

基于小波分析以及 ϵ -SVR 的隧道围岩位移时序预测

马家超^{1,2}, 朱珍德^{1,2}, 朱 姝¹, 王 熙^{1,2}

(1. 河海大学 岩土工程科学研究所, 江苏 南京 210098;

2. 河海大学 岩土力学与堤坝工程教育部重点实验室, 江苏 南京 210098)

摘要: 根据龙滩地下洞室围岩现场长期的观测数据, 利用小波变换阈值去噪法对数据信号的高频系数进行量化处理, 恢复原有的监测信号。实例结果表明, 小波变换阈值去噪法能够有效地去除噪声的影响, 获得围岩变形的真实信号。同时将去噪后的数据用于支持向量机(ϵ -SVR)建立时序分析模型中, 并与 GA-BP 模型进行对比, 结果表明 ϵ -SVR 模型的误差更小, 预测效果更佳。

关键词: 小波去噪; 位移预测; ϵ -SVR; 遗传算法; BP 神经网络

中图分类号: U456.3+2

文献标识码: A

文章编号: 1672-1144(2013)02-0090-05

Prediction for Tunnel Surrounding Rock's Displacement Time Series Based on Wavelet Analysis and ϵ -SVR

MA Jia-chao^{1,2}, ZHU Zhen-de^{1,2}, ZHU Shu¹, WANG Xi^{1,2}

(1. Geotechnical Engineering Research Institute of Hohai University, Nanjing, Jiangsu 210098, China;

2. Key Laboratory of Ministry of Education for Geomechanics and Embankment Engineering, Hohai University, Nanjing, Jiangsu 210098, China)

Abstract: Based on the monitoring data of Longtan underground cavern, the proposed method is used to quantify the high frequency coefficients of data signal by means of the wavelet transform so as to restore the originating monitor signal. The example analysis shows that the wavelet transform could effectively remove the influence of noise and achieve the real signal of the surrounding rock. At the same time, the denoised data are used for the timing analysis model established by ϵ -SVR, which is compared with GA-BP model. The results show that ϵ -SVR model has less error and better predicting effect.

Keywords: wavelet denoising; displacement prediction; support vector regression(ϵ -SVR); genetic algorithm; BP neural network

0 引言

隧道围岩在开挖过程中的位移、变形不是瞬间完成的, 而是随时间不断增长。因此, 根据现场监测的围岩位移变化数据, 预测其发展趋势, 从而达到优化设计、施工和运行的目的, 对现存的相关规范也有重要的意义。而现场监测数据常常受到施工条件、人工活动、气候条件、测量仪器等随机因素的影响, 从而导致时序曲线表现出不同程度的波动和震荡,

因而可以把它看成是被有色噪声污染而形成的不平稳的信号^[1]。这些受影响的信号在图像上往往表现为突变或者尖峰, 而且都位于信号的高频部分。正交小波变换具有很强的去数据相关性, 很容易分离出噪声和其他不需要的信息, 因此先利用正交小波变换将信号进行分解, 得到信号的高频和低频系数, 再利用阈值法去除高频系数中噪声的部分, 最后再使用最大尺度的低频和处理过的高频系数一并进行重构即可得到去噪后的信号(数据)。

收稿日期: 2012-12-13

修稿日期: 2013-01-11

基金项目: 国家重点基础研究发展计划(973)资助项目(2011CB013504); 浙江省重大科技计划重点项目(2009C33049); 江苏省 2011 年度研究生培养科技创新计划项目(CXZZ11-0428)

