基于 SVR 的混沌时间序列预测

孙德山[1,2] 吴今培[3] 侯振挺[1] 肖健华[3]
[1](中南大学铁道校区科学研究所,长沙 410075)
[2](辽宁师范大学数学系,大连 116029)
[3](五邑大学智能所,广东江门 529020)
E-mail:sdeshan@163.com

摘 要 支持向量机是一种基于统计学习理论的新颖的机器学习方法,由于其出色的学习性能,该技术已成为当前国际机器学习界的研究热点。这种方法已广泛用于解决分类和回归问题。论文介绍了支持向量回归算法的各种版本,同时将它们应用到混沌时间序列预测中,并且比较了它们的预测性能,为实际应用合理选择模型提供一定的依据。

关键词 支持向量机 回归 混沌时间序列 核函数

文章编号 1002-8331-(2004)02-0054-03 **文献标识码** A **中图分类号** TP18

Chaotic Time Series Prediction Based on SVR

Sun Deshan[1,2] Wu Jinpei[3] Hou Zhenting[1] Xiao Jianhua[3]
[1](Research Department of Railway Compus,Central South University,Changsha 410075)
[2](Department of Mathematics,Liaoning Normal University,Dalian 116029)
[3](Institute of Intelligent Technology & System,Wuyi University,Jiangmen 529020)

Abstract: Support vector machines (SVM)are a kind of novel machine learning methods,based on statistical learning theory,which have become the hotspot of machine learning because of their excellent learning performance.The method of support vector machines has been developed for solving classification and regression problems.Several versions of support vector regression (SVR)are introduced in this paper,and then apply them to chaotic time series prediction.Their performance of prediction is analyzed,which can provide some foundation about reasonable selecting models in practice.

Keywords: support vector machines,regression,chaotic time series,kernel function

1 引言

Vapnik 等人根据统计学习理论提出的支持向量机学习方法[1],近年来受到了国际学术界的广泛重视,已经成为解决分类和回归问题的有利工具。支持向量机的最大特点是根据结构风险最小化原则,尽量提高学习机的泛化能力。另外,由于支持向量机算法是一个凸优化问题,因此局部最优解一定是全局最优解。这些特点是其它学习算法,如神经网络学习算法所不及的。

支持向量机理论以其较好的泛化能力已经在很多领域得到应用[2-5]。论文分别用支持向量回归的各种版本以及 RBF 网络来预测带有噪声的混沌时间序列,并比较分析了它们的预测性能,从而为实际应用合理选择模型提供一定的依据。

2 支持向量回归模型

支持向量机的一大特点是将非线性问题映射到高维空间的线性问题,然后采用一个核函数来代替高维空间中的内积运算,从而避免了高维运算,巧妙地解决了非线性问题。以下各种模型都是先用一非线性映射 ϕ 把数据映射到一个高维特征空间,然后在高维特征空间进行线性回归,其中涉及到的高维内积运算用核函数 $K(x,y)$ 代替,从而导出非线性回归。关键问题是找一个核函数 $K(x,y)$ 代替 $<\phi(x),\phi(y)>$,已经证明,这样的核函数是存在的[1]。

给定训练样本集 $X=\{(x_1,y_1),\cdots,(x_l,y_l)\}\subset\chi\times R$,设拟合函数为:

$$f(x)=<w,\phi(x)>+b$$

定义损失函数:

$$c(x,y,f(x))=\begin{cases}0 & 若\,|y-f(x)|\le\varepsilon\\ \hat{c}(|y-f(x)|-\varepsilon), & 其它\end{cases}$$

根据结构风险最小化原则,回归问题可归结为最小化函数:

$$\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\hat{c}(\xi_i)+\hat{c}(\xi_i^*))$$

约束为:

$$y_i-<w,\phi(x_i)>-b\le\varepsilon+\xi_i,\quad i=1,\cdots,l$$

$$<w,\phi(x_i)>+b-y_i\le\varepsilon+\xi_i^*,\quad i=1,\cdots,l$$

$$\xi_i,\xi_i^*\ge0,i=1,\cdots,l$$

其中,$C>0$,它对上面两式做出折中。通过引入拉格朗日乘子,可以获得该优化问题的对偶形式:

$$\max\begin{cases}-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j)\\ +\sum_{i=1}^{l}(y_i(\alpha_i-\alpha_i^*)-\varepsilon(\alpha_i+\alpha_i^*)+C(T(\xi_i)+T(\xi_i^*)))\end{cases}\tag{1}$$

