

基于 PCA-SVR 模型的大坝裂缝早期预报研究

王 泉^{1,2,3}, 郑东健^{1,2,3}, 范振东^{1,2,3}, 冯树荣⁴, 肖 峰⁴

(1. 河海大学 水利水电学院, 江苏 南京 210098; 2. 河海大学 水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098; 3. 河海大学 水资源高效利用与工程安全国家工程研究中心, 江苏 南京 210098; 4. 中国电建集团 中南勘测设计研究院有限公司, 湖南 长沙 410014)

摘要:支持向量机拟合模型(SVR)可用于大坝裂缝预报,但自变量间的多重相关性和输入变量的高维数对支持向量机拟合模型的精度影响较大。基于大坝裂缝开合度理论,利用主成份分析法(PCA)提取原样本信息缩减后的主成分作为 SVR 模型的输入量,构建了大坝裂缝开合度的早期预报 PCA-SVR 模型。将该模型应用于某大坝监控资料的分析中,与传统回归模型相比,PCA-SVR 模型具有更高的计算精度和运算效率,并可提前预报裂缝开合度信息,能在实际工程中广泛应用。

关 键 词:裂缝开合度; PCA; SVR; 非线性; 预报模型

中图法分类号: TV698

文献标志码: A

DOI:10.16232/j.cnki.1001-4179.2015.05.007

目前,对大坝监测资料分析方法众多。传统的多元线性回归、逐步回归、稳健回归等方法应用广泛,但当自变量因子较多且存在多重相关性时,传统回归模型的拟合预测效果较差,往往不能满足工程要求。神经网络方法近年来发展迅速,尽管其非线性逼近精度高,但往往只能达到局部极小值。时间序列法不能考虑各自变量因素的影响,同时对监测数据具有较高的要求。支持向量机拟合法(SVR)能在有限样本的情况下获得全局最优解,能克服模型必须是观测数据的线性或非线性组合的缺点,且具有良好的泛化能力^[1],已经在分类、模式识别、回归拟合领域得到广泛应用,尤其是在解决非线性问题时效果显著。主成分分析法(PCA)通过研究原始变量相关矩阵内部结构关系,求解影响研究问题的主成分,能保留原始变量的主要信息,也能达到降维目的。

我国水电工程大坝多采用混凝土材料,裂缝是混凝土坝面临的主要问题之一。坝体内裂缝可能会形成集中渗漏通道,恶化大坝的运行状态,影响坝体的整体安全^[2-3]。裂缝产生机理复杂,监控混凝土坝坝体内裂缝性态的发展,对了解裂缝的产生机理与演变规律

具有重要意义,也是大坝安全监控的重要内容。目前学者在裂缝监控方面已开展了相关研究,雷鹏等融合 BP 和 RS 理论,建立了裂缝监控模型^[4];陈福军等应用近似推理模型和信息分配法,建立了裂缝变形的不确定分析模型^[5]。

本文融合 PCA 和 SVR 方法,建立了大坝裂缝开合度的早期预报模型,利用 PCA 对多维自变量进行主成分提取和降维处理,并把它作为 SVR 的输入变量,经训练输出预测值。

1 PCA-SVR 模型

1.1 混凝土坝裂缝开合度统计模型

混凝土坝浇筑过程中温控不当、地形地质条件、基岩构造等原因会引起坝体开裂,形成集中渗漏通道,乃至威胁大坝整体安全,因此有必要对裂缝进行实时监控。根据裂缝的成因,可将其分为 2 类:人工设置的构造缝和再生缝。Well 在 20 世纪 60 年代初期提出的裂缝开度(COD)概念,为分析建立裂缝开合度统计模型提供了方便。

实际工程中,大坝运行条件复杂,坝体裂缝开合度

受温度、水位、降雨、时效、施工、混凝土性能等多种因素影响^[6]。因此,综合考虑各因素作用,建立裂缝开合度统计模型,其表达式为

$$K = F(\sigma_H + \sigma_T + \sigma_\theta) \\ = K_H(t) + K_T(t) + K_\theta(t) \quad (1)$$

$$K_H(t) = \sum_{i=0}^4 a_i H^i \quad (2)$$

$$K_T(t) = \sum_{i=1}^2 [b_{1i} \sin \frac{2\pi i t}{365} + b_{2i} \cos \frac{2\pi i t}{365}] \quad (3)$$

$$K_\theta(t) = c_1 \theta + c_2 \ln \theta \quad (4)$$

式中, σ_H 、 σ_T 、 σ_θ 分别为应力的水位分量、温度分量、时效分量; $K_H(t)$ 、 $K_T(t)$ 、 $K_\theta(t)$ 分别为裂缝开合度对应的水压分量、温度分量、时效分量; H 为监测日对应的上游水头,即上游水位与坝底高程之差; a_i 为水位分量的拟合系数; b_i 为温度分量回归系数; c_i 为时效分量回归系数; t 为监测日到始测日的累计天数; $\theta = t/100$, 从始测日算起,每增加 1 d, θ 增加 0.01。

