

非线性位移时间序列预测的 进化-支持向量机方法及应用

Study and application of genetic-support vector machine for
nonlinear displacement time series forecasting

赵洪波¹, 冯夏庭²

(1. 绍兴文理学院土木系, 浙江 绍兴 312000; 2. 中国科学院 武汉岩土力学研究所岩土力学重点实验室, 湖北 武汉 430071)

摘要: 提出了一种新的岩土结构位移预测的进化支持向量机方法, 用遗传算法来搜索支持向量机的参数和核函数, 避免了人为选择参数的盲目性, 同时提高了支持向量机的推广预测能力; 利用这种非线性智能预测方法, 滚动预测施工位移变形量, 以便及时调整和优化施工步序, 维护岩土结构的稳定性。将该方法用于卧龙寺边坡变形、三峡永船闸边坡变形预测, 结果表明, 该方法具有科学可靠、实时性的优点, 具有广泛的应用前景。

关键词: 遗传算法; 支持向量机; 边坡; 位移; 预测

中图分类号: TU 45

文献标识码: A

文章编号: 1000-4548(2003)04-0468-04

作者简介: 赵洪波(1971-), 男, 河北邢台人, 中国科学院武汉岩土力学研究所博士研究生, 从事智能岩石力学与工程的研究工作。

ZHAO Hong-bo¹, FENG Xia-ting²

(1. Department of Civil Engineering, Shaoxing College of Arts and Sciences, Shaoxing 312000, China; 2. Institute of Rock and Soil Mechanics, The Chinese Academy of Science, Wuhan 430071, China)

Abstract: The genetic-support vector machine method was used to forecast the nonlinear displacement in geotechnical engineering. It is showed this method can escape from the blindness of man-made choice of the parameter and kernel function of support vector machine. It enhances the efficiency and the capability of forecasting. The engineering case studies indicate that it is scientific and there is an extensive prospect for this real time forecasting.

Key words: genetic algorithm; support vector machine; slope; displacement; forecast

0 前言*

位移是岩土结构演化过程中反馈出的重要信息之一。用监测的位移进行建模可以对岩土结构的未来演化规律、发展趋势等进行预测, 及时掌握岩土结构的变化规律, 在工程上具有重要的意义^[1]。目前用于位移时间序列的主要方法是传统的时间序列分析方法和智能岩石力学的神经网络方法, 传统的时间序列分析方法是利用观测的历史位移建立预测的统计模型, 这种模型无需知道岩土结构中的力学过程和机理, 具有一定的优点; 但是这种统计模型一般当因变量和自变量之间是线性关系或一些简单的函数关系时才使用, 同时当数据太少时不具有统计意义; 智能岩石力学的神经网络方法也是基于大样本的一种学习方法。由于岩土结构的复杂性, 所涉及的工程地质条件及岩体特性参数通常是不确定的、模糊的, 影响岩体系统特性的各要素之间存在着非常复杂的非线性关系, 很难用确定的数学模型描述, 这使传统的方法存在不足。最近发展起来的支持向量机方法是基于统计学习理论的新学习方法, 用它建模不必知道因变量和自变量之间的关系, 通过对样本的学习即可获得因变量和自变量之间

非常复杂的映射关系; 同时, 它是基于小样本的一种学习方法, 不必知道太多的数据即可建模。由于支持向量机的推广预测能力很大程度上, 依赖于支持向量机的核函数和参数, 因此他们的合理确定是至关重要的。遗传算法是一种模拟生物界自然进化过程的优化方法, 具有全局最优性、并行性等优点^[2]。本文将遗传算法和支持向量机进行有机结合, 提出了一种用于岩土工程位移预测预报的智能岩石力学新方法—进化支持向量机方法。

1 非线性位移时间序列预测的进化-支持向量机方法

1.1 非线性位移时间序列的支持向量机表达

支持向量机^[3,4] (Support Vector Machine) 是基于统计学习理论的一种新的通用学习方法, 它是建立在一套较好的有限样本下机器学习的理论框架和通用方法之下, 它既有严格的理论基础, 又能较好地解决小样

* 基金项目: 国家自然科学基金资助项目(50179034); 中国科学院
知识创新工程重要方向资助项目(KJCX2-SW-L1-3)

收稿日期: 2002-10-16

本、非线性、高维数和局部极小点等实际问题,其核心思想就是学习机器要与有限的训练样本相适应。支持向量机中的支持向量是通过解一个凸二次优化问题获得的,它保证找到的解是全局最优解,在支持向量机算法中解二次优化的方法主要有:SMO方法^[5]、内点算法^[4]、子集选择方法等。

