

# 遗传算法优化支持向量机在大坝变形预测中的应用

沈哲辉<sup>1</sup> 黄 腾<sup>1</sup> 沈月千<sup>1</sup> 郑 浩<sup>1</sup>

1 河海大学地球科学与工程学院,南京市佛城西路 8 号,211100

**摘 要:** 建立大坝变形预测的支持向量机模型,并用遗传算法对支持向量机模型的核函数参数、惩罚参数和损失参数进行优化。将同一优化方法不同支持向量机核函数、不同优化方法同种支持向量机核函数进行横向对比,将 BP 神经网络、自回归 AR( $p$ )模型、多元回归分析法和周期函数拟合法进行纵向对比。结果表明,该 GA-SVM(RBF)模型不仅能较好地预测大坝的变形趋势,而且能大幅提高预测精度。

**关键词:** 大坝变形因子;支持向量机;遗传算法;优化;预测

**中图分类号:** P258      **文献标识码:** A

本文用遗传算法(genetic algorithm,GA)<sup>[1]</sup>来优化支持向量机(support vector machines,SVM)<sup>[2-3]</sup>的参数,并建立了大坝变形预测的 GA-SVM 模型。通过实例,对比分析了本文算法、BP 神经网络、自回归 AR( $p$ )模型、多元回归分析法和周期函数拟合法。

## 1 支持向量机模型

支持向量机是一种以结构风险最小化原则为基础的机器学习算法,通过解一个线性约束的二次规划问题得到全局最优解,因而不存在局部极小值的问题,保证了收敛速度<sup>[4]</sup>。

本文采用径向基 RBF 核函数(高斯核函数)。通过参数的选择,其可以适用于任意分布的样本,是目前支持向量机中应用最广泛的一种核函数<sup>[5]</sup>。选择合适的损失参数  $\epsilon$ 、惩罚参数  $C$  和核函数参数  $\sigma$  是支持向量机能成功为研究对象建模的关键。

为了验证 RBF 核函数的优势,本文采用多项式核函数建模进行对比。

## 2 基于 GA-SVM 的大坝变形预测模型

### 2.1 建立训练样本集

结合文献[6-7],把水压分量  $H$ (库水位)、温度分量  $T$ (观测时的气温)、周期分量  $t_1$ 、 $t_2$ 、 $t_3$ (为时间  $t$  的函数)、时效分量  $t$ 、 $\ln t$  等 7 个影响因子作为大坝变形的影响因素。

周期分量的表达式为:

$$t_1 = \sin \frac{2\pi t}{365}, t_2 = \cos \frac{2\pi t}{365}, t_3 = \cos^2 \frac{2\pi t}{365} \quad (1)$$

大坝变形的 SVM 模型表达式为:

$$f = x(t, \ln t, t_1, t_2, t_3, T, H) \quad (2)$$

### 2.2 GA 优化 SVM 参数

采用 GA 优化 SVM 的 3 个参数,其优化过程如下。

1)适应度函数的选择。适应度函数是个体对环境的适应能力的表达,其与目标函数有关。本文的目标函数选为均方误差:

$$MSE = \frac{1}{N} \sum_{i=1}^N (R_0 - R_m)^2 \quad (3)$$

式中, $R_0$  为实测值, $R_m$  为拟合值。

2)初始化 GA 参数。初始化最大进化代数、最大种群数量、选择概率、交叉概率和变异概率等。

3)编码。确定支持向量机 3 个参数的编码区间,设置参数  $C$  的寻优区间为  $(0, 100]$ ,设参数  $\epsilon$  的寻优区间为  $[0, 1]$ ,参数  $1/\sigma$  的寻优区间为  $[0, 1\ 000]$ 。对其进行二进制编码,产生一组代表这 3 个参数的染色体,即为初始种群。

4)遗传算子操作。①选择:是指在群体中选择生命力强(适应度大)的个体,产生新的群体的过程。②交叉:是指两队相互配对的染色体按照某种方式进行部分基因座的交换,从而形成两个新的个体,使种群不断有新鲜个体产生,从而对搜索空间中的新区域进行搜索,防止陷入局部最优。

