

加权支持向量机求解路径算法研究

毛文涛, 闫桂荣, 董龙雷, 张刚

(西安交通大学强度与振动教育部重点实验室, 710049, 西安)

摘要: 针对不同训练样本重要性的差异对模型推广能力的影响, 提出了对各个样本的误差惩罚参数赋予不同权重的加权支持向量机求解路径算法. 根据样本重要性的不同, 利用分段线性插值得到加权系数, 并通过加权系数调整求解路径, 从而改变不同样本在回归模型中的作用. 采用支持向量机加权求解路径算法对圆柱壳结构在不同边界条件下的时、频域响应数据进行预测, 训练样本的重要性通过与测试样本的欧式距离来表达, 结果显示所提算法可减小位移响应在多个评价指标下的预测误差, 提高支持向量回归机的推广能力. 该方法同样适用于其他求解路径算法, 如 λ -路径算法和 ν -支持向量回归路径算法.

关键词: 求解路径; 支持向量回归; 加权系数

中图分类号: TP181 **文献标志码:** A **文章编号:** 0253-987X(2008)10-1226-04

A Weighted Solution Path Algorithm for Support Vector Regression

MAO Wentao, YAN Guirong, DONG Longlei, ZHANG Gang

(MOE Key Laboratory of Strength and Vibration, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Considering the existence of significant difference among the training samples, a weighted solution path algorithm for support vector regression is proposed in which the error penalty parameter on each training sample is weighted differently. The weighting coefficients are obtained by using piecewise linear interpolation based on the difference of samples' importance. Then the solution path is adjusted by the resulting weighting coefficients, such that the effects of different samples in regression model are changed correspondingly. The experiments on the prediction of response data of cylinder structure under two different boundary conditions in time frequency domain, where the importance of each training sample is characterized by the geometrical distance to the test sample, show that the weighted solution path algorithm for support vector regression can reduce prediction errors of displacement response on different evaluation indexes, and that the generalization of regression model is improved. It is noted that the main idea of the proposed algorithm can also be applied to other solution path algorithms, such as λ -path and ν -SVR path algorithms.

Keywords: solution path; support vector regression; weighting coefficient

在支持向量回归(Support Vector Regression, SVR)中, 误差惩罚参数 C 和误差参数 ϵ 对模型推广能力有重要影响. ϵ 表明了系统对估计函数在样本数据点上的误差期望, C 控制着对于估计函数误差大于 ϵ 的样本数据的惩罚. 在实际应用中, 一般需对

待定参数进行线性搜索(一个参数)或网格搜索(多个参数), 因此计算量比较大.

求解路径算法是近年来提出的一种新型的模型选择算法, 同时也是一种全新的训练算法, 它通过直接计算对偶问题的全体解, 即 Lagrange 乘子 $\alpha_i^{(*)}$,

从而可以用较小的计算代价获得最优模型^[1].但是,该算法在构建原始问题的时候,对所有样本采用的是相同的 C 和 ϵ ,即对不同样本的精度和误差惩罚的要求是平等的.在实际应用中^[2-3],有些数据比较重要,要求的训练误差较小,相反则容许有一定大小的误差.因此,在利用求解路径算法解决此类问题时,需对每个样本数据设置不同的 C 或 ϵ ,即对精度或误差惩罚参数进行加权学习.

本文通过对圆柱壳结构在不同边界条件下的时、频域响应数据的学习与预测,验证了加权求解路径算法具有更好的推广能力.

1 支持向量机求解路径算法

求解路径算法的特点是:对于损失函数分别为二次项、分段二次项、线性、分段线性的支持向量机,其对偶问题的全体解 $\alpha_i^{(*)}$ 是分段线性函数^[1,4-6].因此,与现有的训练算法有所不同的是,求解路径算法只需关注分段线性路径中的折点,从而利用线性关系可直接计算 $\alpha_i^{(*)}$,无需求解凸二次型优化问题,所以它在性能上具有优势.

