

文章编号: 1001-0920(2008)06-0660-05

# 基于一类局部固定核的支持向量回归建模

张军峰, 胡寿松

(南京航空航天大学 自动化学院, 南京 210016)

**摘要:** 在考虑样本向量相似性的基础上, 提出一种核函数——高斯余弦核, 并证明了该核是一类局部固定核. 局部固定核同时描述了样本全局和局部结构, 因而高斯余弦核可将样本数据映射到信息描述更为丰富的特征空间. 将该核用于 Logistic 混沌时间序列和煤气炉数据集的支持向量回归建模与预测, 仿真结果表明所提出的方法能有效地提高预测精度, 而且不增加算法的复杂度.

**关键词:** 核; 局部固定核; 支持向量回归; 相似性

**中图分类号:** TP18      **文献标识码:** A

## Modeling of support vector regression based on locally stationary kernels

ZHANG Jun-feng, HU Shou-song

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

Correspondent: ZHANG Jun-feng, E-mail: wufeng7919 @163.com)

**Abstract:** A kernel, denoted as Gauss-Cos, is presented based on considering the similarity of the sample vectors, which is proved as a class of locally stationary kernel. The global and local structures of the samples are described. So the proposed kernel can map the samples into desirable feature space with more information. Finally, it is applied to the modeling and prediction of Logistic chaotic time series and gas furnace data set based on support vector regression (SVR) method. The simulation results indicate that the proposed method can effectively improve the prediction precision, which is not at the price of increasing the algorithmic complexity.

**Key words:** Kernels; Locally stationary kernels; Support vector regression (SVR); Similarity

### 1 引言

系统建模与预测已在控制领域得到广泛的应用. 对于非线性系统, 系统模型的建立并没有统一的方法, 一般运用神经网络加以实现. 然而, 神经网络存在局部极小点、过学习以及结构选择过分依赖于经验等固有缺陷, 严重降低了应用效果. 与常规神经网络基于经验风险最小化原理不同, 由 Vapnik<sup>[1]</sup> 等人创立的支持向量机 (SVM) 遵循了结构风险最小化原理, 而且集成了最大间隔超平面、Mercer 核、凸二次规划、稀疏解和松弛变量等多项技术, 使其可以将非线性问题转化为线性问题, 并得到全局最优解.

支持向量回归<sup>[2]</sup> (SVR) 在系统建模中的应用, 实际上是通过核将数据映射到高维特征空间进行线性回归的学习过程. 因此, 在神经网络应用中的结构选择问题便转变为 SVR 中选择合适的核函数问题.

近年来, 核学习引起了国内外众多学者的广泛关注: 利用量测输入输出数据之间相似性的核排列<sup>[3,4]</sup>, 基于核函数基本性质的核混合<sup>[5,6]</sup>, 通过在核空间上定义再生核 Hilbert 空间 (RKHS) 引出的超越核<sup>[7,8]</sup> 等, 当然还包括诸如文本、结构化数据等特定核的研究. 文献 [5] 将核混合用于多输入单输出 (MISO) 系统建模, 取得了很好的效果, 然而核混合的权重系数难以确定. 为了解决该问题, Lanckriet<sup>[6]</sup> 将标准的 SVM 学习的线性约束下二次规划问题转化为二次约束下的二次规划问题, 提出了多重核学习 (MKL), 然而核的混合必然加重运算负担.

利用 SVR 进行系统建模与预测, 其效果主要取决于高维特征空间中数据分布对原始数据的表达能力. 一般常用的核函数是基于距离的高斯核, 本文在

收稿日期: 2007-04-02; 修回日期: 2007-07-11.

基金项目: 国家自然科学基金重点项目 (60234010); 航空科学基金项目 (05 E52031).

作者简介: 张军峰 (1979 →), 男, 江苏建湖人, 博士生, 从事机器学习、故障预报的研究; 胡寿松 (1937 →), 男, 浙江慈溪人, 教授, 博士生导师, 从事复杂系统的可靠控制等研究.

