

# 用于混沌时间序列预测的组合核函数 最小二乘支持向量机\*

田中大<sup>1)†</sup> 高宪文<sup>2)</sup> 石彤<sup>3)</sup>

1) (沈阳工业大学信息科学与工程学院, 沈阳 110870)

2) (东北大学信息科学与工程学院, 沈阳 110819)

3) (辽宁林业职业技术学院人文社会科学系, 沈阳 110101)

(2013年12月24日收到; 2014年4月3日收到修改稿)

针对混沌时间序列的预测问题, 考虑到单一核函数的最小二乘支持向量机无法明显提高预测精度, 提出了一种组合核函数的最小二乘支持向量机预测模型, 模型中采用多项式函数与径向基函数组合构建核函数. 同时, 还对遗传算法进行了改进, 使之具有更快的收敛速度和更高的精度, 改进的遗传算法适用于解决预测模型中的参数优化问题. 通过典型的 Lorenz 时间序列、Mackey-Glass 时间序列、太阳黑子数时间序列以及具有混沌特性的网络流量时间序列对该模型进行了验证. 仿真结果表明所提出的模型是有效的.

**关键词:** 混沌时间序列, 最小二乘支持向量机, 组合核函数, 改进遗传算法

**PACS:** 05.45.Tp, 05.45.-a, 02.50.Ey

**DOI:** 10.7498/aps.63.160508

## 1 引言

时间序列预测是一个重要的研究课题, 在天气预测<sup>[1]</sup>、电力负荷预测<sup>[2]</sup>、金融股票预测<sup>[3]</sup>、太阳黑子数预测<sup>[4]</sup>等方面有着广泛的应用. 自然界和人类活动中很多的时间序列都是非线性甚至是混沌的, 因此如何精确地进行时间序列特别是非线性时间序列的预测是非常重要的. 目前已经有很多的研究方法应用到时间序列的预测之中, 整体而言可分为单一预测模型和混合预测模型两类.

单一预测模型包括基于线性理论的自回归<sup>[5]</sup>、自回归滑动平均<sup>[6]</sup>、自回归求和滑动平均 (ARIMA)<sup>[7]</sup>等方法, 这些成熟的方法在线性时间序列的预测中得到了大量的应用, 取得了很好的预测效果, 但是对于非线性时间序列的预测存在着局限性. 而非线性预测模型主要包括神经网络 (Elman 神经网络<sup>[8]</sup>、回声状态网络<sup>[9]</sup>、递归神经网络<sup>[10]</sup>、径向基函数 (RBF) 神经网络<sup>[11]</sup>等)、非线性

自适应滤波器<sup>[12]</sup>、灰色模型<sup>[13]</sup>以及支持向量机 (SVM)<sup>[14-16]</sup>等. 虽然已经证明神经网络可以任意精度地逼近任何函数, 但是神经网络由于训练样本数量的局限, 存在着过拟合与易陷入局部最优等缺点, 同时其结构需要人工指定, 这些都限制了神经网络的适用性. 非线性自适应滤波器的非线性耦合项和待定系数使其实现非常困难. 灰色模型只适合原始数据符合指数分布规律且变化不是剧烈的情况. 而 Vapnik<sup>[17]</sup>提出的 SVM 在非线性、小样本以及高维模式识别等问题中有独特的优势, 但是 SVM 的主要缺点在于其模型参数目前还没有统一的确定方法.

混合预测模型就是利用几种预测方法对时间序列进行分析与预测. 单一预测模型都存在各自的局限性, 为了能更全面、更有效地反映时间序列的真实特性, 一些混合预测模型被相继提出, 如模糊神经网络模型<sup>[18]</sup>、经验模式分解<sup>[19]</sup>、误差补偿多元时间序列预测模型<sup>[20]</sup>等. 文献<sup>[21, 22]</sup>的仿真结果

\* 国家自然科学基金重点项目 (批准号: 61034005) 资助的课题.

† 通讯作者. E-mail: tianzhongda@126.com

都表明,混合预测模型能够提高预测度.虽然混合预测模型可提高预测精度,但是预测精度的提高主要在于针对时间序列特性采用合适的预测模型,最终的预测精度还是取决于单个模型.因此,对于时间序列的预测,提高单一预测模型的预测效果是最重要的.

1999年, Suykens 和 Vandewalle<sup>[23]</sup>在SVM基础上提出了最小二乘支持向量机(LSSVM),将SVM中二次规划问题转化为线性方程组求解,降低了求解的复杂度<sup>[24]</sup>,因此非常适合用于时间序列的预测.但是,LSSVM的预测精度受到预测模型采用的核函数以及核函数参数的影响,目前还没有明确的理论来支持如何进行核函数以及核函数参数的选择.为了提高LSSVM的预测精度,针对LSSVM算法中影响预测精度的核函数以及核函数参数的选择问题,本文通过采用全局性核函数与局部性核函数进行组合构建核函数的方法,利用全局性核函数拟合出样本中距离较远的相关性、利用局部性核函数逼近拟合邻近数据域的相关性,从而提高模型的泛化性,同时利用一种改进的遗传算法(IGA)进行参数的优化选择,提高预测的精度.通过3个典型的时间序列以及实际采集的网络流量时间序列进行仿真,验证本文提出方法的有效性.

## 2 LSSVM

LSSVM用如下形式的函数对非线性时间序列进行预测:

$$y = \mathbf{w}^T \Phi(x) + b. \quad (1)$$

非线性函数 $\Phi(\cdot)$ 可将输入空间映射到高维的特征空间,使输入空间的非线性拟合问题变成高维特征空间的线性拟合问题.给定训练集 $\{x_i, y_i\}, i = 1, 2, \dots, N$ ,根据风险最小化原理,回归问题可以表示为约束优化问题,即

$$\min J(\mathbf{w}, \mathbf{e}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (\gamma > 0), \quad (2)$$

$$y_i = \mathbf{w}^T \Phi(x_i) + b + e_i, \quad (3)$$

式中, $x_i$ 为输入集合, $y_i$ 为输出预测集合, $\gamma$ 为正则化参数, $b$ 为常值偏差.可通过建立拉格朗日函数求解上述优化问题,

$$\begin{aligned} L(\mathbf{w}, b, \mathbf{e}, \mathbf{a}) \\ = J(\mathbf{w}, \mathbf{e}) - \sum_{i=1}^N a_i (\mathbf{w}^T \Phi(x_i) + b + e_i - y_i), \end{aligned} \quad (4)$$

式中 $a_i$ 为拉格朗日乘子.对 $\mathbf{w}, b, \mathbf{e}, \mathbf{a}$ 分别求偏微分并化简后可得

$$\begin{bmatrix} 0 & \mathbf{l}^T \\ \mathbf{l} & \Omega + \gamma^{-1} \mathbf{l} \end{bmatrix} \begin{bmatrix} b \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}. \quad (5)$$