称路径为某种参数的全体值对应解的集合,那么根据路径对应的不同参数,求解路径算法具有不同的形式.针对参数 C ,Zhu 等人提出了损失项为一阶范数的支持向量分类机的求解路径算法^[4],Hastie 等人提出了损失项为二阶范数的支持向量分类机求解路径算法^[5],Gunter 和 Zhu 研究了针对参数 C 的 ϵ -SVR 求解路径算法^[6],Wang 则研究了针对参数 ϵ 的 ϵ -SVR 求解路径算法^[7].下面将着重讨论 ϵ 求解路径算法.

为了便于计算 $\alpha_i^{(*)}$,可将 ϵ -SVR 原始问题改写为如下的正则化形式

$$\left. \begin{aligned} \min \quad & \frac{\lambda}{2} \|w\|^2 + \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s. t.} \quad & y_i - (w \cdot \phi(x_i) + w_0) \leq \epsilon + \xi_i \\ & (w \cdot \phi(x_i) + w_0) - y_i \leq \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \quad i = 1, 2, \dots, l \end{aligned} \right\} \quad (1)$$

其中, λ 等价于 $1/C$. 决策函数可表达为 $f(x) = \frac{1}{\lambda} \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + w_0$. 那么,由 Karush-Kuhn-Tucker(KKT)条件,根据不同的 α_i, α_i^* ,定义下述 5 个集合

$$\left. \begin{aligned} R &= \{i: y_i - f(x_i) > \epsilon, \alpha_i = 1, \alpha_i^* = 0\} \\ E_R &= \{i: y_i - f(x_i) = \epsilon, \alpha_i \in [0, 1], \alpha_i^* = 0\} \\ E_C &= \{i: |y_i - f(x_i)| < \epsilon, \alpha_i = 0, \alpha_i^* = 0\} \\ E_L &= \{i: y_i - f(x_i) = -\epsilon, \alpha_i = 0, \alpha_i^* \in [0, 1]\} \\ L &= \{i: y_i - f(x_i) < -\epsilon, \alpha_i = 0, \alpha_i^* = 1\} \end{aligned} \right\} \quad (2)$$

如果 ϵ 求解路径算法从一个较大的 ϵ 出发,随之逐渐减小,将导致样本点在上述 5 个集合中移动.这说明算法的基础就是通过判断 α_i, α_i^* 的变化,找到路径中的折点.称一个点离开或进入 E_R 或 E_L 为一个事件,而事件的发生导致了式(2)变化,因此事件发生将产生路径折点.可见,监视运行过程中的事件,可计算出路径中对应的折点.

首先计算相邻 2 个折点间的线性关系.令 $\alpha_i^l, \epsilon^l, f^l(x)$ 表示第 l 个事件的初值,设 $\epsilon^{l+1} < \epsilon < \epsilon^l$,又因 $E_R \cup E_L$ 中的 $\alpha_i^{(*)}$ 是不定值,则仅需考虑相邻 2 个事件中停留在 $E_R \cup E_L$ 的点即可.由此,通过计算得到

$$\left. \begin{aligned} \alpha_0 &= \alpha_0^l + \lambda(\epsilon^l - \epsilon)c_0 \\ \alpha_i &= \alpha_i^l + \lambda(\epsilon^l - \epsilon)c_i, \quad \forall i \in E_R^l \\ \alpha_j^* &= \alpha_j^{*l} + \lambda(\epsilon^l - \epsilon)c_j, \quad \forall j \in E_L^l \end{aligned} \right\} \quad (3)$$

然而, $c = A_l^{-1} \mathbf{1}^a, A_l = \begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & K_l \end{bmatrix}$.

式(3)表明,2 次事件间的 $\alpha_i^{(*)}$ 是 ϵ 的线性函数,由此可快速求得各点对应的 $\alpha_i^{(*)}$.

