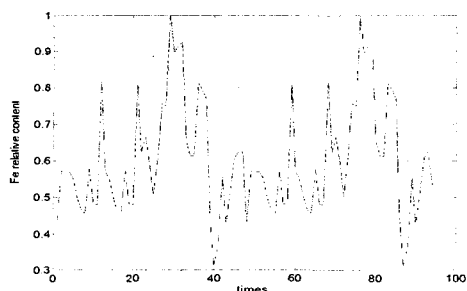


图2 Fe 相对含量时间序列

2.4.1 数据的预处理

首先,对数据进行相对化处理,得到的时间序

列结果如图 2 所示,可以看出 Fe 金属含量数据基本围绕 0.6021 点上下波动,再对数据零化处理得到新的 Fe 相对含量时间序列 $\{x_t\}$ 。本文采用工程处理方法不对其进行正态性检验,认为经过处理之后的时间序列 $\{x_t\}$ 满足平稳、正态要求。

2.4.2 AR 模型与回归支持向量机预测模型的建立

综合 2.2 基于支持向量回归的时间序列预测模型、2.3 自回归模型以及 AR 模型的系数、阶次确定和 AR 模型预测机理,将采集的滑油数据带入,可得分析结果如图 3 所示。

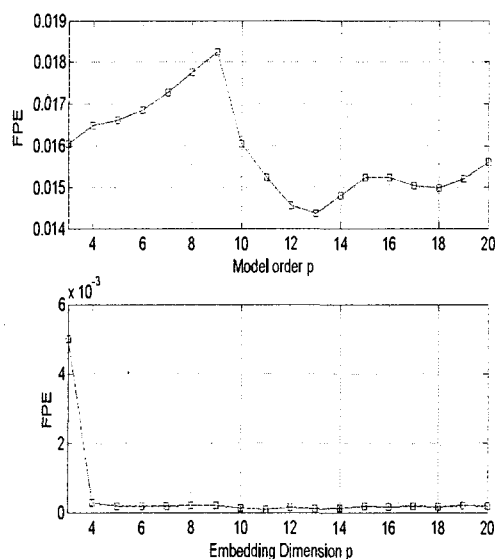


图3 p 与 FPE

从图 3 中可以看出 $P_{AR} = 13$ 时 AR 模型 FPE 值达到最小 1.44×10^{-2} ; 而当 $P_{svm} = 11$ 时 FPE 值达到最小 0.92×10^{-6} , 图中也可以反映出支持向量回归方法具有很高的建模精度。

2.4.3 训练与测试结果比较

将数据分成两组:前 60 次数据作为训练数据,后 34 次作为测试数据,采用 AR 模型、支持向量回归模型分别进行前 1 步~5 步的预测,其中前 1 步~5 步的结果比较分别见图 4~图 8, 1 步~5 步二者的预测平均相对误差值如表 1 所示。其中回归向量机采用高斯径向基函数,且 $\sigma^2 = 0.5$, 常数 $C = 10000$, $\varepsilon = 0.005$ 。

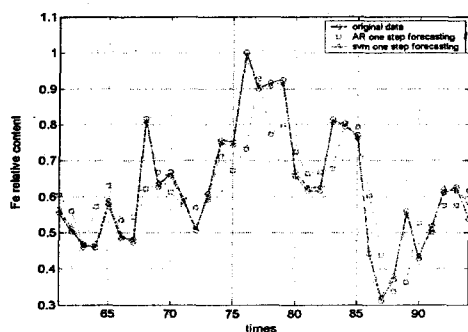


图4 一步提前预测的结果

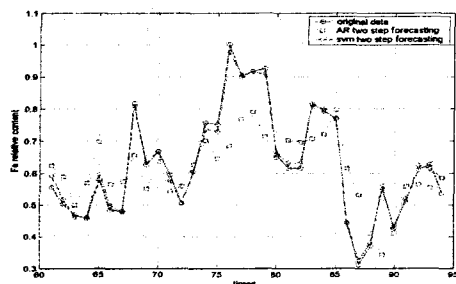


图5 两步提前预测的结果

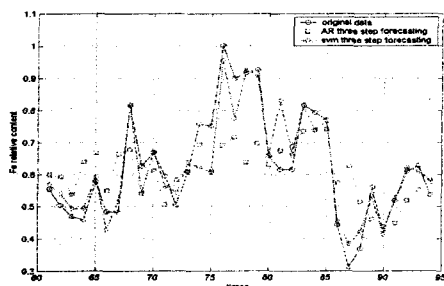


图6 三步提前预测的结果

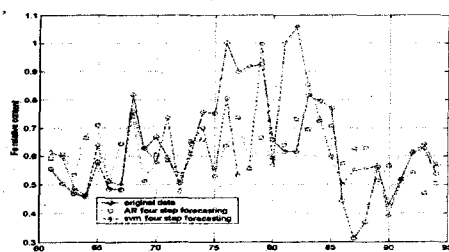


图7 四步提前预测的结果

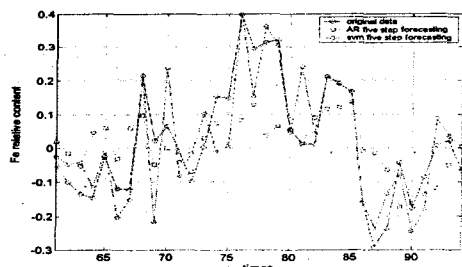


图8 五步提前预测的结果

3 结束语

通过数值仿真和测试分析得出:

1) 支持向量机应用于航空发动机金属含量的时间序列预测是成功的,由表1可以看出随着预测步数的增加 AR 模型以及支持向量机的预测平均相对误差均增加,但支持向量机与 AR 相比仍然具有较高的精度,这说明支持向量机具有很强的推广能力。

表1 预测步数与平均相对误差

预测步数 (forecasting step ahead)	svm	AR
1	0.0028	0.1201
2	0.0042	0.1473
3	0.0164	0.1717
4	0.0845	0.1984
5	0.1818	0.2725

2) 与 AR 模型相比,支持向量回归仅仅取决于支持向量,而支持向量一般小于样本的个数,从而有效的降低了建模的复杂性。在序列预测中,支持向量与反映趋势变化的点紧密相关,因而回归型支持向量机能更好的跟踪时间序列的发展趋势。

3) 在给出金属含量警戒值的情况下,采用支持向量机可以进行较长区间有效预警,从而可以为发动机的监控提供重要的依据。

4) 滑油中包含的金属颗粒是滑油监控的重要信息源,有助于判断磨损的过程类型以及损伤的机制,而且这种监控较经济实惠,随着数据积累量的增多,这种方式也就越来越优越。

参考文献:

- [1] 邓乃扬,田英杰. 数据挖掘中的新方法 - 支持向量机 [M]. 北京科学出版社,2004.
- [2] 马勇,等. 基于支持向量机的软测量建模方法[J]. 信息与控制,2004,33(4).
- [3] 崔伟东,等. 支持向量机研究[J]. 计算工程与应用,2001(1).
- [4] 孙护国,等. 航空发动机滑油系统监控与诊断技术[J]. 航空科学技术,2000(4).
- [5] 陈志英. 航空发动机滑油监视与诊断系统软件研制[J]. 推进技术,1998,19(5).
- [6] 温熙森,等. 模式识别与状态监控[M]. 国防科学技术大学出版社,1997.

[责任编辑 刘敏]