这里, $\mathbf{y} = [y_1, \dots, y_N]$ , $\mathbf{l} = [1, \dots, 1]$ , $\mathbf{a} = [a_1, \dots, a_N]^T$ , $\Omega_{ij} = \Phi^T(x_i) \Phi(x_j) = \mathbf{K}(x_i, x_j)$ ,其中 $i = 1, 2, \dots, N$ , $\mathbf{K}(x_i, x_j)$ 为核函数.对于已知的训练集 $\{x_i, y_i\}$ ,利用(5)式即可求出参数 $\mathbf{a}$ 与 $b$ .同时,当给定实际集合 $x$ 即可结合训练集中的 $x_i$ 用如下公式来求出系统实际输出 $y$ :

$$y = \sum_{i=1}^N a_i \mathbf{K}(x, x_i) + b. \quad (6)$$

设样本时间序列为 $x_1, x_2, \dots, x_N$ ,将序列转化为下列的矩阵形式:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & \dots & x_m \\ x_2 & x_3 & \dots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-m} & x_{N-m+1} & \dots & x_{N-1} \end{bmatrix}, \quad (7)$$

式中 $x_k$ 为 $k$ 时刻的时间序列值, $N$ 为输入时间序列的总长度, $m$ 为嵌入维数.输出预测时间序列可表示为 $(N - m) \times s$ 矩阵 $\mathbf{Y}$ ,即

$$\mathbf{Y} = \begin{bmatrix} x_{m+1} & x_{m+2} & \dots & x_{m+s} \\ x_{m+2} & x_{m+3} & \dots & x_{m+s+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_{N+1} & \dots & x_{m+s+N-1} \end{bmatrix}, \quad (8)$$

式中 $s$ 为预测步数.本文仿真中采用的是一步预测,即 $s = 1$ .

利用上述介绍的 $\mathbf{X}$ 作为输入, $\mathbf{Y}$ 作为输出即可完成LSSVM的建模.

LSSVM预测模型的非线性映射能力是由核函数决定的<sup>[25]</sup>.核函数可分为局部性核函数和全局性核函数两类.常见的核函数有线性核函数、多项式核函数、RBF和Sigmoid核函数.线性核函数为

$$\mathbf{K}(x, x_i) = (x, x_i'); \quad (9)$$

多项式核函数为

$$\mathbf{K}(x, x_i) = [(x, x_i) + 1]^q; \quad (10)$$

RBF为

$$\mathbf{K}(x, x_i) = \exp(-\|x - x_i\|^2 / \sigma^2); \quad (11)$$

Sigmoid 核函数为

$$K(x, x_i) = \tanh(v(x, x_i) + e). \quad (12)$$

核函数实现 LSSVM 预测模型样本从输入空间到特性空间映射. 因此, 由于不同的映射形式, 全局性核函数和局部性核函数在学习能力和泛化能力上差异较大. 对于全局性核函数, 距测试点  $x_i$  越远的点对核函数值产生的作用越大; 对于局部性核函数, 距测试点  $x_i$  越近的点对核函数值产生的作用越大<sup>[26]</sup>. 在众多实际应用中 RBF 都表现出了优异的性质, 但是 RBF 是典型的局部性核函数, 因此以 RBF 为基础并选择其他核函数进行线性组合是提高模型预测精度的一个途径. 本文采用典型的多项式函数 ( $K_{\text{poly}}$ ) 和 RBF ( $K_{\text{RBF}}$ ) 来构建预测模型中的组合核函数  $K_{\text{com}}$ , 即

$$K_{\text{com}} = (1 - \rho)K_{\text{poly}} + \rho K_{\text{RBF}}. \quad (13)$$

由于数据集是有限的, 因此在时间序列的预测过程中序列的未来值只能用有限的历史序列值来预测, 亦即以上描述的嵌入维数  $m$ , 因此组合核函数 LSSVM 预测模型的预测精度受到历史序列数  $m$ 、组合核函数比例  $\rho$ 、正则化参数  $\gamma$ 、核函数参数  $q$  和 RBF 宽度  $\sigma$  这 5 个参数的影响. 本文将利用 IGA 进行参数的寻优, 以提高模型的预测精度.

### 3 IGA

为了进行组合核函数 LSSVM 预测模型中的参数寻优, 本文在递阶遗传算法<sup>[27,28]</sup>基础上对其交

叉和变异方式进行改进, 采用改进的自适应交叉、变异方式进行寻优. 这种改进的自适应递阶遗传算法具有更快的收敛速度与更高的精度. 下面介绍算法的具体设计.

#### 3.1 交叉操作改进

因为递阶染色体中既包括二进制码基因串, 又包括十进制基因串, 故要分别进行交叉操作. 首先按照自适应交叉概率  $p_c$  选取参与交叉的每对个体, 然后再进行交叉操作. 对控制基因, 采用单点交叉; 子群  $P_1(t)$  对每对个体随机确定一个交叉点, 并交换这两个个体的交叉点一侧的基因位; 对参数基因, 采用算术交叉, 传统的交叉方式按下式进行交叉操作:

$$a' = (1 - u)a + ub, \quad b' = (1 - u)b + ua, \quad (14)$$

式中  $u$  为比例因子, 通常取 (0, 1) 之间的一个常数. 由于 (14) 式中  $u$  是任意选取的, 因此交叉方式存在一定的盲目性, 并且容易陷入局部最优, 不能在全局范围内寻优, 这会导致算法的精度降低. 本文改进后的交叉方式, 按下列公式进行操作:

$$a'_1 = (1 - P_c)a + P_c b \quad (f_a < f_{\text{avg}}), \quad (15)$$

$$a'_2 = (1 - P_c)b + P_c a \quad (f_a \geq f_{\text{avg}});$$

$$b'_1 = (1 - P'_c)b + P'_c a \quad (f_b < f_{\text{avg}}), \quad (16)$$

$$b'_2 = P'_c b + (1 - P'_c)a \quad (f_b \geq f_{\text{avg}}).$$

这里,

$$P_c = \begin{cases} p_{c1} - \frac{(p_{c1} - p_{c2})(f_a - f_{\text{avg}})}{f_{\text{max}} - f_{\text{avg}}} & (f_a \geq f_{\text{avg}}), \\ p_{c1} & (f_a < f_{\text{avg}}), \end{cases} \quad (17)$$

$$P'_c = \begin{cases} p_{c1} - \frac{(p_{c1} - p_{c2})(f_b - f_{\text{avg}})}{f_{\text{max}} - f_{\text{avg}}} & (f_b \geq f_{\text{avg}}), \\ p_{c1} & (f_b < f_{\text{avg}}), \end{cases} \quad (18)$$

式中,  $p_{c1} = 0.9$ ,  $p_{c2} = 0.6$ . 由 (18) 式可知, 交叉操作的比例因子是随着个体各自的自适应交叉概率变化的, 因此改进后的交叉操作加快了算法的收敛速度, 同时保证了算法的全局寻优能力.