很容易理解,知道了折点,则折点之间的解可由式(3)快速求得,由此整体路径便可获得.改写决策函数表达式为

$$f(x) = (\epsilon^l - \epsilon)h^l(x) + f^l(x) \quad (4)$$

其中

$$h^l(x) = \sum_{i \in E_R^l} c_i K(x_i, x) - \sum_{j \in E_L^l} c_j K(x_j, x) + c_0$$

由于 ϵ 通道得以缩减,所以算法需监视下述 2 种事件.

(1) $E_R \cup E_L$ 中的点离开,其对应的 $\alpha_i^{(*)}$ 将变成 0 或 1,可由式(3)计算得到 ϵ .

(2) 其他集合的点进入 $E_R \cup E_L$,即满足 $|y_i - f(x_i)| = \epsilon$,可由式(4)计算得到 ϵ .

计算得到的 ϵ 最大值,便是第 $(l+1)$ 个事件对应的 ϵ ,记为 ϵ^{l+1} .同时,更新式(2).依次迭代,便可得到解的整体路径.文献[6-7]对此做了详细的介绍和推导.

2 加权支持向量机求解路径算法

在 ϵ 路径算法中, λ 是可以自由选择的,但是该

值一旦选定,就说明所有样本数据对偏离精度要求的惩罚是统一的,而在实际应用中,样本的重要程度往往不同,这就需要对不同的样本选择不同的惩罚参数。

对式(1)加权,则变为

$$\left. \begin{aligned} \min \quad & \frac{\lambda}{2} ||\mathbf{w}||^2 + \sum_{i=1}^n s_i (\xi_i + \xi_i^*) \\ \text{s. t.} \quad & y_i - (\mathbf{w} \cdot \boldsymbol{\phi}(\mathbf{x}_i) + w_0) \leq \varepsilon + \xi_i \\ & (\mathbf{w} \cdot \boldsymbol{\phi}(\mathbf{x}_i) + w_0) - y_i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \quad i = 1, 2, \dots, l \end{aligned} \right\}$$

式中: s_i 表示第 i 个训练样本对参数 C 的加权系数。利用拉格朗日乘子法求解此问题,便有

$$\begin{aligned} \max_{\alpha, \alpha^*, \beta, \beta^*} \min_{\mathbf{w}, w_0} L_p = & \frac{\lambda}{2} ||\mathbf{w}||^2 + \sum_{i=1}^n s_i (\xi_i + \xi_i^*) - \\ & \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w} \cdot \boldsymbol{\phi}(\mathbf{x}_i) + w_0) - \\ & \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w} \cdot \boldsymbol{\phi}(\mathbf{x}_i) - w_0) - \\ & \sum_{i=1}^n (\beta_i \xi_i + \beta_i^* \xi_i^*) \end{aligned}$$

分别求偏导,则有

$$\left. \begin{aligned} \frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} &= \frac{1}{\lambda} \sum_{i=1}^n (\alpha_i - \alpha_i^*) \boldsymbol{\phi}(\mathbf{x}_i) \\ \frac{\partial L_p}{\partial w_0} = 0 \Rightarrow \sum_{i=1}^n (\alpha_i - \alpha_i^*) &= 0 \\ \frac{\partial L_p}{\partial \xi_i^*} = 0 \Rightarrow s_i - \alpha_i^{(*)} - \beta_i &= 0 \Rightarrow 0 \leq \alpha_i, \alpha_i^* \leq s_i \end{aligned} \right\} \quad (5)$$

由回归决策函数

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b_0$$

可知, α_i 的绝对值越大,相应的样本在模型中的作用就越大。由式(5)可以看到,本算法与 ε 求解路径算法的最大不同在于,加权系数 s_i 决定了 $\alpha_i^{(*)}$ 的上界。它正是通过调整 s_i ,继而调整 $\alpha_i^{(*)}$ 的大小,从而改变样本在决策函数中的作用。也就是说,加权系数体现了根据样本的重要性对推广能力的调整程度。

