基于残差预测修正的局部在线时间序列预测方法

刘大同, 彭　宇, 彭喜元
(哈尔滨工业大学自动化测试与控制研究所, 黑龙江哈尔滨 150080)

摘　要: 对于复杂的非线性和非平稳时间序列预测, 基本的支持向量回归(Support Vecotr Regression, SVR)在线算法无法有效兼顾执行效率和预测精度. 本文首先采用局部 SVR 进行时间序列建模预测, 同步计算在线更新序列数据预测的残差, 并采用 Online SVR 对残差序列进行混沌时间序列预测, 将预测残差值实时补偿到局部 SVR 模型预测输出. 实验结果表明, 新方法在执行效率和预测精度方面较单一 Online SVR 均显著提高.

关键词: 时间序列预测; 在线预测; SVR; 残差
中图分类号: TP391　文献标识码: A　文章编号: 0372-2112 (2008) 12A-081-05

Local Online Time Series Prediction Based on the Residual Compensation with Online SVR

LIU Da-tong, PENG Yu, PENG Xi-yuan
(*Automatic Test and Control Institute, Harbin Institute of Technology, Harbin, Heilongjiang 150080, China*)

Abstract: For complicated nonlinear and nonstationary time series prediction, the precision will be decreased if a faster processing speed is reached in online SVR. A new time series prediction method is proposed. Local SVR is firstly adopted to make prediction of the time series. Then the forecast residual with the real data stream is calculated. Finally, the residual is estimated with online SVR algorithm to compensate the predicted value with the local SVR. Experimental results showed that the proposed method outperformed the Online SVR.

Key words: time series prediction; online prediction; SVR; residual

1　引言

时间序列分析和预测在水文、天气、电力、金融等领域应用十分广泛, 可以为各种决策工作提供必要的参考, 已逐渐成为数据挖掘研究的热点问题之一, 并得到越来越多研究者的关注[1~3].

但是, 随着信息时代知识的爆炸性增长, 许多时间序列的非线性和非平稳性, 使得传统的方法, 如 AR、MA、ARMA 等[4]很难获得令人满意的结果. 因此, 随着人工智能技术的不断发展, 研究者们开始采用神经网络等智能算法进行时间序列的建模、预测以及分析. 但是, 神经网络存在过拟合、训练过程易陷于局部最小、采用经验风险最小原则等缺陷, 其建模预测和分析的效果限制了其广泛的应用[1,3].

近年来, 针对传统方法以及神经网络等模型进行时间序列预测存在的不足, 应用支持向量回归(Support Vector Regression, SVR)进行时间序列预测的研究十分广泛. 由于其具有算法简单、不存在局部最小和维数灾难问题和泛化能力强等优点[5], 在时间序列预测方面取得很好的效果. SVR 建立在统计学习理论和结构风险最小化原理基础上, 能够较好地解决小样本、非线性、高维数和局部极小等实际问题. 但是 SVR 算法存在训练时间较长、模型更新训练计算量大等问题, 不适用于在线问题的处理. 近期出现了很多在线 SVR 算法(比如增量训练算法、AOSVR 等), 可以在线学习和更新, 适合解决时间序列数据特性变化大、模型频繁更新等问题, 从而适合在线问题的分析和预测[6,7].

针对在线时间序列预测问题, 本文提出一种新的在线时间序列预测方法, 首先采用 SVR 进行局部建模, 并采用 Online SVR 进行局部模型预测残差在线预测, 实时修正预测结果, 从而达到提高总体预测精度的目的.

2　算法理论基础

支持向量回归是 Vapnik 等人根据统计学理论提出的一种新的通用学习方法, 它是建立在统计学理论的 VC 维理论和结构风险最小化原理基础上, 以结构风险最小化原理为基础, 具有其它以经验风险最小化原理为基础的算法难以比拟的优越性, 同时它是一个凸二次优化问

收稿日期: 2008-10-07; 修回日期: 2009-01-16

题,能够保证得到的极值解是全局最优解. SVR 综合考虑经验风险与置信风险,具有很好的泛化能力[6,7].