#### 3.2 变异操作改进

变异操作对参数基因和控制基因分别进行变异. 首先按自适应变异概率  $p_m$  选取参加变异的个体, 然后再对选取的个体进行变异操作. 自适应变异概率  $p_m$  可表示为

$$p_m = \begin{cases} p_{m1} - \frac{(p_{m1} - p_{m2})(f - f_{\text{avg}})}{f_{\text{max}} - f_{\text{avg}}} & (f \geq f_{\text{avg}}), \\ p_{m1} & (f < f_{\text{avg}}), \end{cases} \quad (19)$$

式中,  $p_{m1} = 0.1, p_{m2} = 0.001$ .

对控制基因, 采用二进制变异方式对变异位进行取反; 对参数基因, 传统的变异算子是对群体中的个体实施随机变异, 与历史状态和当前状态无关, 也不考虑与最优解的距离. 常用的参数基因的均匀变异方式为

$$x'_k = x_k + u(x_k^1 - x_k^2), \quad (20)$$

式中,  $x_k^1$  和  $x_k^2$  分别是  $x_k$  的上下界;  $u$  为比例因子, 通常  $u$  的取值范围为  $(0, 1)$ . 从 (20) 式可以看出变异过程是随机的且方向也是随机的. 本文采用改进变异算子对个体进行变异操作. 使个体根据种群的迄今最优解和个体自身的迄今最优解变异的方向和幅度进行变异, 即

$$\Delta x_{k,t+1} = r_1(x_{\max,k} - x_{k,t}) + r_2(x_{\max} - x_{k,t}), \quad (21)$$

$$x_{k,t+1} = x_{k,t} + \Delta x_{k,t+1}, \quad (22)$$

式中,  $\Delta x_{k,t+1}$  为变异个体  $x_k$  进化到  $t + 1$  代时的变化幅度,  $x_{k,t}$  为个体  $x_k$  在  $t$  代的值,  $x_{\max,k}$  为个体  $x_k$  在进化过程中所出现的最优值,  $x_{\max}$  为种群的历史最优解,  $x_{k,t+1}$  为变异后个体  $x_k$  的值,  $r_1$  和  $r_2$  为  $[0, 1]$  内的随机数. 变异的个体是向着群体的最优值方向进化的, 这就加快了算法的收敛速度与收敛精度.

### 3.3 验证

为了验证算法改进后的收敛速度和精度, 利用下列 Rosenbrock 函数进行了测试:

$$f(x) = \sum_{i=1}^n 100(x_{i+1}^2 - x_i)^2 + (1 - x_i)^2, \quad (23)$$

式中,  $x$  的取值范围为  $[-2.048, 2.048]$ ,  $N = 10$ .

仿真参数如下: 种群数量为 50, 进化代数为 50, 标准遗传算法 (GA) 交叉概率为 0.8, 变异概率为 0.2, 采用二进制编码, 长度为 10, 其他参数同上. IGA 控制基因采用二进制编码, 下级子种群为对应的参数基因, 采用实数编码, 其他参数如上. 为了消除随机性的影响, 所有算法都运行 10 次, 取平均值作为寻优结果, 图 1 所示为其中一次的适应度进化曲线, 表 1 列出了仿真性能指标. 从图 1 和表 1 可以看出, 本文的 IGA 具有更快的收敛速度和更高的收敛精度.

## 4 组合核函数 LSSVM 预测模型

本文的 IGA 优化的组合核函数 LSSVM 时间序列预测模型建模过程如图 2 所示.

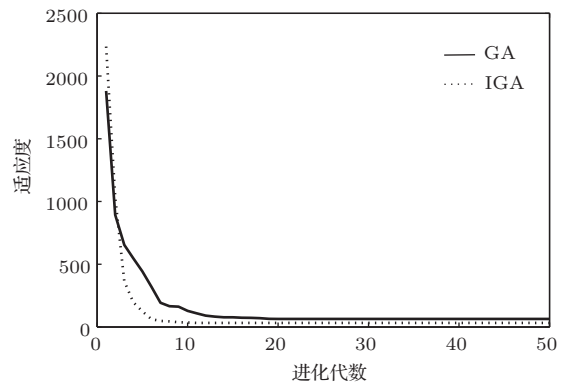


图 1 适应度变化曲线

表 1 仿真结果

	平均适应度	最佳适应度	标准差
GA	30.85	27.74	12.12
IGA	59.27	47.26	27.24

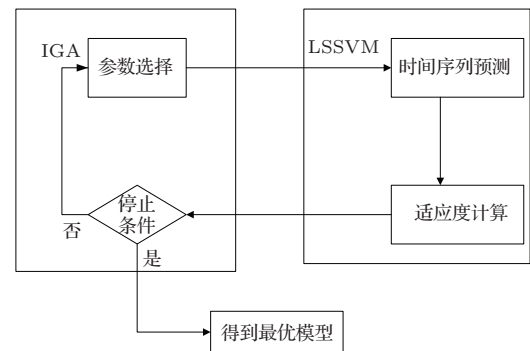


图 2 组合核函数 LSSVM 时间序列预测模型

下面给出建模与寻优步骤.

**步骤 1** IGA 算法参数初始化, 包括种群数、最大迭代次数等. 随机生成初始种群, 将其分为两个子种群, 编码成二级的递阶结构. 上级子种群为控制基因, 采用二进制编码, 下级子种群为对应的参数基因, 采用实数编码. 对每个染色体进行初始化.

**步骤 2** 设定待优化参数  $m, \rho, \gamma, q$  与  $\sigma^2$  取值范围, 这里限定待优化参数的范围如下:  $m \in [1 \ 30], \rho \in [0 \ 1], \gamma \in [0.01 \ 1000], q \in (0 \ 1), \sigma^2 \in [0.01 \ 1000]$ .

**步骤 3** 对时间序列采集样本数据并归一化, 按照 (7), (8) 式构造输入、输出数据, 按照 LSSVM



预测算法进行模型的训练与建模.

