第 27 卷第 3 期
2010 年 3 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 27 No. 3
Mar. 2010

文章编号: 1000−8152(2010)03−0303−07

基于混合最小二乘支持向量机网络模型的非线性系统辨识

陈 杰, 朱 琳

(北京理工大学 信息科学技术学院 自动控制系, 北京 100081)

摘要: 针对基于输入输出数据的非线性系统辨识问题, 提出一种新的混合最小二乘支持向量机(LS-SVMs)网络模型及相应的学习算法. 该算法将系统的辨识问题动态自适应的划分为若干子问题, 将支持向量机(SVM)用于各子模块辨识; 通过分析模型的统计学特性, 给出基于整体框架优化的系统参数辨识方法. 针对系统中参数相关联的特性, 采用期望条件最大化(ECM)算法对其进行条件辨识, 同时结合正则化理论和最小二乘法, 保证各专家模块的结构风险最小化辨识原则. 试验结果表明, 该方法兼具良好的辨识精度和泛化性能.

关键词: 混合专家系统; 最小二乘支持向量机; 非线性系统辨识; 期望条件最大化; 正则化

中图分类号: TP183　　　**文献标识码**: A

New identification approach for nonlinear systems based on the combination network model of least squares and support vector machines

CHEN Jie，ZHU lin

(Department of Automatic control, School of Information Science and Technology,
Beijing Institute of Technology, Beijing 100081, China)

Abstract: A novel combination network model of least squares and support vector machines(MLS-SVMs) and the associate learning algorithm for identifying nonlinear systems based on the input-output data are proposed. In the model, the identification task is dynamically decomposed into several subtasks according to the physical or statistical natures of the problem. The SVMs are applied as learning machines to every subtask. After analyzing the statistical characteristics of the model in the formal characterization, we give an algorithm for training the MLS-SVMs, based on the frame optimizing principle. The expectation conditional maximization(ECM) algorithm is applied to solve the dependence problem of parameters. Regularization theory and least squares method assure the identification principle of minimal construction risk for expert modules. Experiment illustrates good performance of the proposed method by high approximation accuracy and generalization levels.

Key words: combination network model; least squares and support vector machines; nonlinear systems identification; ECM; regularization

1 引言(Introduction)

现代复杂工程系统越来越趋近于多元化、模块化,考虑系统本身的非线性、不确定性等特性, 基于单一模型的系统辨识方法显示出诸多的局限性, 因而将复杂非线性系统辨识问题分解为若干子问题, 以获得更好的精度和适应性以及训练算法的快速收敛性, 成为近年来智能建模中令人关注的研究方向. Jacobs[1]提出的混合专家系统(ME)正是秉承“分而治之”的理念, 可根据系统的输入输出特性动态提取相关重要信息自适应划分辨识问题, 实现对非线性系统的有效逼近, 其良好的适应性和建模能力受到诸多学者的关注, 在信号处理[2]、复杂时间序列建模[3,4]、系统辨识[5,6]等领域都得到成功的应用. 但与非模块化模型相比, ME系统因其模型化结构往往产生较大的误差方差,而使其推广泛化性能大大降低, 特别是存在小样本、高维数的问题.

支持向量机(SVM)因其结构风险最小化的学习原则, 对基于小样本数据的辨识和估计问题具有出色的学习推广性能, 且基于Mercer条件的核函数特性保证了其参数估计问题的全局最优解特性. 若将支持向量机融入混合模块化理念中, 有望实现两者的优势互补. Cao[7]提出利用SOM聚类算法将输入空

收稿日期: 2008−06−12; 收修改稿日期: 2009−05−10.
基金项目: 北京市教育委员会共建重点实验室资助项目(CSYS100070417) .

间分为若干子集, 分别在子集上训练各SVM, 集成输出结果实现复杂时间序列预测. Wei等[8]提出利用混合核函数提高运算速度和辨识能力, Nemmour[9]提出了利用模糊推理将输入空间划分为若干子空间. 上述方法虽然都借鉴了混合和支持向量机的理念, 但都仅单一考虑空间划分或支持向量机核函数混合, 没有从整体上对系统进行分析, 无法实现真正混合意义上整体优化建模.

