第 23 卷 第 1 期 控 制 与 决 策 2008 年 1 月
Vol. 23 No. 1 Control and Decision Jan. 2008

文章编号: 1001-0920(2008)01-0065-05

一种新的回归型约简多分辨率相关向量机

丁二锐[a], 曾 平[a], 丁 阳[b], 王义峰[a]
(西安电子科技大学 a. 计算机学院, b. 电子工程学院, 西安 710071)

摘 要: 提出一种新的稀疏贝叶斯回归算法. 基于相关向量机,首先通过尺度核和小波核构造完备基以提高预测精度;然后利用保局投影对输入矩阵的列进行主成分提取以减少训练时间,从而形成算法的初步模型. 为进一步减小较大规模训练数据集的回归时间压力,算法对训练数据集的分层采样建立了初步模型,进而产生实际较小规模的训练数据集. 实验结果表明,算法在预测精度和鲁棒性上优于传统支持向量机和相关向量机,且其训练时间较相关向量机少.

关键词: 相关向量机回归; 尺度核函数; 小波核函数; 保局投影; 数据采样
中图分类号: TP181; TP301.6 **文献标识码**: A

A novel regression algorithm of reduced multi-resolution relevance vector machine

DING Er-rui[a], *ZENG Ping*[a], *DING Yang*[b], *WANG Yi-feng*[a]
(a. School of Computer Science and Technology, b. School of Electronic Engineering, Xidian University, Xi'an 710071, China. Correspondent: DING Er-rui, E-mail: roger_ting@126.com)

Abstract: To further improve the prediction accuracy and running efficiency, a novel sparse Bayesian algorithm for regression is proposed. Based on relevance vector machine, a set of complete basises are constructed by combing scaling and wavelet kernels to increase the prediction accuracy, and then the principal components of input matrix columns are extracted by using locality preserving projections to reduce the training time, which forms a primary model. To further reduce the time pressure for a larger training data set, the algorithm creates a smaller training data set via the primary model based on a stratified sample. Experimental results of artificial and real data show that the proposed algorithm is superior to traditional support vector machine and relevance vector machine in both prediction accuracy and robustness and has less training time than relevance vector machine.

Key words: Relevance vector machine for regression; Scaling kernel; Wavelet kernel; Locality preserving projections; Data sampling

1 引 言

基于统计学习理论的结构风险最小化原则, Vapnik[1]提出的支持向量机(SVM)通过最小化经验风险和置信范围提高了算法的泛化能力. 与以训练误差最小化为优化目标的方法不同,支持向量机以训练误差作为约束条件,以置信范围最小化作为优化目标,其泛化能力优于神经网络等传统学习方法. 另外,支持向量机的解是全局最优的并且是唯一的,因而,支持向量机一经出现就受到广泛的重视.

然而支持向量机却存在以下缺点[2]: 1) 支持向量数目基本上随训练集的规模线性增长; 2) 其输出为一个点估计,而不是概率的; 3) 在回归中需估计 ε 不敏感参数,因而需进行交叉验证; 4) 其核函数需要满足 Mercer[2]条件. 为摆脱以上问题, Michael[2]提出一种与支持向量机相似的稀疏概率模型,即相关向量机(RVM),其训练是在贝叶斯框架下进行的,可以进行回归估计预测,从而获得预测值的分布,最后得到一个基于核函数的稀疏解. 相比支持向量机,相关向量机克服了上述所有缺点,已被证实在回归精度方面优于前者[2],并成功地应用于混沌时间序列预测[3,4]和电力负荷预测[5]等领域. 然而相关向量机需要不断迭代计算超参数,因而比支持向量机的

收稿日期: 2006-09-29; **修回日期**: 2006-12-25.
基金项目: 国家部委预研基金项目(413160501); 西安电子科技大学研究生创新基金课题(05008).
作者简介: 丁二锐(1980—),男,西安人,博士生,从事统计学习理论的研究; 曾平(1956—),男,重庆人,教授,博士生导师,从事色彩管理、图形图像处理等研究.

