

基于 SVM 的混沌时间序列分析*

赵志宏¹ 杨绍普²

(1. 石家庄铁道学院计算机与信息工程分院, 石家庄 050043) (2. 石家庄铁道学院院办, 石家庄 050043)

摘要 支持向量机是一种基于统计学习理论的新的机器学习方法, 该方法已用于解决模式分类问题. 本文将支持向量机(SVM)用于混沌时间序列分析, 实验数据采用典型地 Mackey - Glass 混沌时间序列, 先对混沌时间序列进行支持向量回归实验; 然后采用局域法多步预报模型, 利用支持向量机对混沌时间序列进行预测. 仿真实验表明, 利用支持向量机可以较准确地预测混沌时间序列的变化趋势.

关键词 时间序列分析, 混沌, 支持向量机

引言

混沌时间序列(或混沌信号)是指对一个混沌系统进行观测采样而得到的一个单变量时间序列^[1]. 混沌时间序列的预测可看作动力学系统研究的“反问题”. “正问题”是给定非线性动力学系统, 研究其相空间中轨道的各种性质. “反问题”是给定相空间中的一串迭代序列(轨道的演化过程)或一组观测序列, 要构造一个非线性映射来表达原系统, 这个映射就可作为预测模型. 因此, 如何构造预测模型是混沌时间序列预测中的一个关键问题, 对混沌时间序列进行建模和预测已成为当前混沌信号处理研究领域的研究热点^[2,3,4].

基于结构化风险最小化方法的统计学习理论是一种专门的小样本统计理论, 它为研究有限样本情况下的统计模式识别, 并为更广泛的机器学习问题建立了一个较好的理论框架, 同时也发展了一种新的模式识别方法 - 支持向量机(Support Vector Machine, 简称 SVM)^[5]. 统计学习理论和支持向量机方法对有限样本情况下模式识别中的一些根本性问题进行了系统的理论研究, 并且在此基础上建立了一种较好的通用算法. 以往困扰很多机器学习方法的问题, 比如模型选择与过学习问题、非线性和维数灾难问题、局部极小问题等, 在这里都得到了很大程度上的解决. 因此, 统计学习理论和支持向量机是机器学习领域的一个重要分支, 已经得到了日益广泛的重视^[6].

本文将支持向量机用于混沌时间序列分析, 分别用支持向量机进行混沌时间序列的回归与预测, 并分析了它们的性能. 仿真结果表明, 支持向量机是一种混沌时间序列分析的工具.

1 支持向量机概述

V. Vapnik 提出的支持向量机理论^[5]因其坚实的理论基础和诸多良好特性在近年获得了广泛的关注. 已经有许多事实证明, 作为支持向量机最基本思想之一的结构化风险最小化原则(Structural Risk Minimization, SRM)要优于传统的经验风险最小化原则(Empirical Risk Minimization, ERM). 不同于 ERM 试图最小化训练集上的误差的做法, SRM 试图最小化 VC 维的上界, 从而使其学习机获得了更好的推广性能, 这恰恰是统计学习理论最重要的目标之一. 支持向量机的主要应用领域有模式识别、函数逼近和概率密度估计等.

支持向量机首先通过用内积函数定义的非线性变换将输入空间变换到一个高维空间, 在这个空间中求(广义)最优分类面. SVM 分类函数形式上类似于一个神经网络, 输出是中间节点的线性组合, 每个中间节点对应一个支持向量.

SVM 是从线性可分情况下的最优分类面发展而来的, 基本思想可用图 1 的两维情况说明.

图 1 中, 实心点和空心点代表两类样本, H 为分类线, H_1, H_2 分别为过各类中离分类线最近的样本且平行于分类线的直线, 它们之间的距离叫做分

类间隔 (margin). 所谓最优分类线就是要求分类线不但能将两类正确分开 (训练错误率为 0), 而且使分类间隔最大. 分类线方程为

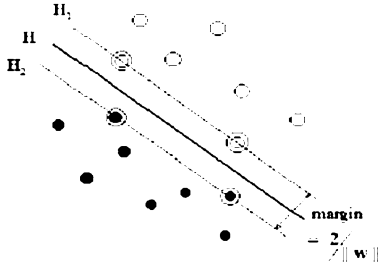


图 1 最优分类面

Fig. 1 optimal hyperplane

$$(w \cdot x) + b = 0 \quad (1)$$

可以对它进行归一化, 使得对线性可分的样本集 $(x_1, y_1), \dots, (x_n, y_n), x_i \in R^d, y_i \in \{-1, +1\}$ 满足:

$$y_i [(w \cdot x_i) + b] \geq 1 \quad (2)$$

此时分类间隔等于 $2/\|w\|$, 使间隔最大等价于使 $\|w\|^2$ 最小. 满足条件(2)且使 $\frac{1}{2}\|w\|^2$ 最小的分类面就叫做最优分类面, H_1, H_2 上的训练样本点就称作支持向量.