1.2 主成分分析法(PCA)

自变量的多重相关性会影响模型的精度^[7-8],使模型在外延时产生严重偏差和不良影响。PCA 是一种通过降维技术把多个变量化为少数几个主成分的多元统计方法。主成分分析的过程实质是对原坐标系进行平移和旋转变换,使新坐标系的原点与样本点集合的重心重合,新坐标系的第一轴与数据变异的最大方向对应,第二轴与第一轴标准正交,并且对应于数据变异的第二大方向,依此类推。利用 PCA 提取影响大坝裂缝的主要成分,这些主成分能够反映原始变量的大部分信息,通常表示为原始变量的线性组合,并且各主成分之间互不相关。PCA 可避免自变量间的多重相关性影响,降低输入变量的维度,并减少计算成本,其计算步骤如下。

(1) 数据标准化处理。标准化处理的目的是使样本点集合的重心与坐标原点重合,而压缩处理可以消除由量纲不同所引起的虚假变异信息,使得分析结果更加合理。

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, p) \quad (5)$$

式中, \bar{x}_j 是 x_j 的样本均值; S_j 是 x_j 的样本标准差。

(2) 计算标准化数据矩阵 X 的协方差矩阵 V , 同时 V 又是 X 的相关系数矩阵。

(3) 求 V 的前 m 个特征值, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, 以及对应的特征向量 a_1, a_2, \dots, a_m , 要求它们是标准正交的。

(4) 求第 h 主成分 F_h , 有

$$F_h = Xa_h = \sum_{j=1}^p a_{hj}x_j \quad (6)$$

式中, X 为标准化处理后的矩阵; a_{hj} 主轴 a_h 的第 j 个分量。

(5) 累积贡献率为

$$Q_m = \frac{\sum_{h=1}^m \text{Var}(F_h)}{\sum_{j=1}^p S_j^2} \quad (7)$$

所以,主成分 F_h 是原变量 x_1, \dots, x_p 的线性组合,组合系数恰好为 a_{hj} 。

1.3 支持向量机拟合法(SVR)

支持向量机是一种新的机器学习方法,其基础是 Vapnik 创建的统计学习理论^[9]。统计学习理论采用结构风险最小化准则,在最小化样本点误差的同时,最小化结构,提高了模型的泛化能力,且没有数据维数的限制。在 SVM 的基础上引入了不敏感损失函数,得到了 SVR。其本质是寻找一个最优分类面使得所有训练样本离该最优分类面的误差最小。

设含有 l 个训练样本的训练集样本对为 $\{(x_i, y_i), i = 1, 2, \dots, l\}$, 其中, x_i 是第 i 个训练样本的输入列向量, $x_i = [x_i^1, x_i^2, \dots, x_i^d]$, $y_i \in R$ 为对应的输出值。设在高维特征空间中建立的线性回归函数为

$$f(x) = \omega \varphi(x) + b \quad (8)$$

定义 ε 线性不敏感损失函数

$$L[f(x), y, \varepsilon] = \begin{cases} 0, & |y - f(x)| < \varepsilon \\ |y - f(x)| - \varepsilon, & |y - f(x)| \geq \varepsilon \end{cases} \quad (9)$$

为求解 ω, b , 将线性回归问题转化为最优化问题,则有

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ s. t. \begin{cases} y_i - w\varphi(x_i) - b \leq \varepsilon + \xi_i \\ -y_i + w\varphi(x_i) + b \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0 \end{cases} \quad (10)$$

式中, $f(x)$ 为回归函数返回的预测值; y 为对应的真实值; $\varphi(x)$ 为非线性映射函数; ε 为不敏感损失函数; ξ_i, ξ_i^* 为松弛变量; C 为惩罚因子。

为求解公式(10),引入 Langrange 函数,并转换为对偶函数

$$\max_{\alpha, \alpha^*} \left[-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon + \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i \right]$$

$$s.t. \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases} \quad (11)$$

求解公式(11)得到最优解为 $\alpha_i = [\alpha_1, \alpha_2, \cdots, \alpha_n], \alpha_i^* = [\alpha_1^*, \alpha_2^*, \cdots, \alpha_n^*]$ 。