基金项目:广东省自然科学基金资助(编号:021349)

约束为：

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0$$

$$\alpha\le C\partial_\xi\hat{c}(\xi)$$

$$\xi=\inf\{\xi|C\partial_\xi\hat{c}\ge\alpha\}$$

$$\alpha,\xi\ge 0$$

其中，$w=\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)\phi(x_i)$，$T(\xi):=\hat{c}(\xi)-\xi\partial_\xi\hat{c}(\xi)$。

最后的解为：

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

α 由二次优化解出，b 通过 KKT 条件解出。

3 不同版本的回归算法

3.1 ε 不敏感函数下的支持向量回归（ε-SVR）

取损失函数为 ε-不敏感函数，即 $\hat{c}(\xi)=|\xi|$，则 $T(\xi)=\xi-\xi\cdot 1=0$，$\partial_\xi\hat{c}(\xi)=1$，因此有：

$$\xi=\inf\{\xi|C\ge\alpha\}$$

所以得 $\alpha\in[0,C]$。这样就得到了在 ε-不敏感函数下的支持向量回归算法，最大化函数：

$$W(\alpha,\alpha^*)=-\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j)+\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)y_i-\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)\varepsilon$$

约束为：

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0$$

$$0\le\alpha_i,\alpha_i^*\le C,i=1,\cdots,l$$

3.2 Huber 损失函数下的支持向量回归（h-SVR）

Huber 损失函数定义为：

$$\hat{c}(\xi)=\begin{cases}\frac{1}{2\sigma}(\xi)^2, & \text{如果}|\xi|\le\sigma\\ |\xi|-\frac{\sigma}{2}, & \text{否则}\end{cases}$$

分两种情况讨论，$0\le\xi\le\sigma$ 和 $\xi\ge\sigma$。对于第一种情况可得：

$$T(\xi)=\tilde{c}(\xi)-\xi\partial_\xi\tilde{c}(\xi)=\frac{1}{2\sigma}\xi^2-\xi\frac{\xi}{\sigma}=-\frac{1}{2\sigma}\xi^2$$

和

$$\xi=\inf\{\xi|C\partial_\xi\tilde{c}\ge\alpha\}=\inf\{\xi|C\frac{\xi}{\sigma}\ge\alpha\}=\sigma C^{-1}\alpha$$

因此

$$T(\xi)=-\frac{1}{2}\sigma C^{-2}\alpha^2$$

对于第二种情况可得：

$$T(\xi)=\xi-\frac{\sigma}{2}-\xi=-\frac{\sigma}{2},\xi=\inf\{\xi|C\ge\alpha\}=\sigma$$

因此 $\alpha\in[0,C]$。两种情况可以合并到一起为：

$$\xi=\sigma,\alpha\in[0,C],CT(\alpha)=-\frac{1}{2}\sigma C^{-1}\alpha^2$$

将各参数代入（1）中，可得 Huber 损失函数下的支持向量回归算法，这里取 $\sigma=0.001$。

3.3 ν-SVR

文献[6]给出一种能够自动获得 ε 参数的方法，该方法把 ε 参数作为一个优化变量。其优化模型为最小化函数：

$$\frac{1}{2}\|w\|^2+C(\sum_{i=1}^{l}(\tilde{c}(\xi_i)+\tilde{c}(\xi_i^*))+l\nu\varepsilon)$$

约束为：

$$y_i-\langle w,\phi(x_i)\rangle-b\le\varepsilon+\xi_i,i=1,\cdots,l$$

$$\langle w,\phi(x_i)\rangle+b-y_i\le\varepsilon+\xi_i^*,i=1,\cdots,l$$

$$\xi_i,\xi_i^*\ge 0,i=1,\cdots,l$$

其中，$0\le\nu\le 1$。采用 ε-不敏感损失函数，可得该问题的对偶形式：

$$\max\ \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j)+\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)$$