由此,可重新定义集合

$$\left. \begin{aligned} R &= \{i: y_i - f(\mathbf{x}_i) > \varepsilon, \alpha_i = s_i, \alpha_i^* = 0\} \\ E_R &= \{i: y_i - f(\mathbf{x}_i) = \varepsilon, \alpha_i \in [0, s_i], \alpha_i^* = 0\} \\ E_C &= \{i: |y_i - f(\mathbf{x}_i)| < \varepsilon, \alpha_i = 0, \alpha_i^* = 0\} \\ E_L &= \{i: y_i - f(\mathbf{x}_i) = -\varepsilon, \alpha_i = 0, \alpha_i^* \in [0, s_i]\} \\ L &= \{i: y_i - f(\mathbf{x}_i) < -\varepsilon, \alpha_i = 0, \alpha_i^* = s_i\} \end{aligned} \right\} \quad (6)$$

可以看出,随着 ε 的递减,样本点将在上述 5 个集合中移动,同时触发事件。由于权值 s_i 为预先计算得到的固定值,因此仅需关注 $E_R \cup E_L$ 即可。在构造相邻 2 个事件间的线性关系时,式(3)及式(4)对于 $E_R \cup E_L$ 依然成立。

如前所述,随着 ε 的减小,2 种事件可能发生。当 $E_R \cup E_L$ 中的样本点进入其他集合时, $\alpha_i^{(*)}$ 将为 0 或 s_i 。因此,样本点进入 $R \cup L$ 的计算式为

$$\varepsilon = (\alpha_i^l - s_i + \lambda \varepsilon^l c_i) / (\lambda c_i) \quad \text{进入 } R$$

$$\varepsilon = (s_i - \alpha_i^{*l} + \lambda \varepsilon^l c_i) / (\lambda c_i) \quad \text{进入 } L$$

当其他集合的样本点进入 $E_R \cup E_L$ 后, $\alpha_i^{(*)}$ 将从 0 或 s_i 变为 $[0, s_i]$,且满足 $|y_i - f(\mathbf{x}_i)| = \varepsilon$,因此有

$$\varepsilon = (y_i - \varepsilon^l h^l(\mathbf{x}) - f^l(\mathbf{x})) / (1 - h^l(\mathbf{x})) \quad \text{进入 } E_R$$

$$\varepsilon = (y_i + \varepsilon^l h^l(\mathbf{x}) + f^l(\mathbf{x})) / (1 + h^l(\mathbf{x})) \quad \text{进入 } E_L$$

ε 的最大值,便是第 $(l+1)$ 个事件所对应的 ε ,即折点处的 ε 。同时,根据 s_i 的大小更新式(6),检查 $\alpha_i^{(*)}$,并将大于 s_i 的 $\alpha_i^{(*)}$ 设置为 s_i 。

3 实验验证

本文以圆柱壳结构为研究对象,以连接不同的振动夹具来模拟不同边界条件,通过动力学仿真获得 2 种力学环境下的冲击响应数据。利用加权支持向量机求解路径算法,对相同载荷、2 种夹具条件下的时域数据进行学习预测。

图 1 为圆柱壳结构有限元模型,其上选取了多个测点,圆柱壳通过与钢材料夹具和镁铝材料夹具连接,来模拟 2 种不同的边界条件。冲击激励点设置在图 1 中的“+”处,方向定义为 X ,即经过同水平面轴线点。

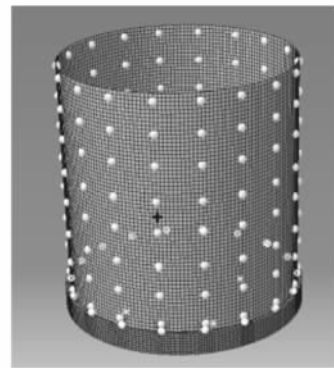


图 1 圆柱壳有限元模型

对于预测效果,采用了 3 种评价指标,见表 1。

在应用加权求解路径算法预测不同载荷下位移响应时,采用了测试样本与训练样本集的欧式距离

表 1 误差评判标准

最大平方误差	$\max(f(\mathbf{x}_i)-y_i ^2)$
平均误差	$\frac{1}{l} \sum_{i=1}^l f(\mathbf{x}_i)-y_i $
均方差	$\frac{1}{l} \sum_{i=1}^l f(\mathbf{x}_i)-y_i ^2$