本文提出了一种新的非线性系统的智能辨识模型–混合最小二乘支持向量机模型(MLS–SVMs), 该模型将支持向量机作为专家模块融入ME系统, 提高ME模型的推广能力、辨识精度. 首先系统地给出了混合最小二乘支持向量机的模型描述, 分析系统的统计学特性, 给出基于极大似然估计理论的全局模型参数辨识方法. 通过引入最小二乘法和正则化理论, 保证各专家训练时的SVM结构风险最小化的原则, 在不损松散性和鲁棒性的同时, 避免了求解凸二次规划问题, 结构简单, 算法简练.

2 混合LS-SVMs网络模型描述(MLS-SVMs model)

定义 1 非线性系统辨识的混合LS–SVMs网络模型描述如下:

$$y(x)=\sum_{j=1}^{M}g_j(x,\beta)y_j(x,\vartheta). \tag{1}$$

其中: $y(x)$为辨识输出结果, $x=\{y_{t-1},y_{t-2},\cdots,y_{t-p},v_{t-1},v_{t-2},\cdots,v_{t-q}\}^{\mathrm T}\in\mathbb{R}^n$为输入向量, v为输入量, q,p为输入输出阶次, M为网络的节点个数.

$y_j(x,\vartheta),\ j=1,2,\cdots,M$为专家网络的各节点输出, ϑ为网络节点的辨识参数, 选取支持向量机作为专家节点的输出模型, 如式(2)所示:

$$y_j=\varphi_j(x)w_j+b_j,\ j=1,2,\cdots,M. \tag{2}$$

$g_j(x,\beta),\ j=1,2,\cdots,M$为门网络的激励函数, β为网络节点的辨识参数. 可根据输入向量x自适应地将输入量划分为总和为1的若干区间, 通过“软竞争”机制选择最适合的描述该观测值的一个或几个专家模块, 类似于神经网络的权值作用. 满足如下约束$g_j(x,\beta)\geqslant 0,\ j=1,\cdots,M,\ \sum\limits_{j=1}^{M}g_j(x,\beta)=1$. 本文选取高斯函数为门激励函数, 如式(3)所示:

$$g_j(x,\beta)=\frac{\gamma_j p(x,v_j)}{\sum\limits_{l=1}^{M}\gamma_l p(x,v_l)}. \tag{3}$$

其中: $\Sigma_j\gamma_j=1,\ \gamma_j\geqslant 0,\ j=1,2,\cdots,M.\ p_j(x,v_j)$为概率密度函数, 如式(4)所示:

$$p_j(x,v_j)=(2\pi)^{-n/2}|\varSigma_j|^{-1/2}\cdot \exp[-\frac{1}{2}(x-\nu_j)^{\mathrm T}\varSigma_j^{-1}(x-\nu_j)]. \tag{4}$$

式中: n为x的维数, ν_j为中心矩, $\varSigma_j$是门函数的方差矩阵.

此时, 门函数$g_j(x,\beta)(j=1,2,\cdots,M)$可表示为$g_j(x,\gamma,\nu,\varSigma),\ j=1,2,\cdots,M.\ \gamma,\nu,\varSigma$为门函数的辨识参数.

3 混合LS–SVMs网络的统计特性分析(Statistic analysis of MLS–SVMs network)

3.1 系统统计特性描述(Statistic characteristic description)

回归估计问题可以描述为: 存在一个概率决策机制, 通过对一系列映射关系的有序概率决策, 实现x和y的映射关系. 在混合LS–SVMs网络模型中, 选定门网络作为此概率决策机制, g_j为每个决策的发生概率, 因而选择概率模型来描述g_j, 如高斯函数, 指数函数等. 专家网络则被认为是一系列的映射分布, 对每个子节点, 存在一个均值为y_j, 方差为σ_j^2的条件概率密度$\pi_j(y|x,\vartheta)$使得x被映射为y.

则整个系统条件概率密度可以描述如下:

$$f(y|x,\theta)=\sum_{j=1}^{M}g_j(x,\beta)\pi_j(y|x,\vartheta). \tag{5}$$

其中: θ为整个系统的所有参数向量, π_j为第j个专家模块的条件概率密度分布, 如式(6)所示:

$$\pi_j(y|x,\vartheta)=(2\pi)^{-1/2}(\sigma_j^2)^2\exp\{-\frac{(y-y_j)^{\mathrm T}(y-y_j)}{\sigma_j^2}\}. \tag{6}$$

其中y_j由式(2)定义. 给定式(5)的条件概率密度函数后, 便可将参数辨识问题转化为极大似然估计问题.