作者简介: 马家超(1987—), 男, 江苏仪征人, 硕士研究生, 研究方向为岩土工程反分析。

选择一个相对精确的预测模型则成了另一个重要的必备条件。人工神经网络(ANN)^[2]和遗传算法(GA)^[3]曾被广泛应用到岩土工程中来。但人工神经网络是基于启发式的,自身存在一定的缺陷,缺乏严格的理论基础,在学习样本有限时,难以保证精度,而学习样本过多时,又会陷入“维数灾难”,泛化能力不高,不能很好的控制训练后的网络推广,有时甚至不能达到全局最优。而支持向量机(support vector machines, SVM)有严格的理论基础,是一种基于结构风险最小化原则的机器学习技术,具有其他以经验风险最小化原则为基础的算法难以比拟的优越性。由于 SVM 算法以统计学理论为基础,能够很好地解决小样本、高维数、非线性等一些列传统算法不可回避的问题,同时它是一个凸二次优化问题,能够保证极值点为全局最优解^[4-6]。鉴于岩土工程反分析问题的特点,本文采用非线性支持向量回归机(support vector regression, ϵ -SVR),对龙滩地下洞室群隧道位移监测数据进行了机器学习、回归和预测,并与 GA-BP 网络模型预测结果进行了对比。

1 小波去噪原理

1.1 离散序列小波分解与重构

利用小波变换处理实际工程中经采样系统得到的离散信号,一般并不使用尺度函数或小波函数来运算,而是直接利用双通道滤波器组进行^[7-8]。一维离散小波变换一般采用 mallat 算法,即先对较大尺度的信号进行小波变换,在选择其中的低频部分在原来尺度的 1/2 尺度上再进行小波变换。Mallat 算法公式为:

$$C_k^1 = \sum_m h_0(m-2k)C_m^0 \quad D_k^1 = \sum_m h_1(m-2k)C_m^0 \quad (1)$$

式中: h_0, h_1 为双通道滤波器组; C_k^1, D_k^1 分别为第一层分解的低频和低频系数。

采用相同滤波器对第一层分解得到的低频系数 C_k^1 再次做相同的分解,得到 C_k^2, D_k^2 ,以此类推可得到多层小波分解系数。

重构是分解的逆过程,即先对 $j+1$ 层上的系数 C_k^{j+1}, D_k^{j+1} 进行二插值,从而恢复二抽取前的序列长度,在通过低频和高频重构滤波器恢复出上一层的低频信号 C_k^j ,把这个过程继续下去,直到得到原始离散信号序列。重构公式为:

$$C_m^j = \sum_k h_0(m-2k)C_k^{j+1} + \sum_k h_1(m-2k)D_k^{j+1} \quad (2)$$

1.2 小波软阈值去噪

正交小波具有很强的去数据相关性能使携带信息的原始信号能量集中在小波域的绝对值较大的系数中,而噪声信号能量相对分散,其系数绝对值较小。基于这一原理,采用阈值的方法,去掉绝对值小于一定阈值的小波系数,保留原始信号系数,从而达到去噪的目的。

关于阈值量化处理,Donoho 提出了两种方法:硬阈值法和软阈值法。

(1) 硬阈值法

当小波系数的绝对值小于阈值时,令其为 0;否则不变。

$$\hat{w}_{j,k} = \begin{cases} w_{j,k} & |w_{j,k}| \geq \lambda \\ 0 & |w_{j,k}| < \lambda \end{cases} \quad (3)$$

(2) 软阈值法

当小波系数绝对值小于阈值时,令其为 0;否则减去阈值。

$$\hat{w}_{j,k} = \begin{cases} \text{sign}(w_{j,k})(|w_{j,k}| - \lambda) & |w_{j,k}| \geq \lambda \\ 0 & |w_{j,k}| < \lambda \end{cases} \quad (4)$$

以上两式中, $w_{j,k}, \hat{w}_{j,k}$ 分别为量化前后的小波系数, $w_{j,k}$ 即指各层的高频系数 $D_{k,j}, j=1,2,\dots,J, J$ 为最大分解层数。 λ 为选定的阈值,其选取方法较多,本文中取 $\lambda = \sigma\sqrt{2\log N}$,其中 N 为离散序列长度, σ 为噪声的标准方差,可利用第一层分解的小波系数来估计^[3],即 $\sigma = \text{median}(|D_{1,k}|)/0.6745$, $\text{median}(\cdot)$ 表示求中值。

