第 27 卷 第 8 期
Vol. 27 No. 8

控 制 与 决 策
Control and Decision

2012 年 8 月
Aug. 2012

文章编号: 1001-0920(2012)08-1273-04

基于奇异值分解及 PRESS 统计的模型结构优化方法

李德才, 韩 敏

(大连理工大学 电子信息与电气工程学部，辽宁 大连 116023)

摘 要: 针对线性参数模型的基函数选择问题, 结合奇异值分解和 PRESS 统计提出一种模型结构优化算法. 通过预先对候选基函数矩阵进行分块操作, 减少非最优列间的重复比较. 在此基础上, 对各子块采用奇异值分解与 PRESS 统计相结合的方法进行选择, 直接以模型的泛化能力作为目标, 自适应地选择基函数. 通过奇异值分解, 在降低候选基函数数量的同时, 使其彼此之间相互正交, 有效地简化了 PRESS 统计的计算复杂度. 仿真结果表明, 所提出的方法能够有效简化模型结构, 并保持较高的预测精度.

关键词: 奇异值分解；模型结构优化；PRESS 统计；稀疏基选择

中图分类号: TP183 **文献标识码:** A

Model structure optimization based on SVD and PRESS statistic

LI De-cai, *HAN Min*

(School of Electronic and Information Engineering，Dalian University of Technology，Dalian 116023，China. Correspondent：HAN Min，E-mail：minhan@dlut.edu.cn)

Abstract: Focusing on the problem of the basis functions selection for the linear-in-the-weights regression models, this paper proposes a model structure optimization algorithm based on singular value decomposition(SVD) and PRESS(predicted residual sums of squares) statistic. By dividing the original candidate matrix into several parts firstly, the comparison among poor candidate regressors is avoided. Based on this operation, the SVD and PRESS statistic are applied to each sub-block for candidate regressors selection. The generalization error of the model is taken as the direct target, and the candidate regressors are selected in an adaptively manner. By using SVD, the number of the candidate regressors is reduced and the regressors are orthogonal with each other, which effectively reduces the computation burden of the PRESS statistic. Simulation results show the effectiveness of the proposed method.

Key words: singular value decomposition；model structure optimization；predicted residual sums of squares statistic；sparse bases selection

1 引 言

分析观测时间序列的演变规律是掌握系统动力学特性的重要手段, 而将具有一定相关性的多个时间序列作为一个整体进行研究, 有利于更好地了解各时间序列的演化规律. 但是, 多个输入变量的引入可能导致模型结构过于复杂, 从而影响算法的效率及精度. 因此, 有必要对模型结构进行有效控制, 在保证其学习能力的同时, 尽可能简化模型结构, 以便合理利用多变量数据所包含的信息.

根据原理不同可以将模型结构简化方法分为构造方法[1]和剪枝方法两类[2]. 文献 [3] 利用正交前向回归结合构造方法有效地控制了模型规模. 在此基础上, [4] 采用 PRESS (predicted residual sums of squares) 统计作为损失函数, 直接优化模型的泛化能力, 根据损失函数的变化情况, 自适应地选择模型基函数. 此外, 可以将正交前向回归方法同 D-optimality[5]准则相结合, 得到具有鲁棒性的简化结构. 然而, 在上述结构简化方法中, 正交前向回归的实现均基于经典的 Gram-Schmidt 正交化方法. 在进行最优基函数的选择过程中, 每选择一个基函数都几乎要将所有候选基函数进行一次正交化处理, 且正交化复杂程度随已选择基函数个数的增加而增大.

针对正交前向回归方法在基函数选择过程中存在的问题, 本文结合奇异值分解和 PRESS 统计提出

收稿日期: 2010-12-29；**修回日期:** 2011-03-28.

基金项目: 国家自然科学基金项目(61074096).

作者简介: 李德才(1983−), 男, 博士生, 从事复杂系统建模及预测的研究；韩敏(1959−), 女, 教授, 博士生导师, 从事神经网络理论及应用等研究.