训练时间长,而且当训练集的规模较大时,其训练时间往往难以忍受. 此外,相关向量机假设采用的核函数为解空间的一个完备基,但这一假设并不合理.

在保持相关向量机优点的基础上,本文通过选取尺度核函数及小波核函数构造解空间的完备基,应用保局投影(LPP)[6]约简输入矩阵列维数,并对训练数据集采样建模得到较小规模的训练数据集,进一步减小训练时间,从而提出了约简多分辨率相关向量机(RMR_RVM)算法. 实验结果表明了算法在预测精度、鲁棒性和训练时间上的优越性.

2 相关向量机的回归原理

对于样本集$(x_1, y_1), \cdots, (x_M, y_M)$,设 y 独立且数据噪声服从方差为$\sigma^2$ 的高斯分布,则相关向量机最后的回归形式$f(x)$ 可表示为

$$f(x) = \beta_0 + \sum_{j=1}^{M} k(x, x_j)\beta_j. \tag{1}$$

其中:$k(x, x_j)$ 为核函数,β_j 为回归系数. 对应系数不为零的向量称为相关向量. 结合式(1),则样本集的似然函数为

$$p(y \mid \beta, \sigma^2) = (2\pi\sigma^2)^{-M/2} \exp\left\{-\frac{1}{2\sigma^2}\|y - \Phi\beta\|^2\right\}. \tag{2}$$

其中:β为回归系数的矩阵向量形式;Φ为$M \times (M+1)$ 输入矩阵,且$\Phi_{nm} = k(x_n, x_m)$,$\Phi_{n1} = 1$.

根据先验概率分布和似然分布,利用贝叶斯公式求得回归系数β的后验概率分布为

$$p(\beta \mid y, \alpha, \sigma^2) = (2\pi)^{-(M+1)/2} |\Sigma|^{-1/2} \times \exp\left\{-\frac{1}{2}(\beta - \mu)^{\mathrm{T}} \Sigma^{-1} (\beta - \mu)\right\}. \tag{3}$$

其中:$\Sigma = (\Phi^{\mathrm{T}} B \Phi + A)^{-1}$,$\mu = \Sigma \Phi^{\mathrm{T}} B y$,$B = \sigma^2 I_M$,$A = \mathrm{diag}(\alpha_0, \alpha_1, \cdots, \alpha_M)$,$\alpha$为超参数.

RVM 中回归系数β的估计值可由式(3) 的后验分布的均值给出,即可通过最大后验(MAP) 估计得到. MAP 估计取决于超参数和噪声方差,两者可通过最大化边缘似然分布得到. 超参数和噪声方差的边缘似然分布如下式所示:

$$p(y \mid \alpha, \sigma^2) = (2\pi)^{-M/2} |B^{-1} + \Phi A^{-1} \Phi^{\mathrm{T}}|^{-1/2} \times \exp\left\{-\frac{1}{2} y^{\mathrm{T}} (B^{-1} + \Phi A^{-1} \Phi^{\mathrm{T}})^{-1} y\right\}. \tag{4}$$

3 约简多分辨率相关向量机

3.1 尺度核函数和小波核函数的结合及构造

在相关向量机中,核函数往往只从一类核函数中选取,如仅选取高斯核函数. 这类核函数通过平移不可能生成平方可积空间 $L_2(R^N)$ 上的一组完备基[7]. 但在大多数实际应用中,要求回归模型能逼近平方可积空间 $L_2(R^N)$ 中的函数[8]. 按照多分辨分析观点,$L_2(R^N)$ 函数将投影到不同尺度下的小波空间和某个较大尺度下的尺度空间,如