**步骤4** 在第  $k$  次迭代中, 对训练样本每组参数进行训练, 计算每组参数的适应度. 确定时间序列实际值与预测值的均方根误差 (RMSE) 为适应度函数,

$$e_{\text{rms}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}.$$

**步骤5** 按照 IGA 算法对个体进行选择、交叉以及变异操作, 产生新的个体.

**步骤6** 保留最佳个体, 结束条件判断. 未结束返回步骤4, 结束则解码末代种群中最佳个体, 输出优化参数.

**步骤7** 用最优参数进行时间序列的预测.

### 5 仿真与实验分析

为了说明本文组合核函数 LSSVM 预测模型的预测精度, 除了 RMSE 外本文同时采用平均绝对误差 (MAE) 和平均绝对百分误差 (MAPE) 两个指标衡量预测效果. MAE 的表达式为

$$e_{\text{ma}} = \frac{1}{N} \sum_{i=1}^N |y_{\text{pre}}(i) - y_{\text{rea}}(i)|. \quad (24)$$

MAPE 的表达式为

$$e_{\text{map}} = \frac{1}{N} \sum_{i=1}^N \{|y_{\text{pre}}(i) - y_{\text{rea}}(i)| / y_{\text{rea}}(i)\} \times 100\%. \quad (25)$$

为了验证本文方法的有效性, 本文对模拟混沌时间序列与实际混沌时间序列分别进行了仿真验证.

#### 5.1 Lorenz 时间序列预测

Lorenz 系统作为一个最经典的三维二次多项式自治系统的混沌模型, 对 Lorenz 系统的研究纵贯整个混沌时间序列预测的发展, 因而研究 Lorenz 系统具有基本而重要的意义. 下面以其产生的时间序列为例来验证模型的精确性. Lorenz 时间序列的表达式为

$$\begin{aligned} \frac{dx}{dt} &= -\alpha(x - y), \\ \frac{dy}{dt} &= -xz + cx - y, \\ \frac{dz}{dt} &= xy - bz. \end{aligned} \quad (26)$$

当选取参数  $\alpha = 10, b = 8/3, c = 28, x(0) = -1, y(0) = 0, z(0) = 1$  时, (26) 式具有混沌特性. 采用四阶 Runge-Kutta 算法生成包含 1000 个数据的混沌时间序列, 采样时间间隔  $\Delta t = 0.05$ , 图 3 为 Lorenz 混沌序列的三维图, 图 4 为产生的 1000 组 Lorenz 序列中的  $x(t), y(t), z(t)$ . 由于 LSSVM 建模不需要大量的样本, 因此选取 701—900 的 200 组数据作为训练样本, 用于建立预测模型, 选取 901—1000 的 100 组数据作为测试样本, 用于检验模型的有效性.

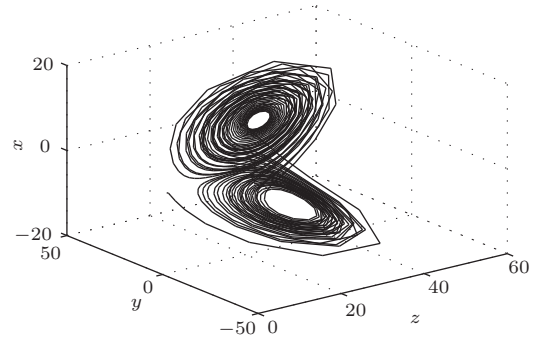


图3 Lorenz 混沌时间序列三维图

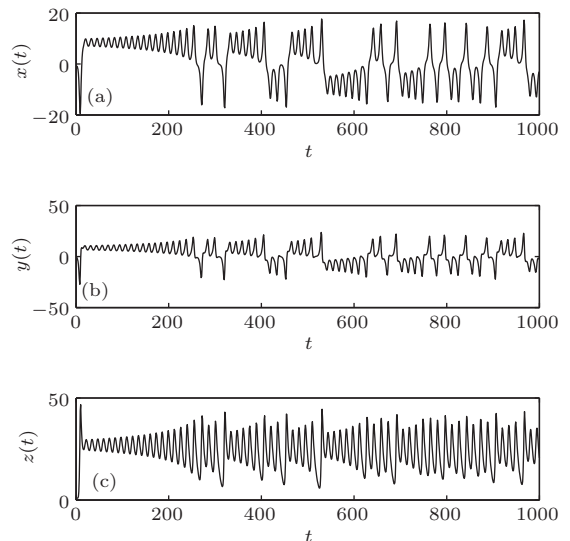


图4 1000 组 Lorenz 序列中的  $x(t), y(t), z(t)$  (a)  $x(t)$ ; (b)  $y(t)$ ; (c)  $z(t)$

按照本文所设定的待优化参数范围, IGA 种群数量为 30, 进化代数数为 100,  $p_{c1} = 0.9, p_{c2} = 0.6, p_{m1} = 0.1, p_{m2} = 0.001$ . 利用 200 组样本数据进行组合核函数 LSSVM 参数的优化, 图 5 给出了适应度变化曲线, 寻优之后的参数如表 2 所列.

在得到最优参数之后, 对 100 组 Lorenz 测试时间序列的  $x(t), y(t), z(t)$  数据进行了预测效果验证, 图 6 所示为预测值与真实值的对比曲线, 图 7 所示为预测误差变化曲线.

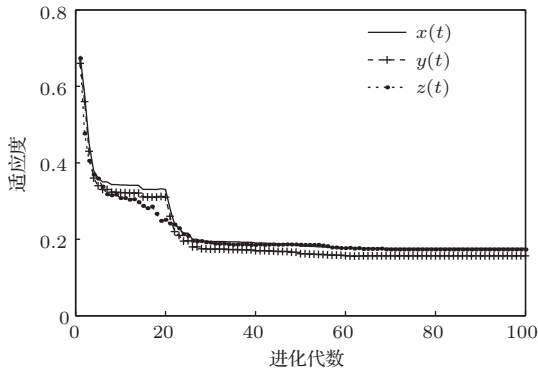


图5 Lorenz 时间序列 IGA 适应度曲线

表2 Lorenz 时间序列寻优后的参数

参数	$x(t)$	$y(t)$	$z(t)$
$m$	27	11	26
$\rho$	0.6506	0.9091	0.9299
$\gamma$	99.3	862.17	609.97
$q$	0.4370	0.5816	0.1603
$\sigma^2$	13.6882	43.9892	6.8436

