

DOI: 10.7652/xjtuxb201306009

一种可自适应分组的多输入多输出支持向量机算法

毛文涛¹, 冯云芝¹, 闫桂荣²

(1. 河南师范大学计算机与信息工程学院, 453007, 河南新乡;

2. 西安交通大学机械结构强度与振动国家重点实验室, 710049, 西安)

摘要: 针对不同输出端之间相关程度的差异对多输入多输出回归模型泛化能力的影响, 提出了一种基于自适应分组的多输入多输出支持向量机算法。该算法基于相关性强的输出端其模型参数也较相似的假设, 首先在多维支持向量机的基础上引入带分组结构的正则项, 进而将上述正则化问题转变为混合 0-1 规划; 其次, 采用交替优化的方法, 使相关性强的输出端在同一个分组内进行独立训练, 最终自适应地识别最优分组结构和模型参数。分别采用仿真数据和圆柱壳振动工程数据对所提算法进行测试, 结果表明, 该算法可有效辨识出输出端的相关度, 与传统算法相比, 该算法可有效提高支持向量机回归模型的泛化能力。

关键词: 支持向量机; 自适应分组; 混合 0-1 规划

中图分类号: TP181 **文献标志码:** A **文章编号:** 0253-987X(2013)06-0050-05

New Support Vector Machine with Multi-Input and Multi-Output Based on Adaptive Grouping

MAO Wentao¹, FENG Yunzhi¹, YAN Guirong²

(1. College of Computer and Information Engineering, Henan Normal University, Xinxiang, Henan 453007, China;

2. State Key Laboratory for Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Recently, multi-dimensional support vector regression (M-SVR) has been a promising tool in solving a wide range of multi-input and multi-output (MIMO) regression problems. However, when some output dimensions have more dependencies than others, M-SVR is are generally hard to obtain impressive performance due to the negative or redundant domain knowledge across all outputs. A structure of adaptive grouping is introduced into the M-SVR to solve the problem. It is assumed that model parameters of related output dimensions are similar to each other, then a new regularization-based M-SVR algorithm is presented by introducing a regularizer with grouping structure. The regularization problem is then converted into a mixed 0-1 programming problem. Alternating optimization is employed for learning related outputs jointly in the same group, and to obtain the optimal grouping structure and model parameters through adaptively identifying the grouping structure. The proposed algorithm is tested empirically on a toy problem as well as a real-life data set. The experimental results show that the performance of the proposed algorithm outperforms the performance of the classical M-SVR and the single-output SVM algorithm.

Keywords: support vector machine; adaptive grouping; mixed 0-1 programming

在系统辨识、状态估计等工程领域中,基于输入输出数据的黑箱建模技术常常需要对多个变量或指标同时进行预测,即输出变量为一个向量,这就是多输入多输出(Multi-Input Multi-Output, MIMO)回归建模问题。与传统的支持向量机(Support Vector Machine, SVM)^[1]等不同,MIMO建模强调同时对多个输出端进行回归建模。由于这种建模方式可以利用多个输出端之间的有用信息,因此往往能取得较好的效果。提高MIMO建模的精度有着明确的工程意义。

目前有多种MIMO建模方法,最常用的做法是将一个MIMO模型按照输出端分解为多个单输出的子模型,采用SVM等算法单独进行回归建模。这种方法只需重构样本,简单明了,但无法利用输出端之间的联系来提高建模效果,同时,重复建模也会增加计算量。尽管传统的多元统计回归法也可用于MIMO建模,但该类方法依赖于数据分布,难以广泛适用。神经网络是另一种有效的方法,但其依赖于初值设定,容易陷入局部极值,导致过学习。与上述方法不同,Pérez-Cruz提出了多维SVM回归机(Multi-dimensional SVM Regression, M-SVR)^[2],该算法充分利用了SVM对小样本数据的处理能力,通过在超球空间重定义 ϵ 不敏感损失函数,采用迭代变权最小二乘法(IRWLS)求解对偶问题,最终得到每个输出端的预测模型,是一种适用于小样本MIMO建模的有效方法,已经成功解决脑电源信号定位、股票投资预测等问题。

