小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems

2008年1月 第1期
Vol.29 No.1 2008

基于信息几何的混沌支持向量机预测

王德吉[1,2],关 柯[1],熊范纶[1]

[1](中国科学院 合肥智能机械研究所,安徽 合肥 230031)
[2](中国科学技术大学 信息科学技术学院,安徽 合肥 230026)
E-mail:wangdejiboy@yahoo.com.cn

摘 要:复杂环境中存在大量的混沌现象,难以用传统的预测方法进行准确预测.针对这一问题,本文利用信息几何理论、支持向量机理论与重构相空间理论,提出混沌支持向量机CSVM,对含有混沌现象的时间序列进行预测;针对混沌环境下核函数难于构造,从信息几何角度,提出在混沌环境下,如何方便准确得进行构造核函数;最后将CSVM应用于Henon混沌系统实验.实验结果表明,误差随嵌入维数变化和延迟时间变化趋于恒定;与BP、RBF和SVM相比,CSVM具有所需支持向量少,收敛速度快,准确性高等特点.

关键词:信息几何;核函数;支持向量机;预测;混沌
中图分类号:TP173　　**文献标识码**:A　　**文章编号**:1000-1220(2008)01-0110-04

Prediction with Chaos SVM Based on Information Geometry

WANG De-ji[1,2], GUAN Ke[1], XIONG Fan-lun[1]
[1](*Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei* 230031, *China*)
[2](*School of Information Science & Technology, University of Science & Technology of China, Hefei* 230026, *China*)

Abstract: There are many chaos phenomenons in complex environment, so it is difficult to predict by the traditional methods. Chaos support vector machine was given to predict time-series with chaos phenomenon to overcome the disadvantages of the traditional methods in this paper based on information geometry, SVM theory and chaos theory. Especially, a new kernel function was introduced into the chaos support vector machine from the perspective of information geometry and thus it is easy to design the kernel function. Finally, the method was applied to Henon chaos system compared with the BP, RBF and SVM. The prediction results indicate that the predictive error changes with the increase of embed dimension and delay time to a constant. And the results also show that the chaos support vector machine is more precise although it requires smaller support vector, and has faster convergence rate, compared with BP、RBF and SVM.

Key words: information geometry; kernel function; SVM; prediction; chaos

1 引 言

复杂环境中存在着大量的混沌时间序列,如电力系统负荷、股票价格和水文数据等.而对这类现象的预测具有很重要的意义,因此吸引了很多科学工作者对此进行研究.目前混沌时间序列的建模和预测已成为混沌信息处理研究领域中近几年来的一个重要研究热点[1].混沌时间序列是一种介于确定性和随机性之间的随机、敏感、充分不规则的非线性动力现象.宏观上表现为无序无律的混乱运动,以及对初值十分敏感的蝴蝶效应,微观上呈现无穷嵌套几何自相似性.运动具有普适规律,是有序和无序、确定性和随机性的中间态,混沌时间序列区别于随机时间序列,它具有短期的可预测性和长期的不可预测性.这是因为混沌信号并不是真正的随机信号,而是由低维非线性动力系统产生的确定的伪随机信号.它与高维的随机信号的区别主要在于:混沌信号可通过重构体现在嵌入空间的一个低维流形上,其重构的轨迹可以预测,而随机信号则不能.

支持向量机在时间序列预测问题研究比较成熟,而关于混沌时间序列的预测问题研究则刚刚起步.本文试图利用信息几何理论、支持向量机(SVM)理论与重构相空间理论研究混沌时间序列预测的问题,提出基于信息几何的混沌支持向量机(CSVM).

