DOI:10.16182/j.cnki.joss.2006.07.014

支持向量机的时间序列回归与预测

董 辉，傅鹤林，冷伍明
（中南大学土木建筑学院，长沙 410075）

摘 要：详细分析了支持向量机用于时间序列预测的理论基础。**采用支持向量机、*RBF* 和 *Elman* 神经网络模型，对仿真时序和工程滑坡变形时序进行了回归与外延预测**。结果表明，在噪声水平较低时， *SVR* 回归效果稍好， *Elman* 与 *RBF* 网络的稳健性相对较差；随着噪声水平增大，两种神经网络的回归精度迅速下降。对于外延预测，两种神经网络仅限于短期的非线性模拟，而泛化性能更好的 *SVR* 在短期具有比较理想的效果，在较长的时间区间里也具有较高的预测精度(7 步预测准确度控制在 83.5%以上)。

关键词：支持向量机；回归； *Elman* 网络；滑坡变形
中图分类号：TP391.9 **文献标识码：**A **文章编号：**1004-731X (2006) 07-1785-04

Support Vector Machines For Time Series Regression and Prediction

DONG Hui, FU He-lin, LENG Wu-ming
(Civil Architectual Engineering College, Central South University, Changsha 410075, China)

Abstract: A method for predicting time series based on support vector machines was proposed. *The time series, including simulated data and landslide deformation data sets, were preformed for regression and prediction by support vector machine, RBF networks, and elman recurrent neural networks. A comparison of these three methods was made based on their predicting ability.* The results show that: when noise level is lower in simulated experiment, support vector machine is perfect relatively, and the Elman and RBF network are of more instability, on the other hand, with the higher noise levels, the greater relative error of two networks models is made. For landslide data sets prediction, the neural networks are limited to predict short term nonlinear time series in terms of their accuracy, whereas support vector machine has a higher precision in the short term and long term.

Keyword: support vector machine; regression; elman recurrent network; landslide deformation

引 言

随着现代数理力学理论和计算机技术的迅速发展，基于人工智能领域中遗传算法和人工神经网络的时间序列预测方法在实践中取得了较好的效果[1]。但是这些方法本身却存在着难以克服的缺陷，在学习样本数量有限时，精度难以保证；样本数量很多时，泛化性能又不高。如何找到一种在有限样本情况下，精度既高同时泛化性能也强的机器学习算法便显得很迫切。支持向量机(Support Vector Machines，SVM)是一种以结构风险最小化原理为基础的新算法，具有其它以经验风险最小化原理为基础的算法难以比拟的优越性，同时由于它是一个凸二次优化问题，能够保证得到的极值解是全局最优解[6]。本文基于这种新的算法，对比 *RBF* 与 *Elman* 两种神经网络算法进行仿真数据回归和工程数据预测研究。

1 支持向量机回归算法(SVR)

支持向量机是 Vapnik 等人根据统计学理论提出的一种新的通用学习方法，它是建立在统计学理论的 *VC* 维理论和结构风险最小化原理基础上，能较好地解决小样本、非线性、高维数和局部极小点等实际问题，被视为替代人工神经网络的较好算法。

支持向量机回归算法：给定 k 个样本数据，其值表示为：$\{x_k, y_k\}$，式中 $x_k \in R^n$ 的 n 维向量，$y_k \in R$ 为相对应的输出变量，回归算法的基本思想是通过一个非线性映射 φ，将数据集映射到高维特征空间 H，并在这个空间进行线性回归。具体的函数形式可表示为：

$$f(x) = (\omega, \varphi(x_k)) + b\ ,\quad \varphi : R^n \to H\ ,\quad \omega \in R^n$$

b 为偏置量。这样，在高维特征空间的线性回归便对应于低维空间的非线性回归，且免去了在高维空间 ω 和 φ 的点积计算。函数回归问题等价于使如下泛函最小：

$$R_{reg}[f] = R_{emp}[f] + 0.5\|\omega\|^2 = \sum_{i=1}^{k} C(e_k) + 0.5\|\omega\|^2$$

式中 $e_k = f(x_k) - y_k$，$C(\cdot)$ 是损失函数。损失函数有多种形式，考虑到线性 ε 不敏感损失函数具有较好的稀疏特性，可保证结果的泛化性，选取损失函数为：

$$|y - f(x)|_\varepsilon = \max\{0, |y - f(x)| - \varepsilon\}$$

取经验风险 $R_{emp}^{\varepsilon}[f] = k^{-1}\sum_{i=1}^{k}|y - f(x)|_\varepsilon$，则求解上式等价于求解如下的优化问题：

$$\min J = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{k}(\xi_i^* + \xi_i)$$
$$s.t.\begin{cases} y_i - (\omega, \varphi(x)) - b \le \varepsilon + \xi_i^* \\ (\omega, \varphi(x)) + b - y_i \le \varepsilon + \xi_i \\ \xi_i^*, \xi_i \ge 0 \end{cases} \qquad (1)$$