2.1 基于 SVR 的时间序列预测

设给定时间序列数据集:

$$T=\left\{(x_1,y_1),\cdots,(x_i,y_i)\right\}\in(X\times Y)^l$$

其中 $x_i\in X=R^n$, $y_i\in Y=R$, $i=1,\cdots,l$. 构造回归函数(预测函数)为:

$$f(x)=W^{\mathrm T}\Phi(x)+b \tag{1}$$

式(1)中, $W^{\mathrm T}\in R^n$, $b\in R$, $\Phi(\cdot)$ 把输入样本从输入空间映射到高维特征空间, w 和 b 通过优化问题求解得到:

$$\min_{w,b}P=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*) \tag{2}$$

式(2)需满足以下条件:

$$((wx_i)+b)-y_i\leqslant\varepsilon+\xi_i,\quad i=1,2,\cdots,l,$$
$$y_i-((wx_i)+b)\leqslant\varepsilon+\xi_i,\quad i=1,2,\cdots,l,$$
$$\xi_i^*\geqslant 0,\quad i=1,2,\cdots,l.$$

这里, ξ 和 ξ^* 为松弛变量, C 为惩罚参数, ε 为不敏感损失函数.

转换上述问题的 Lagrange 优化问题:

$$\min_{\alpha,\alpha^*}\frac{1}{2}\sum_{i}^{l}\sum_{j}^{l}Q_{ij}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)+\varepsilon\sum_{i}^{l}(\alpha_i+\alpha_i^*)-\sum_{j}^{l}y_i(\alpha_i-\alpha_i^*) \tag{3}$$

式(3)需满足以下条件:

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0,$$
$$0\leqslant\alpha_i,\alpha_i^*\leqslant\frac{C}{l},\quad i=1,2,\cdots,l$$

定义核函数 $Q_{ij}=\Phi(x_i)^{\mathrm T}\Phi(x_j)=K(x_i,x_j)$, 对于回归函数(预测函数) $f(x)=W^{\mathrm T}\Phi(x)+b$ 表达为:

$$f(x)=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)K(x_i,x)+b \tag{4}$$

由于 SVR 只考虑了高维特征空间的点积运算,从而解决了因映射函数未知而无法显式表达的问题,按照 KKT(Karush-Kuhn-Tucker)定理,可以得到:

$$\begin{cases} h(x_i)\geqslant\varepsilon, & \theta_i=-C\\ h(x_i)=\varepsilon, & -C<\theta_i=0\\ -\varepsilon\leqslant h(x_i)\leqslant\varepsilon, & \theta_i=0\\ h(x_i)=\varepsilon, & 0<\theta_i=C\\ h(x_i)\geqslant-\varepsilon, & \theta_i=C\end{cases} \tag{5}$$

其中 $\theta_i=\alpha_i-\alpha_i^*$

$$h(x)\equiv f(x_i)-y_i=\sum_{j=1}^{l}Q_{ij}\theta_j-y_i+b$$

根据 $f(x_i)-y_i$ 误差的情况,可分为以下几种情况:

(1) 错误集: $E=\left\{i\,|\,|\theta_i|=C\right\}$;

(2) 支持向量集: $S=\left\{i\,|\,0<|\theta_i|<C\right\}$;

(3) 保留集: $R=\left\{i\,|\,\theta_i=0\right\}$.

普通的离线 SVR 需要通过对已获得数据集进行训练,然后进行预测,而对于实时性较高、数据更新较快的情况,每次更新数据时,都需要对 SVR 模型进行重新训练,效率较差.因此,针对实时应用情况,出现了很多 SVR 在线训练的方法,应用较多的就是增量学习算法(Incremental Algorithm)和减量学习算法(Decremental Algorithm),当更新数据时,并不重新对 SVR 进行训练,而是通过直接增加一个数据或者去掉一个数据,在线动态调整 SVR 结构以及对应训练数据集数据的特性,使得 SVR 仍然可以符合 KKT 条件的要求[8].