2 混沌支持向量机

混沌时间序列描述了混沌系统在一段时间内的动力学行为.Takens在拓扑学方面的研究表明,可以在一个高维的相空间中恢复系统原来的规律,从而混沌时间序列 $x(t)$ 在原则上是可预测的.他证明了在一定条件下,对几乎所有的 τ 和满足特定条件的 m,存在一个光滑映射 $f:\mathbf{R}^m \to \mathbf{R}$,使得:

$$x(n\tau)=f(x((n-1)\tau),\ x((n-2)\tau),\ \cdots,\ x((n-m)\tau)) \quad (1)$$

收稿日期:2006-09-14 收修改稿日期:2007-01-10 基金项目:国家自然科学基金重点项目(69835001)资助 作者简介:王德吉,男,1975年生,博士,助研,研究方向为数据挖掘,决策系统;关 柯,男,1978年生,博士,助研,研究方向为数据融合,决策系统;熊范纶,男,1940年生,教授,研究领域为农业信息化,智能决策.

其中 m 被称作嵌入维,使等式(1)成立的最小的 m 被称为最小嵌入维,τ 被称为延迟时间,文献[2,3]已经讨论了 m,τ 的计算方法,这里不做介绍.

给定长为 N 的时间序列 $x_1,x_2,\cdots,x_N$ 作为训练样本.由(1)式,选择合适的 $m,\tau\in\mathbf{R}$,样本数据可记为:

$$\prod=\left\{((x_{(n-1)\tau},x_{(n-2)\tau},\cdots,x_{(n-m)\tau}),x_{n\tau})\in\mathbf{R}^m\times\mathbf{R}\,|\,m<n\leqslant\frac{N}{\tau}\right\} \tag{2}$$

为了书写简便,记 $\boldsymbol{x}_n=(x_{(n-1)\tau},x_{(n-2)\tau},\cdots,x_{(n-m)\tau}),y_n=x_{n\tau}$.若 $|\prod|=T$,可记为:$\prod=\{(\boldsymbol{x}_i,y_i)\,|\,i=1,2,\cdots,T\}$.这样就可以考虑支持向量机来求其函数.

SVM 是 Vapnik 及其同事在 1992 年的 COLT 会议上首次提出的(Boser,Guyon and Vapnik,1992)最初用于模式识别问题[4].概括地说,SVM 是通过非线性映射将输入空间映射到一个高维特征空间(figure space),在这个空间中构造最优分类超平面的实现过程(Vapnik,1995).所谓最优分类面,就是不但能将所有样本正确分类,而且使训练样本中离分类面最近的点到分类面的间隔(Margin)最大.通过使间隔最大化来实现较好的泛化能力.

支持向量机方法的重要理论基础是统计学习理论的 VC 维理论和结构风险最小化原理(SRM).SVM 根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误的识别任意样本的能力)之间寻求最佳折衷,以求获得最佳的泛化能力.SVM 可以自动寻找对分类有较好区分能力的支持向量,由此构成的分类器可以最大化类与类之间的间隔(Margin).

SVM 方法的几个主要优点可以归纳如下:

· SVM 是专门针对有限样本情况的,其目标是得到现有信息下的最优解,而不仅仅是样本趋于无穷多时的最优解;

· 在 SVM 算法中,训练 SVM 就相当于解决一个线性约束的二次规划问题,因而 SVM 的解是唯一的、全局的、最优的.而传统的神经网络方法可能存在很多局部最小点,因此在训练中很难得到全局最优解;

· SVM 算法将实际问题通过一个非线性变换映射到一个高维的特征空间,在高维的特征空间构造线性判别函数以替换原空间的非线性判别函数.这样既能够保证机器有较好的泛化能力,同时又巧妙的解决了维数问题,其算法复杂度与样本维数无关.

在 SVM 算法中,只要定义不同的核函数,就可以实现多项式逼近、贝叶斯分类器、径向基函数模型、多层感知器网络等许多现有的学习算法.