③变异:将个体染色体编码串中的某些基因座上的基因值用改基因座的其他等位基因来替换,从而形成一个新个体。它能保证遗传算法的有效性,决定遗传算法的局部搜索能力,维持群体的多样性,防止出现早熟现象。④重插入:为了确保最适应个体在每一次传播到下一代的数目保持不变,需要随机重插入固定数量的新个体。

5)遗传算法的终止。由于遗传算法的随机搜索性,找到一个明确的收敛性判别准则是困难的。本文算法终止条件采用预先设定的代数。

用GA对训练集优化,得到最优的SVM 3个参数。再根据未来大坝的7个影响因子,训练SVM模型预测未来大坝的变形。

### 3 实例分析

以福建省某梯级水电站第三级大坝的表面水平位移引张线测点15在2008-01-01~2012-12-01期间每月1日的60期监测数据作为历史数据进行建模,预测2013-01-01~2014-08-01的20期数据。

以大坝的长周期历史变形数据进行建模。7个影响因子之一的时效分量 $t$ 序列较长,例如2008-01-01作为第1天,那么2012-12-01编号则为1431。若 $t$ 直接参与建模,其相对于其他6个影响因子的数量级较大,导致模型过分拟合,预测能力下降。所以,对样本数据进行归一化处理<sup>[7]</sup>:

$$\xi' = 2 \times \frac{\xi - \xi_{\min}}{\xi_{\max} - \xi_{\min}} - 1 \tag{4}$$

式中, $\xi$ 为原始样本数据, $\xi_{\min}$ 、 $\xi_{\max}$ 分别为样本集中的最小值、最大值, $\xi'$ 为归一化后的值。

当迭代代数为100时,MSE基本达到收敛,得到最优的惩罚参数 $C=4.9816$ ,损失参数 $\epsilon=0.1747$ ,核函数参数倒数 $1/\sigma=0.0772$ 。

为验证遗传算法具有更好的优化能力,采用网格优化的方法对支持向量机(RBF核)参数进行优化。

将不同支持向量机核函数相同优化方法的对比和相同支持向量机核函数不同优化方法的对比一起并称为横向对比。GA-SVM(多项式核)法得到的最优的惩罚参数 $C=48.6846$ ,损失参数 $\epsilon=0.4253$ ,核函数参数 $\text{coef}=14.7409$ , $q=6.7651$ ;网格-SVM得到的最优惩罚参数 $C=9.1896$ ,损失参数 $\epsilon=0.1482$ ,核函数参数倒数 $1/\sigma=0.0490$ 。

将本文的GA-SVM(BRF)模型与BP神经网络、自回归AR( $p$ )模型(其中AR( $p$ )模型采用

FPE准则定阶,最终定阶为 $p=4$ )、多元回归分析法和周期函数拟合法<sup>[9]</sup>(经计算得 $a=-0.1106$ , $b=-0.0945$ , $c=6.0628$ )进行纵向对比。

横向对比预测曲线见图1。由图1看出,各条SVM模型预测曲线都与原测数据变化一致,但显然使用RBF核函数的SVM预测曲线更加贴合原测数据,说明高斯核函数优于多项式核函数;GA优化下的SVM预测曲线和网格优化下的SVM预测曲线比较接近,但从表1中可以算出,GA-SVM(RBF)预测相对误差绝对值的平均值为2.30%,网格-SVM(RBF)预测相对误差绝对值的平均值为2.36%,所以GA-SVM(RBF)是优于网格-SVM(RBF)的。

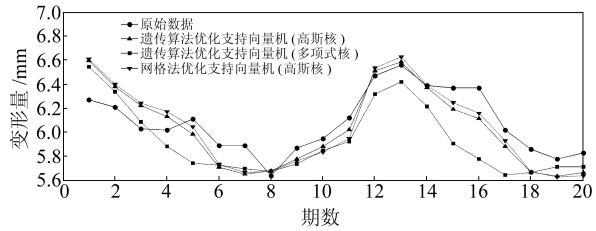


图1 横向对比预测曲线  
Fig.1 Horizontal contrasting prediction curve graph

纵向对比预测曲线见图2。由图2可以看出,GA-SVM(RBF)模型预测结果明显更加贴合原始监测数据,GA-SVM和BP神经网络预测趋势和原测数据基本一致,但AR( $p$ )、多元回归分析法和周期函数拟合法几乎未体现出趋势性。所以对于大坝来说,通过其变形因子对变形量进行建模比用单一变形量建模更加合理。