最后,回归函数为

$$f(x) = w^* \varphi(x) + b^* = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b^* \quad (12)$$

其中, $K(x_i, x) = \varphi(x_i) \varphi(x_j)$ 为核函数。

1.4 PCA-SVR 模型

PCA-SVR 模型利用 PCA 提取影响混凝土坝裂缝的主要影响因素,并把它们作为 SVR 模型的输入变量,通过优化训练计算得裂缝预测值。PCA-SVR 计算流程如图 1 所示。

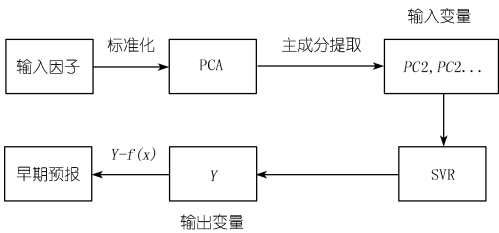


图 1 PCA-SVR 模型计算流程

2 工程实例

2.1 工程概况

锦屏一级水电站属 I 等大(一)型工程,主要水工建筑物为混凝土双曲拱坝。工程于 2004 年开始前期筹建工作,2006 年 12 月 4 日实现大江截流;2009 年 10 月 23 日,大坝首仓混凝土在 14 号坝段(高程 1 580 ~ 1 581.5 m)开始浇筑;2012 年 11 月 30 日,导流洞下闸开始首次蓄水。大坝共有 25 个坝段,坝段间共布置有横缝测缝计 747 支。在 2014 年 7 月 3 日开始第四阶段蓄水,根据监测资料发现 10 号横缝张开较大。本次选取 J10-40 测点 2012 年 11 月 30 日至 2014 年 7 月 9 日共 544 组监测数据用于模型拟合和训练,2014 年 7 月 10 日至 17 日共 8 组监测数据用于预测。

2.2 自变量因子的选取

认真分析 J10-40 测点裂缝监测资料,选取以下 10 个因子作为自变量: $H, H^2, H^3, H^4, \sin \frac{2\pi t}{365}, \sin \frac{4\pi t}{365}, \cos \frac{2\pi t}{365}, \cos \frac{4\pi t}{365}, \ln \theta, \theta$, 随后建立以下因子输入模型

$$y = f(H^i, \sin \frac{2\pi jt}{365}, \cos \frac{2\pi jt}{365}, \ln \theta, \theta) \quad (i = 1 \sim 4, j = 1 \sim 2) \quad (13)$$

2.3 主成分提取

利用 PCA 对自变量进行预处理,提取对因变量解释最强的主成分,详细参数见表 1。当主成分个数为 5 时,累积贡献率达到 99.21%,超过 85%,满足要求。

表 1 主成分计算

主成分 t_i	特征值 λ_i	贡献率/%	累积贡献率/%
1	6.37	57.88	57.88
2	1.72	15.60	73.48
3	1.08	9.83	83.31
4	0.90	8.18	91.48
5	0.85	7.73	99.21

2.4 结果及分析

将 PCA 得到的 5 个主成分作为 SVR 模型的输入量,经 SVR 训练得到较优结果。将 PCA-SVR 与传统的 SVR 模型相比较,结果如表 2 所示。

表 2 模型对比

回归方法	判定系数 R	计算时间 t/s
SVR	0.9321	220
PCA-SVR	0.9961	179

由表 2 分析得,SVR 和 PCA-SVR 拟合精度均能满足要求,但 PCA-SVR 模型拟合精度更高,为 0.996 1,进一步分析得,当观测数据的多重相关性显著时,PCA-SVR 模型的优越性更强;PCA-SVR 的计算时间更短,因为 PCA 把输入变量由 10 维降到 5 维,大大减少了输入数据量,在自变量较多的观测系统中,该方法优越性更显著。

分别运用 PCA-SVR、SVR 方法对 2014 年 7 月 10 日后 7 d 的裂缝开合度进行预测,并与实测值进行对比分析,结果见表 3 所示。

表 3 预测误差对比 mm

时间	实测值	PCA-SVR		SVR	
		预测值	绝对误差	预测值	绝对误差
2014 年 7 月 10 日	7.0296	7.0282	0.0014	7.0281	0.0015
2014 年 7 月 11 日	7.0336	7.0279	0.0057	7.0275	0.0061
2014 年 7 月 12 日	7.0316	7.0275	0.0041	7.0269	0.0047
2014 年 7 月 13 日	7.0296	7.0264	0.0032	7.0260	0.0036
2014 年 7 月 14 日	7.0256	7.0255	0.0001	7.0252	0.0004
2014 年 7 月 15 日	7.0296	7.0243	0.0053	7.0240	0.0056
2014 年 7 月 16 日	7.0236	7.0232	0.0004	7.0223	0.0013
2014 年 7 月 17 日	7.0236	7.0223	0.0013	7.0210	0.0026