$$L_2(R^N) = V_i \oplus W_i \oplus W_{i+1} \oplus W_{i+2} \cdots. \tag{5}$$

其中:V_i 为尺度空间,$W_{i+0,1,2\cdots}$ 为小波空间. 根据多分辨分析理论,文献[9]提出了小波核函数,但小波核函数仅能捕获信号高频部分而会丢失低频信息,而且在样本数据较稀疏的情况下,可能无法逼近频率较丰富的信号. 文献[7]提出一种尺度核函数,但该尺度核函数仅当 i 充分大时,才近似认为其尺度空间 V_i 为平方可积空间 $L_2(R^N)$. 为了同时捕捉高频和低频信息,提议将相关向量机的单个核函数扩展成多个核函数(此处为尺度核函数和小波核函数,记为多分辨率核函数),这样才有可能得到解空间的完备基. 多个核函数[10]可以构造一个多尺度问题的合适解,而且容易设置超参数. 多个核函数可视为一个生成函数词典,即基词典,其对应的解为从词典中估计出的稀疏解.

关于如何从核函数词典中选择核函数,Mallat[11]提出一个逐步构造线性回归函数的算法. 该算法是迭代的,每一步从词典中选择那些最能反映原始信号(即待估计函数) 的基函数. 而 Chen[12]则考虑词典中所有的基函数,力求寻找最好的线性组合估计. 在构造多分辨率相关向量机时,考虑采用 Chen 的方法,因为该方法比较简单,且适合相关向量机的扩展.

考虑核函数词典 $K = [k_1 \cdots k_i \cdots, k_P]$,则式(1)变为

$$f(x) = \beta_0 + \sum_{j=1}^{M}\sum_{i=1}^{P} \beta_{ji} k_i(x, x_j), \tag{6}$$

其中 $k_i(x, x_j)$ 为尺度核函数及不同尺度下的小波核函数. 记式(6) 的矩阵形式为

$$F(X) = \Phi\beta. \tag{7}$$

尺度核函数和小波核函数的构造可依下述定理[7,9]实现:

定理 1 设 $\phi(x)$,$\varphi(x)$ 分别为一维尺度函数和其对应的小波,a 为伸缩因子,其中 $x, a \in R$. 若 $x, z \in R^N$,则平移不变尺度核函数和对应小波核函数分别为

$$k(x, z)_{\mathrm{scaling}} = \prod_{i=1}^{N} \phi(x_i - z_i),$$

$$k(x, z)_{\mathrm{wavelet}} = \prod_{i=1}^{N} \varphi\left(\frac{x_i - z_i}{a}\right). \tag{8}$$

对于具体的 Shannon 尺度函数 $\phi(x)$ 和对应的小波函数 $\varphi(x)$,其形式如下式所示:

$$\phi(x) = \frac{\sin(\pi x)}{x},$$

$$\psi(x) = \frac{\sin(2\pi x) - \sin(\pi x)}{\pi x}. \tag{9}$$

则根据定理 1 构造的尺度核函数和小波核函数分别为

$$k(\mathrm{x},\mathrm{z})_{\mathrm{scaling}} = \prod_{i=1}^{N} \frac{\sin(\pi(x_i - z_i))}{\pi(x_i - z_i)}, \tag{10}$$

$$k(\mathrm{x},\mathrm{z})_{\mathrm{wavelet}} =$$

$$\prod_{i=1}^{N} \frac{\sin\left(2\pi\left(\frac{x_i - z_i}{a}\right)\right) - \sin\left(\pi\left(\frac{x_i - z_i}{a}\right)\right)}{\pi\left(\frac{x_i - z_i}{a}\right)}. \tag{11}$$

构造完上述多分辨率核函数之后,便可通过这些核函数的扩展使最后的回归精度得以提高. 然而发现扩展后式(7) 的输入矩阵 Φ 与式(2) 的 Φ 不同:扩展后 Φ 的大小为 $M \times (P \cdot M + 1)$. Φ 的规模增大了,却使最后的训练时间不减反增,这便增大了算法的训练时间压力.