收稿日期：2005-05-08 **修回日期：**2005-07-19
基金项目：贵州省交通厅建设科技项目“西部地区公路地质灾害监测预报技术研究”(200331880201)
作者简介：董辉(1976-)，男，湖南安乡人，博士，研究方向为 GIS 与智能岩土信息技术。

式中 C 为一正常数，是函数回归模型的复杂度和样本拟合精度之间的折衷，值越大，拟合程度越高；ε 是回归允许的最大误差，控制支持向量的个数和泛化能力，其值越大，支持向量越少。利用对偶原理，同时引入拉格朗日乘子和核函数，将(1)转化为：

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

$$s.t.\begin{cases}\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)=0\\ 0\le\alpha_i,\alpha_i^*\le C\end{cases} \tag{2}$$

此时，$\omega=\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)\varphi(x_i)$

解上述凸二次规划问题得到非线性映射表示：

$$f(x)=\sum_{i=1}^{k}(\alpha_i-\alpha_i^*)K(x_i,x)+b \tag{3}$$

由于支持向量机理论只考虑高维特征空间的点积运算 $K(x_i,x_j)=\varphi(x_i)\cdot\varphi(x_j)$，而不是直接使用函数 φ，从而巧妙地解决了因映射函数 φ 的未知而 ω 无法显式表达的问题。称满足 Mercer 条件的对称函数 $K(x_i,x_j)$ 为核函数，常用的核函数有：

(1) 多项式核函数 $K(x_i,x_j)=(x_i\cdot x_j+1)^d$，$d=1,2,\cdots$；

(2) Sigmoid 核函数 $K(x_i,x_j)=\tanh[b(x_i\cdot x_j)+c]$；

(3) 高斯径向基函数核函数 $K(x_i,x_j)=\exp(-\|x_i-x_j\|^2/2p^2)$。

按照 *KKT* (karush-kuhn-tucker)定理，推导方程

$$\begin{cases}\varepsilon-y_i+f(x_i)=0 & \alpha_i\in(0,C)\\ \varepsilon+y_i-f(x_i)=0 & \alpha_i^*\in(0,C)\end{cases}$$

可求解出偏置 b 值。

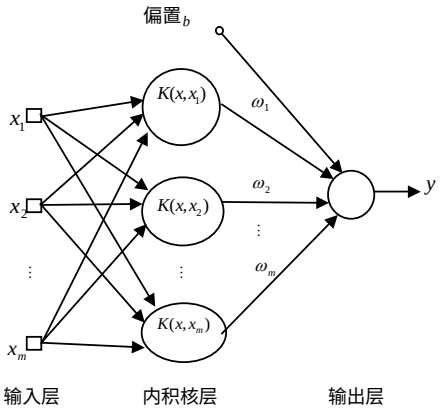

图 1 支持向量机的体系结构

2 算例

为了更好的说明支持向量机的优良特性，本文分别采用仿真数据与现实工程数据进行试验。同时，为了对比试验效果，对两组数据也用神经网络方法进行了回归预测。神经网络方法选用径向基函数(*RBF*)前向型神经网络和 *Elman* 反馈型神经网络。*RBF* 是一种前向型神经网络，它具有结构自适应确定，输出与初始权值无关的优良特性。*Elman* 型神经网络在前向型网络的基础上增加了一个承接层，它主要用来记忆隐含层单元前一时刻的输出值，可以认为是一个一步延时算子。这种神经网络的特点是隐含层的输出通过承接层的延迟与存储，自联到隐含层的输入，这种自联方式使其对历史状态的数据具有敏感性，内部反馈网络的加入增加了网络本身处理动态信息的能力，从而达到了动态建模的目的。

2.1 仿真试验

数据来源　仿真数据集 $\{x_k,y_k\},k=1,2,\cdots$,输入 x 为均匀格网数据，范围在[-10,10]间，输出 y 值通过函数 $\sin c=\sin(x)/x+\delta$ 生成，共 41 个数据，其中 δ 为零均值，方差 σ 的高斯白噪声。表 1 是三种方法相关优化参数。本文使用的 *SVR* 损失函数为 ε 不敏感函数，主要是对 C 和 ε 参数进行优化，具体参数选择方法参考文献[4-5]。核函数选择径向基函数，对于输入为单变量时序，函数的宽度 $p\sim(0.1-0.5)*range(x)$；多变量 d 维时序，$p^d\sim(0.1-0.5)$。*Elman* 的隐含神经元数目需要经过反复试验得出。