一种新型的模型结构简化算法, 称为 OLS-SVD. 其主要思想是: 在基函数的选择过程中, 预先对候选基函数矩阵进行分块操作, 减少非最优列之间的重复比较. 此外, 通过对每个子块进行奇异值分解, 并采用得到的正交矩阵替代候选基函数矩阵, 实现候选基函数的批量正交化, 从而避免基于 Gram-Schmidt 分解的方法中频繁地正交化操作, 以达到简化计算复杂度的目的. 在此基础上, 结合 PRESS 统计方法直接优化模型的泛化能力, 自适应地选择基函数.

2 线性参数模型和 PRESS 统计

2.1 线性参数模型

考虑如下的非线性离散系统:

$$y(k)=f(y(k-1),\cdots,y(k-n_y),u(k-1),\\ \cdots,u(k-n_u))+e(k). \quad (1)$$

其中: $f(\cdot)$ 为待求的映射关系, $u(k)$ 和 $y(k)$ 分别为系统的输入和输出变量, $n_u\in\boldsymbol{N}$ 和 $n_y\in\boldsymbol{N}$ 分别为变量的系统延迟, $e(t)$ 为白噪声. 如果给定一组观测样本 $\{x(k),y(k)\}_{k=1:N}$, 其中 N 为观测样本的数量, 则可以通过建立适当的函数关系逼近 $f(\cdot)$, 即

$$\boldsymbol{y}=\boldsymbol{\varPhi}\boldsymbol{\theta}+\boldsymbol{e}. \quad (2)$$

其中: $\boldsymbol{y}=[y(1),y(2),\cdots,y(N)]^{\mathrm{T}}$, $\varPhi=[\varPhi_1,\varPhi_2,\cdots,\varPhi_{n_M}]$ 为模型的基函数, 且 $\varPhi_i=[\varPhi_i(x(1)),\varPhi_i(x(2)),\cdots,\varPhi_i(x(N))]^{\mathrm{T}}$, $i=1,2,\cdots,n_M$, n_M 为候选基函数的数量, $x(k)=[y(k-1),\cdots,y(k-n_y),u(k-1),\cdots,u(k-n_u)]^{\mathrm{T}}$ 为系统的输入向量; $\boldsymbol{\theta}=[\theta_1,\theta_2,\cdots,\theta_{n_M}]^{\mathrm{T}}$ 为模型的输出权值; $\boldsymbol{e}=[e(1),e(2),\cdots,e(N)]^{\mathrm{T}}$.

2.2 PRESS 统计

对于线性参数模型 (2), 设当前模型仅包含 q 个候选基函数, $\{x(k),y(k)\}_{k=1:N,-k}$ 为从训练数据 $\{x(k),y(k)\}_{k=1:N}$ 中移除第 k 个样本后得到的数据集, 对应的模型输出和残差分别为 $\hat{y}_{q,-k}(k)$ 和 $\varepsilon_{q,-k}(k)$. 因此, 对于包含 n_M 个候选基函数的线性参数模型, 其 PRESS 统计误差可以表示为

$$\varepsilon_{n_M,-k}(k)=y(k)-\hat{y}_{n_M,-k}(k)=\\ \frac{\varepsilon_{n_M}(k)}{1-\varPhi^{\mathrm{T}}(k)(\varPhi^{\mathrm{T}}\varPhi)^{-1}\varPhi(k)}, \quad (3)$$

其中 $\varepsilon_{n_M}(k)=y(k)-\hat{y}_{n_M}(k)$.

3 基于奇异值分解的稀疏前向正交模型

3.1 模型结构设计

由式 (4) 可以看出, 通过最小化 PRESS 统计误差可以得到较为理想的模型结构. 但是, 在式 (4) 中需要计算矩阵 $\boldsymbol{\varPhi}^{\mathrm{T}}\boldsymbol{\varPhi}$ 的逆, 因此求解过程往往伴随较大的计算量. Chen 等人[5]通过正交化方法, 将参数模型 (2) 转化为正交回归模型, 并将其引入基函数的选择过程中. 由于经过正交化处理, 此时候选基函数之间相互正交, 从而可以采用迭代方式计算模型的 PRESS 统计误差, 避免了矩阵求逆过程. 然而, 上述正交化方法在选择基函数时仍存在一些缺陷, 如在进行最优基函数的选择过程中, 每选择一个最优基函数都几乎要将所有候选基函数进行一次正交化处理, 且正交化过程的复杂度随已选择基函数个数的增加而增大; 另一方面, 每选择一个最优基函数都要利用所有未确定的候选基函数计算相应的 PRESS 统计误差, 而且计算复杂度随候选列向量的增加而增大, 从而当候选基函数的数量较大时, 将造成计算资源的浪费.