3.2　维数约简

经多分辨率核函数扩展后的矩阵 Φ 在规模上,其列数目增加了许多,所以解决时间压力的最直观的方法为对矩阵 Φ 的列进行变量综合和维数约简. 保局投影(LPP)[6] 是最近提出的一种新的降维算法,其目的在于发现固有的、以欧几里得方式内嵌的流行结构. 与传统的 PCA 降维算法相比,LPP 保存了输入数据集的局部结构,使得孤立点对降维效果的影响较小.

设有一个位置相似矩阵 S 和输入矩阵 X (此处 $\mathrm{X} = \Phi^{\mathrm{T}}$),则 LPP 可通过求解下式的最小化问题得到最佳投影:

$$\mathrm{w_{opt}} = \arg\min_{\mathrm{w}} \sum (\mathrm{w^T X}_i - \mathrm{w^T X}_j)^2 \mathrm{S}_{ij} = \arg\min_{\mathrm{w}} \mathrm{w^T XLX^T w},$$

$$\mathrm{s.t.}\ \mathrm{w^T XDX^T w} = 1. \tag{12}$$

其中:X_i 为第 i 列输入数据;D 为对角矩阵,且 $\mathrm{D}_{ii} = \sum_j \mathrm{S}_{ij}$;$\mathrm{L} = \mathrm{D} - \mathrm{S}$ 为拉普拉斯算子. 经过简单变换可将式(12) 变为如下特征向量问题:

$$\mathrm{XLX^T w} = \lambda \mathrm{XDX^T w}, \tag{13}$$

其中 λ 为特征值. 由式(12) 和(13) 得到约简后的多核输入矩阵为

$$\Phi_{\mathrm{new1}} = \Phi \mathrm{W_{opt}}. \tag{14}$$

其中:$\mathrm{W_{opt}}$ 为式(13) 中前 L 个特征向量组成的矩阵;L 为主成分数目,可由特征值的能量比决定. 通常情况下,能量比设定为 100 %,这就可以保证约简后的数据集仍能完全反映原来数据集的信息,从而使得回归精度几乎不会降低. 将式(14) 中的 Φ_{new1} 替换式(2) 中的 Φ,并结合式(3) 和(4),便形成了约简多分辨率相关向量机算法的初步模型. 此时 Φ_{new1} 的规模变为 $M \times L$, $L < P \cdot M + 1$.

3.3　数据采样后的训练数据集

维数约简后的输入矩阵 Φ_{new1} 在一定程度上可减轻算法的训练时间压力,但当训练样本集的数目 M 较大时,Φ_{new1} 的行数仍很大,此时行数是影响训练时间的最重要因素. 从相关向量机(本文算法) 的实现可以发现,最后的回归精度很大程度上由相关向量决定,因此可通过数据采样技术来提取潜在的、有可能包含相关向量的较小训练样本集. 假设数据采样之后的训练样本集数目为 M_1,则经过维数约简和数据采样之后输入矩阵 Φ_{new2} 的规模为 $M_1 \times L_1$,其中:$M_1 < M$;L_1 为采样训练样本集维数约简后的主成分数目,且 $L_1 < P \cdot M_1 + 1$. 于是输入矩阵的整体规模降低,可大幅减小算法的训练时间.

在采样过程中主要应考虑两个问题:采样的均匀性和采样率的大小[13]. 采样样本集应能有效代表原样本集,使其能反映原样本集的数据分布. 另外,采样率若太低,采样样本集必然丢失原样本集的某些特质,破坏数据的分布,因此采样率必须不小于某一阈值.

对于第 1 个问题,可采用分层随机采样[14] 技术完成. 该采样技术将总体数据集分成不重叠的子总体,然后在每一层中用简单随机抽样的方法进行抽样. 分层随机采样考虑了异质子总体之间存在的差异,充分反映原始数据集中固有的特性. 分层可通过一些聚类方法实现,如 FCM.

