DOI：10.11830/ISSN.1000-5013.201708019

KPCA-LSSVM 方法在视频时间序列预测中应用

张观东，李军

（兰州交通大学 自动化与电气工程学院，甘肃 兰州 730070）

摘要： 为提高时间序列预测精度及降低预测过程中的计算复杂度，提出一种基于核主成分分析（KPCA）与最小二乘支持向量机（LSSVM）相结合的预测方法. 首先，将输入数据通过核方法映射至高维特征空间；然后，在特征空间上提取有效非线性主元；最终，通过 LSSVM 建立时间序列模型. 为验证 KPCA-LSSVM 方法的有效性，将其应用于交通流及视频流预测中，在同等条件下，与单一的 LSSVM 及神经网络等预测方法进行比较. 实验结果表明：基于 KPCA-LSSVM 建立的模型具有较好的推广性及较高的辨识精度.

关键词： 时间序列预测；交通流量；视频流量；核主成分分析；最小二乘支持向量机

中图分类号： TP 183 **文献标志码：** A **文章编号：** 1000-5013(2018)02-0281-05

Application of KPCA-LSSVM in Video Trace and Time Series Prediction

ZHANG Guandong, LI Jun

(School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: A prediction method based on kernel principal component analysis (KPCA) and least squares support vector machine (LSSVM) is proposed for the prediction of time series that increasing prediction precision and decreasing the computing complexity. Firstly, the input data will be mapped to high-dimensional feature space through kernel method, then the effective nonlinear principal element can be extracted in the feature space, and finally the time series model is established by LSSVM. In order to verify the validity of KPCA-LSSVM method, it is used in traffic flow and video flow prediction, and compared with single LSSVM and neural network in the same condition. The experimental results show that the model based on KPCA-LSSVM has good generalization and high identification accuracy compared with other methods.

Keywords: time series prediction; traffic flow; video flow; kernel principal component analysis; least squares support vector machine

时间序列广泛存在于金融、交通等领域[1-3]. 目前，神经网络[4-6]、支持向量机（SVM）[7-10]及各种智能算法[11-16]在时间序列预测中已取得成功的应用. 然而，前馈神经网络本质为静态神经网络，无法有效地应用于系统的实时辨识中. 递归作为动态神经网络，虽然具有动态记忆能力，但在实际辨识过程中很难取得应用，且容易导致神经网络陷入局部最优、过拟合、泛化能力弱等困局. SVM 遵循结构风险最小化原则[7-10]，具有较好的稀疏性及鲁棒性，但求解凸二次优化的计算量大且复杂. 核主成分分析（KPCA）作

收稿日期： 2017-08-16

通信作者： 李军（1969-），男，教授，博士，主要从事复杂非线性建模与控制的研究. E-mail：lijun691201@mail.lzjtu.cn.

基金项目： 国家自然科学基金资助项目（51467008）

为一种非监督学习方法,可在高维特征空间内有效提取数据的非线性特征.最小二乘支持向量机(LSSVM)作为标准SVM的延伸,可将SVM中求解凸二次优化问题转化为求解线性方程组,极大降低计算复杂度.本文结合KPCA与LSSVM各自的优点,提出一种KPCA-LSSVM的时间序列建模预测方法.

1 KPCA-LSSVM方法

将KPCA-LSSVM应用于时间序列预测过程中.首先,对训练数据进行KPCA处理,提取其非线性主元;然后,通过LSSVM建立混时间序列的模型.

1.1 KPCA方法

首先,将训练数据映射至高维特征空间.然后,在特征空间进行主成分分析(PCA)处理.

训练数据 $\boldsymbol{x}_i(i=1,2,\cdots,l;\boldsymbol{x}_i\in\boldsymbol{R}^n)$ 通过非线性映射 $\varphi:\boldsymbol{x}\in\boldsymbol{R}^n\rightarrow\varphi(\boldsymbol{x})\in\boldsymbol{F}\subseteq\boldsymbol{R}^N$,映射到 N 维特征空间中,构成 $l\times N$ 维的矩阵 $\boldsymbol{X}[\varphi(\boldsymbol{x}_1),\varphi(\boldsymbol{x}_2),\cdots,\varphi(\boldsymbol{x}_l)]^{\mathrm{T}}$.定义核矩阵为

$$\boldsymbol{K}=\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}. \tag{1}$$

式(1)中:$\boldsymbol{K}$ 中元素为 $K_{i,j}=k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\varphi(\boldsymbol{x}_i)^{\mathrm{T}}\varphi(\boldsymbol{x}_j)$.