假设现有一个新的样本 x_C 加入到训练集中,此时需要加入 θ_C, 并计算各个 $\theta_i(i=1,2,\cdots,n)$ 和 θ_C 的更新值 $\Delta\theta_i$ 及 $\Delta\theta_C$ 以使得所有的所有样本均满足 KKT 条件.

$$\Delta h(x_i)=K(x_i,x_c)\Delta\theta_c+\sum_{i=1}^{n}K(x_i,x_j)\Delta\theta_j+\Delta b \tag{6}$$

又因为

$$\theta_c+\sum_{i=1}^{n}\theta_i=0 \tag{7}$$

对集合 S, 可得:

$$\begin{aligned}&\sum_{j\in S}K(x_i,x_j)\Delta\theta_j+\Delta b=-K(x_i,x_c)\Delta\theta_c\\&\sum_{j\in S}\Delta\theta_j=-\Delta\theta_c\\&j\in S\end{aligned} \tag{8}$$

如果定义集合 S 中样本的下标集为 $S\{s_1,s_2,\cdots,s_{l_S}\}$, 则式(8)可以表示为矩阵形式:

$$\begin{bmatrix}0 & 1 & \cdots & 1\\ 1 & K(x_{S_1},x_{S_1}) & \cdots & K(x_{S_1},x_{S_{l_S}})\\ \vdots & \vdots & \ddots & \vdots\\ 1 & K(x_{S_{l_S}},x_{S_{l_l}}) & \cdots & K(x_{S_{l_S}},x_{S_{l_S}})\end{bmatrix}\begin{bmatrix}\Delta b\\ \Delta\theta_{S_1}\\ \vdots\\ \Delta\theta_{S_{l_S}}\end{bmatrix}=-\begin{bmatrix}1\\ K(x_{S_1},x_C)\\ \vdots\\ K(x_{S_{l_S}},x_{S_{l_S}})\end{bmatrix}\Delta\theta_c \tag{9}$$

由上式可得:

$$\begin{bmatrix}\Delta b\\ \Delta\theta_{S_1}\\ \vdots\\ \Delta\theta_{S_{l_S}}\end{bmatrix}=-\begin{bmatrix}0 & 1 & \cdots & 1\\ 1 & K(x_{S_1},x_{S_1}) & \cdots & K(x_{S_1},x_{S_l})\\ \vdots & \vdots & \ddots & \vdots\\ 1 & K(x_{S_{l_S}},x_{S_{l_l}}) & \cdots & K(x_{S_{l_S}},x_{S_{l_S}})\end{bmatrix}^{-1}\begin{bmatrix}1\\ K(x_{S_1},x_C)\\ \vdots\\ K(x_{S_{l_S}},x_{S_{l_S}})\end{bmatrix}\Delta\theta_c=\beta\Delta\theta_c \tag{10}$$

由上面推导可得:

$$\begin{bmatrix}\Delta h(x_{n_1})\\ \Delta h(x_{n_2})\\ \vdots \\ \Delta h(x_{n_{l_N}})\end{bmatrix}$$

$$=\left\{\begin{bmatrix}K(x_{n_1},x_C)\\ K(x_{n_2},x_C)\\ \vdots\\ K(x_{n_k},x_C)\end{bmatrix}+\begin{bmatrix}1 & K(x_{n_1},x_{S_1}) & \cdots K(x_{n_1},x_{S_{l_S}})\\ 1 & K(x_{n_2},x_{S_1}) & \cdots K(x_{n_2},x_{S_{l_S}})\\ \vdots & \vdots & \ddots & \vdots\\ 1 & K(x_{n_{l_N}},x_C & \cdots K(x_{n_{l_N}},x_{S_{l_S}})\end{bmatrix}\beta\right\} \tag{11}$$

由此实现 SVR 的在线学习, 并实现 S、E 和 R 集的更新.