对于一个非线性位移时间序列,通过监测获得其位移随时间变化的一个时间序列 $\{x_i\} = \{x_1, x_2, \dots, x_N\}$,对这个非线性位移序列进行预测,就是要寻找在 $i+p$ 时刻的位移值 x_{i+p} 与前 p 个时刻的位移值 $x_i, x_{i+1}, \dots, x_{i+p-1}$ 的关系,即 $x_{i+p} = f(x_i, x_{i+1}, \dots, x_{i+p-1})$, $f()$ 为一个非线性函数,表示位移时间序列间的非线性关系。

根据支持向量机理论^[3,4],上述的非线性关系可以用支持向量机对 n 个实测位移的学习来获得,也就是通过对 $n-p$ 个位移时间序列 $x_i, x_{i+1}, \dots, x_{i+p-1}, i = 1, \dots, n-p$,的学习,来获得位移时间序列之间的非线性关系

$$f(x_{n+m}) = \sum_{i=1}^{n-p} (\alpha_i - \alpha_i^*) K(X_{n+m}, X_i) + b. \quad (1)$$

式中 $f(x_{n+m})$ 表示第 $n+m$ 时刻的位移值; X_{n+m} 表示 $n+m$ 时刻前 p 个时刻的位移值, $X_{n+m} = (x_{n+m-p}, x_{n+m-p+1}, \dots, x_{n+m-1})$, X_i 表示第 $p+i$ 时刻前 p 个时刻的位移值, $X_i = (x_i, x_{i+1}, \dots, x_{i+p-1})$; $K()$ 是核函数; α, α^* 和 b 是通过解如下的二次规划问题获得的:

$$\begin{aligned} \text{Max: } W(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{i,j=1}^{n-p} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(X_i \cdot X_j) + \sum_{i=1}^k x_{i+p}(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{n-p} (\alpha_i + \alpha_i^*), \\ & \left| \sum_{i=1}^{n-p} (\alpha_i - \alpha_i^*) = 0, \right. \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, n-p. \end{aligned}$$

由上面的论述可以看出,在支持向量机算法中 C 值需要人为的确定,它的选择对预测结果有很大的影响;核函数 $K()$ 对算法的推广预测能力和算法的效率都有很大的影响。通常的方法是采用人工的方法,通过对多个参数的比较来确定较好的参数,这种方法盲目性大、效率低。针对这个问题,本文提出了基于遗传算法的支持向量机方法,采用遗传算法来搜索最佳的支持向量机参数。

1.2 支持向量机参数的进化

支持向量机的参数(核函数和 C)对算法的效率和推广预测能力有很大的影响,他们的选择是建立一个支持向量机模型的重要内容。遗传算法的出现,使其成为可能。遗传算法是模拟生物界自然进化过程的优化方法,具有全局最优性、隐含并行性、算法的高度稳

定性和广泛的可用性等优点。将遗传算法用于搜索支持向量机的核函数和参数 C ,将减少人为选择的盲目性,提高支持向量机的推广预测能力。

本文应用支持向量机对监测到的位移进行建模,以预测未来的位移变化,同时用遗传算法来搜索支持向量机的最优的核函数和参数 C ,其步骤和算法如下。

(1) 对进化支持向量机进行初始化设置,包括设置进化代数、群体规模、核函数和参数 C 的搜索范围、杂交率和变异率等。

(2) 随机产生一组可能的值作为父代,其中每个个体代表一组支持向量机参数(核函数和 C)。

(3) 用每一个个体决定了一个支持向量机模型,用它预测测试样本的位移值,计算每个个体的适应值 $f(x)$,以反映本支持向量机模型的推广预测能力,即

$$f(x) = \min(\max(\frac{|x_i - x_i'|}{x_i'})) \quad (2)$$

式中 x_i 为第 i 个测试样本的位移预测值; x_i' 为第 i 个测试样本的实测值; $i = 1, 2, \dots, n$, n 为测试样本数。

(4) 随机选择两个适应值低于平均适应值的父代个体 i_1, i_2 ,对 i_1, i_2 进行杂交操作,产生两个新的个体。

(5) 按照突变概率随机选择一父代个体进行突变操作,以产生新的个体。

(6) 重复步骤(4)~(5),直到生成一子代群体。

(7) 将父代中最好的个体随机置换子代中一个个体,将子代转化为父代,计算每一个个体的适应值,然后重复步骤(3)~(6),直至得到满意的支持向量机核函数和参数 C (以上算法见图1)。