硬阈值函数在 $\hat{w}_{j,k} = \lambda$ 处是不连续的,容易造成去噪后的图像在起点附近出现明显的 Pseudo-Gibbs 现象^[9]。故本文使用软阈值法。

2 非线性 ϵ -SVR 原理^[10]

给定训练样本 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}, x_i \in R^l, y_i \in R$, 训练一个线性回归模型,线性函数为:

$$y = w \cdot x + b \quad (5)$$

式中: w 为权向量, b 为偏差。假定所有训练样本在 ϵ 精度下无误差地用线性函数拟合,即

$$|y_i - wx_i - b| \leq \epsilon \quad (i = 1, 2, \dots, n) \quad (6)$$

考虑到允许拟合误差的情况,引入松弛因子 $\zeta_i \geq 0, \zeta_i^* \geq 0$,则优化目标变为最小化函数:

$$R(w, \zeta, \zeta^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*)$$

$$\left. \begin{aligned} y_i - wx_i - b &\leq \varepsilon + \zeta_i \\ wx_i + b - y_i &\leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* &\geq 0 \end{aligned} \right\} \quad (7)$$

其中常数 $C > 0$, C 表示对超出误差 ε 的惩罚程度。

引入 Lagrange 函数 $L(w, b, \zeta_1, \zeta_i^*, \alpha_i, \alpha_i^*, \eta_i, \eta_i^*)$, 可得上述优化问题的对偶问题

$$\begin{aligned} \max W(\alpha, \alpha^*) &= -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i x_j) \\ &+ \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \end{aligned} \quad (8)$$

s. t:

$$\left. \begin{aligned} \sum_{i=1}^n (\alpha_i - \alpha_i^*) &= 0 \\ 0 &\leq \alpha_i, \alpha_i^* \leq C_0 \end{aligned} \right\} \quad (9)$$

式中: α_i, α_i^* 为带求解的 l 维支持向量, 从而得到支持向量拟合函数为

$$f(x) = w \cdot x + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (x_i x) + b \quad (10)$$

对于非线性函数回归问题, 通过非线性映射 $x \rightarrow \phi(x)$ 把数据映射到一个高维的特征空间, 然后在该高维特征空间进行线性回归。因此, 非线性问题转化为

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i x) + b \quad (11)$$

式中: $K(x_i x) = \phi(x_i) \phi(x)$ 为核函数, 常用的核函数有多项式核函数、径向基核函数、Sigmoid 核函数等。考虑到隧道问题的较为复杂, 影响因素较多, 因此, 本文选用了非线性核函数即径向基核函数:

$$K(x, z) = \exp(-\|x - z\|^2 / 2\sigma^2) \quad (12)$$

3 工程实例

3.1 工程概况

龙滩水电工程是我国“西电东送”的重点项目之一, 位于广西天峨县境内, 距离天峨县 15 km, 坝址控制流域面积 98 500 km², 占该河总流域面积的 71%。龙滩水电站坝址河谷为较宽坦的“V”型谷, 宽高比约 3.5。河流流向 S30°E, 至坝址处转向 S80°E。枯水期河水面高程 219 m, 水面宽度为 90 m ~ 100 m, 水深为 13 m ~ 19.5 m, 河床两岸均有基岩礁滩裸露。左岸地形较整齐, 右岸受冲沟切割, 地形完整性略逊于左岸, 两岸山顶高程约 600 m, 岸坡坡度 32° ~ 42°, 残坡积层厚 0.5 m ~ 2 m, 局部 8 m ~ 25 m。坝址地层为三叠系下统罗楼组 (T₁₁) 和三叠系中