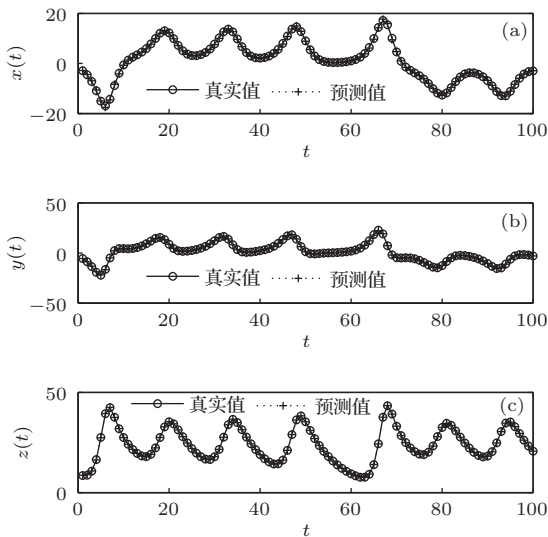


图6 Lorenz 时间序列预测值与真实值对比 (a)  $x(t)$ ; (b)  $y(t)$ ; (c)  $z(t)$

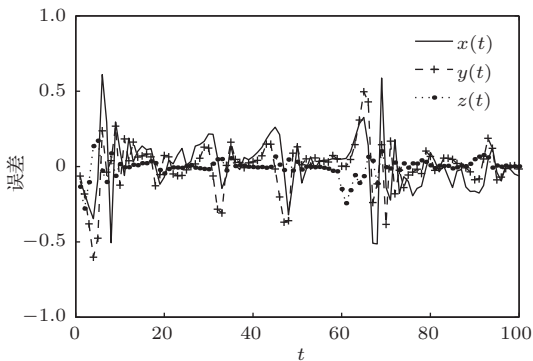


图7 Lorenz 时间序列预测误差曲线

为了进行预测效果的对比, 将本文模型与RBF的LSSVM, ARIMA 预测模型和Elman神经网络进行了比较. RBF-LSSVM利用工具箱中的交叉验证进行参数优化, 得到的RBF-LSSVM参数如表3所列, ARIMA 预测模型通过赤池信息量准则 (Akaike information criterion, AIC) 得到参数如表4所列, Elman神经网络输入层为15, 中间层为20, 输出层为1, 最大迭代次数为3000. 表5、表6和表7分别给出了 $x(t)$ 序列、 $y(t)$ 序列和 $z(t)$ 序列4种预测方法的性能指标对比.

表3 RBF-LSSVM 预测模型参数

参数	$x(t)$	$y(t)$	$z(t)$
$\gamma$	33.9	12.18	24.62
$\sigma^2$	24.11	9.71	10.68

表4 ARIMA 预测模型参数

参数	$x(t)$	$y(t)$	$z(t)$
$p$	3	4	6
$d$	2	1	1
$q$	4	5	4

表5  $x(t)$  序列不同预测方法的性能指标对比

预测模型	$e_{rms}$	$e_{ma}$	$e_{map}$
本文模型	0.1826	0.1338	3.1
RBF-LSSVM	4.6643	1.653	9.36
ARIMA	15.233	5.91	18.32
Elman 神经网络	13.361	4.24	14.56

从以上结果可以看出, 本文的预测模型对于Lorenz时间序列中的 $x(t)$ ,  $y(t)$ ,  $z(t)$ 序列都有很好的预测效果, 性能指标优于其他常见的单一预测模型.

## 5.2 Mackey-Glass 时间序列预测

1977年, Mackey 和 Glass 发现了滞系统中的混沌现象, 时滞混沌系统便引起了人们的广泛关注, 并常常将其作为检验非线性系统模型性能的标准. Mackey-Glass 混沌时间序列由下列的时滞微分方程产生:

$$\frac{dx(t)}{dt} = \frac{0.2x(t - \Delta)}{1 + x^{10}(t - \Delta)} - 0.1x(t), \quad (27)$$

式中 $\Delta$ 为时滞参数. 当 $\Delta > 17$ 时 (27)式呈现混沌特征, 其分形维数近似于2.1,  $\Delta$ 越大系统的混沌特征越明显. 图8给出了当 $\Delta = 30$ 时的1000组 Mackey-Glass 时间序列, 图9为 $\Delta = 30$ 时的相图, 由此可看出该时间序列具有非常复杂的非线性混沌特征.

表6  $y(t)$  序列不同预测方法的性能指标对比

预测模型	$e_{rms}$	$e_{ma}$	$e_{map}$
本文模型	0.163	0.109	2.76
RBF-LSSVM	6.325	5.365	10.23
ARIMA	12.364	7.854	15.32
Elman 神经网络	9.632	6.998	11.58

表7  $z(t)$  序列不同预测方法的性能指标对比

预测模型	$e_{rms}$	$e_{ma}$	$e_{map}$
本文模型	0.061	0.036	0.24
RBF-LSSVM	2.365	1.523	2.56
ARIMA	7.514	3.458	6.89
Elman 神经网络	6.784	3.365	4.52

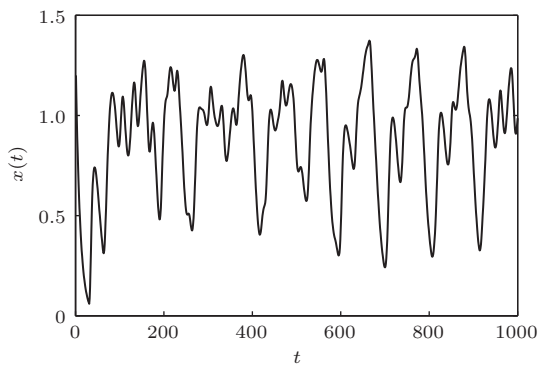


图8 Mackey-Glass 时间序列

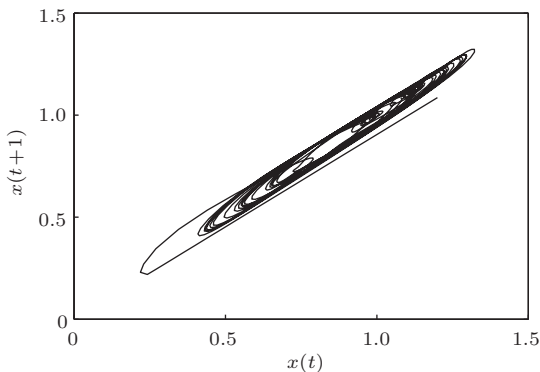


图9 Mackey-Glass 时间序列的相图

同样选取701—900的200组数据作为训练样

本, 用于建立预测模型, 选取901—1000的100组数据作为测试样本, 用于检验模型的有效性. 仿真参数的选取与Lorenz时间序列相同. 图10给出了适应度变化曲线, 寻优之后的预测模型参数如表8所列.

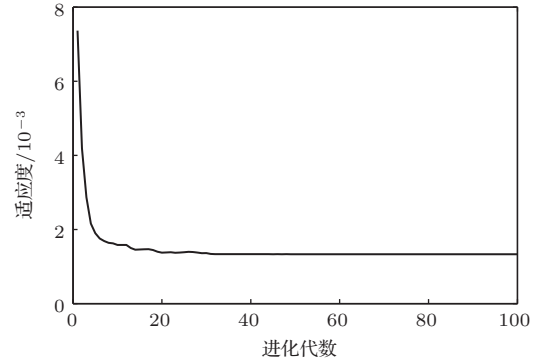


图10 Mackey-Glass 时间序列 IGA 适应度曲线

表8 Mackey-Glass 时间序列寻优后的参数

参数	优化结果
$m$	18
$\rho$	0.8319
$\gamma$	714.56
$q$	0.9306
$\sigma^2$	286.4132

利用优化得到的预测模型, 对100组 Mackey-Glass 测试样本序列进行了预测模型的验证, 图11所示为预测值与真实值的对比曲线, 图12所示为预测误差曲线.