表 1 预测方法参数优化设置

	高斯噪声(零均值，方差 $\sigma=0.1$)		
	SVR	Elman	RBF
隐层神经元	—	9	自适应
C	25	—	—
ε	0.1	—	—
核函数(p)	0.03125	—	—
分布密度	—	—	6
隐层 TF	—	sigmoid	radbas
输出层 TF	—	linear	linear

结果分析图 2 给出了试验的回归结果，从图中可以看出，*SVR* 模型除第一个波谷误差较大外，其他回归点与原函数拟合较好，尤其表现在波形的首尾处的拟合要明显好于后两种方法，*RBF* 网络回归在首尾的误差最大，这主要是由于首尾两个样本点的权值训练存在偏差造成，此外，两个波谷处的误差也是三种方法中最大的。*Elman* 反馈网络回归，在两个波谷处的拟合精度最高，但同样存在首尾误差较大的缺点，而在最后一个波峰处，拟合精度已不能接受。表 2 给出三种方法的均方差和相对误差对比。本文对不同水平的噪声下回归也做了试验，发现当无噪声或低噪时，*SVR* 与 *Elman* 网络相差不大，后者有时甚至优于前者；当噪声水平较高时，*SVR* 仍能保持较好的精度要求，而 *RBF* 与 *Elman* 的回归精度降低较快，这是由于后两种方法的稳健性相对较差的缘故。

表 2 误差对比

	MSE	相对误差 $_{max}$	相对误差 $_{min}$
SVR	0.047385	4.79061	0.000368
RBF	0.055463	8.67426	0.006852
Elman	0.055147	10.7993	0.002854

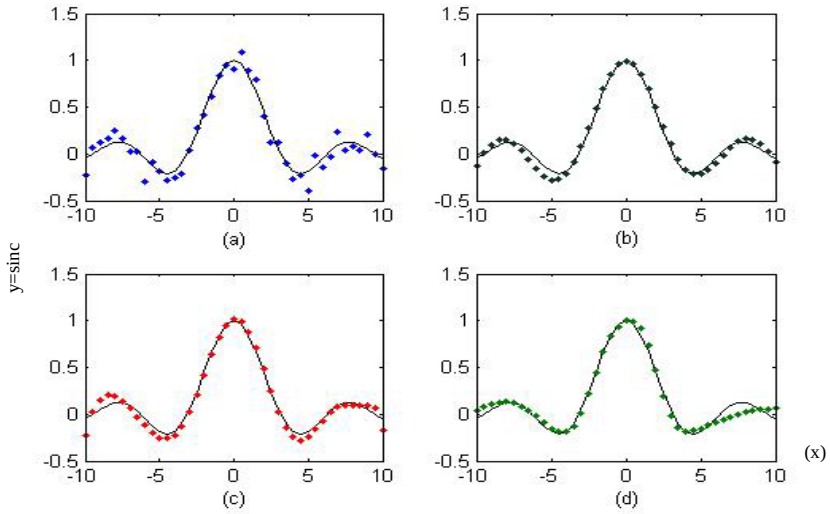

图 2　(a)真值与带高斯白噪声(0 均值,方差 $\sigma = 0.1$)的原始数据集; (b)SVR 方法回归;
(c)RBF 前向神经网络回归; (d)Elman 反馈神经网络回归。实线为真值，点为回归点集

2.2 工程算例

近年来，随着社会经济的发展和西部大开发，大型工程建设项目日渐增多，在水电工程、矿山开采、高速公路、铁路建设等领域，滑坡灾害发生频繁、强度增大、日趋严重，每年都要造成重大的经济损失和人员伤亡，其危害性已严重影响到工程的安全建设和经济效益。

滑坡是一个受地质条件、地下水、地震和人类工程活动等多种因素影响而发展演化的多维非线性动力系统。对滑坡这种地质灾害进行定量预测一直是滑坡研究的核心内容，然而由于诱发因素的随机性和不可控制性，使对滑坡作出准确可靠的预报显得十分困难。位移作为滑坡变形破坏的重要反馈信息，同时关联着其他难以测定的影响滑坡的因素。由于其实测数据在工程实际中较易获得，目前，滑坡的时间预报多数方法都是基于这种反映滑坡动态变形的实测位移时间序列，通过对观察到的数据序列建立恰当的动态模型来预测位移变形趋势，从而及时评估滑坡体结构稳定状态的变化情况。