约束为：

$$\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0$$

$$\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)\le C\nu l$$

$$0\le\alpha_i,\alpha_i^*\le C,i=1,\cdots,l$$

该模型能够自动确定参数 ε 的值，并且可以通过调整参数 ν 来控制支持向量数[6]。

3.4 最小平方支持向量回归（Least Square Support Vector Regression，LSSVR）[7,8]

最小平方支持向量回归是将标准回归中的不等式约束化成等式约束而得到的。优化模型为：

$$\min\frac{1}{2}\|w\|^2+\gamma\frac{1}{2}\sum_{i=1}^{l}\xi_i^2$$

约束为：

$$y_i=w^T\phi(x_i)+B+\xi_i,i=1,\cdots,l$$

定义拉格朗日函数为：

$$L=\frac{1}{2}\|w\|^2+\gamma\frac{1}{2}\sum_{i=1}^{l}\xi_i^2-\sum_{i=1}^{l}\alpha_i(w^T\phi(x_i)+B+\xi_i-y_i)$$

根据 KT 条件，得：

$$\frac{\partial L}{\partial w}=0\to w=\sum_{i=1}^{l}\alpha_i\phi(x_i)$$

$$\frac{\partial L}{\partial B}=0\to\sum_{i=1}^{l}\alpha_i=0$$

$$\frac{\partial L}{\partial\xi_i}=0\to\alpha_i=\gamma\xi_i,i=1,\cdots,l$$

$$\frac{\partial L}{\partial\alpha_i}=0\to w^T\phi(x_i)+B+\xi_i-y_i=0,i=1,\cdots,l$$

这是一个线性方程组，由最小二乘法解出 α 和 β 的值。于是拟合函数为：

$$f(x)=\sum_{i=1}^{l}\alpha_iK(x_i,x)+B$$

该模型是基于最小二乘法的，所以运算速度快。它的缺点是支持向量数几乎等于训练样本数，因此该算法已失去稀疏性。

4 举例

下面用 ε-SVR、h-SVR、ν-SVR、LSSVR 和 RBF 网络分别

对混沌时间序列进行预测，然后比较分析它们的预测效果。核函数取为：

$$K(x_i,x_j)=\exp(-\frac{\|x_i,x_j\|^2}{2\rho^2}),\rho^2=0.75$$

其它参数取值分别列于表1到表3的表头中，其中 m 为嵌入维数。测试指标采用均方误差：

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

其中 X_i 为实际值，$\hat{X}_i$ 为预测值，K 为测试样本的数量。

混沌时间序列分别为 Logistic 映射、hénon 映射和 Mackey-Glass 方程，都取其中300个样本，前280个作为训练集，后20个作为测试集。

(1)Logistic 映射由式 $X_{n+1}=0.4X_n(1-X_n)$ 产生，取初值为0.1得到300个数据。

(2)hénon 映射由式 $X_{n+1}=1-1.4X_n^2+0.3X_{n-1}$ 产生，取初值[0 0]获得300个数据。

(3)Mackey-Glass 方程由下式产生：

$$\frac{dx}{dt}=\frac{0.2x(t-\tau)}{1+x(t-\tau)^{10}}-0.1x(t)$$

当 $\tau\geq17$ 时，产生的数据是混沌的。该数据可从网上下载获得(http://www.boltz.cs.cmu.edu/bench.html)，从中任取300个数据。

每次预测都采用离待测试输入样本较近的一些样本作为训练集，选择标准为欧氏距离小于 $\delta(>0)$ 的样本。各种版本的支持向量回归算法及 RBF 网络对不同噪声水平的混沌时间序列的预测结果如表1至表3。

表1 Logistic **预测结果**(C=500，v=0.5，γ=500000，ε=0.0001，δ=0.04，m=1)

模型	无噪声情况	加入 N(0,0.001)噪声	加入 N(0,0.01)噪声
ε-SVR	3.1127e-004	0.0032	0.0275
h-SVR	3.9780e-004	0.0030	0.0292
v-SVR	0.0022	0.0041	0.0330
LSSVR	5.4881e-004	0.0030	0.0279
RBF	9.1783e-010	0.0030	0.0256