定义 2 给定输入输出样本集$\{x^k,y^k\}_{k=1}^N$, 系统的似然函数可定义如下:

$$L(\theta,x,y)=\prod_{k=1}^{N}f(y^k|x^k,\theta)=\prod_{k=1}^{N}\sum_{j=1}^{M}g_j(x^k,\beta)\pi_j(y^k|x^k,\vartheta). \tag{7}$$

3.2 基于完备数据的似然函数(Likelihood function based on complete data)

Dempster等提出的EM算法, 因其可靠的全局收敛和简便运算过程, 被大量的用于非完备数据条件下的极大似然估计问题. 为利用EM算法实现模型的估计问题, 首先定义指示变量z_j, 当y^k由第j个专家结果输出时z_j^k取值为1, 否则为0. 用指示变量集z表示所有的缺失数据z_j.

定义 3　基于缺失数据集z和给定输入输出样本集$\{x^k, y^k\}_{k=1}^N$, 系统的似然函数定义如下:

$$L(\theta, x, y, z) = \prod_{k=1}^{N} f(y^k|x^k, \theta) = \prod_{k=1}^{N}\sum_{j=1}^{M} z_j(g_j(x^k, \beta)\pi_j(y^k|x^k, \vartheta)). \tag{8}$$

为计算方便, 取对数似然函数:

$$l(\theta, x, y, z) = \log(L(\theta, x, y, z)) = \sum_{k=1}^{N}\log(f(y^k|x^k, \theta)). \tag{9}$$

由缺失数据z_j定义可得

$$f(y^k|x^k, \theta) = (g_j(x^k, \beta)\pi_j(y^k|x^k, \vartheta))^{z_j}. \tag{10}$$

定义 4　则系统的对数似然函数可表示为

$$l(\theta, x, y, z) = \sum_{k=1}^{N}\sum_{j=1}^{M} z_j \log(g_j(x^k, \beta)\pi_j(y^k|x^k, \vartheta)). \tag{11}$$

4 基于ECM的参数估计(Parameters estimation based on ECM)

4.1 基于EM的估计方法(Parameters estimation based on EM)

给定2.2部分中的对数目标似然函数定义后, 模型的参数估计问题可通过EM算法的E步和M步交替迭代解得, 具体如下.

E步　求$l(\theta, x, y, z)$的期望$Q(\theta, \theta(t))$:

$$Q(\theta, \theta(t)) = \mathrm{E}(l(\theta, x, y, z)|y^k, x^k) = \sum_{k=1}^{N}\sum_{j=1}^{M}\mathrm{E}[z_j|y^k, x^k]\log(g_j(x^k, \beta)\pi_j(y^k|x^k, \vartheta)) = \sum_{k=1}^{N}\sum_{j=1}^{M}\mathrm{E}[z_j|y^k, x^k][\log g_j(x^k, \beta) + \log(\pi_j(y^k|x^k, \vartheta))] = Q_\beta + Q_\vartheta. \tag{12}$$

当不可观测变量z_j与完备数据条件下对数似然函数成线性关系时, 条件期望中的不可观测变量z_j可直接用后验条验概率表示[4]. 因而式(12)中z_j即可用所有参数固定情况下(参数选取上个迭代所得的参数)的后验概率来代替, 即

$$\mathrm{E}[z_j|y^k, x^k] = h_j^k(t) = \frac{g_j(x^k, \beta)\pi_j(y^k|x^k, \vartheta)}{\sum_{l=1}^{M} g_l(x^k, \beta)\pi_l(y^k|x^k, \vartheta)}. \tag{13}$$

M步　求使期望最大的参数$\theta(t+1)$.

参数向量θ由门模块参数β和专家模块参数ϑ组成, 由式(12)可知, 门模块参数β只影响Q_β, 而专家模块参数ϑ只影响Q_ϑ, 则M步可被分成如下2个独立的最优化问题.