本文选取卧龙寺新滑坡 5 号裂缝的变形时间序列[1]作为本文算例的原始数据。将数据分为两部分，其中前 25～59 时步的 35 个滑坡位移监测值，用来训练学习机器，进行相关参数的估计，其余的 60～66 时步的 7 个数据用来验证预测模型的有效性。

2.2.1 时序相空间重构

为了降低建模误差，对原始数据首先进行均值零处理以及数据的归一化，然后根据 *Takens* 理论进行相空间重构，即将一维的时间序列转化成矩阵形式以获得数据间的关联关系，从而挖掘到尽可能多的信息量。为了使重构的相空间能较充分而细致的反映系统运动特征，恰当的选取嵌入维 m 和延迟时间 τ 的大小是相空间重构的关键。这里取延滞时间间隔 τ 为位移观测间隔 Δt，并取 Δt 为单位 1(即 $\Delta t = 1$ 天)。建立滑动时间窗口(自相关输入) $x_t = (x_{t-1}, x_{t-2}, \cdots x_{t-m})$ 与输出 $y_t = \{x_t\}$ 之间的映射关系 $f: R^m \to R$。对于嵌入维 m 的确定，本文采用最终误差预报准则评价模型的预测误差，以当误差最小时来确定 m。经过计算确定 m 值取 5 最为合理，这样原始的一维时间序列变形后得到用于预测学习的样本 $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_n)$。

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n-m} & x_{n-m+1} & \cdots & x_{n-1} \end{bmatrix}, \quad Y = \begin{bmatrix} x_{m+} \\ x_{m+2} \\ \vdots \\ x_n \end{bmatrix}$$

2.2.2 滑坡预测模型

规范数据样本后，即可进行基于支持向量机的学习训练，其回归函数可表示如下：

$$y_t = \sum_{i=1}^{n-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i \cdot \bar{x}_t) + b, \quad t = m+1, \cdots, n$$

注意到 $\bar{x}_{n-m+1} = \{x_{n-m+1}, x_{n-m+2}, \cdots, x_n\}$ 没有利用，故可得到第 $n+1$ 点的预测值：

$$y_{n+1} = \sum_{i=1}^{n-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i \cdot \bar{x}_{n-m+1}) + b$$

得到 $\bar{x}_{n-m+1} = \{x_{n-m+1}, x_{n-m+2}, \cdots, x_n\}$ 后，可进一步得到一个样本数据：

$$\bar{x}_{n-m+2} = \{x_{n-m+2}, x_{n-m+3}, \cdots, x_n, \hat{x}_{n+1}\}$$

其中 $\hat{x}_{n+1}$ 是第 $n+1$ 点的预测值，如此递推，得到 l 步的支持向量机预测模型：

$$y_{n+l} = \sum_{i=1}^{n-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i \cdot \bar{x}_{n-m+l}) + b$$

2.2.3　预测结果

根据预测模型，逐步预测滑坡的最后 7 天变形数据，

这里每预测一步后，将增加的预测值添加到训练样本中，同时保持总训练样本数不变。支持向量机预测模型选取径向基函数(*RBF*)作为核函数，内部参数 p、惩罚常数 C 和 ε 通过交叉验证或格网搜索优化确定，也可用穷举法在 *Matlab* 中编程估算。*RBF* 前向神经网络采用 6×1 的输入与输出层结构，隐含层神经元采用自适应调节。*Elman* 反馈神经网络中间层神经元数目为 35(反复试算)。三种方法计算预测结果如表 3 和图 3(其中 y 为实测值)所示。

表 3 样本预测对照表

时段序号	实测值	SVR		RBF		Elman	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
60	30.0	29.83	0.57	29.92	0.27	30.07	0.23
61	31.0	30.59	1.32	30.95	0.16	31.36	1.16
62	32.0	32.70	2.18	31.58	1.31	32.18	0.56
63	33.0	34.73	5.24	34.24	3.76	34.46	4.42
64	42.0	40.35	3.93	36.67	12.7	36.63	12.8
65	47.0	44.57	5.17	36.98	21.3	39.37	16.2
66	61.0	50.96	16.5	46.51	23.8	42.99	29.5