结合支持向量机原理和混沌理论分析,可以考虑如下形式的近似函数预测:

$$f(x,c)=<\mathbf{c},\Phi(\mathbf{x})>+b=\sum_{i=1}^{d}c_i\varphi_i(x)+b \tag{3}$$

其中 d 特征空间的维数,是待定参数,$\Phi(x)$ 为输入空间 S 到特征空间 F 非线性映射,c 为权值向量,b 为偏差 (Bias).b,c 可以通过使下面的误差表达式取最小值求出:

$$R(c)=\frac{1}{T}\sum_{i=1}^{T}|y_i-f(x_i,y_i)|_\epsilon \tag{4}$$

$|y_i-f(x_i,y_i)|_\epsilon$ 被称作 ϵ 截断误差函数:

$$|y_i-f(x_i,y_i)|_\epsilon=\begin{cases}0,|y_i-f(x_i,y_i)|\leqslant\epsilon\\|y_i-f(x_i,y_i)|-\epsilon,|y_i-f(x_i,y_i)|>\epsilon\end{cases} \tag{5}$$

Vapnik 给出了映射 $f:\mathbf{R}^m\to\mathbf{R}$ 的形式[4]:

$$\mathrm{f}(\boldsymbol{x},\alpha,\alpha^*)=\sum_{i=1}^{T}(\alpha_i^*-\alpha_i)K(x,x_i)+b \tag{6}$$

其中 $0\leqslant\alpha_i,\alpha_i^*\leqslant\frac{1}{\lambda},\sum_{i=1}^{T}(\alpha_i^*-\alpha_i)=0,\lambda$ 是可调参数.参数 α_i 和 α_i^* 可通过最大化下面的二次方程式求出:

$$R(\alpha_i,\alpha_i^*)=-\epsilon\sum_{i=1}^{T}(\alpha_i^*+\alpha_i)+\sum_{i=1}^{T}y_i(\alpha_i^*-\alpha_i)-\frac{1}{2}\sum_{i=1}^{T}\sum_{j=1}^{T}(\alpha_i^*-\alpha_i)(\alpha_j^*-\alpha_j)K(x_i,x_j) \tag{7}$$

$K(x,y)$是支持向量机的核函数,它描述了 d 维特征空间的内积,核函数的选择必须服从 Mercer 条件.最常用的几类核函数:多项式核、高斯核、双曲核等.但对于混沌环境下很难选择核函数,为此我们提出从信息几何的角度构造核函数.

3 基于信息几何的核函数设计

核函数和核方法的研究是 SVM 理论研究中极具特色、非常活跃的一个研究分支.核函数可隐含地将输入空间的数据映射到高维特征空间,通过非线性变换转化为某个高维空间的线性问题.

一般核函数的选择以及 Lagrange 算子的优化都未考虑环境数据的影响,对于复杂环境下具有混沌现象的数据预测很不准确,而信息几何的能很好解决这个问题.

信息几何是采用(Riemann 流形上的)微分几何方法来研究统计学的理论.自 1975 年 Efron 首先在统计学中采用微分几何方法以来,许多统计学家在这方面进行了大量的工作.特别是由于甘利俊一(S. Amari)和 Zhu Haiyu 等人的杰出工作,使得信息几何理论得到学术界的广泛关注,成为统计学中一个令人瞩目的新分支,并在许多领域得到了大量应用[5].

设 m 维样本空间上随机变量 $\boldsymbol{X}$ 的概率分布(参数)簇 $S=\{p(\boldsymbol{X}|\theta)\,|\,\theta\in\Theta\}$,其中 θ 为该分布簇的参数向量,Θ 为 n 维欧式空间 $\mathbf{R}^n$ 的一个开集.在 p 满足一些正则条件的情况下,S 形成一个微分流形,称为统计流形,θ 称为统计流形的自然坐标[6].

在自然坐标 θ 下,Fisher 信息矩阵成为此概率分布所对应的流形 S 的 Riemann 度量.事实上,从保持充分统计量变换下度量不变的意义上说,Fisher 信息矩阵是统计流形上唯一合适的 Riemann 度量.与欧式空间的距离不通,该度量具有相对坐标变化的不变性,从而在很大程度上体现了样本分布的内在特征.