M-SVR要取得良好效果需满足一个前提,即MIMO模型的所有输出端之间都存在密切联系^[3],但在实际工程中,不同输出端之间的相关度往往存在较大差异,某些输出端之间的依赖性更强。例如,在综合力学环境预测中,多个物理测点采集的响应数据存在内在联系(由同一刚体结构生成),可以采用MIMO方式建模,但由于一个机械结构不同部位的固有属性(振动模态频率、振型等)存在差异,导致响应信号的相关程度也会有不同,此时直接采用M-SVR建模,往往无法取得满意的效果。这是由于相关度不高的输出端之间存在大量冗余的、甚至负面的信息,无助于提高、甚至会减弱建模效果。如果在M-SVR中能将关联度高的输出端分为一组进行建模,则可以提高MIMO的建模效果。事实上,本文的出发点正是源自上述工程应用的需求。受文献[4]的启发,本文基于相关性强的输出端其模型参数也较相似的假设,将分组结构引入M-SVR,并采用混合0-1规划和交替优

化的解决方案,自适应计算得到最优分组结构和模型参数,最终通过仿真数据和圆柱壳振动数据验证了所提算法的有效性。

1 M-SVR简介

给定训练样本集合 $\{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)\} \subset \mathbf{R}^d \times \mathbf{R}^Q$,文献[2]将标准SVR的原始形式扩展为

$$\min_{\mathbf{w}, \mathbf{b}} L_p = \frac{1}{2} \sum_{j=1}^Q \|\mathbf{w}^j\|^2 + C \sum_{i=1}^n L(u_i) \quad (1)$$

式中: $L(u) = \begin{cases} 0 & u < \epsilon \\ u^2 - 2u\epsilon + \epsilon^2 & u \geq \epsilon \end{cases}; u_i = \|\mathbf{e}_i\| = (\mathbf{e}_i^T \mathbf{e}_i)^{1/2}, \mathbf{e}_i^T = \mathbf{y}_i^T - \phi^T(\mathbf{x}_i) \mathbf{W} - \mathbf{b}^T, \phi(\cdot)$ 为原始空间到特征空间的映射函数, $\mathbf{W} = [\mathbf{w}^1, \dots, \mathbf{w}^Q], \mathbf{b} = [\mathbf{b}^1, \dots, \mathbf{b}^Q]^T$ 。

通过一阶泰勒展开式,得到式(1)的二次型近似

$$L'_p(\mathbf{W}, \mathbf{b}) = \frac{1}{2} \sum_{j=1}^Q \|\mathbf{w}^j\|^2 + \frac{1}{2} \sum_{i=1}^n a_i u_i^2 + \Omega \quad (2)$$

式中: Ω 为与 $\mathbf{W}$ 和 $\mathbf{b}$ 无关的高阶项;

$$a_i = \begin{cases} 0 & u_i^k < \epsilon \\ \frac{2C(u_i^k - \epsilon)}{u_i^k} & u_i^k \geq \epsilon \end{cases}, k \text{ 为迭代次数。}$$

文献[2]采用了基于迭代的IRWLS算法求解最小化式(2)的最优值。该算法通过在前一次迭代结果的基础上线性搜索变量的下降方向,从而逐次收敛至最优值。由于特征空间中 $\mathbf{w}^j = \sum_i \phi(\mathbf{x}_i) \beta^j = \Phi^T \boldsymbol{\beta}^j$,因此对式(1)的求解转换为对模型参数 $\boldsymbol{\beta}$ 和 $\mathbf{b}$ 的求解,算法如下^[2]。

Step1 设定初值 $k=0, \boldsymbol{\beta}^k=\mathbf{0}, \mathbf{b}^k=\mathbf{0}$,计算 u_i^k 和 a_i 。