统板纳组 (T_{2b})。

由于主厂房 C 断面范围主要是三叠系下版纳组 T_{2b}²⁹⁻³⁷ 层泥板岩为主的砂岩、板岩互层岩体, 层间错动较发育, 同时受 F₁、F₁₂ 等陡倾角断层切割, 围岩的类别主要是 III₁ 类, 还有少量的 IV 类围岩。由于断层的切割, 严重的破坏了层状岩体的连续性。泥板岩中密集发育的劈理, 使得软硬相间的层状结构岩体完整性变差。随着开挖下降, 特定块体变形呈下滑趋势, 同时也牵动上部岩体变形, 剪切变形较大。因此选用 C 断面多点位移计 Mc⁴-4 (桩号: HL0+051.625, 位置: 上游墙, 高程: 234.300) 测得的时间序列来进行预测分析。多点位移计 Mc⁴-4 沿水平孔径方向 4 个锚头深度分别为: 2 m、6 m、12 m、25 m。(C 断面多点位移计布置如图 1 所示)。

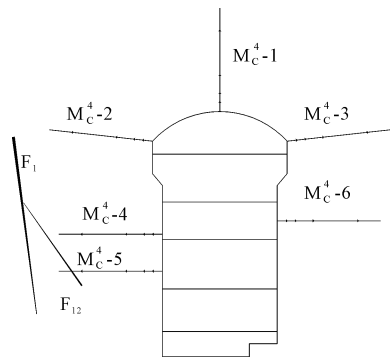


图 1 C 断面多点位移计布置图

3.2 数据处理

时序分析要求等间隔监测, 龙滩水电站监测频率大约为 7 天一次。由于现场监测数据在一定程度上受到外界的干扰, 为了得到更加精确的预测数据, 先用 db4 小波对实测数据进行 3 层分解, 得到各层的低频和高频系数, 再使用软阈值法对各层的高频系数进行去噪, 最后使用最大尺度上的低频和处理后的高频系数进行重构, 得到最终去噪后的数据。经去噪后的时序曲线 (如图 2) 比去噪前更光滑, 波动更小, 去噪结果比较理想。

3.3 样本构造与参数设置

将实测数据前 40 组数据作为学习样本, 后 6 组数据用来预测和检验。前 40 组数据中, 顺次将每 8 个数据作为输入值, 第 9 个作为输出值来训练网络。GA-BP 网络训练结束后, 为了避免数据跳跃, 提高预测精度, 用 33 ~ 40 组数据预测第 41 个数据, 并将其作为下一组预测数据, 以此类推直到预测出第 46 个数据为止。