从几何的观点看,非线性映射 $\Phi(x)$是一个子流形,它定义了从输入空间 S 到特征空间 F 的一个嵌入.一般 F 为再生核 Hilbert 空间(RKHS),RKHS 是 Hilbert 空间的子空间,因此,可以在 S 空间中引入一个黎曼度量 $G_{i,j}$,这个黎曼度量可以用核函数 $\boldsymbol{K}(\boldsymbol{x},\boldsymbol{x}')$近似地表示为

$$G_{i,j}=\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j'}K(x,x')\,|x'=x \tag{8}$$

其中 σ 为归一化参数，此时黎曼度量为

$$G_{i,j}(x)=\frac{\delta_{i,j}}{\sigma^2} \tag{9}$$

$$\text{其中 } \delta_{i,j}=\begin{cases}1,\ i=j\\0,\ i\neq j\end{cases} \tag{10}$$

下面在核函数中引入一个保角映射 $D(x)$，于是得新的核函数为：

$$\tilde{K}(x,x')=D(x)D(x')K(x,x') \tag{11}$$

由(11)式，可得到变换后的黎曼度量：

$$\tilde{G}_{i,j}(x)=\frac{\partial D(x)}{\partial x_i}\frac{\partial D(x)}{\partial x'_j}+[D(x)]^2G_{i,j}(x) \tag{12}$$

基于上述讨论，如果适当选取一个保角映射，就可以在保持原来空间拓扑结构不变的情况下，对非线性数据中的重要样本点附近的区域实现有效放大，从而提高预测效果.

定理:新的核函数 $\tilde{K}(x,x')$满足 Mercer 定理，可以作为 SVM 中的核函数.

证明:设 $\boldsymbol{A}$ 为 $\boldsymbol{R}^N$ 的紧子集，$h(x)\in L_2(A)$.

由 $\tilde{K}(x,x')=D(x)D(x')K(x,x')$，易见是连续并且对称.

因为 $D(x)>0$，因此必存在一个正数 β，使得 $D(x)\geqslant\beta>0$，从而

$$\iint_{A\times A}\tilde{K}(x,x')h(x)h(x')dxdx'=$$

$$\iint_{A\times A}D(x)D(x')K(x,x')h(x)h(x')dxdx'\geqslant$$

$$\beta^2\iint_{A\times A}K(x,x')h(x)h(x')dxdx'\geqslant 0 \tag{13}$$

由此可见 $\tilde{K}(x,x')$半正定.

新的核函数 $\tilde{K}(x,x')$满足 Mercer 定理，因此可作为 SVM 中核函数.

4 实验

根据上述理论，把 CSVM、标准 SVM、BP 网络和 RBF 网络用于混沌时间序列建立预测模型，比较分析它们的预测效果.

采用一步预测方法，将待测试点输入训练好的回归模型中，进而对下一时间点进行预测. 测试指标采用均方误差：

$$MSE=\sqrt{\frac{1}{l}\sum_{i=1}^{l}(X_i-\hat{X}_i)^2} \tag{14}$$

其中 X_i 为实际值，$\hat{X}_i$ 为预测值，l 为测试和样本的数量.

混沌时间序列采用 Henon 映射产生，Henon 混沌系统有别于 Hamilton 系统相空间容积不变的系统(相空间维数在运动中保持不变)，Henon 在运动时相空间维数低于原来相空间维数，相空间收缩到吸引子(Attactors)上，是耗散系统中混沌的独特现象. 采用 Henon 映射产生混沌时间序列，其具有复杂的非线性混沌效应和自相似性. 加之能量不守恒，耗散系统特点的影响，其混沌时间预测模型建立较困难. 采用基于信息几何的混沌支持向量预测模型，可以在对模型未知的情况下，由较小的训练样本集对未知的 Henon 序列预测. 取其中 360 个数据，前 280 个作为训练集，后 80 个作为测试集. 它的产生式如下：

$$x_{n+1}=1-1.4x_n^2+0.3x_{n-1} \tag{15}$$

取初值[0,0]，产生 360 个数据.

在基于信息几何的 CSVM 预测中，相空间重构的关键在于重构参数 τ 和 m 的选取. 延迟时间 τ 的计算采用平均互信息法，嵌入维数 m 的计算可采用伪邻近点法. 在不同延迟时间内的互信息值和不同维数时的伪临近点分数分别如图 1、图 2 所示.