高斯核的基础上引入数据之间的角度余弦函数,提出一类局部固定核——高斯余弦核 (Gauss-Cos Kernel). 高斯余弦核可以增强数据在高维特征空间中的表达能力,从而提高 SVR 系统建模与预测的效果. 本文首先简要介绍支持向量回归与核排列;然后提出高斯余弦核;最后将其应用于 Logistic 混沌时间序列和煤气炉数据集的建模与预测,取得了良好的效果,进而验证了方法的有效性.

## 2 支持向量回归与核排列

### 2.1 支持向量回归

对于回归问题,假设给定的样本

$$S = \{ (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \}. \quad (1)$$

其中:  $x_i \in \mathbf{R}^m, y_i \in \mathbf{R}$

首先用一个非线性映射 将数据映射到一个高维特征空间;然后在高维空间中对新样本  $x_{new}$  进行线性回归

$$f(x_{new}) = W^T(x_{new}) + b. \quad (2)$$

定义 不敏感损失函数

$$\begin{cases} |y_i - f(x_i)| & \leq \epsilon \\ 0, & |y_i - f(x_i)| > \epsilon \end{cases} \quad (3)$$

根据结构风险最小化原则,并引入松弛变量  $\xi_i$  和  $\xi_i^*$ ,回归问题在一阶范数软间隔约束下可归结为如下优化问题:

$$\begin{aligned} \min_{w, b, \xi_i, \xi_i^*} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*). \quad (4) \\ \text{s.t.} & y_i - w \cdot (x_i) - b \leq \xi_i, \\ & w \cdot (x_i) + b - y_i \leq \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n. \end{aligned}$$

应用拉格朗日乘子法求解该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如下:

$$\begin{aligned} \max_{\alpha, \beta, \gamma_i, \gamma_i^*} & - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_j)(\beta_i - \beta_j) K(x_i, x_j) - \\ & \sum_{i=1}^n (\alpha_i + \beta_i) y_i (\alpha_i - \alpha_j). \quad (5) \\ \text{s.t.} & \sum_{i=1}^n (\alpha_i - \beta_i) = 0, \\ & 0 \leq \alpha_i, \beta_i \leq C, i = 1, 2, \dots, n. \end{aligned}$$

其中  $K(x_i, x_j)$  为满足 Mercer 条件<sup>[1]</sup>的核函数.

由于 SVM 的稀疏性,二次规划(5)的解中只有少量样本的系数  $(\alpha_i - \beta_i)$  不为 0,系数不为 0 所对应的样本  $(x_i, y_i)$  就是支持向量 (SV). 设支持向量个数为  $[n_{sv}]$ ,因而回归函数形式为

$$f(x_{new}) = \sum_{i=1}^{n_{sv}} (\alpha_i - \beta_i) K(x_{new}, x_i) + b. \quad (6)$$

### 2.2 核排列

核选择在基于核的机器学习问题上具有举足轻重的作用. Cristianini 等<sup>[3]</sup>通过度量核函数与目标之间的相似度提出了核排列的方法,用以解决支持向量分类学习中的模型选择问题;然后又将其推广到回归领域<sup>[4]</sup>. 对于给定的样本  $S$ ,SVR 中核排列如下:

$$A(S, K_1, K_2) = \frac{K_1, K_2 - F}{\sqrt{K_1, K_1 - F, K_2, K_2 - F}}. \quad (7)$$

其中:  $K_1, K_2 - F = \sum_{i,j=1}^n K_1(x_i, x_j) K_2(x_i, x_j)$ ,  $K_1$  为 SVR 中待选核,  $K_2 = yy^T$  为秩等于 1 的目标核,  $y_i = y_i - \bar{y}, \bar{y}$  表示均值.

文献[4]证明了核排列中  $A(S, K_1, K_2)$  值越大,基于核函数  $K_1$  的 SVR 的泛化性能越好. 因此,本文主要运用  $A(S, K_1, K_2)$  来衡量下面提出的高斯余弦核较一般的高斯核是否具有优越性.

## 3 高斯余弦核

### 3.1 核的定义与性质

定义 1(Mercer 核的等价定义)<sup>[9]</sup> 称  $K(x, z)$  是 Mercer 核,如果  $K(x, z)$  是  $\mathbf{x}$  上的连续对称函数,  $\mathbf{x}$  是  $\mathbf{R}^m$  上的紧集,且  $K(x, z)$  关于任意的  $x, z$  的 Gram 矩阵半正定.