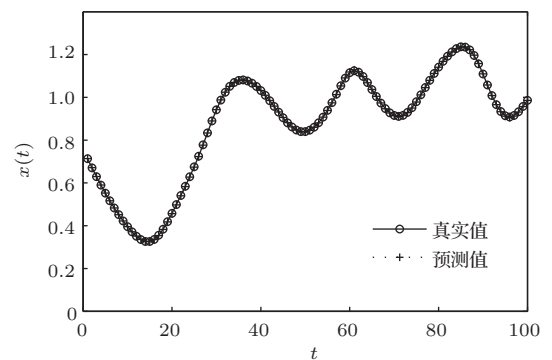


图11 Mackey-Glass 时间序列预测值与真实值的对比

为了进行预测效果的比较, 同样将本文模型与RBF的LSSVM, ARIMA 预测模型和Elman神经网络进行了对比. RBF-LSSVM 参数为 $\gamma = 47.23$ ,  $\sigma^2 = 5.58$ , ARIMA 预测模型通过AIC准则得到 $p = 3, d=1, q=2$ , Elman神经网络输入层为15,

中间层为20, 输出层为1, 最大迭代次数为3000. 表9列出了4种方法的性能指标的对比.

从以上结果可以看出, 本文模型对于Mackey-Glass时间序列的预测效果好于其他模型, 表明本文模型能够高精度地估计Mackey-Glass序列的变化趋势.

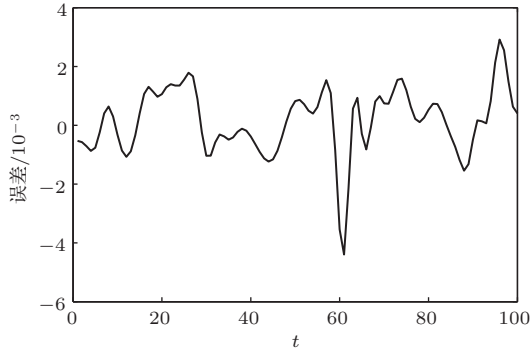


图12 Mackey-Glass时间序列预测误差曲线

表9 Mackey-Glass时间序列不同预测方法性能指标对比

预测模型	$e_{rms}$	$e_{ma}$	$e_{map}$
本文模型	0.113	0.18	0.17
RBF-LSSVM	4.231	5.35	3.23
ARIMA	12.365	8.69	6.21
Elman神经网络	10.878	6.36	5.99

### 5.3 太阳黑子数时间序列预测

太阳黑子数时间序列具有强烈的非线性混沌特性, 因此很多文献都将其用于检验非线性预测模型的预测精度. 本文选取河南省三门峡水文站从1700年到2003年的304组年太阳黑子数和黄河径流量两变量数据作为本文模型的仿真数据, 利用前254组数据建模, 后50组数据进行验证, 304组太阳黑子数据如图13所示. 利用IGA算法进行最优参数的求取, 得到适应度曲线如图14所示, 优化后的参数如表10所列.

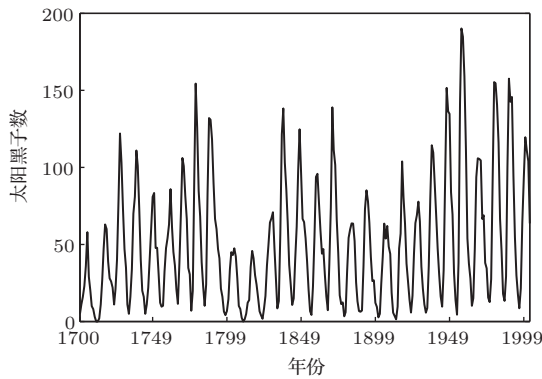


图13 太阳黑子时间序列

利用得到的最优参数对50组太阳黑子测试样本进行了预测验证, 得到图15所示的太阳黑子数预测值与真实值的对比曲线, 图16所示为预测误差曲线.

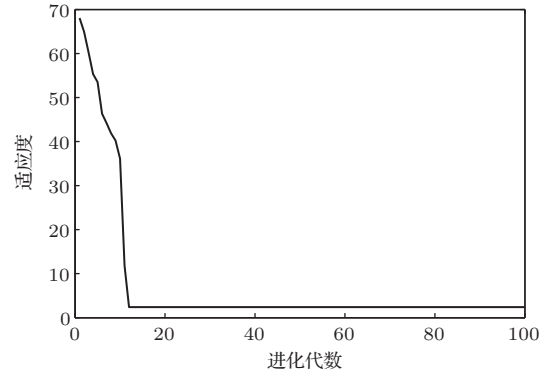


图14 太阳黑子时间序列IGA适应度曲线

表10 太阳黑子序列寻优后的参数

参数	优化结果
$m$	22
$\rho$	0.774
$\gamma$	869.99
$q$	0.433
$\sigma^2$	13.6862

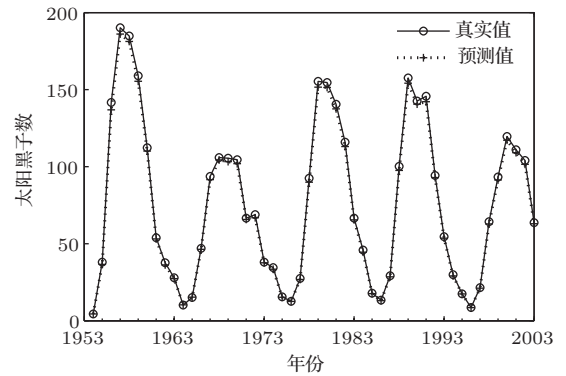


图15 太阳黑子时间序列预测值与真实值对比

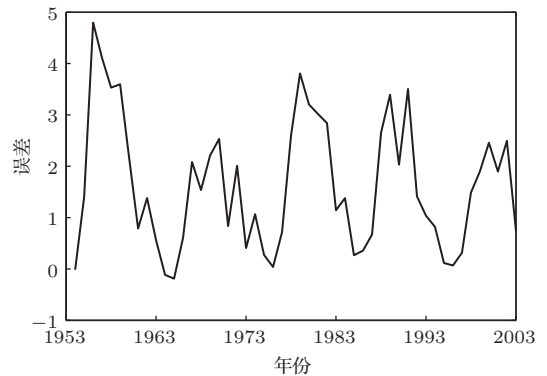


图16 太阳黑子时间序列预测误差曲线



作为对比, RBF-LSSVM 参数为  $\gamma = 38.23$ ,  $\sigma^2 = 2.03$ , ARIMA 预测模型通过 AIC 准则得到  $p = 2, d = 1, q = 2$ , Elman 神经网络输入层为 15, 中间层为 20, 输出层为 1, 最大迭代次数为 3000. 表 11 给出了 4 种方法的性能指标的对比.