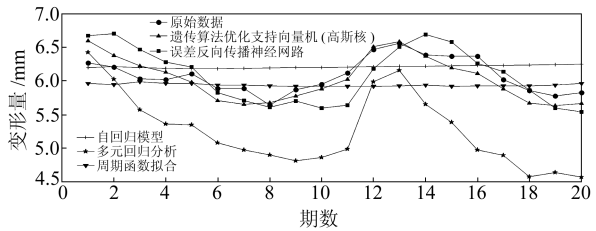


图2 纵向对比预测曲线  
Fig.2 Longitudinal contrasting prediction curve graph

由表1计算出GA-SVM(RBF)预测相对误差绝对值的平均值为2.30%,BP神经网络预测相对误差绝对值的平均值为3.70%,AR( $p$ )预测相对误差绝对值的平均值为4.03%,多元回归分析预测相对误差绝对值的平均值为13.88%,周期函数拟合法预测相对误差绝对值的平均值为3.59%。

综合横向对比和纵向对比,GA-SVM(RBF)模型不仅较好地预测了变形的趋势,而且预测相对误差也最小。可见,GA-SVM(RBF)模型是一种比较优秀的大坝变形数据建模预测方法。

表 1 预测数据对比  
Tab.1 Predicted data contrasting

时间	监测 值 /mm	GA-SVM (RBF)		GA-SVM (多项式核)		网络-SVM (RBF)		BP 神经 网络		多元回 归分析		AR( <i>p</i> )		周期函 数拟合	
		预测 值 /mm	相对误 差 /%	预测 值 /mm	相对误 差 /%	预测 值 /mm	相对误 差 /%	预测 值 /mm	相对误 差 /%	预测 值 /mm	相对误 差 /%	预测 值 /mm	相对误 差 /%	预测 值 /mm	相对误 差 /%
2013-01-01	6.270	6.597	−5.22	6.547	−4.42	6.612	−5.45	6.672	−6.41	6.423	−2.44	6.199	1.13	5.967	4.83
2013-02-01	6.210	6.381	−2.76	6.339	−2.08	6.400	−3.06	6.705	−7.97	6.026	2.96	6.216	−0.10	5.946	4.26
2013-03-01	6.030	6.223	−3.20	6.087	−0.95	6.246	−3.56	6.466	−7.23	5.574	7.57	6.200	−2.82	5.989	0.67
2013-04-01	6.020	6.131	−1.85	5.885	2.25	6.173	−2.55	6.283	−4.37	5.359	10.98	6.189	−2.81	5.964	0.94
2013-05-01	6.110	5.984	2.06	5.746	5.96	6.046	1.05	6.208	−1.61	5.349	12.45	6.181	−1.17	5.963	2.41
2013-06-01	5.890	5.711	3.04	5.728	2.74	5.736	2.62	5.829	1.04	5.080	13.75	6.183	−4.97	5.942	−0.88
2013-07-01	5.890	5.651	4.05	5.698	3.27	5.666	3.80	5.707	3.10	4.978	15.48	6.187	−5.04	5.941	−0.87
2013-08-01	5.640	5.679	−0.69	5.675	−0.63	5.678	−0.67	5.607	0.58	4.904	13.05	6.193	−9.81	5.927	−5.08
2013-09-01	5.870	5.780	1.53	5.736	2.28	5.761	1.86	5.701	2.88	4.813	17.80	6.199	−5.60	5.919	−0.83
2013-10-01	5.950	5.884	1.11	5.851	1.66	5.835	1.93	5.598	5.93	4.862	18.29	6.204	−4.27	5.919	0.53
2013-11-01	6.120	6.025	1.56	5.923	3.21	5.947	2.83	5.644	7.79	4.988	18.49	6.208	−1.44	5.918	3.30
2013-12-01	6.470	6.512	−0.65	6.318	2.35	6.536	−1.02	6.182	4.45	5.992	7.39	6.213	3.98	5.918	8.53
2014-01-01	6.560	6.585	−0.39	6.419	2.15	6.627	−1.02	6.508	0.79	6.155	6.17	6.217	5.23	5.925	9.68
2014-02-01	6.390	6.372	0.29	6.217	2.70	6.391	−0.05	6.694	−4.75	5.655	11.50	6.221	2.64	5.938	7.07
2014-03-01	6.370	6.192	2.80	5.907	7.27	6.248	1.92	6.583	−3.34	5.389	15.40	6.226	2.27	5.919	7.09
2014-04-01	6.370	6.116	3.98	5.777	9.31	6.157	3.33	6.263	1.68	4.978	21.85	6.230	2.20	5.926	6.97
2014-05-01	6.020	5.885	2.24	5.644	6.24	5.930	1.50	6.138	−1.95	4.894	18.71	6.234	−3.56	5.927	1.55
2014-06-01	5.860	5.671	3.22	5.665	3.32	5.673	3.19	5.863	−0.05	4.575	21.93	6.239	−6.47	5.927	−1.15
2014-07-01	5.780	5.635	2.51	5.714	1.13	5.628	2.62	5.596	3.18	4.639	19.74	6.243	−8.02	5.942	−2.81
2014-08-01	5.830	5.665	2.83	5.714	1.99	5.640	3.26	5.539	4.99	4.574	21.54	6.248	−7.17	5.963	−2.28