1) 对门模块参数β, 解决如下的优化问题:

$$\beta(t+1) = \arg\max\Big(\sum_{k=1}^{N}\sum_{j=1}^{M} h_j^k(t)\log(g_j(x^k, \beta))\Big). \tag{14}$$

式(14)优化问题可通过求解下述参数的一步运算得到, 具体过程参见文献[6]:

$$\gamma_j(t+1) = \frac{1}{N}\sum_{k=1}^{N} h_j^k(t), \tag{15}$$

$$\nu_j(t+1) = \frac{1}{\sum_{k=1}^{N} h_j^k(t)}\sum_{k=1}^{N} h_j^k(t)x^k, \tag{16}$$

$$\sum_j(t+1) = \frac{\sum_{k=1}^{N} h_j^k(t)[x^k - \nu_j(t)][x^k - \nu_j(t)]^{\mathrm{T}}}{\sum_{k=1}^{N} h_j^k(t)}. \tag{17}$$

2) 对专家模块参数ϑ, 考虑专家模块的独立性, 此处单独讨论每个专家模块参数ϑ_j, 解决如下优化问题:

$$\vartheta_j(t+1) = \arg\max\Big(\sum_{k=1}^{N} h_j^k(t)\log(\pi_j(y^k|x^k, \vartheta))\Big). \tag{18}$$

将式(6)代入式(18)中得

$$\vartheta_j(t+1) = \arg\max\Big[-\sum_{k=1}^{N}\frac{h_j^k(t)}{2\sigma_j^2}(y^k - y_j^k)^2 - \Big(\frac{1}{2}\log 2\pi - \log\sigma_j\Big)\sum_{k=1}^{N} h_j^k(t)\Big]. \tag{19}$$

4.2 基于ECM的关联参数估计(Associated parameters estimation based on ECM)

显然式(19)中的σ_j^2和y_j中的参数估计并不相互独立,无法得到直接的显式解, 此时, 借鉴ECM估计方法中的概念, 将式(19)的优化问题分为以下两步:

CME–步 1　在参数σ_j^2固定为$\sigma_j^2(t)$时, 估计y_j中的参数, 即求解如下优化问题:

$$y_j(t+1) = \arg\max\Big[-\sum_{k=1}^{N}\frac{h_j^k(t)}{2\sigma_j^2(t)}(y^k - y_j^k)^2 - \Big(\frac{1}{2}\log 2\pi - \log\sigma_j(t)\Big)\sum_{k=1}^{N} h_j^k(t)\Big]. \tag{20}$$

式(20)中的$-(\frac{1}{2}\log 2\pi - \log\sigma_j(t))\sum_{k=1}^{N} h_j^k(t)$为一常数, 则式(20)简化成如下表达式:

$$\vartheta_j(t+1) = \arg\max\Big[-\sum_{k=1}^{N}\frac{h_j^k(t)}{2\sigma_j^2}(y^k - y_j^k)^2\Big]. \tag{21}$$

其中: y_j由式(2)定义, 待辨识的参数为w_j和b_j, 显然从式(22)估计w_i和b存在不适定性, 为克服不适定性问题, 采用Tikhonov[12]正则化方法处理. 式(22)可转化为如下优化问题:

$$\min_{\mu_j\in H} \mathrm{E}[\mu_j] \equiv$$
$$\min_{\mu_j\in H} C\sum_{t=1}^{N}(\phi_j^k(y^k-y_j^k)^2)+\frac{1}{2}\Omega(y_j). \quad (22)$$

其中: $\phi_j^k=\dfrac{h_j^k(t)}{\sigma_j^2}$为引入的正则化算子或镇定泛函, 它起到圆滑(压制奇异性)的作用. $C>0$为正则化参数或折衷系数, 协调误差范数和镇定泛函间关系. 令$e^k=y^k-y_j^k$, 则式(21)问题转化为约束条件下的问题:

$$\min_{\mu_j\in H} \mathrm{E}[\mu_j]\equiv \min_{\mu_j\in H} C\sum_{t=1}^{N}(\phi_j^k(e^k)^2)+\frac{1}{2}w^{\mathrm{T}}w, \quad (23)$$

$$\text{s.t. } y^k=y_j^k+e^k=w_j\varphi_j(x^k)+b_j+e^k. \quad (24)$$

则将上述优化问题转化到对偶空间, 得如下Lagrange函数:

$$L(w_j,b_j,e_j;\alpha_j)=$$
$$C\sum_{t=1}^{N}(\phi_j^k(e_j^k)^2)+\frac{1}{2}w_j^{\mathrm{T}}w_j+$$
$$\sum_{t=1}^{N}\alpha^k(y^k-w_j\varphi_j(x^k)-b_j^k-e_j^k). \quad (25)$$