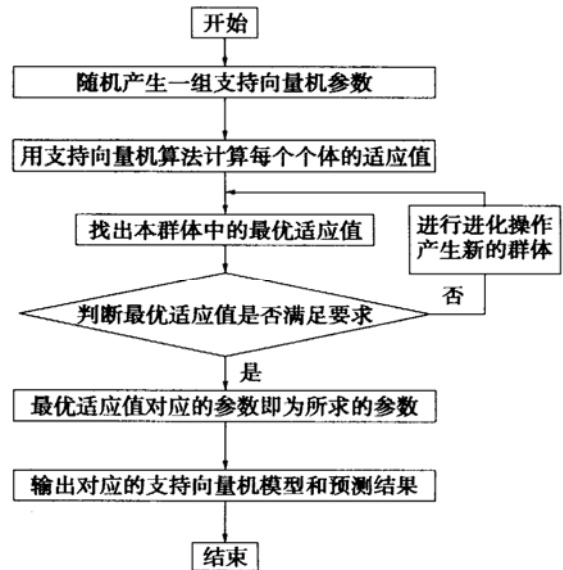


图1 进化-支持向量机算法

Fig.1 The algorithm of genetic support vector machine

1.3 滚动预测方法

在利用进化支持向量机方法进行位移时间序列预

测时,为了充分利用最新的信息,提高预测的准确性,采用了滚动预测的方法。其基本思想是,假设要对时间序列 $\{x_t\}$ 的进行预测,最佳历史点数为 p ,预测的步数为 m (p, m 根据实际问题确定);目前已经获得 n 个时间序列 $\{x_0, x_1, \dots, x_{n-1}\}$,滚动预测的第一步是用 n 个时间序列的 $\{x_i, x_{i+1}, \dots, x_{i+p-1}, x_{i+p}\}$ ($i=0, 1, 2, \dots, n-p-1$) $n-p$ 组时序预测 n 时刻后的 $\{x_n, x_{n+1}, \dots, x_{n+m-1}\}$ m 个时序;随着后面 m 个时序的获得,用 m 个新的时序替代前面的 $\{x_0, x_1, \dots, x_{m-1}\}$ m 个时序进行下一步的预测,得到下一次的 m 个预测值,依次类推。

从以上可以看出,进化支持向量机方法既利用了支持向量机的非线性映射、有限样本和较好的推广预测能力,又利用了遗传算法的全局优化特性;而滚动预测方法每次利用了最新的观测数据。因此,它作为一种新的非线性智能识别方法,具有广泛的应用前景。

2 实例分析

2.1 卧龙寺新滑坡^[6]

卧龙寺新滑坡是一个塬边黄土滑坡,1971 年发现裂缝后于该年 3 月 11 日起开始对其变形进行观测(见表 1),5 月 5 日凌晨 3 时 15 分产生剧滑。本文用支持向量机对监测位移进行滚动预测,后 22 个时步检验支持向量机的预测能力,选择位移序列最佳历史点数为 10。由于支持向量机的核函数和参数 C 直接影响预测效果,所以用遗传算法搜索核函数和参数 C ,搜索空间为:核函数分别为多项式核函数、RBF 核函数和 Sigmoid 核函数, C 为 1~500。在进化计算过程中,采用实数编码方式,遗传操作为部分重组和均匀变异,设置群体规模为 50,实测值和预测值的对比见图 2。由图 2 可以看出,通过支持向量机建立了位移时间序列之间的关系后,再用滚动预测的方法进行预测,除 1971 年 5 月 6 日的位移预测值外均具有很高的精度;虽然在 1971 年 5 月 6 日预测值有很大的突变,但预测精度较差,原因是此时发生了滑坡,位移变化不再符合以前的位移变化规律,这方面工作有待进一步研究。

表 1 卧龙寺滑坡监测数据

Table 1 Measured data of Wolongso slope

日期 /(年-月-日)	位移 /mm	日期 /(年-月-日)	位移 /mm	日期 /(年-月-日)	位移 /mm
1971-03-15	1.0	1971-03-16	1.5	1971-03-17	1.7
1971-03-18	2.5	1971-03-19	3.2	1971-03-20	4.0
1971-03-21	4.4	1971-03-22	5.1	1971-03-23	5.9
1971-03-24	6.3	1971-03-25	7.0	1971-03-26	7.3
1971-03-27	7.8	1971-03-28	8.2	1971-03-29	8.4
1971-03-30	8.7	1971-03-31	9.0	1971-04-01	9.2