2.2　基于 SVR 的混沌时间序列预测

混沌时间序列预测是建立在 Takens 提出的嵌入定理和相空间重构理论基础上的, 其基本思想是系统中的任一分量的演化是由与之相互作用着的其它分量所决定的, 这些相关分量的信息就隐含在任一分量的发展过程中, 可以从仅仅与时间相关的混沌数据中提取和恢复出系统原来的规律, 实质上表达为高维空间下的一种轨迹.

根据状态空间重构理论, 假设观测到的混沌时间序列为$\{x(k), k=1,2,\cdots,N\}$, 则在状态空间中重构的一点状态矢量可表示为:

$$\boldsymbol{X}(k)=[x(k),x(k+\tau),\cdots,(k+(m-1)\tau)]^{\mathrm{T}} \tag{12}$$

其中 m 为嵌入维数, τ 为延迟时间.

由 Takens 定理可知, 混沌序列的预测重构本质上是一个动力系统的逆问题, 对于一步预测也就是建立

$$\boldsymbol{X}(k+1)=F(\boldsymbol{X}(k)),\quad k=1,2,\cdots,N \tag{13}$$

其中 $F(\cdot)$ 为预测重构模型[9].

3　基于残差预测修正的局部在线时间序列预测

3.1　算法基本思路

对于复杂的非线性、非平稳的时间序列来说, 使用单一模型进行预测,往往无法获得满意的效果[10].

针对上述单一在线 SVR 模型预测复杂时间序列效果问题, 本文提出一种融合局部 SVR 模型与 Online SVR 模型的时间序列预测方法. 算法首先采用离线 SVR 对时间序列数据进行分段局部建模, 同步采用 Online SVR 模型对局部 SVR 训练误差构成的序列进行的初始训练. 对于更新时间序列数据, 采用局部 SVR 进行预测, 并利用 Online SVR 算法进行残差的在线预测, 并对局部 SVR 预测值实时修正, 从而实现对于时间序列的在线实时预测. 算法的结构示意图如图 1 所示.

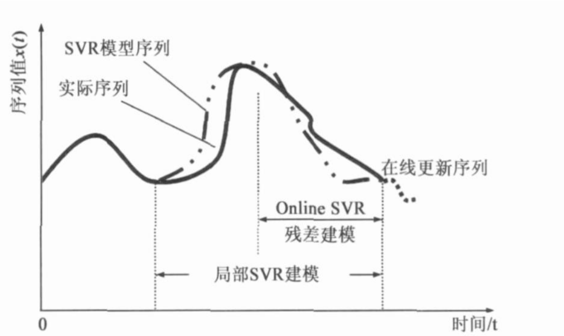


图1　基于残差预测修正的局部在线模型算法示意图

从图 1 中可以看到, 对于一定邻域内的时间序列进行预测, 算法进行了一次局部 SVR 建模和 Online SVR 建模, 对于在线数据的更新, 局部 SVR 模型并未进行更新, 而是采用 Online SVR 的残差预测进行在线更新, 相比于复杂时间序列, 进行混沌时间序列预测时其嵌入维数也会相对较大, 而残差序列作为一种模型等效的系统误差, 其建模的嵌入维数也会降低, 所以进行残差建模的序列数据集规模可以相对较小. 因此, 相对于 Online SVR 直接进行复杂时间序列预测, 本文所提出方法的执行效率能够得到提升. 同时由于算法进行了局部 SVR 模型预测结果的残差建模和预测修正, 所以从精度上相比 Online SVR 直接预测也会有所提高.