第 2 个问题可采用 Chernoff Bounds[13] 方法确定最小采样数据量,即

$$M_{\min} = tM + \frac{M}{|u_{\min}|}\lg\left(\frac{1}{\delta}\right) + \frac{M}{|u_{\min}|}\sqrt{\lg\left(\frac{1}{\delta}\right)^2 + 2t|u_{\min}|\lg\left(\frac{1}{\delta}\right)}. \tag{15}$$

其中:$M_{\min}$ 为最小采样数据量;M 为整个数据库包含的数据量,显然 $M_1 \geq M_{\min}$;$u_{\min}$ 为最小类包含的数据量;t,δ 为实参数,其中 $0 \leq t \leq 1$.

基于以上论述,为较准确地提取含有相关向量的候选训练样本,本文算法中较小规模训练样本集的生成采用一种基于初步模型的数据采样方法,其步骤如下:

1) 通过分层随机采样对训练样本集进行采样,记采样数目为 M_1(M_1 亦为最终的采样数目);

2) 对采样样本集建立初步模型,将初步模型得到的相关向量记入新的训练样本集,其数目为 M_2;

3) 对非相关向量的其余训练样本,应用初步模型进行预测,并将预测误差从小到大依次排列,选择前面的 $M_1 - M_2$ 个记入新的训练样本集;

4) 采样完成并得到新的训练样本集(包含 M_1 个数据).

上述数据采样方法会尽可能地提取相关向量,使得采样后的训练样本集在数目较小的情况下仍能反映原来样本集的回归信息. 得到新的训练样本集后,经 LPP 维数约简可得到最终的输入矩阵 Φ_{new2},再结合式(3) 和(4) 便形成了约简多分辨率相关向量机算法的最终算法模型.

4 实验及分析

为充分比较本文算法与支持向量机、相关向量机的性能差异,分别将 3 种算法应用于混沌时间序列预测、波士顿住房问题和色彩再现这些非线性回归问题. 支持向量机和相关向量机的核函数采用高斯核,其中高斯核的宽度参数均取 0.25. 高斯核[7,9]是应用最广泛的核函数之一. 约简多分辨率相关向量机的核函数分别采用式(10) 和(11) 的 Shannon 尺度函数和小波函数. 为简单起见,约简多分辨率相关向量机的小波核函数的伸缩因子 a 均取为 1[9]. 特征值的能量比设定为 100 %. 式(15) 的参数选择如文献[13],即 $t = 0.3$, $\delta = 0.9$. 本文算法初步模型和最终模型的最大迭代次数分别选为 5 和 100. 测试方法采用普遍应用的预测泛化能力方法[2-5].

4.1 混沌时间序列预测

混沌现象是自然界普遍存在的一种不规则运动,是一种来源于非线性动力系统内部的、本身固有的复杂动态行为. 近年来,混沌时间序列的建模和预测已成为混沌信号处理研究领域中一个非常重要的研究方向[3]. Logistic 是混沌动力系统中研究比较透彻的一种混沌序列,其形式为

$$\theta_{n+1} = k\theta_n(1 - \theta_n), \tag{16}$$

本文同文献[3] 一样针对其进行预测. 式(16) 中参数 k 在 $3.57 < k < 4$ 情况下,系统呈混沌状态. 实验选取 $k = 3.7$, $\theta_0 = 0.41$,嵌入维数为 7,时间序列数目为 300,选择前 250 个作为训练集,后 50 个作为测试集. 采样后的训练样本集数据为 154 个. 预测的结果及训练时间如表 1 所示.