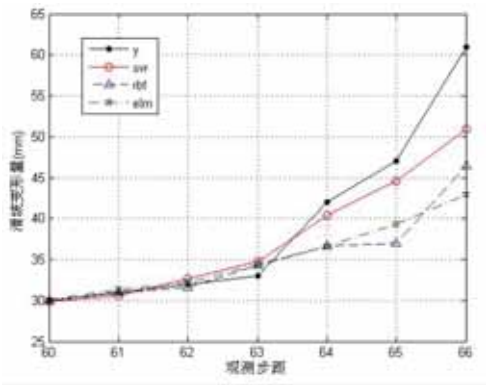

图 3 svr ,rbf ,elman 方法预测滑坡变形对照图

2.2.4 数据分析

从图表中可以看出，对滑坡变形时间序列数据的外推预测，支持向量机前 6 步预测结果的相对误差控制在 6%内，性能相当不错，仅在第 7 步误差较大，这是由于滑坡已处于临滑突变阶段，数据不再具有指导性，但即便如此，83.5%的准确性仍能满足工程要求。由此可见，*SVR* 预测方法在短期具有非常理想的效果，在较长的时间区间里(7 步)也具有较高的预测精度。*RBP* 与 *Elman* 神经网络在前 4 步预测中也表现出较好的效果，相对误差比 *SVR* 还小，但 4 步后的预测误差则下降较快，准确性在整体上不如 *SVR*。注意到 *RBF* 第 6 步预测相对误差较大，这是由于网络稳健性相对较差和样本数据较少造成的。相比 *RBF*，*Elman* 网络具有较好的稳健性，但它在预测时序某段有上升幅值时，预测值却没有相应增大(如图 2(d)、图 3)，这可能是由于网络隐含层加入了时序内部间的相关性信息的原因。同时，*Elman* 网络的隐含层神经元数目需要经过反复大量的试算，且随机赋予的初始权值也极大的影响了网络的性能。

通过对本次滑坡位移的外推预测，依据变形速率的变化情况，重点观察其速率较大的时步点(图 3 中的第 5、6 时步)，并且在结合其他滑坡判别方式(如专家系统)的同时，对滑坡现状进行理论评估，即卧龙新寺滑坡在第 6 时步，已达临滑突变阶段，应及时进行滑坡预警及安全防护措施。滑坡实际过程印证了评估的正确性。

总之，根据滑坡岩体结构位移变化的历史监测值，预测未来演化规律，可以及时了解岩体结构的稳定状态以及进行滑坡稳定性控制，从而达到减灾防灾的目的。

3 结论

(1) 本文对三种方法进行了仿真回归试验和工程滑坡变形预测。对于仿真回归，当噪声水平较低或无噪时，*SVR* 回归效果稍好，但有时 *Elman* 网络更好于 *SVR*，*RBF* 网络的稳健性则相对较差；随着噪声水平增大，两种神经网络的回归精度也迅速下降。对于滑坡变形外延预测，当样本数较少时，两种神经网络仅限于短期的非线性模拟(4 步)，而 *SVR* 方法在短期或较长的时间区间里(7 步)同样具有较高的预测精度。

(2) 支持向量机回归(*SVR*)是以结构风险最小化为基础的统计方法，它解决了神经网络方法中易陷入局部最小值、精度与泛化不可调和的矛盾。*SVR* 与神经网络在中、短期预测中相差不大，但在长期预测中，前者更为稳健。

(3) *RBP* 和 *Elman* 两种神经网络算法是建立在经验风险最小理论上，只有当样本数得到保证，给出合理的初始权值与隐含层的神经元数目时，网络对非线性行为才能较好的模拟。

(4) 支持向量机通过寻找有代表性的关键点(支持向量)，并以少的支持向量代替原始数据样本建立预测模型，使模型的外延预测能力增强，但算法的缺点是计算时间较长，相关的改进算法可参考文献[2]。

参考文献：

[1] 冯夏庭. 智能岩石力学导论[M]. 北京: 科学出版社, 2000.
[2] A J Smola, B Scholkopf. A tutorial on support vector regression[D]. RoyalHolloway College, University of London, UK, 1998.
[3] U Thissen, R van Brakel, A P de Weijer, et al. Using support vector machines for time series prediction[J]. Chemometrics and Intelligent Laboratory Systems(S0899-7667), 2003, 69: 35-49.
[4] Cherkassky V, Ma Y. Comparison of model selection for regression[J]. Neural Computation(S0169-7439), 2003, 15(7): 1691-1714.
[5] Smola A, Mtirata N, Scholkopf B, Muller K. Asymptotically optimal choice of ε-loss for support vector machines[C]// proceedings of ICANN, 1998.
[6] V Vapnik. The Nature of Satistical Learning Theory[M]. Springer Verlag, 1995.
[7] V Vapnik. Satistical Learning Theory [M]. Wiley, 1998.