由表 3 分析得,PCA-SVR 的预测残差的绝对值小于 SVR 预测残差的绝对值,故 PCA-SVR 的预测效

果优于 SVR。

PCA-SVR 与 SVR、传统的稳健回归、逐步回归的预测曲线对比见图 2。从图中可以看出:逐步回归和稳健回归预测曲线的趋势是逐渐偏离实测曲线,误差有扩大趋势;PCA-SVR 与 SVR 相比,预测误差更小,且误差有收敛趋势。

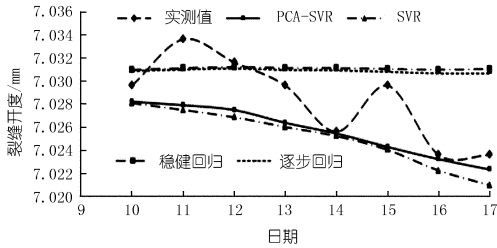


图 2 2014 年 7 月模型预测曲线对比

3 结 语

(1) 大坝裂缝开合度受多种因素影响,采用 PCA 方法提取影响因变量的主成分,避免了自变量之间的多重相关性对模型精度的影响,同时通过降低 SVR 的输入维数,能大大节省计算时间,提高模型的计算效率。

(2) SVR 利用非线性映射将样本从原始空间映

射到高维线性空间,很好地解决了大坝裂缝的非线性问题。

(3) 与其他模型相比,PCA-SVR 模型的预测精度和效率较高,能提前预报大坝裂缝开合度,故能广泛运用于工程实际中。

参考文献:

[1] 姜涛男,梁冰.基于 PSO-SVM 的大坝渗流监测时间序列非线性预报模型[J].水利学报,2006,37(3):331-335.
[2] 沈长松.水工建筑物[M].北京:中国水利水电出版社,2008.
[3] 罗福海,张保军,夏界平.水布垭大坝施工期面板裂缝成因分析及处理措施[J].人民长江,2011,(1):50-53.
[4] 雷鹏,肖峰.基于 RS-BP 融合的裂缝监控模型[J].长沙理工大学学报:自然科学版,2011,8(2):45-50.
[5] 陈福军,塔娜,赵二峰,等.混凝土坝裂缝变形的不确定性分析模型[J].黑龙江水专学报,2009,36(1):54-56,60.
[6] 吴中如.水工建筑物安全监控理论及其应用[M].南京:河海大学出版社,1990.
[7] 许后磊,冯茂静,杨阳,等.因子相关性对大坝监测模型精度的影响探究[J].水电能源科学,2009,27(5):77-80.
[8] 吴云芳,李珍照,徐帆,等.因子相关性对两种大坝监测模型的影响分析[J].武汉大学学报:工学版,2004,37(1):36-40.
[9] Manoj K. Blast-induced ground vibration prediction using support vector machine[J]. Eng. Comput., 2011, (27):193-200.

(编辑:胡旭东)

Study on early warning of dam crack based on PCA-SVR Model

WANG Quan^{1,2,3}, ZHENG Dongjian^{1,2,3}, FAN Zhendong^{1,2,3}, FENG Shurong⁴, XIAO Feng⁴

(1. College of Water Conservancy & Hydropower Engineering, Hohai University, Nanjing 210098, China; 2. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China; 3. National Engineering Research Center of Water Resources Efficient Utilization and Engineering Safety, Hohai University, Nanjing 210098, China; 4. Hydrochina Zhongnan Engineering Corporation, Changsha 410014, China)

Abstract: The SVR model can be used in dam crack prediction, but the multiple correlations among independent variables and high dimension of input variables have large impact on the accuracy of support vector machine fitting model. On the basis of dam crack opening theory, the Principal Component Analysis method is adopted to extract the information of original sample, and the obtained main components are used as the input variable of SVR model, which can reduce the computational cost. The PCA-SVR model was established. This new model is applied to the analysis of monitoring data, and the results show that the PCA-SVR model has higher calculation accuracy and efficiency, and can predict the dam opening information compared with traditional regression methods.

Key words: crack opening degree; PCA; SVR; nonlinear; prediction model