Step2 利用下式计算得到 $\boldsymbol{\beta}^k$ 和 $\mathbf{b}^k$

$$\begin{bmatrix} \mathbf{K} + \mathbf{D}_a^{-1} & \mathbf{1} \\ \mathbf{a}^T \mathbf{K} & \mathbf{1}^T \mathbf{a} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^j \\ \mathbf{b}^j \end{bmatrix} = \begin{bmatrix} \mathbf{y}^j \\ \mathbf{a}^T \mathbf{y}^j \end{bmatrix}, j=1, \dots, Q$$

式中: $\mathbf{K}$ 为核矩阵; $\mathbf{a} = [a_1, \dots, a_n]^T; (\mathbf{D}_a)_{ij} = a_i \delta(i-j)$ 。计算得到下降方向 $\mathbf{P}^k = \begin{bmatrix} \mathbf{W}^k - \mathbf{W}^k \\ (\mathbf{b}^k - \mathbf{b}^k)^T \end{bmatrix}^T$ 。

Step3 利用线性搜索法,计算得到 $\boldsymbol{\beta}^{k+1}$ 和 $\mathbf{b}^{k+1}$,并进而得到 u_i^{k+1} 和 a_i 。如不满足终止条件,则返回Step2。

收敛后所得到的 $\boldsymbol{\beta}^j$ 和 $\mathbf{b}^j$ 即为第 j 个输出端的模型参数,进而可计算输出值。

2 基于自适应分组的MIMO SVM算法

在式(1)中,正则项为所有模型参数的 Frobe-

nus 范数,经验风险项为所有输出端的 ϵ 不敏感损失函数。M-SVR 的这种优化目标建立在所有输出端存在联系的假设之上,目的是提高模型的信息含量。但是,这种假设过于严格。在实际工程应用中,经常存在一部分输出端的相关度远大于其他输出端的情况。此时,若直接引入 M-SVR 建模,相关度不高的输出端所提供的信息往往是冗余的、甚至会起到负面作用。基于相关性强的输出端其模型参数也较相似的假设,本节将所有输出端按相关度划分为多个分组,每组成员模型参数的 Frobenius 范数作为 M-SVR 新的正则项,并将优化目标转换为混合 0-1 规划问题,最终采用交替优化的方法计算最优模型参数和分组结构。

2.1 优化目标

设 $\mathbf{W}=[\boldsymbol{\beta}^1, \boldsymbol{\beta}^2, \dots, \boldsymbol{\beta}^Q]$ 为所有输出端的模型参数矩阵, G 为分组个数, $\mathbf{W}_g$ 为第 g 个组的输出端的参数矩阵。根据上述思路,式(1)可扩展为

$$\mathbf{W}^* = \arg \min \sum_{i=1}^n L(u_i) + \lambda \sum_g \|\mathbf{W}_g\|_F^2 \quad (3)$$

式中: $\|\cdot\|_F$ 为 Frobenius 范数; $L(u)$ 的计算同式(1)。

设对角阵 $\mathbf{A}_g \in \mathbf{R}^{Q \times Q}$ 为分组标识矩阵,其中对角线元素 $q_{gi} \in \{0, 1\}$ 用来标识第 i 个输出端是否属于第 g 组,则 $\mathbf{W}_g = \mathbf{W} \mathbf{A}_g$ 。式(3)等价于如下优化问题

$$\left. \begin{aligned} \min & \sum_g \sum_{i=1}^n L_g(u_i) + \lambda \sum_g \|\mathbf{W} \mathbf{A}_g\|_F^2 \\ \text{s. t. } & \sum_g \mathbf{A}_g = \mathbf{I} \text{ 且 } q_{gi} \in \{0, 1\} \end{aligned} \right\} \quad (4)$$

式(4)中的约束条件保证了每个输出端只能分配到一个组。很明显,如果 $G=1$,即所有输出端分到同一个组,式(4)变为传统的 M-SVR;如果 $G=Q$,则式(4)变为多个传统 SVM 算法的叠加。