2 支持向量回归

设给定的训练样本为:

$$\{(x_1, y_1), \dots, (x_k, y_k)\} \subset R^n \times R$$

首先用一个非线性映射 ϕ 把数据映射到一个高维特征空间, 然后在高维特征空间中进行线性回归^[7], 设回归函数为:

$$f(x) = \langle w, \phi(x) \rangle + b \quad (3)$$

优化问题是最小化

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i + \xi_i^* \quad (4)$$

约束为:

$$f(x_i) - y_i \leq \xi_i^* + \varepsilon, \quad i = 1, \dots, k$$

$$y_i - f(x_i) \leq \xi_i + \varepsilon, \quad i = 1, \dots, k$$

$$\xi_i^*, \xi_i \geq 0, \quad i = 1, \dots, k$$

式(4)中第一项使函数更为平坦, 从而提高泛化能力, 第二项则为减小误差, 常数 C 对两者做出折中.

这是一个二次优化问题, 其对偶问题为:

$$\max -\frac{1}{2} \sum_{i,j=1}^k (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \phi(x_i), \phi(x_j) \rangle + \sum_{i=1}^k (\alpha_i - \alpha_i^*) y_i - \sum_{j=1}^k (\alpha_j + \alpha_j^*) \varepsilon$$

约束为:

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

$$0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = 1, \dots, k$$

解这个二次优化, 可以得到 α 的值, w 的表达式为:

$$w = \sum_{i=1}^k (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x_j) \rangle + b$$

于是回归函数 $f(x)$ 的表达式为:

$$f(x) = \sum_{i=1}^k (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x_j) \rangle + b$$

可以看到, 在上面的优化中需要计算高维特征空间中的内积运算, 如果找一个核函数 $K(x, y)$ 代替高维空间中的内积, 就可以避免复杂的高维计算问题. 已经证明 $\langle \Phi(x), \Phi(y) \rangle$, 对称函数 $K(x, y)$ 只要满足 Mercer 条件即可满足要求, 这样就避免了明确知道 $\Phi(x)$, 从而巧妙地解决了在高维空间中的运算.

3 基于 SVM 的混沌时间序列分析

用 SVM 对混沌时间序列 Mackey - Glass 方程产生的数据进行回归与预测.

Mackey - Glass 方程定义如式(5)所示:

$$\frac{dx}{dt} = \frac{0.2x(t-\tau)}{1+x(t-\tau)^{10}} - 0.1x(t) \quad (5)$$

当 $\tau \geq 17$ 时, 产生的数据是混沌的. 该方程的离散数据可从网上下载获得, 从中取 1200 个数据, 所得数据如图 2 所示. 从图 2 中可以看到, 此混沌数据的变化比较剧烈, 要想准确地进行回归与预测比较困难.

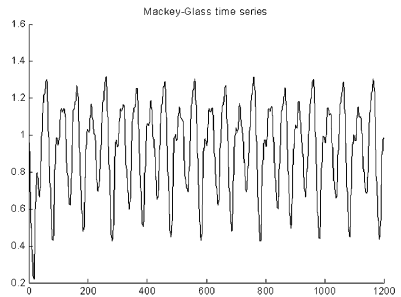


图 2 Mackey - Glass 时间序列的前 1200 个数据

Fig. 2 The first 1200 data of Mackey - Glass time series

3.1 基于 SVM 的混沌时间序列回归

首先利用支持向量机对 Mackey - Glass 时间序列进行回归实验, 核函数使用高斯核函数, C 取 10.

用 SVM 对 Mackey - Glass 混沌时间序列的前 1200 个数据进行回归的结果如图 3 所示。

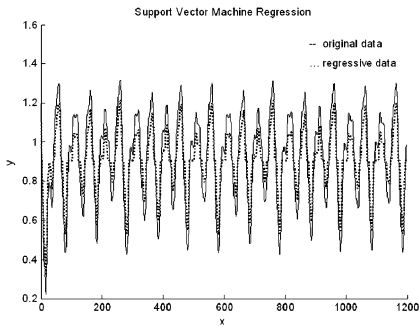


图 3 对 1200 个数据进行回归的结果

Fig. 3 Regression result of 1200 data

从图 3 中可以看到,利用 SVM 进行混沌时间序列的回归,尽管在峰值处的差别比较大,但回归结果的变化趋势与真实数据基本吻合。

3.2 基于 SVM 的混沌时间序列预测

根据相空间重构理论利用 SVM 对 Mackey - Glass 方程混沌时间序列进行预测. 预测模型采用局部法多步预报模型 (Local - Region Multi - steps Forecasting Model). 核函数使用高斯核函数。