针对前述正交化选择方法中存在的问题, 本文结合奇异值分解和 PRESS 统计提出一种新的模型结构优化方法. 其具体实现过程如下所述.

假设由所有基函数组成的候选矩阵为 $\boldsymbol{\varPhi}_{\mathrm{all}}$. 首先将 $N\times n_{M_{\mathrm{all}}}$ 维矩阵 $\boldsymbol{\varPhi}_{\mathrm{all}}$ 按列等分为 d 个子块, 其中 d 为自定义参数, 可以根据观测数据的规模人为设定. 分块后得到的一个子块包含 $n_M(n_M\leqslant n_{M_{\mathrm{all}}})$ 个列向量, 若矩阵的秩为 p, 则对其进行奇异值分解, 有

$$\tilde{\boldsymbol{\varPhi}}=\boldsymbol{U}\begin{bmatrix}\boldsymbol{\varSigma} & O\\ O & O\end{bmatrix}\boldsymbol{V}^H. \quad (4)$$

其中: $\boldsymbol{\varSigma}=\mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_p)$; $\boldsymbol{U}$ 为 N 行 N 列的酉矩阵, 对应于矩阵 $\tilde{\boldsymbol{\varPhi}}^{\mathrm{T}}\tilde{\boldsymbol{\varPhi}}$ 的 N 个正交特征向量; $\boldsymbol{V}$ 为 M 行 M 列的酉矩阵. 对矩阵 $\tilde{\boldsymbol{\varPhi}}^{\mathrm{T}}\tilde{\boldsymbol{\varPhi}}$ 的特征值进行排序, 仅保留代表主要信息的前 r 个特征值, 并使其满足

$$\sum_{i=1}^{r}\sigma_i^2\Big/\sum_{j=1}^{p}\sigma_j^2>\eta_0, \quad (5)$$

其中 $\eta_0(0<\eta_0<1)$ 为自定义可调参数. 将矩阵 $\boldsymbol{U}$, $\boldsymbol{V}$ 和 $\boldsymbol{\varSigma}$ 进行分块, 有

$$\boldsymbol{V}=[\boldsymbol{V}_1|\boldsymbol{V}_2], \quad (6)$$

$$\boldsymbol{U}=[\boldsymbol{U}_1|\boldsymbol{U}_2], \quad (7)$$

$$\boldsymbol{\varSigma}=[\boldsymbol{\varSigma}_1|\boldsymbol{\varSigma}_2]. \quad (8)$$

其中: $\boldsymbol{U}_1$ 为 N 行 r 列, $\boldsymbol{U}_2$ 为 N 行 $(N-r)$ 列, $\boldsymbol{V}_1$ 为 M 行 r 列, $\boldsymbol{V}_2$ 为 M 行 $(M-r)$ 列的正交矩阵, $\boldsymbol{\varSigma}_1$ 为 r 行 r 列, $\boldsymbol{\varSigma}_2$ 为 $(p-r)$ 行 $(p-r)$ 列的对角阵.

根据式 (6)∼(8), 式 (4) 可以简化为

$$\tilde{\boldsymbol{\varPhi}}_r=\boldsymbol{U}_1\boldsymbol{\varSigma}_1\boldsymbol{V}_1^H, \quad (9)$$

其中 $\boldsymbol{U}_1=[\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_r]$. 令样本输出为 $\boldsymbol{y}$, 所要求取的权值矩阵为 $\boldsymbol{W}$, 可以得到线性参数模型 (2) 的输出表达式为

$$\boldsymbol{y}=\tilde{\boldsymbol{\varPhi}}\boldsymbol{\theta}+\boldsymbol{e}\approx\tilde{\boldsymbol{\varPhi}}_r\boldsymbol{\theta}+\boldsymbol{e}=\boldsymbol{U}_1\boldsymbol{\varSigma}_1\boldsymbol{V}_1^H\boldsymbol{\theta}+\boldsymbol{e}. \quad (10)$$