表 11 太阳黑子时间序列不同预测方法性能指标对比

预测模型	$\epsilon_{rms}$	$\epsilon_{ma}$	$\epsilon_{map}$
本文模型	2.065	0.8257	2.02
RBF-LSSVM	10.683	4.98	5.56
ARIMA	20.256	14.45	8.54
Elman 神经网络	12.362	13.32	7.20

通过太阳黑子数时间序列的仿真对比可知, 本文的模型减小了预测值与真实值之间的偏差, 可更好地描绘太阳黑子的动力学特性.

### 5.4 网络流量时间序列预测

网络流量的准确预测是网络管理与维护的重要问题, 很多学者研究认为可将网络流量看作具有混沌特性的时间序列 [29,30], 因此通过实际采集的网络流量数据进一步验证本文方法的有效性. 采集了 300 组中国联通网络通信有限公司辽宁省分公司 3G 核心网某一路由器的流入数据, 利用前 200 组进行建模训练, 后 100 组进行验证. 300 组网络流量数据如图 17 所示.

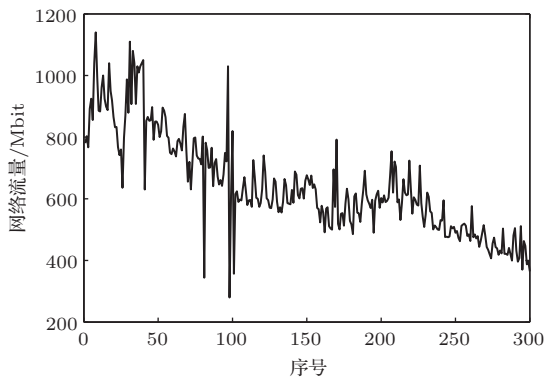


图 17 网络流量时间序列

通过 IGA 算法进行参数优化, 得到的适应度曲线如图 18 所示, 表 12 列出了最终的参数优化结果.

利用最优模型参数对 100 组网络流量测试样本进行了预测仿真, 得到的预测值与真实值对比曲线如图 19 所示, 预测误差变化曲线如图 20 所示.

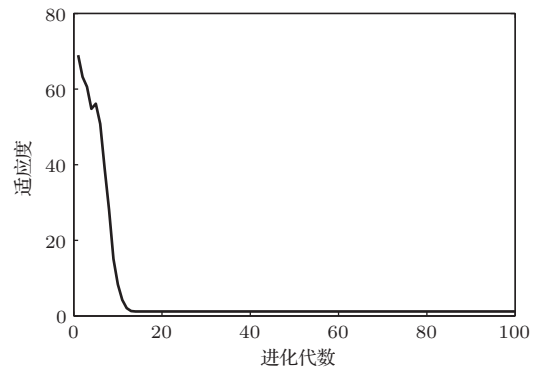


图 18 网络流量时间序列 IGA 适应度曲线

表 12 网络流量时间序列寻优后的参数

参数	优化结果
$m$	23
$\rho$	0.899
$\gamma$	221.89
$q$	0.3177
$\sigma^2$	4.886

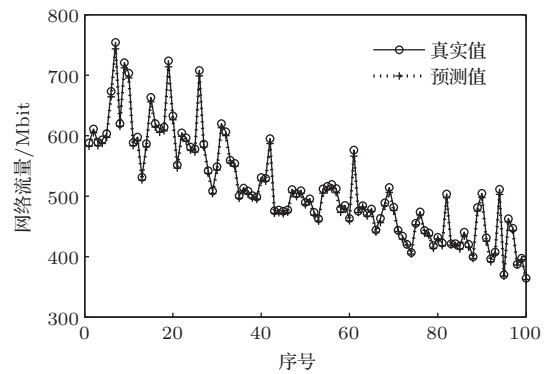


图 19 网络流量时间序列预测值与真实值对比

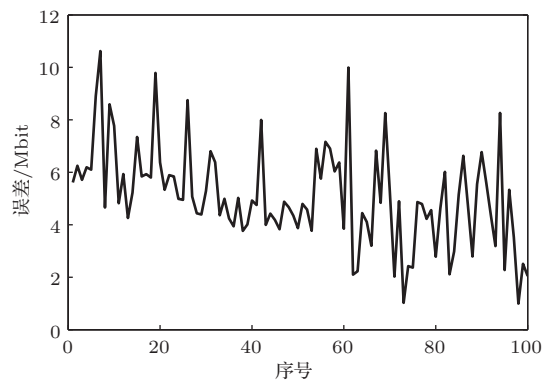


图 20 网络流量时间序列预测误差曲线

与 Lorenz 时间序列仿真类似, 选择其他几个预测模型进行对比. RBF-LSSVM 参数为  $\gamma = 113.57$ ,  $\sigma^2 = 2.36$ , ARIMA 预测模型通过 AIC 准则得到  $p = 2, d = 1, q = 3$ , Elman 神经网络输入层为 15,

中间层为20, 输出层为1. 表13列出了4种方法的性能指标的对比.

表13 网络流量时间序列不同预测方法性能指标对比

预测模型	$e_{rms}$	$e_{ma}$	$e_{map}$
本文模型	5.4204	5.08	1.96
RBF-LSSVM	17.32	11.85	4.56
ARIMA	21.23	23.23	9.31
Elman 神经网络	16.21	18.25	8.57

通过对实验结果的对比分析可知, 相对于其他预测模型, 本文方法对具有混沌特性的网络流量具有更强的非线性拟合能力和更高的预测精度.

## 6 结 论

LSSVM作为一种经典的非线性预测模型在混沌时间序列预测中得到了广泛的应用, 核函数决定了其预测效果. 由于每个核函数都有其各自的优缺点, 它们所构造的模型性能也不同, 同时选择不同的核函数时相应的模型在学习能力和泛化能力上差异较大. 针对该问题, 本文提出一种组合核函数LSSVM预测模型, 为了解决预测模型中最优参数的确定问题而提出一种IGA, 通过4种典型的混沌时间序列的仿真实验表明, 本文提出的预测模型具有更好的预测效果. 未来的研究工作要将本文的预测模型与相空间重构进行结合. 虽然文献[31]通过实验表明相空间重构参数对于SVM的预测精度影响不大, 但是却能减少离线建模的训练量, 因此下一步的工作重点是将时间序列进行相空间重构之后利用本文模型进行时间序列的预测研究.