3.2　算法流程

基于残差预测修正的局部在线时间序列预测方法的基本流程如下:

(1)选取时间序列数据局部域(在线数据邻域), 进行序列数据的相空间重构, 将其映射到状态空间中;

(2)采用局部 SVR 算法对重构时间序列数据进行初始化建模;

(3)对在线更新的数据样本(记为 $\boldsymbol{X}(k)$)进行一步预测, 预测值记为 $\widetilde{\boldsymbol{X}}(k+1)$;

(4)计算上一步预测值的残差序列(记为 $\mathrm{var}(k)=\widetilde{\boldsymbol{X}}(k)-\boldsymbol{X}(k)$);

(5)对残差序列进行相空间重构, 采用 Online SVR 进行在线建模预测残差, 预测值记为$\mathrm{var}(k+1)$;

(6)在线修正局部 SVR 模型预测值, 记为 $\widetilde{\boldsymbol{P}}(k+1)$, $\widetilde{\boldsymbol{P}}(k+1)=\widetilde{\boldsymbol{X}}(k+1)+\mathrm{var}(k+1)$;

(7)数据样本在线更新,继续第(3)~(6)步, 实现对复杂时间序列的局部在线预测;

(8)当更新数据超过选定建模预测领域, 更新局部 SVR 模型, 重复步骤(1)~(7).

分析算法的流程, 可见它实质上是与去趋势、去周期的时间序列预测模型相类似的. 但是, 对于复杂时间序列数据, 本算法并未人为地进行数据的趋势和周期分解, 并没有影响和破坏序列数据的自身特征, 消除了

基于先验知识分解序列数据所引起的序列特征的变化, 也降低了引入序列数据新特征的风险.

4 实验验证

为验证算法的有效性, 利用太阳黑子数据(1700~1970)进行实验验证, 序列数据如图 2 所示.

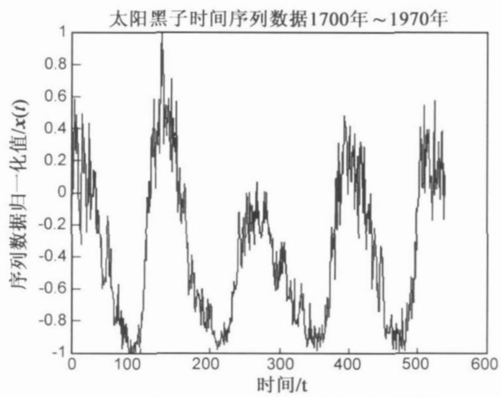


图2 归一化太阳黑子时间序列数据

结合太阳黑子数据更新特性, 实验选取 1~200 个数据样本点进行局部 SVR 初始化建模, 算法采用 RBF 核函数, SVR 参数选取参考文献[12], 预测的后 100 个数据序列的残差, 如图 3 所示.

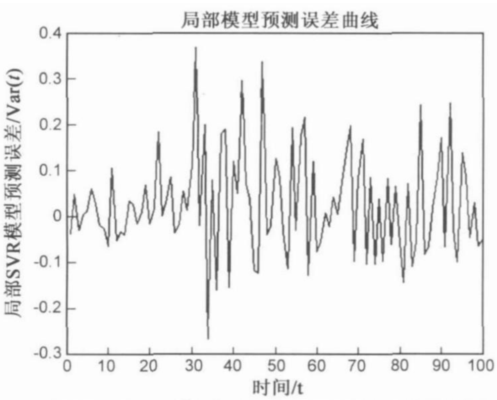


图3 基于局部SVR模型预测太阳黑子时间序列的残差

从图 3 局部 SVR 模型预测的残差来看, 部分样本点预测的效果较差, 可以采用混沌时间序列预测的方法进行预测修正. 因此, 在局部 SVR 建模预测时, 同步计算模型预测残差序列, 利用 Online SVR 进行残差序列在线预测, 进行预测值的实时修正.

算法中 Online SVR 模型仍然选取 RBF 核函数, 核函数参数 $p=15$, $C=100$, $\varepsilon=0.01$, 由于残差序列显示了较强的局部变化特性, 因此选取相同的嵌入维数, 以局部 SVR 建模初始训练数据残差进行 Online SVR 建模(实验中初始化残差数据集长度选为 70), 图 4 为进行残差预测在线修正的预测序列值与实际序列(Sunspot 数据 201-300)的对比情况.