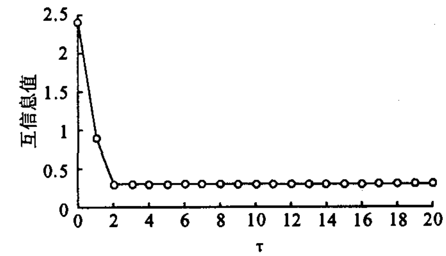

图 1 在不同延迟时间内的互信息值

Fig. 1 Relation between the mutual information values and different delay times

从图 1 可以看出，当延迟时间小于最小延迟时间($\tau=2$)，随延迟增加其预测的互信息值迅速减小；当大于最小延迟时间，延迟时间的增加对预测性能没有明显的改善.

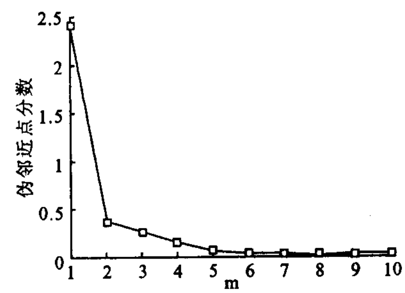

图 2 不同嵌入维数时的伪临近点分数

Fig. 2 Relation between the embedding dimensions and pseudo neighboring points

从图 2 可以看出，小于最小嵌入维数($m=5$)的嵌入，随维数提高其预测的伪临近点分数迅速减小，超出最小嵌入维数的嵌入，嵌入维数的增加对预测性能没有明显的改善.

从图 1 和图 2 知道，相空间重构参数取 $\tau=2,m=5$.

在CSVM 和标准SVM 模型中，取参数 $c=2000,\varepsilon=0.0001$，核函取为

$$K(x,x')=\exp(-\frac{\|x-x'\|^2}{2\rho^2})=\exp(-\frac{\|x_i-x_j\|^2}{2\times 0.75}) \tag{16}$$

将保角映射取为：

$$D(x)=\frac{1}{n}\sum_{i=1}^{n}\exp(-\frac{\|x-o_i\|^2}{\beta_i^2}) \tag{17}$$

其中 n,o_i,β_i 分别为分类点数目，第 i 类的中心与对应的宽度.

在 BP 网络中，我们采用三层结构，隐含层采用五个单

元.需要说明的是,如果增加网络的层数及隐层单元的个数,有时不但不能提高网络性能,还往往会降低网络的推广能力,因此只能根据实际经验来选择网络结构.

RBF 网络是一种特殊的三层前向网络,输入到隐单元的权值固定为 1,只有隐单元到输出单元间的权值为可调,隐单元的作用函数采用径向基函数.混沌序列的预测结果如图 3、图 4、图 5、图 6 和图 7 所示.