根据定义 1,可知核具有如下性质:

设  $K_1$  和  $K_2$  是  $\mathbf{x}$  上的核,  $\mathbf{x} \subseteq \mathbf{R}^m$ . 设常数  $a > 0, p(x)$  是系数全为正数的多项式,则下面的函数均是核:

$$K(x, z) = K_1(x, z) + K_2(x, z), \quad (8)$$

$$K(x, z) = aK_1(x, z), \quad (9)$$

$$K(x, z) = K_1(x, z) K_2(x, z), \quad (10)$$

$$K(x, z) = p(K_1(x, z)). \quad (11)$$

定义 2<sup>[10]</sup> 称  $K(x, z)$  是固定核,若该核函数满足  $K(x, z) = K(x - z)$ ,即平移不变性.

显然,固定核由滞后矢量  $(x - z)$  确定,是关于  $(x - z)$  距离与角度的函数. 因此,固定核有时被称作各向异性的固定核. 若固定核仅由滞后矢量  $(x - z)$  的距离(即  $|x - z|$ ) 确定,则该固定核称为等方固定核. 高斯核函数即是一典型的等方固定核.

定义 3<sup>[10]</sup> 称  $K(x, z)$  是局部固定核,若

$$K(x, z) = K_1\left(\frac{x+z}{2}\right) K_2(x - z).$$

其中:  $K_1$  是非负函数,  $K_2$  是固定核.

通过考察定义可知,若  $K_1$  为正常数,则局部固定核退化为固定核,即固定核是局部固定核的一个特例. 选择  $(x + z)/2$  作为  $K_2$  的变量,是因为该变量可以反映样本  $x$  与  $z$  的中心信息. 若假设  $K_2(0) =$

1, 则  $K(x, x) = K_1(x) K_2(0) = K_1(x)$ , 即  $K_1(x)$  描述了样本全局的结构. 另一方面, 由于固定核具有平移不变性,  $K_2(x - z)$  描述了样本局部的结构. 因此, 局部固定核可以同时描述样本全局和局部的结构, 从而增强样本在特征空间的表示能力, 达到提高 SVM 学习效果的目的.

### 3.2 Gauss-Cos 核

在将 SVR 用于建模或预测时, 相似性度量起着相当重要的作用. 以高斯核为例, 样本向量之间的相似性就是通过 Euclidean 距离来度量的. 然而, 两个样本向量之间的相似性度量不一定局限于距离函数, 它们夹角的余弦函数

$$S(s_i, x) = \cos(\theta_i) = \frac{s_i \cdot x}{\|s_i\| \|x\|}, \quad (12)$$

也是衡量相似性的有效测度. 若  $\cos(\theta_i) > \cos(\theta_j)$ ,  $\forall j \neq i$  成立, 则认为未知样本  $x$  与样本  $s_i$  最相似.

对给定的样本点  $x_i$  进行如下预处理:

$$x_i = \frac{x_i - \bar{x}}{\|x_i - \bar{x}\|}, \quad y_i = \frac{y_i - \bar{y}}{\|y_i - \bar{y}\|}. \quad (13)$$

其中:  $\bar{x}$  表示输入样本向量的均值向量,  $\bar{y}$  表示输出样本均值,  $i = 1, 2, \dots, n$ . 此时线性核函数即可视为样本之间夹角的余弦函数, 因为

$$K(x_i, x_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|} = \cos(\theta_{ij}). \quad (14)$$

此时, 定义一个如下所示的函数:

$$K_{GC}(x_i, x_j) = \frac{1 + x_i \cdot x_j}{2} \cdot \exp\left(-\frac{\|x_i - x_j\|^2}{2}\right). \quad (15)$$

**命题 1**  $K_{GC}$  是核函数, 且是一个局部固定核.

**证明** 由于

$$\begin{aligned} K_{GC}(x_i, x_j) &= \frac{1 + x_i \cdot x_j}{2} \cdot \exp\left(-\frac{\|x_i - x_j\|^2}{2}\right) = \\ &= \frac{1}{2} \left[ \exp\left(-\frac{\|x_i - x_j\|^2}{2}\right) + x_i \cdot x_j \cdot \exp\left(-\frac{\|x_i - x_j\|^2}{2}\right) \right], \end{aligned}$$

根据核的性质(8)~(10)可知, 函数  $K_{GC}(x_i, x_j)$  是核函数.