对上式各变量求导, 并整理得到如下线性方程组:

$$\begin{bmatrix}0 & 1^{\mathrm{T}}\\ 1 & M_j+\varPhi_j\end{bmatrix}\begin{bmatrix}b_j\\ \alpha_j\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix}. \quad (26)$$

其中:

$$y=[y_1,y_2,\cdots,y_N],\ \alpha_j=[\alpha_{j1},\alpha_{j2},\cdots,\alpha_{jN}],$$
$$\varPhi_j=\mathrm{diag}\{(C\phi_j^1)^{-1},(C\phi_j^2)^{-1},\cdots,(C\phi_j^N)^{-1}\},$$
$$\phi_j^k=h_j^k(t)/\sigma_j^2(t),\ 1^{\mathrm{T}}=[1,1,\cdots,1]^{\mathrm{T}};$$

M_j为核矩阵, 其第r列和第s列的元素为

$$M_j(r,s)=K(r,s)=\varphi_j(x^r)^{\mathrm{T}}\varphi_j(x^s),$$
$$s,t=1,2,\cdots,N,$$

其中核函数$K(r,s)$为满足Mercer定理的函数. 本文选用3种核函数: 线性函数$K(x_r,x_s)=x_r^{\mathrm{T}}x_s$, 高斯径向基函数(RBF)

$$K(x_r,x_s)=\exp(-(x_r-x_s)^{\mathrm{T}}(x_r-x_s)/\sigma^2)$$

和指数径向基函数(ERBF)

$$K(x_r,x_s)=\exp(-|x_r-x_s|/2\sigma^2).$$

则每个专家模块参数ϑ_j的估计问题, 可通过求解线性方程组解决, 避免原求解二次规划问题, 线性方程组与单一加权LS–SVM的线性方程组形式上完全一致, 保证了本文所提训练方法贯穿支持向量机的训练理念, 与加权LS–SVM训练过程和方法一致, 同时由于权值$C\phi_j^k$的引入, 解决了最小二乘引入带来的弥散性和鲁棒性丧失的问题[10,11].

CME-步 2 在y_j中参数固定为更新参数时, 估计σ_j^2. 可得

$$\sigma_j^2(t+1)=\frac{\sum\limits_{k=1}^{N}h_j^k(t)(y^k-y_i(t+1))^2}{\sum\limits_{k=1}^{N}h_j^k(t)}. \quad (27)$$

至此整个M步中门网络函数中的参数值和专家网络中的参数值都已得到修正.

4.3 辨识步骤(Identification steps)

Step 1 初始化门函数. 随机选择某些点作为高斯曲线的质心, 将方差设置为训练集的方差.

Step 2 初始化专家模块. 利用K–means将例程分为M类, 其中M是专家的个数; 训练各专家j通过m类中的一个(求解线性方程组(25)), 假定$\phi_j^k=1, j=1,2,\cdots,M$.

Step 3 计算式(12)的期望$Q(\theta,\theta(t))$, θ采用上次迭代所得的新参数.

Step 4 如果$Q(\theta,\theta(t))$变化值小于某个给定的值(文中为0.001)或迭代次数大于给定值(文中为100次), 结束训练, 否则继续;

E步 对每对(x^k,y^k), 依据当前的参数值, 计算后验概率z_j^k; 计算$\phi_j^k=\dfrac{z_j^k}{\sigma_j^2}$;

MG-步 依式(17)～(19)计算门函数参数;

CME-步 1 对任意j, 在参数σ_j^2固定为$\sigma_j^2(t)$时, 求解线性方程组(25);

CME-步 2 对任意j, 在y_j中参数固定为更新参数时, 依式(26)计算σ_j^2;

Step 5 $t=t+1$, 返回Step 3.

5 应用实例(Example)

作者结合系统的时延特性、输入和输出变量间的非线性特性、系统的阶段响应特性等, 选取文献[14]中具有代表性的两个非线性系统, 其中仅获取系统的输入输出数据对结构进行辨识, 并对系统的输入数据做归一化处理. 每个实验采用3组数据, 其中1组用于系统辨识, 1组用于辨识验证, 另1组用于最终结果的测试. 每组实验重复10次, 记录其均值和方差. 文中给出3种实验结果数据: 1) 各组实验中的最好结果; 2) 10次实验的方差和均值; 3) 各种方法

的比较结果.