2.2 优化求解

由最优化理论可知,式(4)为混合 0-1 约束问题,通常难以取得全局最优解。本节采用交替优化的思路,依次固定模型参数 $\mathbf{W}$ 或分组结构 $\mathbf{A}_g$,并优化其余变量。当 $\mathbf{A}_g$ 固定时,可通过标准 M-SVR(式(1))直接计算最优 $\mathbf{W}$;当 $\mathbf{W}$ 固定时,可通过最小化 $\sum_g \|\mathbf{W} \mathbf{A}_g\|_F^2$ 求解最优 $\mathbf{A}_g$ 。

由于 $R(\mathbf{A}) = \sum_g \|\mathbf{W} \mathbf{A}_g\|_F^2$ 不是 $\mathbf{A}_g$ 的元素 q_{gi} 的凸函数,因此该式无法直接求解。对于 $q_{gi} \in \{0, 1\}$ 来说, $R(\mathbf{A})$ 与 $P(\mathbf{A}) = \sum_g \|\mathbf{W} \mathbf{A}_g^{1/2}\|_F^2$ 是等价的。受文献[4]启发,存在如下定理。

定理 1 若 $\{\mathbf{A}_g^*\}$ 为下列问题的最优解或局部极值

解,则 $\mathbf{A}_g^*$ 的元素 $q_{gi}^* \in \{0, 1\}$

$$\left. \begin{aligned} \min & P(\mathbf{A}) = \sum_g \|\mathbf{W} \mathbf{A}_g^{1/2}\|_F^2 \\ \text{s. t. } & \sum_g \mathbf{A}_g = \mathbf{I} \text{ 且 } 0 \leq q_{gi} \leq 1 \end{aligned} \right\} \quad (5)$$

证明 由矩阵论可知

$$\begin{aligned} & \sum_g \|\mathbf{W} \mathbf{A}_g^{1/2}\|_F^2 = \\ & \sum_g \text{trace}((\mathbf{W} \mathbf{A}_g^{1/2})^T (\mathbf{W} \mathbf{A}_g^{1/2})) = \\ & \sum_g \text{trace}((\mathbf{A}_g^{1/2})^T \mathbf{W}^T \mathbf{W} \mathbf{A}_g^{1/2}) = \\ & \sum_g \sum_i \|\mathbf{w}_i\|_2^2 q_{gi} \end{aligned} \quad (6)$$

式中: $\mathbf{w}_i$ 表示矩阵 $\mathbf{W}$ 的第 i 列元素。由此可见,式(6)实质上是关于 q_{gi} 的线性规划,约束条件为 $q_{gi} \geq 0$ 且 $\sum_g q_{gi} = 1$ 。根据线性规划理论可知,线性规划的最优解对应约束条件所界定的凸多面体(可行域)的顶点^[5]。由于式(6)的约束条件为凸,若使 $P(\mathbf{A})$ 最小,则需 $q_{gi}^* \in \{0, 1\}$ 。定理得证。

定理 1 的意义在于提供了一种优化的方式来计算最优 $\mathbf{A}_g$,即证明了可由初值属于 $[0, 1]$ 的 q_{gi} 收敛至 $q_{gi}^* \in \{0, 1\}$,具体的求解方法可采用梯度下降法。 $P(\mathbf{A})$ 关于 $\mathbf{A}_g$ 的梯度可计算如下

$$\begin{aligned} \frac{\partial P(\mathbf{A})}{\partial q_{gi}} &= \frac{\partial \sum_g \|\mathbf{W} \mathbf{A}_g^{1/2}\|_F^2}{\partial q_{gi}} = \\ & \frac{\partial \sum_g \text{trace}((\mathbf{W} \mathbf{A}_g^{1/2})^T (\mathbf{W} \mathbf{A}_g^{1/2}))}{\partial q_{gi}} = \\ & \sum_g 2(\mathbf{W} \mathbf{A}_g^{1/2}) \frac{\partial (\mathbf{W} \mathbf{A}_g^{1/2})}{\partial q_{gi}} = \\ & 2 \sum_g (\mathbf{W} \mathbf{A}_g^{1/2}) \left[\mathbf{W} \left(\frac{\partial \mathbf{A}_g^{1/2}}{\partial q_{gi}} \right) \right] \end{aligned} \quad (7)$$

