

多重核学习非线性时间序列故障预报

张军峰, 胡寿松

(南京航空航天大学 自动化学院, 江苏 南京 210016)

摘要: 针对非线性时间序列故障预报问题, 提出了多重核学习故障预报方法. 利用多重核学习可以减少支持向量的个数, 提高预测性能. 而且在多重核学习定义的混合核空间中运用减聚类能够提取正常原型. 最后, 将本文提出的方法应用于连续搅拌釜式反应器的故障预报, 仿真结果表明该方法能够提高故障预报的准确性与实时性.

关键词: 故障预报; 多重核学习; 支持向量回归; 减聚类

中图分类号: TP273 **文献标识码:** A

Nonlinear time series fault prediction by multiple kernel-learning

ZHANG Jun-feng, HU Shou-song

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 210016, China)

Abstract: A novel fault prediction method based on multiple kernel-learning is proposed for fault prediction in nonlinear time series. In the support-vector regression, the multiple kernel-learning will reduce the number of support vectors, and improve the performance of the prediction model. Furthermore, the normal prototypes could be extracted by conducting subtractive clustering on the mixed kernel space defined by multiple kernel-learning. The proposed method is applied to a continuous stirred-tank reactor(CSTR) for fault prediction. Simulation results indicate that this method predicts faults quickly and accurately.

Key words: fault prediction; multiple kernel-learning; support vector regression; subtractive clustering

1 引言(Introduction)

随着科技进步, 一些可靠性要求高的系统, 希望对系统劣化趋势作出预测, 这就是故障预报. 它是增强故障早期发现能力, 提高实时性的重要手段.

Tse和Atherton^[1]采用回归神经网络对香港一家化工厂的冷却塔鼓风机进行故障预报. 然而神经网络存在局部极小点以及结构选择依赖于经验等缺陷, 降低了其应用效果. 由Vapnik等人^[2]创立的支持向量机(SVM)遵循了结构风险最小化原理, 并且可以将非线性问题转化为线性问题并能得到全局最优解. 文[3]利用基于支持向量回归(SVR)的时间序列预测与聚类相结合的方法构建了故障预报系统, 取得了不错的效果.

众所周知, SVR在时间序列预测中的应用效果取决于核函数及相应参数的选择. 近年来迅速发展的多重核学习(MKL)^[4~6]为解决该问题提供了新途径. 本文针对文[3]提出的故障预报系统, 引入多重核学习, 可以提高故障预报系统的准确性和实时性.

2 故障预报系统(Fault prediction system)

基于多重核学习的故障预报系统由3部分组成: 时间序列预测, 正常原型提取以及相似性度量.

2.1 时间序列预测(Time series prediction)

对于回归问题, 假设给定的样本点:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_l, y_l), \mathbf{x}_i \in \mathbb{R}^n, y_i \in \mathbb{R}.$$

根据结构风险最小化原理, 引入松弛变量 $\xi_i^{(*)} > 0$, 回归问题可归结为如下优化问题:

$$\begin{cases} \min_{\mathbf{w}, \xi, \xi^*} \langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^l (\xi_i + \xi_i^*), \\ \text{s.t. } \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle + b - y_i \leq \varepsilon + \xi_i, \\ y_i - \langle \mathbf{w}, \Phi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^*. \end{cases} \quad (1)$$

应用拉格朗日乘子法求解这个具有线性不等式约束的二次规划问题, 可得到该优化问题的对偶形式:

$$\begin{cases} \max_{\beta} 2\mathbf{h}^T \beta - \beta^T Q(K) \beta, \\ \text{s.t. } 0 \leq \beta \leq C, \beta^T \gamma = 0. \end{cases} \quad (2)$$

其中:

$$Q(K) = \begin{bmatrix} K & -K \\ -K & K \end{bmatrix}, \beta = \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix}, \gamma = \begin{bmatrix} e \\ -e \end{bmatrix},$$

$$h = \begin{bmatrix} -\varepsilon e + y \\ -\varepsilon e - y \end{bmatrix}, e = [1, \dots, 1]^T \in \mathbb{R}^l.$$