又由式(12)可知,  $\|x_i\| = 1$ , 其中  $i = 1, 2, \dots, n$ , 则

$$\begin{aligned} \frac{1 + x_i \cdot x_j}{2} &= \frac{\|x_i\|^2 + \|x_j\|^2 + 2 x_i \cdot x_j}{4} = \\ &= \frac{\|x_i + x_j\|^2}{4}, \end{aligned}$$

即

$$K_{GC}(x_i, x_j) = \frac{\|x_i + x_j\|^2}{4} \cdot \exp\left(-\frac{\|x_i - x_j\|^2}{2}\right).$$

于是, 由局部固定核的定义可知, 核函数  $K_{GC}(x_i, x_j)$  是一个局部固定核.

$K_{GC}$  同时具备基于距离和角度的相似性测度, 且一部分因子为高斯核, 另一部分含有余弦的信息, 故称作高斯余弦核.  $K_{GC}$  作为一个局部固定核, 不仅可以描述样本全局和局部的结构, 而且相对于混合核<sup>[5,6]</sup>, 并没有引入额外的待定参数.

## 4 仿真实验

将本文提出的高斯余弦核应用于 Logistic 混沌时间序列和煤气炉数据集的 SVR 建模与预测, 并与普通的高斯核以及核混合的结果进行对比. 以均方根误差(RMSE)作为性能指标来衡量不同方法对训练数据的拟合效果以及对测试数据的预测效果.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y(i) - \hat{y}(i))^2}. \quad (16)$$

其中:  $y(i)$  和  $\hat{y}(i)$  分别为样本的输出实际值和预测值,  $n$  表示样本的数目.

在使用训练样本进行学习时, 首先应对样本进行式(13)所示预处理, 对数据的预测结果也要进行相应的反处理. 以下实验均在 PC 机 (AMD 1.8 GHz, 256 Mb 内存) 上 Matlab 6.5 运行环境下完成. 该仿真实验所涉及的优化问题(基于高斯核或高斯余弦核的 SVR 是线性约束下二次规划问题, 基于混合核的多重核学习 SVR 是二次约束下二次规划问题)都是利用 MOSEK 优化软件包, 基于内点法<sup>[11]</sup>求解, 可在 <http://www.mosek.com> 下载 MOSEK 及其使用手册.

### 4.1 Logistic 混沌时间序列

Logistic 时间序列

$$x(k+1) = a \cdot x(k) \cdot (1 - x(k)), \quad (17)$$

设  $a = 4$ . 此时系统成为混沌状态, 初始值为 0.1, 产生 1 000 个数据. 选取其中 150 个数据作为训练样本, 另 500 个数据作为测试样本. 仿真中选取嵌入维数为 4, 即输入输出数据为

$$\begin{aligned} x_i &= [x(i-4) \quad x(i-3) \quad x(i-2) \quad x(i-1)], \\ y_i &= x(i). \end{aligned}$$

SVR 学习中所需的参数如下: 不敏感参数  $\gamma$ , 正则化因子  $C$ , 高斯核与高斯余弦核的核参数  $\sigma$  均由基于交叉验证的网格搜索方法获得. 仿真结果如表 1 所示. 其中: 核混合由高斯核与线性核组成,  $A(St_r)$  表示训练数据的核排列值,  $A(St_s)$  表示测试数据的

表 1 Logistic 混沌时间序列仿真结果

核函数	高斯核	高斯余弦核	核混合
正则化因子 $C$	100	50	50
不敏感参数	0.05	0.05	0.05
核参数	0.45	0.6	0.75
拟合 RMSE	0.026 2	0.025 4	0.024 7
预测 RMSE	0.030 7	0.028 4	0.027 6
$A(\text{Str})$	0.074 6	0.113 2	0.121 0
$A(\text{Sts})$	0.051 6	0.087 4	0.101 5
$/nsv/$	35	30	32
时间/s	0.5	0.5	1.1

核排列值.