表2 Henon **预测结果**(C=500，v=0.5，γ=500000，δ=0.3，ε=0.0001，m=2)

模型	无噪声情况	加入 N(0,0.001)噪声	加入 N(0,0.01)噪声
ε-SVR	1.4957e-004	0.0027	0.0238
h-SVR	2.6048e-004	0.0026	0.0213
v-SVR	0.0015	0.0038	0.0239
LSSVR	6.5906e-004	0.0026	0.0206
RBF	8.1769e-006	0.0038	0.0863

从上面的例子可以发现，SVR 的各种版本无论对无噪声数据，还是对含有噪声的数据都具有较好的预测效果，这就说明它们的稳健性能都较好。其中，ε-SVR 的稳健性最好，其次是 h-SVR 和 LSSVR。LSSVR 是基于最小二乘法求解的，因此它的特点是速度快，其缺点是已失去稀疏性。v-SVR 虽然能够自动确定 ε 的值，并且参数 v 能够决定支持向量的多少，但支持向量的多少对预测精度也有很大影响，而且自动获得的 ε 值并不能完全同噪声水平相匹配，因此如何选择最佳的 v 值并不容易。

表3 Mackey-Glass **预测结果**(C=2000，v=0.7，δ=0.06，m=3，γ=500000，ε=0.0001)

模型	无噪声情况	加入 N(0,0.001)噪声	加入 N(0,0.01)噪声
ε-SVR	0.0071	0.0134	0.0375
h-SVR	0.0113	0.0143	0.0385
v-SVR	0.0125	0.0143	0.0369
LSSVR	0.0080	0.0161	0.0648
RBF	0.1279	0.5373	1.1861

RBF 网络在预测不含噪声或是噪声很小的数据时具有较高的预测精度，而在噪声水平较高或是数据较复杂时，它的预测性能相对较低，因此它的稳健性相对较差。

5 结束语

文中用支持向量回归的各种版本以及 RBF 网络对混沌时间序列进行了预测，同时比较了它们的预测性能。在实际应用中，应根据噪声水平的大小以及对预测精度、运算速度等不同要求来合理地选择模型。另外，如何合理地选择各模型中的参数仍需进一步研究。(收稿日期：2003年3月)

参考文献

1.Vapnik V N.Statistical Learning Theory [M].New York,Wiley,1998
2.Smola A J,Scholkopf B.A tutorial on support vector regression[R].NeuroCOLT TR NC-TR-98-030,Royal Holloway College University of London,UK,1998
3.Rodrigo Fernandez.Predicting time series with a local support vector regression machine[C].In:ACAI 99,1999
4.Burges C J C.A Tutorial on Support Vector Machines for Pattern Recognition[J].Knowledge Discovery and Data Mining,1998;2(2)
5.Cortes C,Vapnic V.Support-vector networks[J].Machine Learning,1995;20(3):273~297
6.B Schilkopf,P Bartlett et al.Support vector regression with automatic accuracy control.Proceedings[C].ICANN 98,1998
7.J A K Suykens,J Vandewalle.Least squares support vector machine classifiers[J].Neural Processing Letters,1999;9(3):293~300
8.J A K Suykens,J Vandewalle.Recurrent Least squares support vector machines[J].IEEE Transactions on Circuits and Systems-I,2000;7(47):1109~1114

(上接25页)

2.谢和平，薛秀谦.分形应用中的数学基础与方法[M].北京：科学出版社，1998：12~96
3.林鸿溢，李映雪.分形论——奇异性探索[M].北京理工大学出版社，1994：24~312
4.K J Falconer.The Geometry of Fractal Sets[M].London,Cambridge University Press,1985:1~43
5.王东生，曹磊.混沌、分形及其应用[M].合肥：中国科学技术大学出版社，1995：65~120
6.刘华杰.分形艺术[M].长沙：湖南科学技术出版社，1998；24~158
7.何华灿，王华，刘永怀等.泛逻辑学原理[M].北京：科学出版社，2001：1~256
8.董远，胡光锐.图像分形维数计算技术[J].计算机应用与软件，2001；18(6)：61~65
9.耿宏运等.Delphi 6 组件大全[M].北京：电子工业出版社，2002-07