则核矩阵 $\boldsymbol{K}$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 为

$$l\,\boldsymbol{\Sigma}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}. \tag{2}$$

式(2)中:$\boldsymbol{\Sigma}$ 中的元素为 $\boldsymbol{\Sigma}_{s,t}=\dfrac{1}{l}\sum\limits_{i=1}^{l}\varphi(\boldsymbol{x}_i)_s\varphi(\boldsymbol{x}_i)_t$,其中,$s,t=1,\cdots,N$.

对 $\boldsymbol{K}$ 进行中心化处理,即

$$\hat{\boldsymbol{K}}=\boldsymbol{K}-\frac{1}{l}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{K}-\frac{1}{l}\boldsymbol{K}\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}+\frac{1}{l^2}(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{e})\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}. \tag{3}$$

式(3)中:$\boldsymbol{e}$ 为元素均为1的列向量.

对 $\boldsymbol{\Sigma}$ 及 $\hat{\boldsymbol{K}}$ 进行特征值分解,可得

$$\hat{\boldsymbol{K}}=\boldsymbol{V}\boldsymbol{\Lambda}_l\boldsymbol{V}^{\mathrm{T}},\qquad l\,\boldsymbol{\Sigma}=\boldsymbol{U}\boldsymbol{\Lambda}_N\boldsymbol{U}^{\mathrm{T}}. \tag{4}$$

式(4)中:正交矩阵 $\boldsymbol{V}$ 的列 $\boldsymbol{v}_i$ 为 $\hat{\boldsymbol{K}}$ 的特征向量;正交矩阵 $\boldsymbol{U}$ 的各列形成的向量 $\boldsymbol{u}_i$ 为 $l\,\boldsymbol{\Sigma}$ 的特征向量.

由于 $\hat{\boldsymbol{K}}$ 与 $l\,\boldsymbol{\Sigma}$ 对称,则 $l\,\boldsymbol{\Sigma}$ 的任意特征向量 $\boldsymbol{u}$ 与特征值 λ 可变为 $\hat{\boldsymbol{K}}$ 所对应的特征向量 $\boldsymbol{X}\boldsymbol{u}$ 与特征值 λ.令 $t=\mathrm{rank}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=\mathrm{rank}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\leqslant\min(N,l)$,则 $\boldsymbol{U}$ 的前 t 列特征向量构成的矩阵 $\boldsymbol{U}_t$ 可表示为

$$\boldsymbol{U}_t=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}_t\boldsymbol{\Lambda}_t^{-1/2}. \tag{5}$$

式(5)中:假定 $\hat{\boldsymbol{K}}$ 与 $l\,\boldsymbol{\Sigma}$ 的前 t 个非零特征值是按降序排列的.

由式(5)可知:$l\,\boldsymbol{\Sigma}$ 的第 j 个特征向量 $\boldsymbol{u}_j$ 具有一种相应的对偶表示,即可由核矩阵 $\hat{\boldsymbol{K}}$ 的相应特征向量 $\boldsymbol{v}_j$ 乘以尺度化因子系数 $\lambda_j^{-1/2}$,可得

$$\boldsymbol{u}_j=\lambda_j^{-1/2}\sum_{i=1}^{l}(\boldsymbol{u}_j)_i\varphi(\boldsymbol{x}_i)=\sum_{i=1}^{l}\alpha_i^j\varphi(\boldsymbol{x}_i),\qquad j=1,\cdots,t. \tag{6}$$

式(6)中:向量 $\boldsymbol{u}_j$ 的对偶变量 $\alpha^j=\lambda_j^{-1/2}\boldsymbol{v}_j$;$\boldsymbol{v}_j$,$\lambda_j$ 分别为 $\hat{\boldsymbol{K}}$ 的第 j 个特征向量及对应的特征值.