由表 1 可看出,高斯余弦核相对于普通高斯核可以在保持优化算法复杂度的前提下(优化时间相等),提高模型预测的精度.其中,拟合的 RMSE 性能指标下降了 3.1%,预测的 RMSE 性能指标下降了 7.5%.这些成果是在减少支持向量的基础上取得的,这是由于高斯余弦核可以同时描述样本全局和局部结构.基于核混合的多重核学习支持向量回归(MKL-SVR)虽然提高了模型预测的精度,但其代价是支持向量数目增加,优化速度减慢,而且若加大训练样本的规模,这种代价将更加昂贵.训练数据和测试数据的核排列值也反映了基于高斯余弦核的核矩阵比普通高斯核更相似于目标核.

#### 4.2 煤气炉数据集

煤气炉数据集<sup>[12]</sup>由 296 对输入-输出数据构成,如图 1 所示.输入数据表示空气流入速度,输出

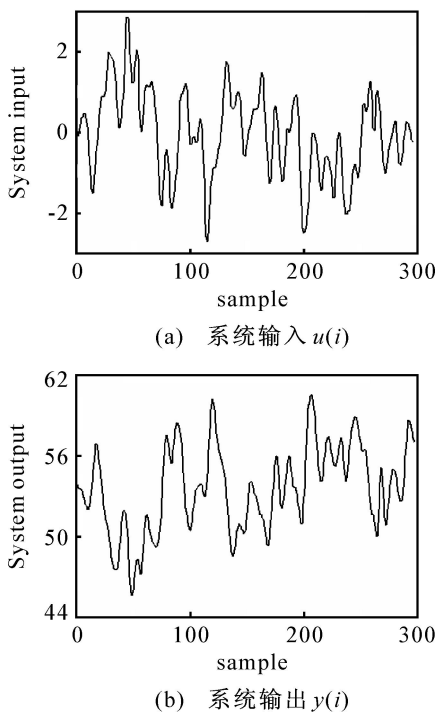


图 1 煤气炉数据集

数据表示煤气炉排放气体中  $\text{CO}_2$  的浓度.

为了构建煤气炉模型,可设  $y(i)$  为系统的期望输出,系统的输入向量

$$x_i = [y(i-1), y(i-2), y(i-3), u(i-1), u(i-2), u(i-3)]. \quad (18)$$

如此可获得 293 个输入输出样本,选取其中下标为奇数的样本作为训练样本,下标为偶数的样本作为测试样本. SVR 学习中所需的参数均由基于交叉验证的网格搜索方法获得,仿真结果如表 2 所示.

表 2 煤气炉数据集仿真结果

核函数	高斯核	高斯余弦核
正则化因子 $C$	100	50
不敏感参数	0.05	0.05
核参数	0.4	0.6
拟合 RMSE	0.208 5	0.238 4
预测 RMSE	0.351 2	0.324 5
$A(\text{Str})$	0.419 6	0.448 6
$A(\text{Sts})$	0.424 0	0.455 1
$nsv$	68	61
时间/s	0.5	0.5

由表 2 可看出,基于高斯余弦核与普通高斯核的 SVR 相比,优化时间相等,拟合的 RMSE 性能指标前者不如后者,其原因在于后者拥有更多的支持向量.然而,预测的 RMSE 性能指标前者比后者下降了 7.6%,支持向量减少了 7 个.由此可见,高斯余弦核的引入对于支持向量回归的建模与预测是卓有成效的,而且不以存储内存、训练时间以及泛化性能为代价.另一方面,从训练数据和测试数据的核排列值看,由高斯余弦核生成的核矩阵与目标核更为相似,以此也可得出高斯余弦核优于普通高斯核的结论.

#### 5 结 论

基于统计学习理论的支持向量机,通过核函数将输入向量映射到一个高维特征空间,从而实现数据线性可分,然后利用线性学习器.因此,支持向量机的学习能力往往取决于核函数的选择和设计.新兴发展的核混合由于集成了多个核函数,可实现输入向量在特征空间较好的分布.然而,计算和存储核需要额外的开销,而且引入权系数必然导致算法复杂度的增加.

本文在集合向量之间相似性的两个测度——距离和角度余弦,定义了高斯余弦核,并证明该核函数属于局部固定核.由于局部固定核的固有特性,使本文提出的高斯余弦核在建模与预测中显示出良好的性能.由仿真实验可以看出,该核在不牺牲算法复杂

度为代价的前提下,可有效地提高预测精度,且支持向量的个数相对于普通高斯核有所减少,增强了模型的泛化性能.