局部法多步预报模型是基于相空间重构理论对混沌时间序列进行预测的一个模型. 相空间重构理论认为系统任意分量的演化是由与之相互作用的分量决定,因此可以从某一分量的一批时间序列数据中提取和恢复出系统原来的规律。

取前 400 个数据作为训练集,后 20 个数据作为测试集,实验结果如图 4 所示. 取前 800 个数据作为训练集,后 20 个数据作为测试集,实验结果如图 5 所示。

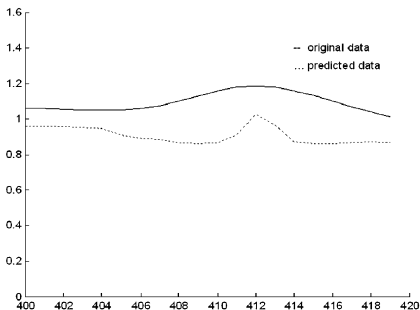


图 4 取前 400 个数据进行训练,后 20 个数据进行测试的结果

Fig. 4 Training of 400 data, and testing of the following 20 data

从仿真结果可以看到,利用 SVM 进行预测,虽然预测的结果有误差,但是预测的数据的变化趋势与原混沌数据较一致。

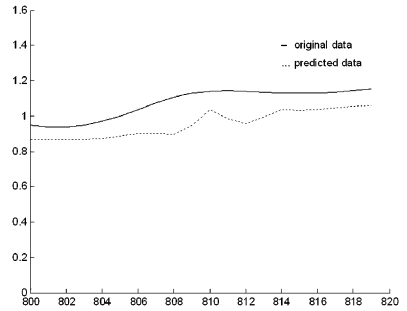


图 5 取前 800 个数据进行训练,后 20 个数据进行测试的结果

Fig. 5 Training of 800 data, and testing of the following 20 data

4 结论

利用支持向量机对 Mackey - Glass 混沌时间序列进行了回归与预测实验,从实验结果可以看到,支持向量机可以较准确地预测出 Mackey - Glass 混沌时间序列的变化趋势,为了更准确地利用支持向量机进行混沌时间序列的分析,还需要做更多地工作。

参 考 文 献

- 1 王永生,范洪达,尚崇伟,刘振. 混沌时间序列的神经网络预测研究. 海军航空工程学院学报,2008,23(1): 21 ~ 25,32 (Wang Yongsheng, Fan Hongda, Shang Chongwei, Liu Zhen. Chaotic Time Series Forecasting Based on Neural Networks. *Journal of Naval Aeronautical Engineering Institute*, 2008, 23(1): 21 ~ 25, 32 (in Chinese))
- 2 Farmer J D, Sidorowich J J. Predicting chaotic time series. *Phys Rev Lett*, 1987, 59: 845 ~ 848
- 3 Maguire L Roche B, Mcgininity T M. Predicting a chaotic time series using a fuzzy neural network. *Information Sciences*, 1998, 112: 125 ~ 136
- 4 Kugiumtzis D, Lingjarde O C, Christophersen N. Regularized local linear prediction of chaotic time series. *Physica D*, 1998, 112: 344 ~ 360
- 5 V Vapnik. *The Nature of Statistical Learning Theory*. Springer - Verlag, New York, 1995
- 6 蒋丰,冯奇,汪玉. 基于支持向量机的舰艇抗冲击生命力分析. 动力学与控制学报. 2004, 2(2): 80 ~ 83 (Jiang Feng, Feng Qi, Wang Yu. Analysis of Anti - Impacted Vital Force of Warships Based on SVM. *Journal of Dynamics and Control*, 2004, 2(2): 80 ~ 83 (in Chinese))
- 7 孙德山,吴今培,肖健华. SVR 在混沌时间序列预测中的应用. 系统仿真学报, 2004, 16(3): 519 ~ 520, 524 (Sun

Deshan, Wu Jinpei, Xiao Jianhua. The application of SVR to prediction of chaotic time series. *Journal of System Simula-*

tion, 2004, 16(3): 519 ~ 520, 524 (in Chinese))

CHAOTIC TIME SERIES ANALYSIS BASED ON SUPPORT VECTOR MACHINE *

Zhao Zhihong¹ Yang Shaopu²

(1. *Computer and information engineering department, Shijiazhuang Railway Institute, Shijiazhuang 050043, china*)

(2. *Shijiazhuang Railway Institute, Shijiazhuang 050043, china*)

Abstract Support vector machine (SVM) is a kind of novel machine learning methods based on statistical learning theory, which has been developed to solve pattern classification problems. This paper applied support vector machine to chaotic time series analysis. The experimental data was Mackey – Glass chaotic time series. First, the support vector regression method was used on the chaotic time series regression problem. Then, Local – Region Multi – steps Forecasting Model was used with support vector machine to predict the chaotic time series. Simulation results show that SVM could predict the trend of chaotic time series correctly.

Key words time series analysis, chaos, support vector machine