表 1 混沌时间序列预测误差及训练时间

算 法	平均误差	最大误差	标准差	CPU Time/ s
SVM	0.032 1	0.196 8	0.056 2	1.953 1
RVM	0.024 2	0.229 8	0.053 3	22.722 7
RMR_RVM	0.005 6	0.039 6	0.008 1	9.764 0

4.2 波士顿住房问题

波士顿住房问题数据集从 UCI 机器学习库(Machine Learning Repository, http://www.ics.uci.edu/~mlearn/databases/housing/) 中获取. 它是检验算法回归性能的基准数据集,包含 506 条 14 维变量的数据. 一般通过前 13 维变量来估计最后一维变量. 实验选择其中的 481 个数据作为训练集,其余的作为测试集. 采样后的训练样本集数据为 285 个. 最后的预测结果及训练时间如表 2 所示.

表 2 波士顿住房问题预测误差及训练时间

算 法	平均误差	最大误差	标准差	CPU Time/ s
SVM	5.759 9	15.664 2	5.237 6	19.000 0
RVM	4.250 0	16.232 8	4.461 7	55.269 5
RMR_RVM	2.916 5	12.198 0	3.478 8	34.689 9

4.3 色彩再现

随着成像设备如显示器、打印机的不断普及,如何能够准确一致地再现设备之间的色彩具有重要的理论意义和实用价值. 然而这些设备因各自不同的色域和自身的色彩特性等原因而不能真实再现彩色图像,而且准确的色彩再现会因涉及到不同设备和再现过程的内在非线性而比较困难[15].

色彩管理通过引入与设备无关的色彩空间(如 Lab 空间) 将色彩再现问题转化为成像设备的色彩校正问题. 实验将 3 种算法(针对目的色彩空间的每一维分别建立算法模型) 应用于打印机的色彩校正问题. 首先通过打印机打印样本集色块,然后利用测色设备测出样本集色块 Lab 值,这样就将 CMY 值和 Lab 值之间的非线性关系对应起来. 测试用的打印机为 HP designjet 800PS 大幅面喷墨打印机. 分光测色仪为 GetagMacbeth SpetroScan Transmission 分光测色仪. 样本集数目为 $11 \times 11 \times 11 = 1\,331$,均匀选取其中的 2/3(887) 样本作为训练样本,其余的作为测试样本. 采样后的训练样本集数据均为 510 个. 最后的预测结果和算法训练时间如表 3 所示.

表 3 色彩再现预测误差及训练时间

算 法	平均误差	最大误差	标准差	CPU Time/ s
SVM	0.141 3	0.531 3	0.075 8	355.406 3
RVM	0.130 2	0.749 8	0.128 9	5 346.057 3
RMR_RVM	0.089 6	0.300 9	0.048 0	852.035 2

从上述 3 个实验结果可以发现,相关向量机较支持向量机的精度(平均误差) 高,这与文献[2] 的结论是一致的,但相关向量机的训练时间却较长. 本文算法在精度(平均误差和最大误差) 和鲁棒性(标准差) 上均优于支持向量机和相关向量机. 另外,由于本文算法采用数据采样和保局投影技术,其训练

时间虽仍大于支持向量机,但较相关向量机已大幅下降,从而提高了算法的运行效率.

5　结　　语

在核算法领域中,相关向量机在很多方面都优于广泛流行的支持向量机,但却存在训练时间较长等问题. 与支持向量机和相关向量机相比,本文算法在保持相关向量机优点的基础上,进一步发展了算法,具有以下特点:

1) 尺度核函数和小波核函数的多核结合为算法提供了一个完备的基词典,使得最后回归精度得以提高,但同时增加了算法的训练时间压力.

2) 为了对多核输入矩阵的列进行综合并达到降维目的,LPP 的应用使其成为可能,而且 LPP 减小了孤立点的影响,在一定程度上缓解了多核带来的训练时间压力.

3) 针对较大规模数据的训练,算法提出对训练数据集进行采样并建立初步模型得到较小规模的训练数据集,从而进一步减小了算法训练时间压力.

4) 上述 3 种措施各司其职,共同提高了算法的性能. 完备基函数词典用来提高算法的有效性;维数约简和数据采样技术提高了算法的求解效率. 其中:前者综合了输入矩阵列的信息,后者提取了输入矩阵行的必要信息(相关向量信息). 两者协作生成一个小规模的输入矩阵,从而使算法在提高精度的同时还提高了运行效率.