进一步定义 $\boldsymbol{g}=\boldsymbol{\varSigma}\boldsymbol{V}_1^H\boldsymbol{\theta}$, 则式 (9) 可以整理为

$$\boldsymbol{y}=\boldsymbol{U}_1\boldsymbol{g}+\boldsymbol{e}. \quad (11)$$

对比式 (2) 和 (11) 可以看出, 虽然两者形式上相似, 但在结构上却存在较大不同. 在式 (11) 中, 通过奇异值分解以及矩阵变换, 将原始候选矩阵替换为酉矩阵 $\boldsymbol{U}_1$, 实现了候选矩阵基函数的批量正交化, 避免了最优基函数选择过程中频繁的正交化操作. 另一方面, 在保证信息量损失最小的条件下, 通过奇异值分解实现了对候选矩阵的特征提取, 得到的约化矩阵 $\boldsymbol{U}_1$ 为 N 行 r 列, 较 N 行 $n_{M_{\rm all}}$ 列的原始候选矩阵 $\boldsymbol{\varPhi}_{\rm all}$ 具有更少的候选列向量, 从而减少了无关列向量之间的重复比较, 简化了正交化选择基函数的过程.

采用留一法交叉检验衡量模型的泛化能力. 与式 (4) 相似, 得到相应的 PRESS 统计误差为

$$\varepsilon_{r,-k}(k)=y(k)-\hat{y}_{r,-k}(k)=\\ \frac{\varepsilon_r(k)}{1-\boldsymbol{U}_1^{\rm T}(k)(\boldsymbol{U}_1^{\rm T}\boldsymbol{U}_1)^{-1}\boldsymbol{U}_1(k)}=\frac{\varepsilon_r(k)}{\eta_r(k)},\tag{12}$$

其中 r 为线性参数模型 (2) 中候选基函数的数目, 正交矩阵 $\boldsymbol{U}_1$ 的列向量满足

$$\boldsymbol{u}_i^{\rm T}\boldsymbol{u}_j=\begin{cases}0,\ i\neq j;\\ \kappa,\ i=j;\end{cases}\\ i,j=1,2,\cdots,n_M.\tag{13}$$

假设正交化选择过程中得到的子模型包含 n 个候选基函数, 定义均方 PRESS 统计误差为

$$J_n={\rm E}[\varepsilon_{n,-k}^2(k)]={\rm E}\Big[\Big(\frac{\varepsilon_n(k)}{\eta_n(k)}\Big)^2\Big]=\frac{1}{N}\frac{\varepsilon_n^2(k)}{\eta_n^2(k)}.\tag{14}$$

其中模型残差 $\varepsilon_n(k)$ 为

$$\varepsilon_n(k)=y(k)-\sum_{i=1}^{n}u_i(k)g_i=\\ \varepsilon_{n-1}(k)-u_n(k)g_n,\tag{15}$$

$$g_i=\boldsymbol{u}_i^{\rm T}\boldsymbol{y}/(\boldsymbol{u}_i^{\rm T}\boldsymbol{u}_i).\tag{16}$$

类似地, PRESS 统计的误差权重 $\eta_n(k)$ 为

$$\eta_n(k)=1-\sum_{i=1}^{n}\frac{u_i^2(k)}{\boldsymbol{u}_i^{\rm T}\boldsymbol{u}_i}=\eta_{n-1}(k)-\frac{u_n^2(k)}{\boldsymbol{u}_n^{\rm T}\boldsymbol{u}_n}.\tag{17}$$

3.2 计算复杂度分析

为了表明本文所提方法在计算效率方面的有效性, 采用基于 MGS 与线性参数模型相接合的正交选择算法 (OLS-MGS) 的计算复杂度与本文算法进行比较. OLS-MGS 方法的计算复杂度可以表示为

$$\underbrace{Nkp(k-1)-N\frac{k(k-1)(2k-1)}{3}}_{\rm MGS}+\\ \underbrace{3N(2p-k+1)k}_{\rm leave-one-out}.\tag{18}$$