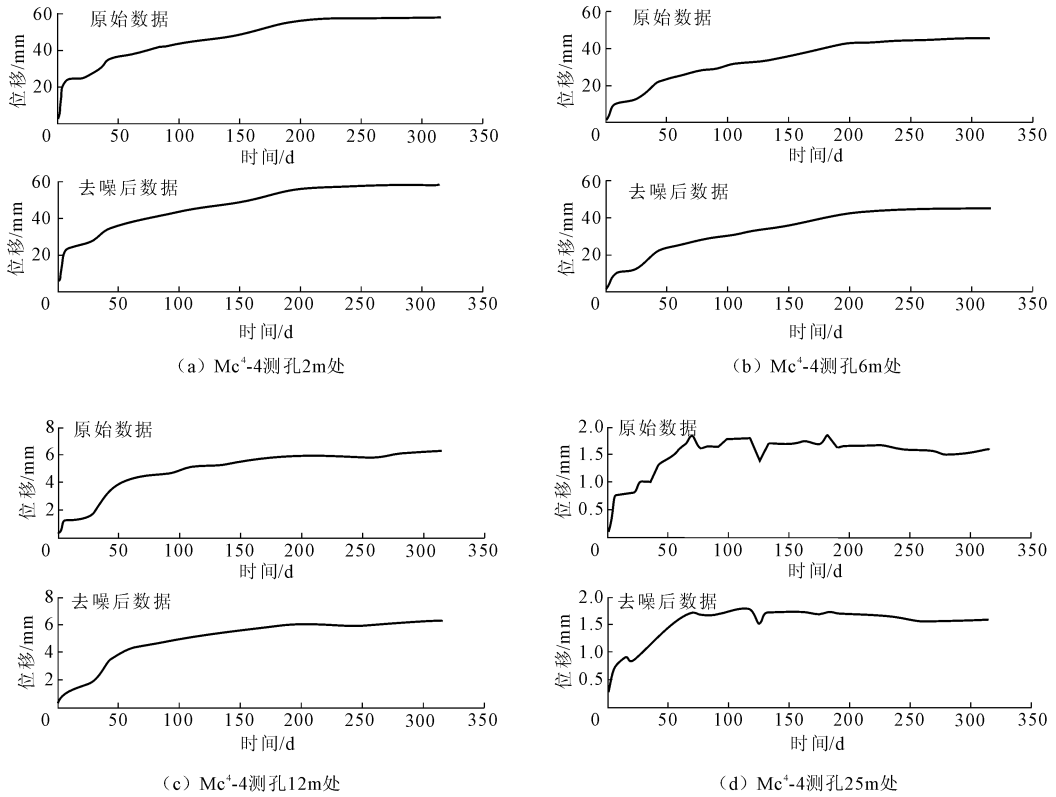


图 2 实测位移时序数据去噪前后对比

ϵ -SVR 计算模型选用径向基核函数,待定参数为 $C, \gamma = 1/2\sigma^2, \epsilon$,其中 ϵ 控制拟合函数误差的大小,从而控制支持向量个数和泛化能力,本文取 $\epsilon = 0.01$ 较为合适,而参数 C, γ 则选用留一法获得^[11-12]。以 1# 数据为例,首先将 γ 值固定,改变 C 值,比较其对预测数据平均相对误差的影响(如图 3);再将 C 值固定,改变 γ 值,并比较其对预测数据平均相对误差的影响(如图 4)。最终各点处参数值如表 1 所示。

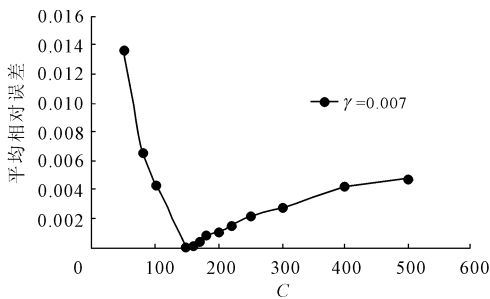


图 3 C 值对平均相对误差的影响

3.4 结果对比

将去噪后的 GA-BP 模型预测结果与 ϵ -SVR 模型的预测结果比较(表 2)。结果表明, ϵ -SVR 模型的预测效果基本好于 GA-BP 的预测效果,说明在隧道围岩位移时序分析中是 ϵ -SVR 模型的泛化

能力要略优于 GA-BP 网络模型,而经过去噪后数据的预测又有了更高的精度,完全满足工程要求,两者的结合具有很强的实践性。

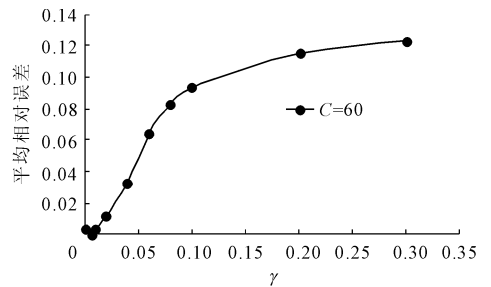


图 4 γ 值对平均相对误差的影响

表 1 各点参数值

参数	1 #	2 #	3 #	4 #
C	150	1400	700	50
γ	0.007	0.003	0.0008	0.008

4 结论

(1) 将小波变换应用于龙滩地下洞室群围岩监测数据中,利用双通道滤波器组对监测数据进行三层小波分解,再利用软阈值法对分解得到的高频系数进行去噪,最后降低频系数和处理已有的高频系数一起进行重构得到去噪后的曲线,结果表明软阈