由于核的混合可以使数据信息在特征空间得到更为充分的表示, 有助于提高学习性能, 因此考虑具有如下形式的混合核, 其中 μ_j 可视为权重系数:

$$K = \sum_{j=1}^m \mu_j K_j. \quad (3)$$

若假设式(3)中混合核的权重系数大于零, 由核函数的性质^[1]可知 K 是对称半正定矩阵, 即 K 是有效的核函数, 其中核矩阵 K_i 可按式(4)进行标准化:

$$\bar{K}_j(\mathbf{x}_p, \mathbf{x}_q) = \frac{K_j(\mathbf{x}_p, \mathbf{x}_q)}{\sqrt{K_j(\mathbf{x}_p, \mathbf{x}_p) \cdot K_j(\mathbf{x}_q, \mathbf{x}_q)}}. \quad (4)$$

假设问题(2)中的核矩阵 K 属于核空间 Ω , 且 Ω 满足条件:

$$\Omega = \left\{ \sum_{j=1}^m \mu_j K_j \mid \mu_j > 0, K_j \succeq 0, \text{tr}(K) = c \right\},$$

则问题(2)可以转化为以下的二次约束下二次优化(QCQP)问题:

$$\begin{cases} \max_{\beta, t} 2h^T \beta - ct, \\ \text{s.t. } 0 \leq \beta \leq C, \beta^T \gamma = 0, \\ t \geq \frac{1}{l} \beta^T Q(\bar{K}_j) \beta. \end{cases} \quad (5)$$

该QCQP问题可以通过优化软件包MOSEK基于内点法求解^[7], 且QCQP问题与QP问题求解算法的复杂度相当^[8]. 而式(5)在训练SVR的同时解决了模型选择的困扰, 显然较之标准SVR, MKL-SVR具有优越性. 此时, SVR的回归函数具有如下形式:

$$f(\mathbf{x}_{\text{new}}) = \sum_{i=1}^{|sv|} \sum_{j=1}^m \mu_j (\alpha_i - \alpha_i^*) K_j(\mathbf{x}_i, \mathbf{x}_{\text{new}}) + b,$$

其中 $|sv|$ 表示支持向量的个数.

使用支持向量机进行时间序列预测, 其关键在于如何重构线性空间, 找到输入输出的对应关系. 若实际时间序列为 $[v_1, v_2, \dots, v_n, \dots, v_t, \dots]$, 应用预测模型, 可以利用样本 $\mathbf{v}_1 = [v_1, v_2, \dots, v_n]^T$ (n 表示嵌入维数), 得到对 v_{n+1} 的预测值 $\hat{v}_{n+1}$:

$$\hat{v}_{n+1} = f(\mathbf{v}_1) = \sum_{i=1}^{|sv|} \sum_{j=1}^m \mu_j (\alpha_i - \alpha_i^*) K_j(\mathbf{x}_i, \mathbf{v}_1) + b.$$

进一步采用滚动多步的预测策略, 可以提高故障预报系统的实时性.

2.2 原型提取(Prototype extraction)

时间序列的正常原型对于故障预报至关重要, 若提取的正常原型越能反映系统在正常情况下的特征, 故障预报便越准确. 核聚类由于是在高维特征空间实现聚类, 其聚类效果优于传统的方法^[9]. 另外, MKL-SVR为时间序列定义了反映正常时间序列信息的混合核空间. 基于此, 本文在该混合核空间上采用减聚类算法提取系统的正常原型.

传统的减聚类算法, 通常是通过计算数据样本在输入空间中基于Euclidean距离的高斯函数, 获得每个数据样本的密度值, 从而实现数据聚类. 而核减聚类^[10]的距离计算如下式所示:

$$\|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|_2^2 = K(\mathbf{x}_i, \mathbf{x}_i) + K(\mathbf{x}_j, \mathbf{x}_j) - 2K(\mathbf{x}_i, \mathbf{x}_j). \quad (6)$$

然而, 由于本文是在MKL-SVR定义的混合核空间上采用减聚类算法, 此时式(6)变为

$$\begin{aligned} \|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|_2^2 &= \sum_{k=1}^m \mu_k K_k(\mathbf{x}_i, \mathbf{x}_i) + \sum_{k=1}^m \mu_k K_k(\mathbf{x}_j, \mathbf{x}_j) - \\ &2 \sum_{k=1}^m \mu_k K_k(\mathbf{x}_i, \mathbf{x}_j). \end{aligned} \quad (7)$$

将式(7)替代减聚类中的Euclidean距离, 算法流程不变, 可以实现时间序列的混合核减聚类, 提取出 c 个正常原型: $\{\mathbf{x}_{c1}, \dots, \mathbf{x}_{cc}\}$.