考虑式(6),若定义 $\boldsymbol{U}_d$ 是特征空间中由前 d 个特征向量 $\boldsymbol{u}_d$ 所张成的子空间,训练数据 $\varphi(\boldsymbol{x})$ 在该子空间上的 d 维向量投影为

$$P_{U_d}(\varphi(\boldsymbol{x}))=(\boldsymbol{u}^{\mathrm{T}}\varphi(\boldsymbol{x}))_{j=1}^{d}=[(\sum_{i=1}^{l}\alpha_i^j\varphi(\boldsymbol{x}_i),\varphi(\boldsymbol{x}))]_{j=1}^{d}=(\sum_{i=1}^{l}\alpha_i^jk(\boldsymbol{x}_i,\boldsymbol{x}))_{j=1}^{d}. \tag{7}$$

式(7)中:$(\cdot)$是内积符号;α_i^j 为第 j 个主元的第 i 个元素.将特征值 λ 按照降序排列,可选取前 d 个主元($d\leqslant l$).文中将使用最常见的高斯核函数,即

$$k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp\{-\parallel\boldsymbol{x}_i-\boldsymbol{x}_j\parallel/(2\delta^2)\}. \tag{8}$$

式(8)中:δ 为核函数半径.

1.2 LSSVM方法

假设有训练样本$(\boldsymbol{x}_i,\boldsymbol{y}_i)(i=1,2,\cdots,M)$,则SVM回归表达式为

$$y=f(\boldsymbol{x})=(\boldsymbol{\omega},\varphi_1(\boldsymbol{x}))+b. \tag{9}$$

式(9)中:$\boldsymbol{\omega}$ 为权系数矩阵;b 为偏差;φ_1 为非线性映射函数.则其结构风险 J 为

$$J = \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} + C\sum_{i=1}^{M}(\xi_i + \xi_i^*). \tag{10}$$

式(10)中：ξ_i, ξ_i^* 为松弛变量；C 为惩罚系数. 由结构风险最小化原则，有

$$\left.\begin{aligned} &\min J = \min_{\boldsymbol{\omega},\xi_i,\xi_i^*}\left(\frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\omega} + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right),\\ &\text{s. t.}\begin{cases} y_i - (\boldsymbol{\omega}, \varphi_1(\boldsymbol{x}_i)) - b \leqslant \varepsilon + \xi_i,\\ (\boldsymbol{\omega}, \varphi_1(\boldsymbol{x}_i)) + b - y_i \leqslant \varepsilon + \xi_i,\\ \xi_i \geqslant 0, \quad \xi_i^* \geqslant 0. \end{cases} \end{aligned}\right\} \tag{11}$$

式(11)中：ε 为允许拟合误差. 由拉格朗日乘子法、对偶原理及核方法可将式(11)转换为

$$\left.\begin{aligned} &\max_{\alpha,\eta}\boldsymbol{W}(\alpha,\eta) = -\frac{1}{2}\sum_{i,j=1}^{M}(\alpha_i - \eta_i)(\alpha_j - \eta_j)\boldsymbol{K}_1(\boldsymbol{x}_i, \boldsymbol{x}_j) - \varepsilon\sum_{i=1}^{M}(\alpha_i + \eta_i) + \sum_{i=1}^{M} y_i(\alpha_i - \eta_i),\\ &\text{s. t.}\begin{cases} \sum_{i=1}^{M}(\eta_i - \alpha_i) = 0,\\ 0 \leqslant \alpha_i, \quad \eta_i \leqslant C, \quad i = 1, \cdots, M. \end{cases} \end{aligned}\right\} \tag{12}$$

式(12)中：α, η 为拉格朗日乘子；$\boldsymbol{K}_1$ 为核矩阵. 由式(12)可求得 α, η 最优解为 α_i^*，η_i^*，其偏差 b 可由库恩塔克条件(KKT)求得.

由上述分析可知：在标准的 SVM 中，其优化求解过程复杂、繁琐. LSSVM 中将式(12)中的约束优化问题转化为如下线性方程组，即

$$\begin{bmatrix} 0 & \boldsymbol{e}_1^{\mathrm{T}} \\ \boldsymbol{e}_1 & \boldsymbol{Q} + \boldsymbol{I}/C \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix}. \tag{13}$$

式(13)中：$\boldsymbol{y} = [y_1, \cdots, y_M]^{\mathrm{T}}$；$\boldsymbol{e}_1 = [1, \cdots, 1]^{\mathrm{T}}$；$\boldsymbol{Q}_{i,j} = \boldsymbol{K}_1(\boldsymbol{x}_i, \boldsymbol{x}_j)$；$\boldsymbol{I}$ 为单位矩阵.