### 参考文献(References)

- [1] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995.
- [2] Smola A J, Schölkopf B. A tutorial on support vector regression[R]. London: University of London, 1998.
- [3] Cristianini N, Shawe-Taylor J, Elisseeff A. On kernel target alignment [C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2002: 367-373.
- [4] Kandola J, Shawe-Taylor J, Cristianini N. On the extensions of kernel alignment[R]. Cumberland Lodge: NeuroCOLT, 2002.
- [5] 朱燕飞, 伍建平, 毛宗源, 等. MISO 系统的混合核函数 LS-SVM 建模[J]. 控制与决策, 2005, 20(4): 417-420.  
(Zhu Y F, Wu J P, Mao Z Y, et al. Modeling of LS-SVM based on mixtures of kernels for MISO systems [J]. Control and Decision, 2005, 20(4): 417-420.)
- [6] Lanckriet G R G, Cristianini N, Bartlett P, et al. Learning the kernel matrix with semidefinite programming [J]. J of Machine Learning Research, 2004, 5: 27-72.
- [7] Ong C S, Smola A J, Williamson R C. Hyperkernels [C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2003: 495-502.
- [8] Ong C S, Smola A J, Williamson R C. Learning the kernel with hyperkernels [J]. J of Machine Learning Research, 2003, 3: 1001-1029.
- [9] Schölkopf B, Smola A J. Learning with kernels [M]. Cambridge: MIT Press, 2002.
- [10] Genton M G. Classes of kernels for machine learning: A statistical perspective [J]. J of Machine Learning Research, 2003, 3: 1001-1029.
- [11] Anderson E D, Anderson A D. The MOSEK interior point optimizer for linear programming [C]. High Performance Optimization. Berlin: Springer, 2000: 197-232.
- [12] Box G E P, Jenkins G M, Reinsel G C. Time series analysis, forecasting and control [M]. 3rd Edition. New Jersey: Prentice Hall, 2005.

(上接第 654 页)

### 参考文献(References)

- [1] Lian F L, Moyne J, Tilbury D. Analysis and modeling of networked control systems: MIMO case with multiple time delays [C]. Proc of the American Control Conf. Arlington, 2001, 6: 4306-4312.
- [2] Zhang W, Branicky M S, Phillips S M. Stability of networked control systems [J]. IEEE Control Systems Magazine, 2001, 21(1): 84-99.
- [3] Yong J. Stabilization of linear systems by time-delay feedback controls [J]. Quarterly Applied Mathematics, 1987, 36(2): 377-388.
- [4] Yong J, Arapostathis A. Stabilization of discrete-time linear systems with a time delay in the feedback loop [J]. Int J Control, 1988, 48(4): 1475-1485.
- [5] Tian Y P, Zhu J. Full characterization on limitation of generalized delayed feedback control for discrete-time systems [J]. Physica D, 2004, 198(3): 248-257.
- [6] Zhu J, Tian Y P. Necessary and sufficient conditions for stabilizability of discrete-time systems via delayed feedback control [J]. Physics Letters A, 2005, 343(1): 95-107.
- [7] Zhu J, Tian Y P. Nonlinear recursive delayed feedback control for chaotic discrete-time systems [J]. Physics Letters A, 2003, 310(4): 295-300.
- [8] Yamamoto S, Hino T, Ushio T. Dynamic delayed feedback controller for chaotic discrete-time systems [J]. IEEE Trans on Circuits and Systems — I: Fundamental and Applications, 2001, 48(6): 785-789.
- [9] Yamamoto S, Hino T, Ushio T. Recursive delayed feedback control for chaotic discrete-time systems [C]. Proc of the 40th IEEE Conf on Decision and Control. Orlando, 2001: 2187-2192.
- [10] Pyragas K. Continuous control of chaos by self-controlling feedback [J]. Physics Letters A, 1992, 170(4): 421-428.
- [11] Ushio T. Limitation of delayed feedback control in nonlinear discrete-time systems [J]. IEEE Trans on Circuits and Systems — I: Fundamental and Applications, 1996, 43(9): 851-856.