2.3 相似性度量(Similarity measure)

对于时间序列 $\mathbf{x}_i, \mathbf{x}_j$, 定义两者的相似性为

$$S(\mathbf{x}_i, \mathbf{x}_j) = 1 - \exp(-D_\Phi/2\gamma_b^2), \quad (8)$$

其中: $D_\Phi = \|\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_j)\|_2^2$, γ_b 为减聚类算法参数. 则 $S(\mathbf{x}_i, \mathbf{x}_j)$ 值小表明 $\mathbf{x}_i, \mathbf{x}_j$ 相似, 反之不相似.

假设预测序列为 $\hat{\mathbf{v}}_t$, 则预测序列与正常原型的相似性可以定义为

$$\text{Sim}(\hat{\mathbf{v}}_t) = \min(S(\hat{\mathbf{v}}_t, \mathbf{x}_{ci})), i = 1, \dots, c. \quad (9)$$

因为 $\hat{\mathbf{v}}_t = [v_t, \hat{v}_{t+1}, \dots, \hat{v}_{t+n-1}]^T$, 若 $\text{Sim}(\hat{\mathbf{v}}_t)$ 的值大于阈值 θ , 则可判断系统在 $t+n-1$ 时刻发生故障.

阈值 θ 可以通过原型的类内数据的相似性获得, 假设 I_{ci} 对应原型为 $\mathbf{x}_{ci}$ 的聚类集合, 那么

$$\theta = (1.5 \sim 2) \max(S(\mathbf{x}_j, \mathbf{x}_{ci})), \mathbf{x}_j \in I_{ci}, \quad (10)$$

其中: $i = 1, \dots, c, j = 1, \dots, |I_{ci}|$, $|I_{ci}|$ 表示 I_{ci} 聚类集合中数据的个数.

3 仿真结果(Simulation results)

以连续搅拌釜式反应器(CSTR)作为故障预报的实例, 验证本文方法. CSTR的表达式:

$$\begin{cases} \frac{dC_A}{dt} = \frac{q}{V}(C_{Af} - C_A) - k_0 \exp(-\frac{E}{RT})C_A, \\ \frac{dT}{dt} = \frac{q}{V}(T_f - T) + \frac{-\Delta H}{\rho C_p} k_0 \exp(-\frac{E}{RT})C_A + \\ \frac{UA}{V\rho C_p}(T_c - T). \end{cases} \quad (11)$$

各变量含义及在标称状态下的参数请参见文[3].

采用欧拉离散法将系统离散化, 采样间隔为 $dt = 0.2 \text{ min}$. 反应器的控制率采用最简单的基于状态反馈的数值PID控制率, 进行浓度设定点的跟踪控制. 对系统设置一个缓变故障, 从第150步(30 min)开始, 进料流速沿指数曲线上升:

$$q(k) = q(150) + \exp(k - 150/50).$$

在应用本文提出的方法预报故障时, 采用下列参数: 嵌入维数为6; 混合核函数为两个高斯核, 且核参数分别为0.5和1.5; $\varepsilon = 0.075$; $C = 1000$. 为了验证该方法的优越性, 与文[3]提出的方法进行对比, 此时: 嵌入维数为6; 核函数为高斯核, 且核参数为1(基于交叉验证的网格搜索方法获得); $\varepsilon = 0.05$; $C = 1000$. K-均值聚类中, 聚类个数选为5. 两种方法对应的故障预报的结果如图1和图2所示.