4 结 语

1)根据大坝变形影响因子与大坝变形之间的关系,用支持向量机对历史数据进行建模,并用具有较好的全局搜索能力以及良好的可扩展性等优点的遗传算法优化支持向量机参数,得到最优的大坝预测支持向量机模型。此优化方法避免了凭经验或试算法选取支持向量机的三参数,从而引起欠学习或过学习的现象。

2)在大坝变形预测领域,支持向量机采用的RBF核函数明显优于多项式核函数。

3)在大坝变形预测领域,采用遗传算法对支持向量机的三参数进行全局寻优,比传统的网格寻优有优势。

4)对于大坝长周期数据进行建模分析,对数据进行归一化是极其必要的。

5)对于大坝来说,通过其变形因子对变形量进行建模比用单一变形量建模更加合理。

6)对比BP神经网络、自回归AR(*p*)模型、多元回归分析法和周期函数拟合法,本文方法不仅更好地预测了变形趋势,而且能大幅提高预测精度。

参考文献

[1] 雷英杰. MATLAB 遗传算法工具箱及应用[M]. 西安:西安电子科技大学出版社,2014(Lei Yingjie. MATLAB Genetic Algorithm Toolbox and Application[M]. Xi'an: Xidi-

an University Press, 2014)  
[2] 吕开云. 黄河小浪底水利枢纽大坝变形预测方法研究与分析[D]. 北京:中国矿业大学,2012(Lü Kaiyun. Research and Analysis about the Deformation Forecasting Methods of Yellow River Xiaolangdi Water Hydropower Dam[D]. Beijing: China University of Mining and Technology, 2012)  
[3] 翟信德. 基于支持向量机的大坝变形预测研究[D]. 合肥:合肥工业大学,2007(Zhai Xinde. Research on Dam Displacement Forecasting Model Based on Support Vector Machine[D]. Hefei: Hefei University of Technology, 2007)  
[4] 唐亮贵,程代杰. 基于小波的支持向量机预测模型及应用[J]. 计算机科学,2006,33(3):202-204(Tang Lianggui, Cheng Daijie. A Forecasting Model via Support Vector Machines Based on Wavelets[J]. Computer Science, 2006, 33(3):202-204)  
[5] 姜慧研,王晓丹,周晓杰,等. 基于SVR的回转窑烧成带温度软测量方法的研究[J]. 系统仿真学报,2008,20(11):2 951-2 955(Jiang Huiyan, Wang Xiaodan, Zhou Xiaojie, et al. Study on Soft Sensor for Temperature of Burning Zone Based on SVR[J]. Journal of System Simulation, 2008, 20(11): 2 951-2 955)  
[6] 吴中如. 水工建筑物安全监控理论及其应用[M]. 南京:河海大学出版,1990(Wu Zhongru. Safety Monitoring Theory & Its Application of Hydraulic Structures[M]. Nanjing: Hohai University Press, 1990)  
[7] 陈伟. 遗传算法与神经网络在大坝安全监测中的应用研究[D]. 西安:长安大学,2009(Chen Wei. Genetic Algorithms and Artificial Neural Networks Apply in Dam Security Monitoring[D]. Xi'an: Chang'an University, 2009)