其中: N 为训练样本的数量, p 为候选基函数的数量, k 为有效基函数的数量. 而在本文提出的 OLS-SVD 算法中, 为了提高算法的计算效率, 采用 Simon 等人提供的 Propack 工具箱实现奇异值分解, 其中奇异值分解过程的计算复杂度为

$$2m^2(N+p).\tag{19}$$

相应的 SVD-OLS 算法的计算复杂度为

$$\underbrace{3N(2m-k+1)k}_{\rm leave-one-out}+\underbrace{2m^2(N+p)}_{\rm SVD}.\tag{20}$$

其中: N 为训练样本的数量, m 为奇异值分解得到的约化矩阵中候选向量的数量, k 为有效基函数的数量. 通过计算复杂度的比较可知, 奇异值分解可避免基函数选择过程中频繁的正交化操作, 并降低了候选基函数的数量; 因此在计算复杂度方面较 OLS-MSG 方法具有一定的优势.

4 仿真实例

4.1 SinC 函数

采用径向基函数网络拟合如下的数值方程:

$$y(x)=\begin{cases}\dfrac{\sin(x)}{x},\ x\neq 0;\\ 1,\ x=0.\end{cases}\tag{21}$$

其中选择高斯核函数作为基函数, 其形式为

$$\varPhi_i(x)=\exp\Big(-\frac{||x-c_i||^2}{2\sigma^2}\Big),\tag{22}$$

c_i 和 σ^2 分别为高斯核函数的中心及核宽度. 为了表明训练样本规模对于正交选择算法的影响, 分别针对不同规模的训练样本进行 5 组实验, 实验结果如表 1 所示. 在每组实验中, 训练样本 $\{x(k),y(k)\}$ 由式 (21) 得到, 其中输入变量 x 服从 $(-10,10)$ 的均匀分布. 此外, 在 $(-10,10)$ 之间等间隔地采集 200 个服从函数关系 $f(x)$ 的数据, 并依次对模型的泛化效果进行验证. 为了体现实验条件的真实性, 在训练样本上叠加均值为零标准差为 0.2 的高斯噪声. 将每一个训练样本 $x(k)$ 视为 RBF 网络的一个候选中心, 因此候选基函数的数量同训练样本的规模相同. 此外, 经过多次实验确定最优核宽度为 $\sigma^2=10$.

表 1　OLS-MGS 和 OLS-SVD 对 SinC 数据的比较结果

方法	训练时间	训练误差	测试误差	训练样本数
OLS-MGS	2.931 2	0.042 9	0.042 4	200
OLS-SVD	0.232 0	0.040 8	0.042 1	200
OLS-MGS	12.003 1	0.028 7	0.029 4	400
OLS-SVD	0.812 5	0.029 0	0.029 2	400
OLS-MGS	26.575 0	0.033 1	0.033 0	600
OLS-SVD	2.112 5	0.030 5	0.031 0	600
OLS-MGS	47.712 5	0.023 5	0.024 4	800
OLS-SVD	3.968 8	0.023 7	0.023 8	800
OLS-MGS	73.509 4	0.017 2	0.017 9	1000
OLS-SVD	6.802 4	0.017 9	0.018 5	1 000

为了验证所提算法的有效性, 采用基于 MGS 与线性参数模型相接合的正交选择算法 (OLS-MGS) 以及本文提出的 OLS-SVD 算法对实验数据进行仿真, 对应的结果如表 1 所示.

由表 1 可见, 本文方法在训练和测试误差方面同基于 MGS 的正交化选择方法具有相似的结果; 但在训练时间方面, 本文方法在计算效率方面具有较明显的优势, 且训练时间随训练样本数目的增加变化相对平缓.

4.2 非线性动态控制系统

考虑如下的非线性动态方程[5]:

$$y(k)=\frac{y(k-1)y(k-2)y(k-3)u(k-2)(y(k-3)-1)}{1+y^2(k-2)+y^2(k-3)}+\frac{u(k-1)}{1+y^2(k-2)+y^2(k-3)},\tag{23}$$

其中系统输入 $u(k)$ 为服从 $[-1,1]$ 均匀分布的随机信号. 由式 (23) 采集 400 个含噪声信号, 其中前 200 组样本用于训练, 后 200 组样本用于模型检验. 为了将均值为零标准差为 0.05 的高斯噪声叠加至模型输出, 选择 RBF 网络对上述非线性系统进行拟合, 其中网络基函数选为

$$\varPhi_i(\boldsymbol{x}(k))=\|\boldsymbol{x}(k)-\boldsymbol{c}_i\|^2\log(\|\boldsymbol{x}(k)-\boldsymbol{c}_i\|),\tag{24}$$