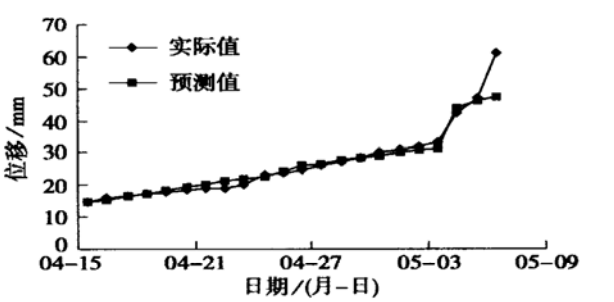


图 2 位移预测值和观测值的关系曲线

Fig. 2 Comparison of measured and predicted displacements

2.2 三峡永久船闸边坡岩体变形预测

长江三峡水利枢纽工程永久船闸位于长江左岸,边坡最高达 170 m,一般坡高为 100~160 m,边坡开挖完成后的长期变形大小直接影响船闸闸门等结构的功能和运行安全。施工中,为了解船闸高边坡在施工和运行期的工作性状,进行了安全监测(见表 2~4)。本文利用进化支持向量机,对三闸首(17-17 断面)的 TP/BM11GP01、TP/BM26GP02 和 TP/BM27GP02 三测点^[7]的变形进行预测,预测值和实测值的对比如图 3~5 所示。

表 2 监测点 TP/BM11GP01 位移观测数据

Table 2 Measured displacement of monitoring point TP/BM11GP01

日期 /(年-月-日)	s /mm	日期 /(年-月-日)	s /mm	日期 /(年-月-日)	s /mm
1996-01-15	0	1996-02-15	-0.19	1996-03-15	0
1996-04-15	0.11	1996-05-15	-0.48	1996-06-15	0.24
1996-07-15	-0.12	1996-08-15	0.95	1996-09-15	2
1996-10-15	1.81	1996-11-15	2.51	1996-12-14	3.35
1997-01-14	4.37	1997-02-15	4.04	1997-03-12	4.2
1997-04-15	4.35	1997-05-16	4.57	1997-06-15	5.61
1997-07-14	5.68	1997-08-16	6.07	1997-09-15	6.44
1997-10-17	7.32	1997-11-18	8.3	1997-12-10	9.18
1998-01-14	9.66	1998-02-14	9.65	1998-03-16	9.09
1998-04-12	10.67	1998-05-16	9.98	1998-06-14	12.1
1998-07-13	13.79	1998-08-10	15.46	1998-09-16	14.22
1998-10-08	15.89	1998-09-16	14.22	1998-11-9	16.56
1998-12-10	18.58	1999-01-08	20.99	1999-02-04	22.03
1999-03-09	23.60	1999-04-09	23.46	1999-05-15	24.72
1999-06-15	25.06	1999-07-15	25.24		

注: s 为位移监测值

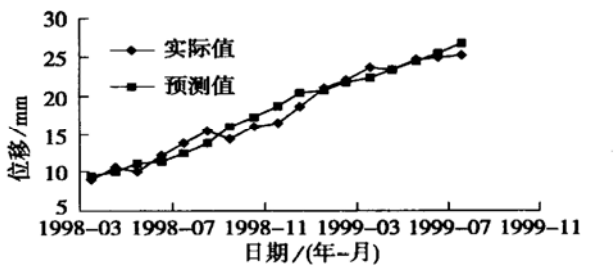


图 3 TPBM11GP01 测点位移预测值和观测值的关系曲线

Fig. 3 Comparison of measured and predicted displacements at monitoring point TP/BM11GP01

表3 监测点TP/BM11GP02 位移观测数据

Table 3 Measured displacement of monitoring point TP/BM11GP02

日期 /(年-月-日)	s /mm	日期 /(年-月-日)	s /mm	日期 /(年-月-日)	s /mm
1996-11-15	0.00	1996-12-14	0.05	1997-01-14	0.79
1997-02-15	0.40	1997-03-12	1.13	1997-04-15	1.69
1997-05-16	1.46	1997-06-15	1.37	1997-07-14	1.94
1997-08-16	2.32	1997-09-15	2.53	1997-10-17	3.72
1997-11-18	4.18	1997-12-10	5.18	1997-01-14	5.71
1998-02-14	6.26	1998-03-16	7.73	1998-04-12	8.43
1998-05-16	7.54	1998-06-14	10.99	1998-07-13	10.16
1998-08-10	12.33	1998-09-16	14.00	1998-10-08	14.45
1998-11-09	16.05	1998-12-10	17.52	1999-01-08	20.11
1999-02-04	19.24	1999-03-09	22.41	1999-04-09	22.45
1999-05-15	23.75	1999-06-15	23.24	1999-07-15	22.95