由上述分析可知，最终通过 LSSVM 建立的模型为

$$y = f(x) = \sum_{i=1}^{M}(\alpha_i^* - \eta_i^*)\boldsymbol{K}_1(\boldsymbol{x}_i, \boldsymbol{x}) + b. \tag{14}$$

LSSVM 的核函数 k_1 可采用线性核，其形式为

$$k_1(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j. \tag{15}$$

2　时间序列预测

对于时间序列$\{y(t)\}$，由相空间重构可将其 k 阶非线性模型表示为

$$y(t+1) = f(y(t), y(t-1), \cdots, y(t-k+1)) + \varepsilon(t). \tag{16}$$

式(16)中：$f(\cdot)$为非线性映射函数；$\varepsilon(t)$为模型不确定项或高斯白噪声信号；$y(t)$为 t 时刻值.

将 KPCA-LSSVM 应用于交通流及视频流预测，为衡量文中所提方法的有效性，在同等条件下，将文中所提方法的预测结果与径向基函数(RBF)神经网络、SVM 及单一的 LSSVM 预测结果进行比较. 为评价预测的精度，采用均方根误差(RMSE)及平均绝对误差(MAE)作为评价指标，其表达式为

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \Big/ n}, \qquad \mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{N}|\hat{y}_i - y_i|. \tag{17}$$

2.1　交通流量预测

实验选取了某地区提供的交通流数据，数据采集时间为 2011 年 3 月 1－5 日. 5 d 内共采集了 336 个数据点，即每隔 15 min 采集一次，其流量如图 1 所示. 图 1 中：n 为样本数；T_{f} 为交通流. 实验参数的选取：嵌入维数 $k=5$；δ 核函数半径为 1；LSSVM 中的正则化参数 $C^2 = 8.4$. 实验过程中，前 250 组数据作为训练数据，余下 86 组作为测试数据. 不同方法的预测效果图，如图 2 所示. 由图 2 可知：KPCA-LSSVM 的预测效果最佳.

为进一步衡量 KPCA-LSSVM 的预测效果，在同等条件下，研究不同预测方法的 RMSE 值和 MAE 值，如表 1 所示. 由表 1 可知：基于 KPCA-LSSVM 的交通流预测精度最高，优势明显.