对应的模型输入向量为

$$\boldsymbol{x}(k)=[y(k-1)\ \ y(k-2)\ \ y(k-3)\ \ u(k-1)\ \ u(k-2)]^{\mathrm{T}}.\tag{25}$$

同样地, 将每一个训练样本 $\boldsymbol{x}(k)$ 视为 RBF 网络的一个候选中心, 则网络中共包含 200 个候选基函数. 此外, 由于候选基函数的数量较少, 矩阵的分块数 d 取值为 1. 采用奇异值分解对式 (2) 中的候选矩阵进行处理, 得到 66 个正交候选基函数, 其中阈值 η 取为 0.99. 结合前向选择算法和 PRESS 统计对候选基函数做进一步选择, 最终确定最优子模型的基函数数量为 30.

为了验证本文方法的有效性, 分别与文献 [5] 中所采用的 OLS-PRESS, LROLS-PRESS, LROLS-MSE, RVM 和 CLS 五种方法的实验结果进行比较. 其中 OLS-PRESS, LROLS-PRESS, LROLS-MSE 三种方法基于 MGS 分解实现线性参数模型的正交化. 相应的比较结果如表 2 所示.

表 2 6 种方法在模型规模以及精度方面的比较结果

方法	规模	训练误差	PRESS 统计	测试误差
OLS-PRESS	51	2.280×10^{-3}	3.864×10^{-3}	5.187×10^{-3}
LROLS-PRESS	31	3.192×10^{-3}	3.706×10^{-3}	5.892×10^{-3}
LROLS-MSE	42	1.883×10^{-3}	3.067×10^{-3}	4.872×10^{-3}
RVM	42	1.598×10^{-3}	2.577×10^{-3}	4.935×10^{-3}
CLS	49	3.940×10^{-3}	7.607×10^{-3}	5.580×10^{-3}
OLS- SVD	30	2.527×10^{-3}	3.517×10^{-3}	5.210×10^{-3}

由于本文方法采用奇异值分解对矩阵 $\boldsymbol{\varPhi}$ 进行预处理, 通过选择适当的阈值 η 可以得到较低维数的正交候选矩阵, 在此基础上结合前向选择算法对候选基函数做进一步选择, 可以得到更为稀疏的模型结构. 经过正交选择过程, 最终选择 30 个基函数构成最优子集, 较除 LROLS-PRESS 方法外的其余 4 种方法具有更为稀疏的模型结构. 另一方面, 虽然本文方法在奇异值分解过程中忽略了较小奇异值对应的特征向量, 但由于较小特征值可视为对噪声的描述, 得到的正交基函数矩阵仍保留了原始候选矩阵 $\boldsymbol{\varPhi}$ 中的主要信息, 从而在稀疏化模型结构的同时仍能保证模型精度.

5 结 论

针对线性参数模型的基函数选择问题, 结合奇异值分解和 PRESS 统计提出了一种新的模型结构优化方法. 实例仿真结果表明, 本文提出的方法能够有效简化模型结构并保持较高的预测精度.

参考文献(References)

[1] Feng G, Huang G B, Lin Q, et al. Error minimized extreme learning machine with growth of hidden nodes and incremental learning[J]. IEEE Trans on Neural Networks, 2009, 20(8): 1352-1357.

[2] Rojas I, Pomares H, Bernier J L, et al. Time series analysis using normalized PG-RBF network with regression weights[J]. Neurocomputing, 2002, 42(1/2/3/4): 267-285.

[3] Mao K Z, Billings S A. Algorithms for minimal model structure detection in nonlinear dynamic system identification[J]. Int J of Control, 1997, 68(2): 311-330.

[4] Chen S, Hong X, Harris C J, et al. Sparse modeling using orthogonal forward regression with PRESS statistic and regularization[J]. IEEE Trans on Systems, Man and Cybernetics, 2004, 34(2): 898-911.

[5] Hong X, Chen S. M-estimator and D-optimality model construction using orthogonal forward regression[J]. IEEE Trans on Systems, Man and Cybernetics, 2005, 35(1): 155-162.