注: s 为位移监测值



图4 TPBM11GP02 测点位移预测值和观测值的关系曲线

Fig. 4 Comparison of measured and predicted displacements at monitoring point TP/BM11GP02

表4 监测点TP/BM27GP02 位移观测数据

Table 4 Measured displacement of monitoring point TP/BM27GP02

日期 /(年-月-日)	s /mm	日期 /(年-月-日)	s /mm	日期 /(年-月-日)	s /mm
1995-11-15	0.00	1995-12-15	1.07	1996-01-15	1.87
1996-02-15	2.21	1996-03-15	1.74	1996-04-15	1.73
1996-05-15	1.34	1996-06-15	1.25	1996-07-15	1.68
1996-08-15	1.73	1996-09-15	2.20	1996-10-15	2.98
1996-11-15	3.17	1996-12-14	3.59	1997-01-14	4.23
1997-02-15	4.28	1997-03-12	4.87	1997-04-15	5.52
1997-05-16	5.24	1997-06-15	4.73	1997-07-14	4.97
1997-08-16	5.35	1997-09-15	5.48	1997-10-17	6.10
1997-11-18	6.52	1997-12-10	7.49	1998-01-14	8.12
1998-02-14	8.57	1998-03-16	10.1	1998-04-12	10.78
1998-05-16	10.93	1998-06-14	12.82	1998-07-13	12.53
1998-08-10	14.12	1998-09-16	13.64	1998-10-08	15.55
1998-11-09	17.36	1998-12-10	18.53	1999-01-08	19.96
1999-02-04	20.24	1999-03-09	22.78	1999-04-09	22.23
1999-05-15	24.00	1999-06-15	23.25	1999-07-15	23.53

注: s 为位移监测值

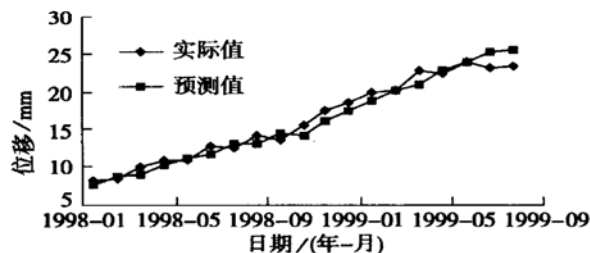


图5 TPBM27GP02 测点位移预测值和观测值的关系曲线

Fig. 5 Comparison of measured and predicted displacements at monitoring point TP/BM27GP02

3 结 论

(1) 采用进化支持向量机方法,可以避免人为选择支持向量机模型的盲目性,提高了支持向量机预测的精度和推广预测能力。

(2) 用进化支持向量机方法,可不断根据新的监测资料对位移进行滚动预测,这种预测具有实时性和较高的精度,为岩土工程的设计和施工提供很好的途径。

(3) 在进行位移时间序列预测时,预测的时步点数对预测的精度具有很大的影响,有必要对其进行优化,这一工作有待于进一步的研究。

(4) 进化支持向量机作为智能岩石力学研究的新方法,由于它的优良特性,不仅适合于非线性位移时间序列预测,还适合于岩土力学与工程中的很多方面,因此基于统计学习理论的支持向量机方法在岩土力学与工程中具有广泛的应用前景。

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. 111-116.
- [2] 刘勇, 康立山, 陈毓屏. 非数值并行算法(第二册)——遗传算法[M]. 北京: 科学出版社, 1997.
- [3] Burge CJC. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, (2): 121-167.
- [4] Alex J Smola, Bernhard Schoelkopf. A Tutorial on Support Vector Regression[R]. NeuroCOLT2 Technical Report Series, 1998.
- [5] John C Platt. Sequential Minimal Optimization: A Fast Algorithm for training Support Vector machines[R]. Technical Report, 1998.
- [6] 黄润秋, 许强. 工程地质广义系统科学分析原理及应用[M]. 北京: 地质出版社, 1997. 42-43.
- [7] 张治强. 岩体力学参数与变形特征的智能识别研究[D]. 沈阳: 东北大学, 2000. 113-119.