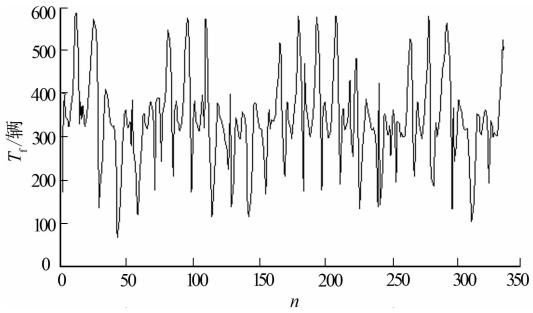


图1 交通流迹序列

Fig. 1 Traffic flow trace sequence

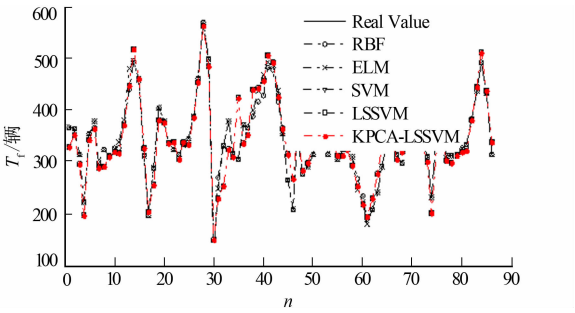


图2 不同方法交通流预测结果

Fig. 2 Traffic flow prediction results in different methods

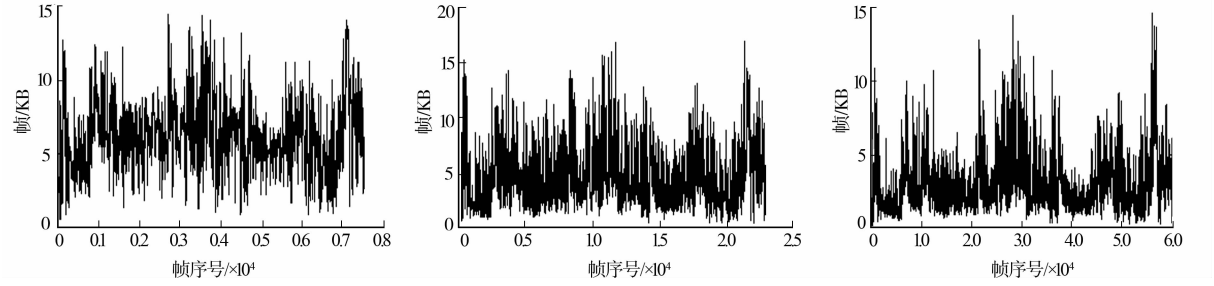
表1 不同方法的预测精度

Tab. 1 Prediction accuracy of different methods

评价指标	预测精度				
	RBF	ELM	SVM	LSSVM	KPCA-LSSVM
MAE	0.007 4	0.006 2	0.004 1	0.002 9	0.001 7
RMSE	0.093 7	0.082 1	0.024 2	0.015 3	0.009 4

2.2 视频流量预测

实验视频迹来源于 http://www-tkn.ee.tu-berlin.de/research/trace/trace.html. 文中选取数据库中的 Die Hard Ⅲ序列进行预测,该视频采用 MPEG-4 压缩标准,在该标准下,Die Hard Ⅲ共包含 I-VOPs,P-VOPs,B-VOPs 3 个迹序列,如图 3 所示. 图 3 中:I-VOPs 序列中共有 7 500 个数据点;P-VOPs 序列中共有 22 500 个数据点;B-VOPs 序列中共有 59 998 个数据点.

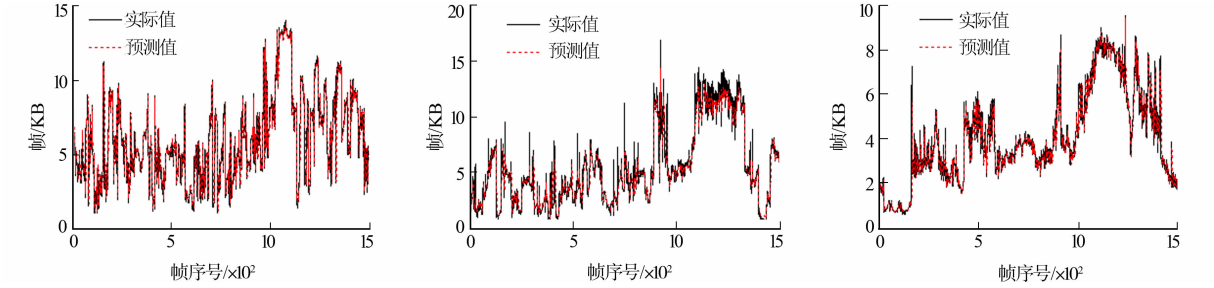


(a) I-VOPs (b) P-VOPs (c) B-VOPs

图3 不同视频迹序列

Fig. 3 Different video trace sequences

实验过程中,I-VOPs 序列取其前 1 000 组作为训练数据,余下 1 500 作为测试数据;P-VOPs 序列选取 2 000 组数据作为训练数据,取 1 500 组作为测试数据;B-VOPs 序列取 2 200 组作为训练数据,取 1 500 组作为测试数据. 实验参数的选取:嵌入维数 $k=8$;核函数半径为 1;正则化参数 $C^2=4.2$. I-VOPs,P-VOPs,B-VOPs 迹序列的预测效果图,如图 4 所示.



(a) I-VOPs (b) P-VOPs (c) B-VOPs

图4 不同视频迹序列预测结果

Fig. 4 Prediction results of different video trace sequences

为衡量 KPCA-LSSVM 的预测效果,在同等条件下 3 种序列不同预测方法的 RMSE 值和 MAE

值，如表 2 所示. 由表 2 可知：在 3 种序列上，基于 KPCA-LSSVM 的预测精度均具有明显优势.