从图 4 中可以看到, 基于 Online SVR 的残差修正的局部在线时间序列预测模型能够有效地预测太阳黑子时间序列. 为全面比较算法的预测性能, 采用两种性能评价标准: 平均绝对误差(Mean Absolute Error, MAE)和规范化均方根误差(Normalized Root Mean Square Error, NRMSE), MAE 的定义为:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|x(i)-\tilde{x}(i)\right| \tag{13}$$

NRMSE 的定义式为:

$$NRMSE=\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[x(i)-\tilde{x}(i)\right]^{2}}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[x(i)-\bar{x}\right]^{2}}} \tag{14}$$

其中 n 为预测集数据个数, $x(i)$ 为真实值, $\tilde{x}(i)$ 为预测值, $\tilde{x}$ 为序列均值[11].

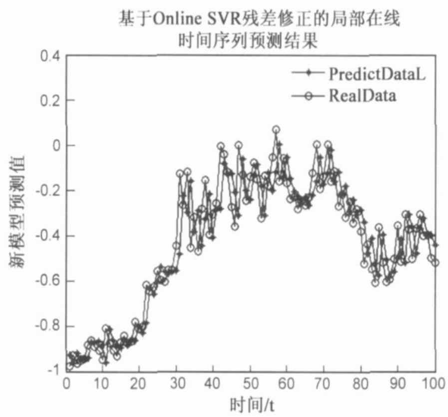


图4 基于在线残差修正的SVR局部模型预测太阳黑子时间序列

对太阳黑子数据进行分段局部建模在线预测, 进行 10 次实验的平均结果如下表 1.

表 1 算法性能比较

序号	预测数据	算 法	MAE	NRMSE	执行时间
1	201-300	局部SVR+Online SVR 残差预测	0.099	0.437	268.4s
		Online SVR	0.112	0.467	2293.8s
2	301-400	局部SVR+Online SVR 残差预测	0.111	0.328	340.9s
		Online SVR	0.210	0.402	2691.6s
3	401-500	局部SVR+Online SVR 残差预测	0.125	0.391	256.6s
		Online SVR	0.231	0.684	2605.8s

从表 1 可以看出, 使用 Online SVR 残差修正的局部在线时间序列预测方法, 其平均绝对误差和规范化均方误差均小于采用单一 Online SVR 模型的误差值, 算法精度得到提高. 由于残差序列的建模长度小, 算法执行速度效率比单一 Online SVR 方法提高约 5~10 倍.

由于局部 SVR 只在序列邻域进行一次建模, 因此执行时间较短, 而对于 Online SVR 直接进行复杂且非线性强的时间序列预测, 必须采用较大规模的初始训练

集,因此,在每次进行在线模型更新时,涉及的矩阵运算量大,所以算法执行时间也长.相比来说,由于残差序列相对被预测序列自身,其复杂度降低,故而可以采取较小的嵌入维数、较小的建模数据长度.所以,算法实时性比直接采用 Online SVR 方法要好.算法精度方面,相比于单一Online SVR 直接预测算法,本文提出的算法性能有所提高,究其原因,由于单一 Online SVR 方法进行预测,存在不敏感损失,加上时间序列数据的复杂特性,预测后所剩残差体现为更微观、更高频的变化特性,采用单一的支持向量机核函数,无法有效对待预测数据进行精确的一一逼近.

本文提出的算法相比于Online SVR 方法,算法实时性较好.融合局部 SVR 算法,实现模型在线更新和预测,因此,比普通离线 SVR 方法适合处理在线预测问题.

5 结论

采用基于 Online SVR 残差修正的局部在线预测方法,通过对局部 SVR 建模核函数单一特性的补偿,在保证算法预测精度基础上,对时间序列预测结果进行在线修正,既保证了时间序列预测算法的执行效率,同时,又有效提高了预测的精度.实验研究证明其预测效能好于单一的 Online SVR 方法.但是,对于不同特性的时间序列,方法在残差修正 Online SVR 的参数选取方面、与局部 SVR 预测模型参数关系以及局部建模邻域规模选择等方面,还需要进一步研究.