5.1 验证条件(Validated conditions)

1) 混合LS–SVMs模型中的专家节点个数的选取. 实验验证中为避免因专家节点个数过多使得每个子专家仅集中于过小的一个输入区域而导致的过学习问题. 在实验时尽可能选择少量的专家模型对系统进行逼近, 本文专家节点个数的选取通过使系统BIC值达到最小而得到. 实例中节点个数为2.

2) SVM的正则算子系数和核函数σ^2的选取. 在文中为了避免由于支持向量基参数的不同选取造成LS–SVM和混合LS–SVMs的结果无法比较, 作者根据经验选定LS–SVM和混合LS–SVMs中的SVM正则算子系数均为$C = 10^4$, RBF和ERBF的核函数均为$\sigma^2 = 1$.

3) 训练终止条件. 文中采用文献[13]中的early stopping策略, 即当似然条件期望之差小于预先设定阈值或者迭代次数大于设定值时, 迭代终止得到辨识模型.

5.2 实例 1(Example 1)

给定如下2阶离散非线性系统[14], 其输出$y(t)$和输入$u(t)$之间的关系用下列差分方程描述

$$y(t+1) = \frac{y(t)y(t-1)[y(t)+2.5]}{1+y^2(t)+y^2(t-1)} + u(t). \quad (28)$$

选择式中$y(t), y(t-1), u(t)$为输入变量, 估计输出值$y(t+1)$, 其中$u(t)$是在$[-2,2]$内均匀分布的随机变量. 生成1000个样本数据, 其中500个样本用来训练系统, 500个用来验证系统. 随后采用两个输入变量$u_{1t} = \sin(2\pi t/25)$, 采样100次和$u_{2t} = 1.5\times\sin(2\pi t/30)$, 采样120次对系统进行测试,采用均方误差作为性能评估标准.

图1给出了测试时基于ERBF核函数的最佳输出结果.

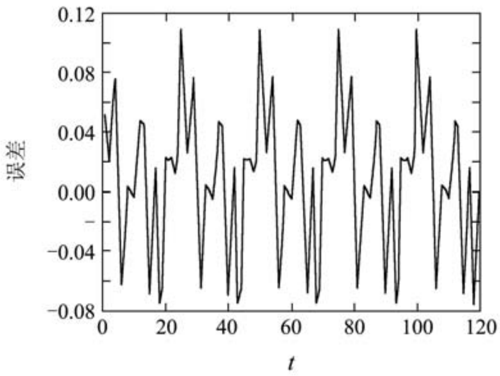

图 1 例1误差曲线

Fig. 1 Error of example 1

为了比较, 在相同条件下, 同时对LS–SVMs和其他ME结构进行验证, 表1中给出了几种算法的均方差. 其中第1列为所采用的方法, 第2列为针对基于500个样本数据训练好的模型, 采用样本验证数据中的另外500个数据验证得到的结果(10次实验的均方差均值和方差), 第3列为采用输入u_{1t}作为输入数据得到的结果, 第4列为采用u_{2t}作为输入数据得到的结果.

表 1 实例1算法均方差对比表

Table 1 MSD versus of example 1

所用方法	均方差均值		
	验证集	测试1(输入u_{1t})	测试2(输入u_{2t})
ME-AR	0.032871	0.037945	0.030092
LS-SVM-RBF	0.013366	0.012895	0.010562
LS-SVM-ERBF	0.010763	0.010321	0.010503
ME-MLP	0.008032	0.009124	0.008882
MLS-SVMs-linear	* 0.004516	* 0.004769	* 0.004656
MLS-SVMs-RBF	* 0.003285	* 0.003152	* 0.003321
MLS-SVMs-ERBF	* 0.003331	* 0.003721	* 0.003509

从表1可以看出, 在验证集和两次测试中, 不论核函数如何取值, 混合LS–SVMs模型都可以给出比常规LS–SVMs模型和MEs模型更精确的结果. 则可以看出混合LS–SVMs的泛化能力, 混合LS–SVMs占据了训练阶段表现最好的前3个模型, 在测试集中依然保持良好的辨识效果和拟合性能, 足以证明其良好的泛化能力.