表 2　不同方法不同视频迹序列预测精度

Tab. 2　Prediction accuracy of different video trace sequences with different methods

迹序列	评价指标	预测精度				
		RBF	ELM	SVM	LSSVM	KPCA-LSSVM
I-VOPs	MAE	0.037	0.332	0.174	0.092	0.049
	RMSE	0.592	0.386	0.142	0.094	0.061
P-VOPs	MAE	0.068	0.045	0.038	0.027	0.016
	RMSE	0.741	0.682	0.363	0.041	0.033
B-VOPs	MAE	0.084	0.073	0.069	0.048	0.026
	RMSE	0.627	0.582	0.541	0.339	0.171

3　结束语

结合 KPCA 及 LSSVM 各自优点，提出一种 KPCA-LSSVM 组合预测的新颖方法. 该方法利用 KPCA 在特征空间上提取非线性主元. 通过 LSSVM 建立时间序列模型，验证文中方法的有效性. 将其应用于交通流及视频流时间迹序列的预测中，表明 KPCA-LSSVM 组合预测具有较好的预测精度.

参考文献：

[1] 李海林，杨丽彬. 时间序列数据降维和特征表示方法[J]. 控制与决策，2013，28(11)：1718-1722.

[2] SUN Baiqing，GUO Haifeng，KARIMI H R，*et al*. Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series[J]. Neurocomputing，2015，151(3)：1528-1536. DOI：10.1016/j.neucom.2014.09.018.

[3] BHATTACHARYA D，KONAR A，DAS P. Secondary factor induced stock index time-series prediction using self-adaptive interval type-2 fuzzy sets[J]. Neurocomputing，2015，171(C)：551-568. DOI：10.1016/j.neucom.2015.06.073.

[4] 李哲敏，许世卫，崔利国，等. 基于动态混沌神经网络的预测研究：以马铃薯时间序列价格为例[J]. 系统工程理论与实践，2015，35(8)：2083-2091. DOI：10.12011/1000-6788(2015)8-2083.

[5] 李瑞国，张宏立，范文慧，等. 基于改进教学优化算法的 Hermite 正交基神经网络混沌时间序列预测[J]. 物理学报，2015，64(20)：104-116. DOI：10.7498/aps.64.200506.

[6] 李松，刘力军，刘颖鹏. 改进 PSO 优化 BP 神经网络的混沌时间序列预测[J]. 计算机工程与应用，2013，49(6)：245-248. DOI：10.3778/j.issn.1002-8331.1108-0081.

[7] 李松，罗勇，张铭锐. 遗传算法优化 BP 神经网络的混沌时间序列预测[J]. 计算机工程与应用，2011，47(29)：52-55.

[8] 赵志宏，杨绍普. 基于 SVM 的混沌时间序列分析[J]. 动力学与控制学报，2009，7(1)：5-8.

[9] 郑永康，陈维荣，戴朝华，等. Gaussian 小波 SVM 及其混沌时间序列预测[J]. 控制工程，2009，16(4)：468-471.

[10] 刘婷婷，史久根，韩江洪. 基于 SVM 的瓦斯体积分数混沌时间序列预测[J]. 合肥工业大学学报(自然科学版)，2009，32(8)：1150-1153. DOI：10.3969/j.issn.1003-5060.2009.08.008.

[11] MIRANIAN A，ABDOLLAHZADE M. Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series prediction[J]. IEEE Transactions on Neural Networks and Learning Systems，2013，24(2)：207-218. DOI：10.1109/TNNLS.2012.2227148.

[12] 韩敏，许美玲，任伟杰. 多元混沌时间序列的相关状态机预测模型研究[J]. 自动化学报，2014，40(5)：822-829.

[13] 韩敏，许美玲. 一种基于误差补偿的多元混沌时间序列混合预测模型[J]. 物理学报，2013，62(12)：106-112.

[14] 赵永平，张丽艳，李德才，等. 过滤窗最小二乘支持向量机的混沌时间序列预测[J]. 物理学报，2013，62(12)：113-121. DOI：10.7498/aps.62.120511.

[15] 张森，肖先赐. 混沌时间序列全局预测新方法：连分式法[J]. 物理学报，2005，54(11)：5062-5068.

[16] 莫小琴，李钟慎. 混沌时间序列的 LSSVM 预测方法[J]. 华侨大学学报(自然科学版)，2014，35(4)：373-377. DOI：10.11830/ISSN.1000-5013.2014.04.0373.

(责任编辑：钱筠　　英文审校：吴逢铁)