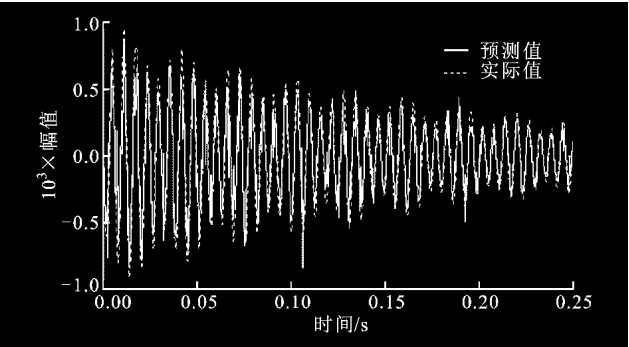
作为样本重要性的评判标准,距离越小,训练样本的重要性越大,加权 s_i 越大. 根据距离的大小,从小到大重新排列训练样本,并根据距离的变化情况,将样本分段加权. 如果段与段的权值差值较大,则段内的权值使用线性插值计算. 设段内权值上限为 σ_u ,下限为 σ_l ,样本数为 m ,则插值公式为 $s(i)=\sigma_l+(m-i)(\sigma_u-\sigma_l)/(m-1)$.

为了验证加权求解路径算法的效果,分别采用了 ϵ -SVR、 ϵ 求解路径算法和加权 ϵ 求解路径算法进行了学习和预测,结果如图 2 所示. 实验中,采用径向基核函数,核参数 $\epsilon=0.2$, $C=100$,并参照文献[2-3]将权值的最大值设为 1,最小值设为 0.1,由此得到了时域冲击数据的时间历程为 2 s.

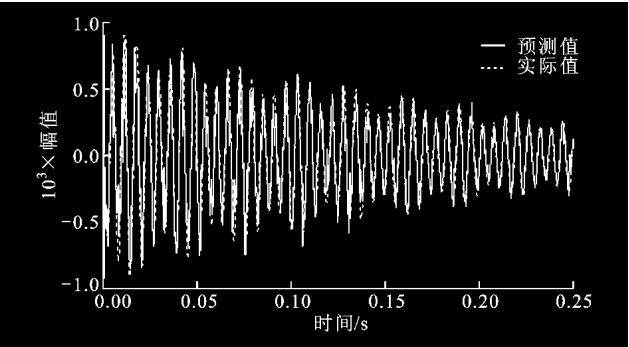
表 2 为时域下的误差数据. 可以看出,加权 ϵ 求解路径算法在 3 种指标下的预测误差均明显减小. 这表明,该算法可增进系统的预测效果,提高回归模型的推广能力. 频域下加权 ϵ 路径求解算法有着类似的优化效果,限于篇幅,这里不再详细介绍.

表 3 为不同权值上限加权 ϵ 路径求解算法的预测误差. 可以看出,随着权值的增加,2 种指标下的预测误差都在增加,这在 SVM 理论上是可以解释的. 因为权值的增大使得误差惩罚参数变大,与置信区间范围相关的 $\|\mathbf{w}\|^2/2$ 的权重相应变小,从而使

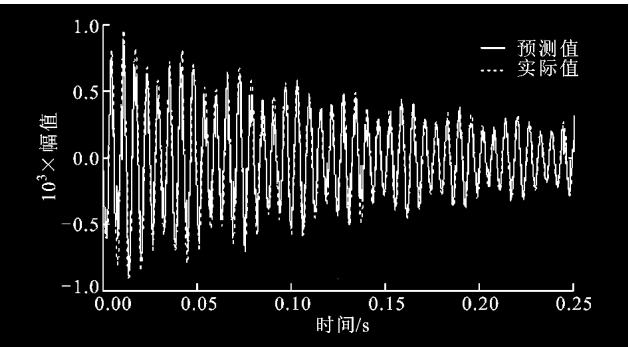
系统推广能力减弱. 因此,最优权值的设定需要进一步研究.