5.3 实例 2 (Example 2)

给定如下离散非线性系统[14], 其输出$y(t)$和输入$u(t)$之间的关系用下列差分方程描述:

$$y(t+1) = \frac{y(t)}{1+y^2(t)} + u^3(t). \quad (29)$$

选择式中$y(t), u(t)$为输入变量, 估计输出值$y(t+1)$. 选定$u(t)$是在$(-1,1)$内均匀分布的随机变量. 生成1000个样本数据, 其中500样本用来训练系统, 500个用来验证系统. 采用如式(31)的输入变量来对系统进行测试. 图2给出了测试时基于RBF核函数的混合LS–SVMs的最佳输出结果.

$$u(t) = \begin{cases} \sin(2\pi t/250), & 如 0 \leqslant t \leqslant 500, \\ 0.8 \cdot \sin(2\pi t/250) + 0.8 \cdot \sin(2\pi t/25), & 如 t > 500. \end{cases} \tag{30}$$

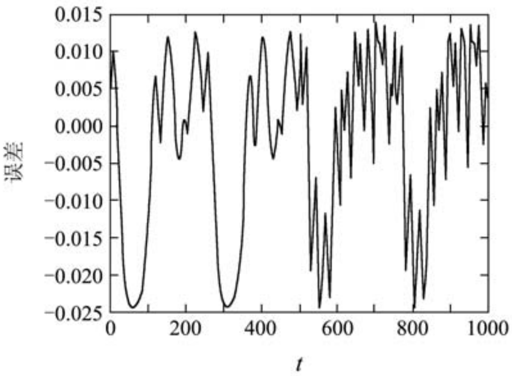

图 2 例2误差曲线

Fig. 2 Error of example 2

表 2 实例2算法均方差对比表

Table 2 MSD versus of example 2

所用方法	均方差	
	验证集	测试集
ME-AR	0.042369	0.041061
LS-SVM-RBF	0.014695	0.028963
LS-SVM-ERBF	0.020983	0.030690
ME-MLP	0.026542	0.015982
MLS-SVMs-linear	* 0.020210	* 0.01003
MLS-SVMs-RBF	* 0.012353	* 0.00669
MLS-SVMs-ERBF	* 0.010366	* 0.00313

从表2可以看出, 虽然混合LS–SVMs模型和LS–SVM模型在样本验证集误差相差不是很大, 但在测试集上却表现出了比LS–SVM模型好得多的辨识效果, 测试集的控制量被设计成两个不同的控制阶段和控制量, 影响系统的跟踪和输出结果. 从上个实验中可以看出, 正是因为混合LS-SVMs模型可以自适应地根据输入变量选择相应的SVM专家节点在不同的输入阶段, 使得混合LS–SVMs具有更好的泛化能力. 此外, 在文中试验中对于SVM的RBF/ERBF核函数并不是通过优化策略选择的, 而是直接选定的. 如果能对δ^2和C进行优化, 必然会获得更优的结果[15].

5.4 实例 3(Example 3)

为了进一步验证文中所提方法的有效性, 将其应用于高精度温控系统的模型辨识中, 很明显该系统是一个典型的非线性系统, 该模型的输入为: 输入电流I, 现在温度T_1, T_2, 其中T_1为环境中温度高的点, T_2为相对温度较低的一个点. 选定100组输入输出数据作为验证数据, 图3给出了输出的误差曲线. 从图中可以看出文中所提方法能够有效的预测出实际输出, 得到满意的预测效果.

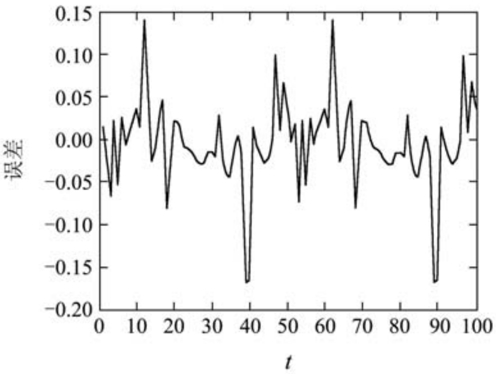

图 3 输出误差曲线

Fig. 3 Estimation error

从上述实验中可以看出, 混合LS–SVMs具有更好的鲁棒性和适应性, 尤其是在对各种不同的系统辨识要求时, 都给出了很好的辨识能力和辨识精度. 从整体上明显优于只考虑单一特性的辨识模型.