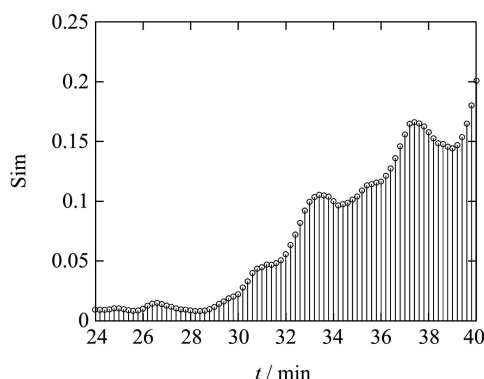


图1 基于多重核学习的故障预报

Fig. 1 Fault prediction by MKL

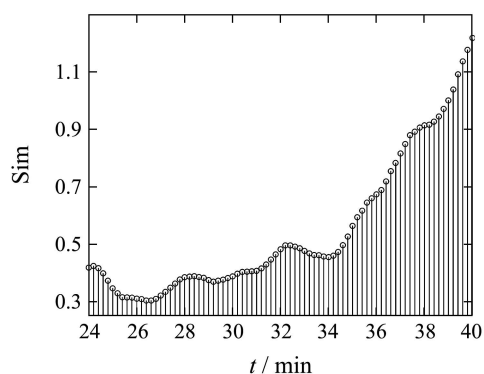


图2 基于SVR和K-均值聚类的故障预报

Fig. 2 Fault prediction by SVR and K-means clustering

多重核学习故障方法中, 运用MKL-SVR对正常时间序列建模获得8个支持向量, 在混合核空间上进行减聚类可提取3个正常原型, 并根据式(10)得到阈值为: $\theta = 0.1$. 则由图1可知, 在33 min时准确预报故障. 而基于SVR与K-均值聚类的方法中, 需要13个支持向量与5个正常原型, 且此时阈值为: $\theta = 0.6$. 则由图2可知, 在35 min时准确预报故障.

本文的故障预报方法由于引入多重核学习, 较之文[3]提前2 min预报了故障. 而且无论是支持向量的个数(8/13)还是正常原型的数目(3/5), 使本文的方法具有更为广阔的应用前景.

4 结论(Conclusion)

支持向量回归应用于故障预报系统, 其关键在于预测模型中支持向量的个数以及提取的正常原型的个数, 若支持向量或正常原型的个数较多, 难以满足故障预报的实时性. 本文引入多重核学习可以有效地解决上述问题, 提高故障预报系统的实时性. 而且, 混合核的使用可以提高故障预报系统的准确性.

参考文献(References):

- [1] TSE P W, ATHERTON D P. Prediction of machine deterioration using vibration based fault trends and recurrent neural networks[J]. *Journal of Vibration & Acoustics*, 1999, 121(7): 355 – 362.
- [2] VAPNIK V N. *The Nature of Statistical Learning Theory*[M]. Berlin: Springer, 1995.
- [3] 张军峰, 胡寿松. 基于聚类和支持向量机的非线性时间序列故障预报[J]. *控制理论与应用*, 2007, 24(1): 64 – 68.
(ZHANG Junfeng, HU Shousong. Nonlinear time series fault prediction based on clustering and support vector machines[J]. *Control Theory & Applications*, 2007, 24(1): 64 – 68.)
- [4] LANCKRIET G, BIE T D. A statistical framework for genomic data fusion[J]. *Bioinformatics*, 2004, 20(16): 2626 – 2635.
- [5] LANCKRIET G, CRISTIANINI N. Learning the kernel matrix with semidefinite programming[J]. *Journal of Machine Learning Research*, 2004, 5: 27 – 72.
- [6] SONNENBURG S, RÄTSCH G. Large scale multiple kernel learning[J]. *Journal of Machine Learning Research*, 2006, 7: 1 – 18.
- [7] ANDERSON E D, ANDERSON A D. The MOSEK interior point optimizer for linear programming[C]//*High Performance Optimization*. Berlin: Springer, 2000, 33: 197 – 232.
- [8] BOYD S, VANDERBERGHE L. *Convex Optimization*[M]. Cambridge: Cambridge University Press, 2004.
- [9] KIM D W, LEE K Y. Evaluation of the performance of clustering algorithms in kernel-induced feature space[J]. *Pattern Recognition*, 2005, 38(4): 607 – 611.
- [10] KIM D W, LEE K Y. A kernel-based subtractive clustering method[J]. *Pattern Recognition Letters*, 2005, 26(7): 879 – 891.

作者简介:

张军峰 (1979—), 男, 博士研究生, 目前研究方向为支持向量机、故障检测与故障预报, E-mail: wufeng7919@163.com;

胡寿松 (1937—), 男, 教授, 博士生导师, 研究方向复杂系统的可靠控制, E-mail: hushousong@nuaa.edu.cn.