式中: $\partial \mathbf{A}_g^{1/2} / \partial q_{gi}$ 对角线的第 i 个元素为 $1/2 q_{gi}^{1/2}$,其余为 0。此外,为简化计算,与文献[4]类似,式(5)可转换为无约束的优化问题,即引入无约束变量 α_{gi} ,使得

$$g_{gi} = e^{\alpha_{gi}} / A^*, \quad A^* = \sum_g e^{\alpha_{gi}}$$

此时, $\partial P(\mathbf{A}) / \partial \alpha_{gi}$ 即为式(7)与 $\partial q_{gi} / \partial \alpha_{gi}$ 的乘积。需要注意的是,定理 1 并没有保证 q_{gi}^* 一定为全局最优解。在实际应用中,可随机设定多组初始值,选取结果最优的一组。算法可描述如下。

Step1 初始化分组矩阵 $\mathbf{A}$ 、正则化参数 λ 以及步长 η ;

Step2 固定 $\mathbf{A}$ 不变,对于每一个分组 g ,分别调用标准 M-SVR(式(1)),求得模型参数矩阵 $\mathbf{W}_g$;

Step3 固定 W_g 不变,利用式(7)与 $\partial q_{gi}/\partial \alpha_{gi}$ 计算 $\partial P(\mathbf{A})/\partial \alpha_{gi}$,并采用最速下降法计算最优分组 $\mathbf{A}^*$;

Step4 若两次迭代中 $\mathbf{A}$ 或 W_g 的差值低于预设阈值,则算法结束,否则转至 Step2。

3 试验验证

本文采用两种 MIMO 数据集验证所提算法的有效性。第 1 种是仿真数据,目的是检验所提算法能否识别预设的分组结构;第 2 种是圆柱壳振动工程数据,目的是检验所提算法的预测效果。本文采用标准 M-SVR 和最小二乘 SVM(LS-SVM)^[6] 做对比测试,其中后者是对每个输出端独立建模。为便于记录,本文所提算法简称 GM-SVR。3 种算法都采用 RBF 核函数 $K(x_i, x_j) = \exp(\|x_i - x_j\|^2 / \sigma^2)$,核参数 $\sigma = 0.2$,正则化参数也保持一致;M-SVR 的正则参数 C 取值 50,LS-SVM 的正则参数 γ 和 GM-SVR 的参数 λ 都等于 $1/2C$ 。在训练前,所有样本被线性归一化到 $[-1, 1]$ 之间。

3.1 仿真数据

根据式(8)

$$\left. \begin{aligned} y_1 &= 4w_1 \sin x_1 - 2w_2 \frac{\sin(\pi x_2)}{\pi x_2} + 5 + n_{i1} \\ y_2 &= 3w_1 \sin x_1 - 3w_2 \cos x_2 + 2 + n_{i2} \\ y_3 &= \frac{-5w_1 \sin(\pi x_1)}{\pi x_1} + 4w_2 \sin x_2 + 1 + n_{i3} \end{aligned} \right\} (8)$$

生成符合独立同分布条件的 MIMO 样本集。输入为两维,均匀采样至 $[-1, +1]$ 。输出为 30 维,按照如下 3 个基准模型划分为 3 组,每组各有 10 个输出端,其中各组的模型参数(w_1, w_2)相差较大,同组内的参数则基本一致,具体数值为:第 1 组靠近(1,1),第 2 组靠近(0.8,0.8),第 3 组靠近(0.5,0.5)。 n_{ij} 是均值为 0、标准差为 0.5 的高斯噪声。共生成 50 个样本,其中 30 个用于训练,20 个用于测试。首先,测试 GM-SVR 能否辨识上述数据内在的分组结构。图 1 给出了矩阵 $\mathbf{A}$ 的分布图。