6 结论(Conclusion)

文中提出了一种新的非线性系统的辨识模型, 该模型将ME模型和支持向量机巧妙的融于同一辨识框架中, 给出了基于整体框架优化的系统参数辨识方法, 针对专家网络参数相关联的特性, 采用ECM算法对其参数进行条件辨识, 同时结合正则化规则约束专家模块的目标函数, 得到无论从性能和复杂度上都与加权最小二乘支持向量机参数辨识问题相一致的线性方程组. 通过对不同非线性系统的试验验证, 混合LS–SVMs具有更好的鲁棒性和适应性, 尤其是在对泛化能力和辨识精度都有很高要求时, 能给出很好的辨识效果.

参考文献(References):

[1] JORDAN M, JACOBS R. Hierarchical mixtures-of-experts and the EM algorithm[J]. *Neural Computation*, 1994, 6(2): 181 – 214.

[2] KURNIK R, OLIVER J, WATERHOUSE S. Application of mixture of experts algorithm for signal processing in a noninvasive glucose monitoring system[J]. *Sensor and Actuators B*, 1999, 60(1): 19 – 26.

[3] CARVALHO A, TANNER M. Mixtures-of-experts of autoregressive time series[J]. *IEEE Transactions on Neural Network*, 2005, 12(1): 39 – 56.

[4] SHU K N, GEOFFREY J M. Using the EM algorithm to train neural networks: misconceptions and a new algorithm for multiclass classification[J]. *IEEE Transactions on Network*, 2004, 15(3): 738 – 749.

[5] LIMA C A M, COELHO A L V, ZUBEN F J V. Mixture of experts applied to nonlinear dynamic systems identification: a comparative study[C] //*Proceedings of the VII Brazilian Symposium on Neural Networks*. Los Alamitos, CA, USA: IEEE Computer Society, 2002: 162 – 167.

[6] FENG S, CHEN J. Low-angle reflectivity modeling of land clutter[J]. *IEEE Transaction on Geoscience and Remote Sensing Letter*, 2006, 3(2): 254 – 258.

[7] CAO L J, ZHANG J G. A mixture of support vector machines for time series forecasting[J]. *Neural Network World*, 2006, 16(5): 381 – 397.

[8] WEI L W, LI J P, CHEN Z Y. Credit risk evaluation using support vector machine with mixture of kernel[C] //*7th International Conference on Computational Science, ICCS 2007*. Beijing: Springer, 2007: 431 – 438.

[9] WNEMMOUR H C Y. Fuzzy integral for a rapid mixture of support vector machines[C] //*Proceedings of the International Joint Conference on Neural Networks (IJCNN 2005)*. Montreal, QC, Canada: IEEE, 2005: 901 – 906.

[10] LIMA C A M, COELHO A L V, ZUBEN F J V. Hybridizing mixtures of experts with support vector machines Investigation into nonlinear dynamic systems[J]. *Information Sciences*, 2007, 177(10): 2049 – 2074.

[11] SUYKENS J A K, DE BRABANTER J, LUKAS L, et al. Weighted least squares support vector machines: robustness and sparse approximation[J]. *Neurocomputings*, 2002, 48(10): 85 – 105.

[12] TIKHONOV A N, ARSENIM V Y. *Solutions of Ill-posed Problems*[M]. New York: Halsted Press, 1977.

[13] HAYKIN S. *Neural Networks—A Comprehensive Foundation*[M]. Upper Saddle River, NJ: Prentice-Hall, 1999.

[14] RUANO A E. *Intelligent Control Systems Using Computational Intelligence Techniques*[M]. Piscataway, NJ, USA: IEE, 2005.

[15] ANAYA J J, ULLATE L G, FRITSCH C. A method for real-time deconvolution[J]. *IEEE Transactions on Instrumentation and Measurement*. 1992, 41(3): 413 – 419.

作者简介:

陈 杰 (1965—), 男, 教授, 博士, 主要从事智能控制与系统辨识方向的研究, E-mail: chenjie@bit.edu.cn;

朱 琳 (1983—), 女, 博士研究生, 主要从事非线性系统辨识方向的研究, E-mail: jennybit@gmail.com.

下 期 要 目