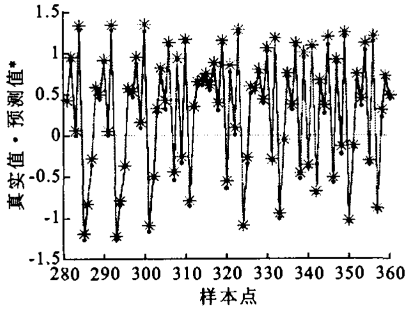

图 3　混沌 SVM 预测值与真实值比较

Fig. 3　The predicting values by chaos SVM

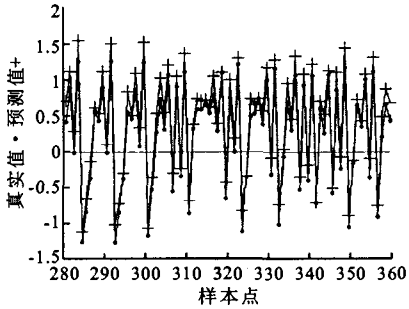

图 4　BP 预测值与真实值比较

Fig. 4　The predicting values by BP

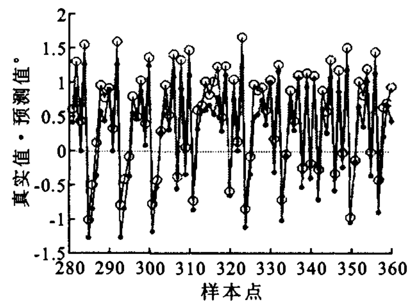

图 5　RBF 预测值与真实值比较

Fig. 5　The predicting values by RBF

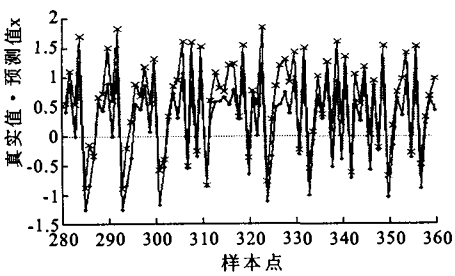

图 6　SVM 预测值与真实值比较

Fig. 6　The predicting values by SVM

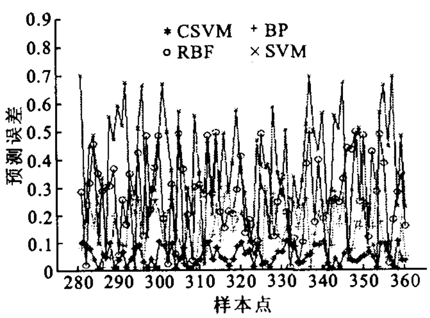

图 7　各种预测方法预测误差比较

Fig. 7　The predicting errors by different approaches

表 1　各种预测方法性能比较

Table 1　Comparative performances by different predicting approaches

预测方法	支持向量数	MSE	预测时间
BP	204	0.0080	153.8
RBF	250	0.0037	128.7
SVM	196	0.0350	131.5
CSVM	95	0.0031	108.5

从图 3—图 7 可以看出,CSVM 在对混沌背景下 CSVM 的预测能力要好于传统的预测方法,且使用信息几何算法选择的核函数的SVM预测性能比一般核函数有了明显提高.

从表 1 可以看出:CSVM 需要的支持向量数比 SVM、RBF 和 BP 少一半,但误差却反而降低了,即精度提高了;预测时间上比原来有大的提高,约提高 16.5%.

5　结　论

本文利用信息几何理论、支持向量机理论与重构相空间理论研究了混沌时间序列的预测问题,并以 Henon 混沌时间序列为样本,对比了 CSVM 和同 BP、RBF、SVM 等传统预测方法的预测能力.实验表明,针对混沌时间序列的预测问题,CSVM 在预测性能(精度、效率、支持向量个数)上明显优于传统的预测方法.

References:

[1] Yu Shu, Ma Jun-hai, Liu Zeng-rong. The state space reconstruction technology of different kinds of chaotic data obtained from dynamical system[J]. Acta Mechanica Sinica, 1999, 1(15):82-92.

[2] He G, Wang D. Simulation study on the transition between chaos and order motion in the traffic flow[J]. China Civil Engineering J, 2003, 36(7): 53-56.

[3] Ma Jun-hai. The state space reconstruction of different kinds of chaotic data obtained from economic dynamical system[A]. In: Proceedings of 99 International Conference On Management Science & Engineering[C]. Yichang, China: 1999,11:106-110.

[4] Vapnik V. The nature of statistical learning theory[M]. New York: Wiley, 1998.

[5] Amari S, Wu Si. An information-geometrical method for improving the performance of support vector machine classifiers [A]. In: Proccedings of International Conference on Artificial Neural Networks'99 [C]. London: IEEE Conference Publication, 1999, 470: 85-90.

[6] Sun Yan-feng, Zhang Wen-li, Gu Xiao-jiong, et al. Optimal partition algorithm of RBF neural network and its applications to stock price prediction[A]. Shi Zhong-zhi, He Qing, eds, In: Proceedings of International Conference on Intelligent Information Technology[M]. Beijing: Posts & Telecom Press, 2002: 448-454.