图 1 中,最终优化得到的 $\mathbf{A}$ 的元素 q_{gi} 为靠近 0 或 1 的小数,根据 q_{gi} 的值指代不同的分组。由图 1 可以看出,当 G 预设为 3 时,本文所提算法基本上辨识出了分组结构(除了两个输出端),相反,若 G 设得过大或过小,尽管仍有相当多的输出端被错误地划分,但所得结果依然能大致识别出预置的分组结构。这表明利用混合 0-1 规划的方式,将自适应分组转化为优化问题是有效的。

由于 q_{gi} 的优化结果依赖于初始值,因此重复 30

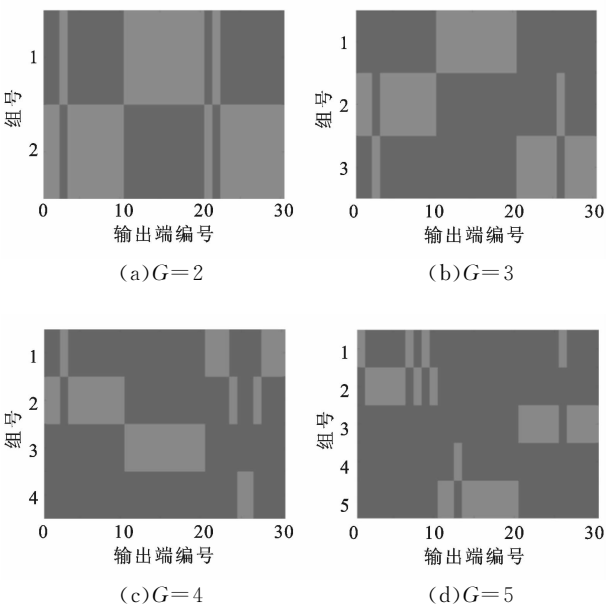


图 1 不同 G 初始值对应的分组矩阵 $\mathbf{A}$ 的分布图

次实验,表 1 给出了对应不同分组数目的 30 组均方根误差(E_{RMS})的平均值。这里考虑了两种特殊情况: $G=Q$ 和 $G=1$ 。前者意味着每个输出端独立训练,后者即为标准 M-SVR。由表 1 可知: $G=1$ 时预测误差明显低于 $G=Q$,这意味着 MIMO 建模方式可以有效利用输出端之间的信息;当 $G=3$ 时,预测误差最低,这表明对相关度高的输出端进行独立训练,可以避免其他输出端的干扰,有效提高预测精度。

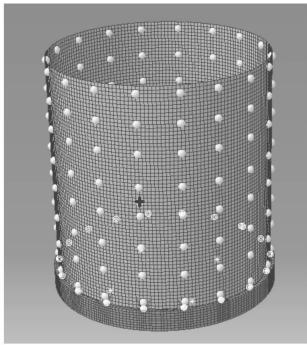
表 1 不同分组的均方根误差

G	Q	1	2	3	4
E_{RMS}	0.982	0.648	0.412	0.383	0.468

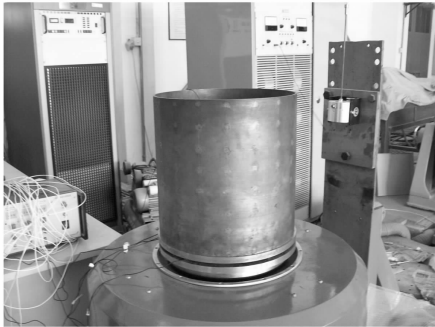
3.2 圆柱壳振动数据

圆柱壳是机械设计、航天航空领域常用的标准件。利用圆柱壳振动数据进行力学环境预示是一种常用的方法^[7],它的基本原理是认为同一结构在不同边界条件下的响应信号(如加速度响应)具有内在的相似性,因而可利用机械结构在一种边界条件下的响应信号,预测同一结构在另一种边界条件下的响应,从而降低试验成本,缩短设计周期。由于篇幅所限,此处不再给出基于 SVM 的力学环境预示的具体理论介绍。本文搭建了圆柱壳随机振动试验系统,并通过 LMS Virtual Lab 软件构建了有限元仿真模型,如图 2 所示。