表 2 位移时序预测结果对比表

点号	时间 /d	GA - BP		ϵ - SVR	
		残差 /mm	百分比 /%	残差 /mm	百分比 /%
1 #	280	-0.0225	-0.04	-0.0141	0.02
	287	-0.0629	-0.11	-0.0470	0.08
	294	-0.0574	-0.10	-0.0538	0.09
	301	-0.0151	-0.03	-0.0178	0.03
	308	-0.0027	-0.05	0.0571	0.10
	315	0.0045	0.01	0.1774	0.31
2 #	280	0.1224	0.28	0.0565	0.13
	287	0.2963	0.67	0.1345	0.30
	294	0.3835	0.86	0.1411	0.32
	301	0.4041	0.90	0.0892	0.20
	308	0.4275	0.96	0.0315	0.07
	315	0.4124	0.92	-0.0533	0.12
3 #	280	0.0426	0.71	0.0647	1.08
	287	0.1415	2.33	0.1471	2.42
	294	0.2128	3.47	0.2020	3.29
	301	0.2347	3.82	0.2084	3.39
	308	0.2415	3.92	0.2159	3.51
	315	0.2302	3.74	0.2124	3.45
4 #	280	0.0116	0.76	-0.0171	1.13
	287	0.0202	1.32	-0.0143	0.94
	294	0.0247	1.61	-0.0121	0.79
	301	-0.0255	1.66	-0.0088	0.57
	308	0.0088	0.56	0.0000	0.00
	315	0.0297	1.89	0.0137	0.87

值去噪使得去噪后的曲线较原始曲线更为光滑、失真小,达到了去噪效果。

(2) 基于结构风险最小化原则的 ϵ - SVR 模型的泛化能力优于基于风险最小化原则的 GA - BP 神经网络模型,同时在隧道围岩位移预测方面前者也

有着不错的表现,因此,在位移时序分析中有一定的推广价值。

参考文献:

- [1] 梁桂兰,徐卫亚,淡小龙,等.岩石高边坡安全监测数据的小波变换去噪处理[J].岩石力学与工程学报,2008,27(9):1837-1844.
- [2] 张代远.神经网络新理论与方法[M].北京:清华大学出版社,2006.
- [3] 邓建辉,李焯芬,葛修润.BP网络和遗传算法在岩石边坡位移反分析中的应用[J].岩石力学与工程学报,2001,20(1):1-5.
- [4] 赵洪波,冯夏庭.位移反分析的进化支持向量机研究[J].岩石力学与工程学报,2003,22(10):1618-1622.
- [5] 黄亚东,张土乔,俞亭超,等.公路软基沉降预测的支持向量机模型[J].岩土力学,2005,26(12):1987-1990.
- [6] 刘开云,乔春生,滕文彦.边坡位移非线性时间序列采用支持向量机算法的智能建模与预测研究[J].岩土工程学报,2004,26(1):57-61.
- [7] 熊智新,胡慕伊,陈朝霞,等.离散小波变换算法剖析及其通用程序实现[J].计算机工程与设计,2007,28(20):4856-4859.
- [8] 姜越川.公路边坡变形监测数据的小波阈值去噪方法[J].铁道科学与工程学报,2011,8(4):46-49.
- [9] 魏明波.基于正交小波软阈值的图像去噪算法[J].内江科技,2010,31(5):25-26.
- [10] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer - Verlag, 1995.
- [11] Xie Z J, Wong H, Ip W C, et al. Artificial neural network and its application to financial forecasting[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2001, 37(3): 421-425.
- [12] Gold C, Sollich P. Model selection for support vector machine classification[J]. Neurocomputing, 2003, 55(1/2): 221-249.