感谢中国联合网络通信有限公司辽宁省分公司集团客户响应中心的顾斌工程师为本文提供3G核心网路由器的网络流量数据.

## 参考文献

- [1] Wang S W, Feng J, Liu G 2013 *Math. Comput. Model.* **58** 677
- [2] Hong W C 2009 *Energy Convers. Manage.* **50** 105
- [3] Zhao X J, Shang P J, Wang J 2013 *Phys. Rev. E* **87** 805
- [4] Chowdhury P, Dwivedi B N 2011 *Sol. Phys.* **270** 365
- [5] Ding J, Han L L, Chen X M 2010 *Math. Comput. Model.* **51** 527
- [6] Toque C, Terraza V 2011 *Commun. Statist.-Theor. Methods* **40** 1533

- [7] Khashei M, Bijari M 2011 *Appl. Soft Comput.* **11** 2664
- [8] Li P, Li Y, Xiong Q, Chai Y, Zhang Y 2014 *Int. J. Elec. Power* **55** 749
- [9] Li D C, Han M 2011 *Acta Phys. Sin.* **60** 108903 (in Chinese) [李德才, 韩敏 2011 物理学报 **60** 108903]
- [10] Zhang J S, Xiao X C 2000 *Chin. Phys. Lett.* **17** 88
- [11] Zhang J F, Hu S S 2007 *Acta Phys. Sin.* **56** 713 (in Chinese) [张军峰, 胡寿松 2007 物理学报 **56** 713]
- [12] Zhang J S, Xiao X C 2001 *Chin. Phys.* **10** 390
- [13] Li G D, Masuda S, Nagai M 2013 *Int. J. Adv. Manuf. Tech.* **67** 2343
- [14] Li P, Tan Z, Yan L, Deng K 2011 *Min. Sci. Tech.* **21** 557
- [15] Wang G L, Yang P C, Mao Y Q 2008 *Acta Phys. Sin.* **57** 714 (in Chinese) [王革丽, 杨培才, 毛宇清 2008 物理学报 **57** 714]
- [16] Li H C, Zhang J S 2005 *Chin. Phys. Lett.* **22** 2776
- [17] Vapnik V N 1999 *IEEE Trans. Neural Netw.* **10** 1045
- [18] Tan W, Wang Y N, Zhou S W, Liu Z R 2003 *Acta Phys. Sin.* **52** 2205 (in Chinese) [谭文, 王耀南, 周少武, 刘祖润 2003 物理学报 **52** 2205]
- [19] Yang Y F, Ren X M, Qin W Y, Wu Y F, Zhi X Z 2008 *Acta Phys. Sin.* **57** 6139 (in Chinese) [杨永锋, 任兴民, 秦卫阳, 吴亚锋, 支希哲 2008 物理学报 **57** 6139]
- [20] Han M, Xu M L 2013 *Acta Phys. Sin.* **62** 120510 (in Chinese) [韩敏, 许美玲 2013 物理学报 **62** 120510]
- [21] Song T, Li H 2012 *Acta Phys. Sin.* **61** 080506 (in Chinese) [宋彤, 李菡 2012 物理学报 **61** 080506]
- [22] Bodyanskiy Y, Vynokurova O 2013 *Inform. Sci.* **220** 170
- [23] Suykens J A K, Vandewalle J 1999 *Neural Process Lett.* **9** 293
- [24] Pan J S, Hong M Z, Zhou Q F, Cai J Y, Wang H Z, Luo L K, Yang D Q, Dong J, Shi H X, Ren J L 2009 *BMC Biotechnol.* **9** 52
- [25] Han X X, Xie G, Han X M, Xie K M 2012 *Contr. Decis.* **27** 953 (in Chinese) [韩晓霞, 谢刚, 韩晓明, 谢克明 2012 控制与决策 **27** 953]
- [26] Smits G F, Jordaan E M 2002 *Proceedings of the International Joint Conference on Neural Networks* (Piscataway: IEEE) p2785
- [27] Wang S, Yang X J, Wei C J 2006 *Proceedings of the World Congress on Intelligent Control and Automation* (Piscataway: IEEE) p3014
- [28] Toledo C F M, Oliveira L, Franca P M 2014 *J. Comput. Appl. Math.* **261** 341
- [29] Guo T, Lan J L, Li Y F, Jiang Y M 2013 *J. Electron. Inform. Technol.* **35** 2220 (in Chinese) [郭通, 兰巨龙, 李玉峰, 江逸茗 2013 电子与信息学报 **35** 2220]
- [30] Meng Q F, Chen Y H, Feng Z Q, Wang F L, Chen S S 2013 *Acta Phys. Sin.* **62** 150509 (in Chinese) [孟庆芳, 陈月辉, 冯志全, 王枫林, 陈珊珊 2013 物理学报 **62** 150509]
- [31] Cui W Z, Zhu C C, Bao W X, Liu J H 2005 *Acta Phys. Sin.* **54** 3009 (in Chinese) [崔万照, 朱长纯, 保文星, 刘君华 2005 物理学报 **54** 3009]

# Combination kernel function least squares support vector machine for chaotic time series prediction\*

Tian Zhong-Da<sup>1)†</sup> Gao Xian-Wen<sup>2)</sup> Shi Tong<sup>3)</sup>

1) (*College of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China*)

2) (*College of Information Science and Engineering, Northeastern University, Shenyang 110819, China*)

3) (*Department of Humanities, Liaoning Forestry Vocational-technical College, Shenyang 110101, China*)

( Received 24 December 2013; revised manuscript received 3 April 2014 )

## Abstract

Considering the problem that least squares support vector machine prediction model with single kernel function cannot significantly improve the prediction accuracy of chaotic time series, a combination kernel function least squares support vector machine prediction model is proposed. The model uses a polynomial function and radial basis function to construct the kernel function of least squares support vector machine. An improved genetic algorithm with better convergence speed and precision is proposed for parameter optimization of prediction model. The simulation experimental results of Lorenz, Mackey-Glass, Sunspot-Runoff in the Yellow River and chaotic network traffic time series demonstrate the effectiveness and characteristics of the proposed model.

**Keywords:** chaotic time series, least squares support vector machine, combination kernel function, improved genetic algorithm

**PACS:** 05.45.Tp, 05.45.-a, 02.50.Ey

**DOI:** [10.7498/aps.63.160508](https://doi.org/10.7498/aps.63.160508)

---

\* Project supported by the Key Program of the National Natural Science Foundation of China (Grant No. 61034005).

† Corresponding author. E-mail: [tianzhongda@126.com](mailto:tianzhongda@126.com)