图 2 中,壳体表面共设 8 圈测点,每圈放置 18 个加速度传感器。其中,图 2a 用于产生冲击时域数据,图 2b 用于产生随机振动数据,经过 Burge 法平



(a)有限元模型



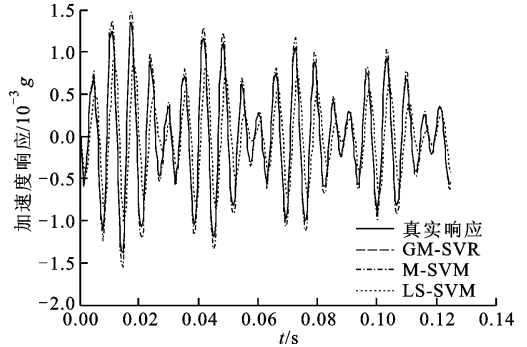
(b)随机振动试验系统

图 2 圆柱壳振动系统构成

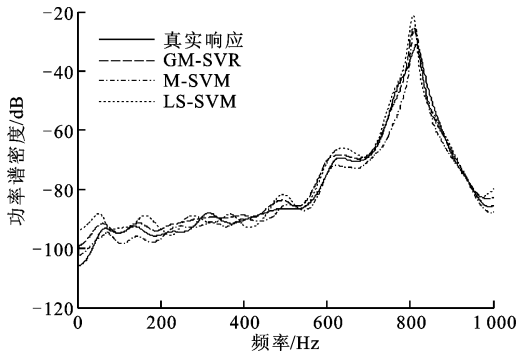
滑去噪后,转换为频域功率谱密度(PSD)数据。试验采用镁铝合金夹具的响应数据预测钢夹具的响应,具体试验参数可参见文献[7]。如果同时对壳体多个测点的响应信号进行预测,则力学环境预示模型就演变为一个 MIMO 模型。由动力学理论可知,由于呼吸模态的影响,不同测点的振动模态频率、振型存在差异。为直观显示,选择第 4 圈第 1 个测点为对象,图 3 给出了 3 种算法的预测效果对比。

由图 3 可知,GM-SVR 比 M-SVR 和 LS-SVM 更贴近真实信号曲线,尤其是在主峰处。这表明本文所提算法可有效识别相关度高的测点,从而更好地利用测点间信息,提高模型泛化性能。同时,M-SVR 的建模效果优于 LS-SVM,这同样表明 MIMO 建模的优势。图 3a 为仿真数据,不含噪声,因而预测曲线与真实曲线较为贴近。其他测点也有类似的对比效果,限于篇幅不再给出。表 2 给出了所有测点预测误差的平均值,可以看出,GM-SVR 和 M-SVR

的预测误差都远低于 LS-SVM,同时,GM-SVR 在两种误差指标上都取得了最低的误差。



(a)时域冲击数据



(b)频域随机数据

图 3 单一测点 3 种算法预测效果对比

4 讨论与展望

建立 MIMO 模型的关键在于有效利用不同输出端之间蕴含的领域知识,而在某些情况下,部分输出端之间的相关度要远高于其他输出端,这将导致相关度不高的输出端之间存在冗余或者负面的信息。本文在传统的 M-SVR 的基础上,引入自适应分组的概念,将相关度高的输出端置于同组进行正则化训练,不同组之间则无须交叉,所得到的优化目标为混合 0-1 规划,最终采用交替优化的方法,自适应识别最优分组结构和计算模型参数,并通过仿真数据和圆柱壳振动数据集验证了该算法的有效性。该方法主要通过划分正则项进行分组运算,因此也适用于其他形式的 MIMO SVM 算法。

表 2 3 种算法的平均预测误差

误差	时域冲击数据			频域随机数据		
	LS-SVM	M-SVR	GM-SVR	LS-SVM	M-SVR	GM-SVR
均方根误差	1.23×10^{-4}	2.27×10^{-5}	1.51×10^{-5}	6.39×10^{-3}	3.17×10^{-3}	2.45×10^{-3}
平均相对误差/%	12.93	5.72	4.21	26.23	16.12	13.25

(下转第 72 页)