下一步的工作是将本文算法应用于更多的实际问题中.

参考文献(References)

[1] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of statistical learning: Data mining, inference and prediction[M]. New York: Springer-Verlag, 2001.

[2] Michael E Tipping. Sparse Bayesian learning and the relevance vector machine[J]. J of Machine Learning Research, 2001, 1(3): 211-244.

[3] 张旭东, 陈锋, 高隽, 等. 稀疏贝叶斯及其在时间序列预测中的应用[J]. 控制与决策, 2006, 21(5): 585-588.
(Zhang Xu-dong, Chen Feng, Gao Jun, et al. Sparse Bayesian and its application to time series forecasting[J]. Control and Decision, 2006, 21(5): 585-588.)

[4] Quinonero-Candela Joaquin, Hansen Lars Kai. Time series prediction based on the relevance vector machine with adaptive kernels[C]. IEEE Int Conf on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2002: 985-988.

[5] 刘遵雄, 张德运, 孙钦东, 等. 基于相关向量机的电力负荷中期预测[J]. 西安交通大学学报, 2004, 38(10): 1005-1008.
(Liu Zun-xiong, Zhang De-yun, Sun Qin-dong, et al. Mid-term electric load prediction based on the relevant vector machine[J]. J of Xi'an Jiaotong University, 2004, 38(10): 1005-1008.)

[6] He Xiao-fei. Locality preserving projections[D]. Chicago: The University of Chicago, 2005.

[7] 张莉, 周伟达, 焦李成. 尺度核函数支撑矢量机[J]. 电子学报, 2002, 30(4): 527-529.
(Zhang Li, Zhou Wei-da, Jiao Li-cheng. Scaling kernel function support vector machine[J]. Acta Electronica Sinica, 2002, 30(4): 527-529.)

[8] 胡丹, 肖建, 车畅. 尺度核支持向量机及在动态系统辨识中的应用[J]. 西南交通大学学报, 2006, 41(4): 460-465.
(Hu Dan, Xiao Jian, Che Chang. Support vector machine with scaling kernel and its application in dynamic system identification[J]. J of Southwest Jiaotong University, 2006, 41(4): 460-465.)

[9] Zhang Li, Zhou Wei-da, Jiao Li-cheng. Wavelet support vector machine[J]. IEEE Trans on Systems, Man and Cybernetics, 2004, 34(1): 34-39.

[10] Vincent Guigue, Alain Rakotomamonjy, Stephane Canu. Kernel basis pursuit[C]. ECML. Berlin: Springer-Verlag, 2005: 146-157.

[11] Mallat S, Zhang Z. Matching pursuits with time-frequency dictionary[J]. IEEE Trans on Signal Processing, 1993, 41(12): 3397-3415.

[12] Chen S, Donoho D, Sauners M. Atomic decomposition by basis pursuit[J]. SIAM J on Scientific Computing, 1998, 20(1): 33-61.

[13] 范玉刚, 李平, 宋执环. 基于样本取样的 SMO 算法[J]. 信息与控制, 2004, 33(6): 665-669.
(Fan Yu-gang, Li Ping, Song Zhi-huan. A sampling-based SMO algorithm[J]. Information and Control, 2004, 33(6): 665-669.)

[14] 郑吉平, 秦小麟. 数据挖掘中采样技术的研究[J]. 系统工程与电子技术, 2005, 27(11): 1946-1949.
(Zheng Ji-ping, Qin Xiao-lin. Research on sampling technology in data mining[J]. Systems Engineering and Electronics, 2005, 27(11): 1946-1949.)

[15] Vrhel M J. Color imaging: Current trends and beyond[C]. IEEE Int Conf on Image Processing. Vancouver: IEEE Press, 2000: 513-516.