(a) ϵ -SVR



(b) ϵ 路径求解算法



(c)加权 ϵ 路径求解算法

图 2 圆柱壳时域位移响应预测效果图

表 2 时域下 3 种算法的预测误差

算法	最大平方误差	平均误差	均方差
ϵ -SVR	$8.300\ 8\times10^{-7}$	$8.666\ 0\times10^{-5}$	$1.772\ 4\times10^{-8}$
ϵ 路径求解算法	$5.399\ 8\times10^{-7}$	$6.643\ 3\times10^{-7}$	$1.139\ 8\times10^{-8}$
加权 ϵ 路径求解算法	$4.880\ 5\times10^{-7}$	$5.067\ 0\times10^{-7}$	$1.077\ 1\times10^{-8}$

表 3 不同权值上限加权 ϵ 路径求解算法的预测误差

σ_u	1	3	5	10	20
最大平方误差	$4.880\ 5\times10^{-7}$	$5.011\ 9\times10^{-8}$	$5.011\ 9\times10^{-7}$	$5.916\ 2\times10^{-7}$	$9.749\ 9\times10^{-7}$
均方差	$1.077\ 1\times10^{-8}$	$1.216\ 2\times10^{-8}$	$1.264\ 5\times10^{-8}$	$1.350\ 4\times10^{-8}$	$1.446\ 2\times10^{-8}$

4 讨论与展望

建立支持向量机最优决策模型的关键在于提高模型推广能力,而在模型参数中,误差惩罚参数 C 和精度参数 ϵ 对最优回归模型的影响较大. 本文在 ϵ 求解路径算法的基础上,根据样本重要性的差异,对误差惩罚参数进行加权,从而改变了模型的推广能力,实验也证明了此方法具有较好的改善效果. 需要强调的是,本文所提方法也适用于其他参数下的求解路径算法,比如 λ -路径算法^[6]、 ϵ -SVR 路径算法.

参考文献:

- [1] ROSSET S, ZHU J. Piecewise linear regularized solution paths[J]. *Annals of Statistics*, 2007, 35(3): 1012-1030.
- [2] 杜树新,吴铁军. 回归型加权支持向量机方法及其应用[J]. *浙江大学学报(工学版)*, 2004, 38(3): 302-306.
DU Shuxin, WU Tiejun. Weighted support vector machines for regression and its application[J]. *Journal of Zhejiang University (Engineering Science)*, 2004, 38(3): 302-306.
- [3] TOY F E H, CAO L J. Modified support vector machines in financial time series forecasting [J]. *Neurocomputing*, 2002, 48(1/4): 847-861.
- [4] ZHU J, ROSSET S, HASTIE T, et al. 1-norm support vector machines [C]// *Advances in Neural Information Processing Systems*. Cambridge, USA: MIT Press, 2004: 49-56.
- [5] HASTIE T, ROSSET S, TIBSHIRANI R, et al. The entire regularization path for the support vector machine[J]. *Journal of Machine Learning Research*, 2004, 5: 1391-1415.
- [6] GUNTER L, ZHU J. Efficient computation and model selection for the support vector regression [J]. *Neural Computation*, 2007, 19(6): 1633-1655.
- [7] WANG G, YEUNG D Y, LOCHOVSKY F H. Two dimensional solution path for support vector regression [C]// *Proceedings of the 23rd International Conference on Machine Learning*. New York, USA: ACM Press, 2006: 993-1000.

(编辑 苗凌)