参考文献:

[1] 董辉,傅鹤林,冷伍明.支持向量机的时间序列回归与预测[J].系统仿真学报,2006,18(7):1785-1788.
Dong Hui, Fu He-lin, Leng Wu-ming. Support vector machines for time series regression and prediction[J]. Journal of System Simulation, 2006, 18(7): 1785-1788. (in Chinese)

[2] 田翔,邓飞其.精确在线支持向量回归在股指预测中的应用[J].计算机工程,2005,31(22):18-20.
Tian Xiang, Deng Feiqi. Application of accurate online support vector regression in stock market index forecasting[J]. Computer Engineering, 2005, 31(22): 18-20. (in Chinese)

[3] 吴琼,杨以涵,刘文颖.基于最小二乘支持向量机的电力系统暂态稳定在线预测[J].中国电机工程学报,2005,27(25):38-43.
Wu Qiong, Yang Yi-han, Liu Wen-ying. Electric power system transient stability on-line prediction based on least squares support vector machine[J]. Proceedings of the CSEE, 2005, 27(25): 38-43. (in Chinese)

[4] Jianqing Fan, Qiwei Yao. Nonlinear Time Series: Nonparametric and Parametric Methods[M]. USA: Springer, 2003. 10-27, 89-120.

[5] K R Müller, A J Smola, G Rätsch, B Schölkopf, J Kohlmorgen, V Vapnik. Predicting time series with support vector machines[A]. Proceeding of ICANN'97[C]. Lausanne, Switzerland, 1997. 999-1004.

[6] Debasish Basak, Srimanta Pal, Dipak Chandra Patranabis. Support vector regression[J]. Neural Information Processing-Letters and Reviews, 2003, 11(10): 203-224.

[7] Alex J. Smola, Bernhand Schölkopf. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3): 199-222.

[8] Junshui Ma, James Theiler. Simon perkins. Accurate on-line support vector regression[J]. Neural Computation, 2003, 15(11): 2683-2703.

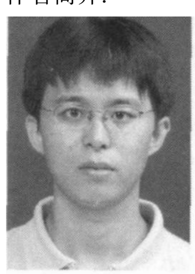
[9] Pavel Laskov, Christian Gehl, Stefan Krüger, Klaus-Robert Müller. Incremental support vector learning: analysis, implementation and applications[J]. Journal of Machine Learning Research, 2006, 7: 1909-1936.

[10] 叶美盈,汪晓东,张浩然.基于在线最小二乘支持向量机回归的混沌时间序列预测[J].物理学报,2005,54(6):2568-2573.
Ye Mei-ying, Wang Xiao-dong, Zhang Hao-ran. Chaotic time series forecasting using online least squares support vector machine regression[J]. Acta Phys. Sin. 2005, 54(6): 2568-2573. (in Chinese)

[11] 彭喜元,王军,彭宇.一种新型时间序列多分辨预测模型研究[J].电子学报,2007,35(11):2146-2149.
Peng Xi-yuan, Wang Jun, Peng Yu. A novel multi-scale predictor for comlpex time series[J]. Acta Electronica Sinica. 2007, 35(11): 2146-2149. (in Chinese)

[12] Jacek M. leski. On support vector regression machines with linguistic interpretation of the kernel matrix[J]. Fuzzy Sets and Systems 2006, 157(8): 1092-1113.

作者简介:



刘大同 男,1982 年 9 月出生于黑龙江明水,哈尔滨工业大学自动化测试与控制系统博士研究生,主要从事自动测试与仿真、故障诊断和预测以及时间序列分析研究工作.

E-mail: ldatong@163.com

彭 宇 男,教授、博士生导师,1973 年6 月生于陕西西安,哈尔滨工业大学自动化测试与控制研究所副所长,主要研究方向为测试诊断技术、无线传感器网络技术和数据挖掘技术等.

E-mail: pengyu@hit.edu.cn