基于NSVM的核空间训练数据减少方法

王 晓，刘小芳

(四川理工学院计算机学院 四川 自贡 643000)

【摘要】针对核空间中大数据集的计算代价高问题，提出用NSVM方法减少分类器的训练数据。先用NSVM、核主成分分析(KPCA)和贪婪KPCA分别从全部训练数据中提取训练分类器的子集；再用子集训练分类器，并用训练和测试数据的错分率对分类结果进行评价。在两个数据集和两种分类器中，用KPCA提取的子集训练的分类器的分类性能弱于NSVM和贪婪KPCA，但用贪婪KPCA提取的子集训练的分类器的泛化能力弱于NSVM。仿真结果表明，用NSVM方法提取的子集训练的分类器，不仅保证了分类器的泛化能力，也降低了分类算法的计算复杂度。

关 键 词 分类器; 贪婪核主成分分析; 核主成分分析; 非线性支持向量机; 支持向量; 训练数据
中图分类号 TP391 文献标志码 A doi:10.3969/j.issn.1001-0548.2013.04.012

Nonlinear Support Vector Machine for Training Data Reduction in Kernel Space

WANG Xiao and LIU Xiao-fang

(School of Computer Science, Sichuan University of Science and Engineering Zigong Sichuan 643000)

Abstract Aiming at the high computational cost issue for large data sets in kernel space, the non-linear support vector machine (NSVM) is proposed to reduce training data of classifier. First, a subset of training classifier is extracted from full training data by using NSVM, kernel principal component analysis (KPCA), and greedy kernel principal component analysis (GKPCA), respectively. Then, the classifier is trained by those subsets, respectively. Finally, the classification results are evaluated by the error rate of the training and test data. The classification performance of the classifier trained by the subsets from the KPCA method is inferior to those of from the NSVM and the GKPCA methods, but the generalization of the classifier trained by the subset from the GKPCA method is inferior to those of from the NSVM method for two data sets through two the classifiers. Simulation results indicate that the classifier trained by the subset from the NSVM method not only ensures the generalization ability of classifier, but also reduces the computational complexity of the classification algorithm.

Key words classifier; greedy kernel principal component analysis; kernel principal component analysis; non-linear support vector machine; support vectors; training data

近十年来，核空间理论(简称核理论)迅速成为模式识别和机器学习领域的一个重要分支。核理论采用核技巧，通过非线性函数 $\boldsymbol{\Phi}$ 将原始数据空间的数据集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}\subset\boldsymbol{R}^q$ 映射到特征空间 $\boldsymbol{H}$，再在特征空间 $\boldsymbol{H}$ 中进行其线性变换。根据核理论，已有的线性理论算法可以通过核技巧扩展其非线性理论算法[1-2]。如非线性SVM和广义判别分析等都是通过核技巧进行非线性扩展而得到的，这些核方法显示了很强的非线性处理能力[3-4]。由于核方法要存储大小为 $n\times n$ 的核矩阵 $\boldsymbol{K}$，n 为数据集的样本数目。当样本数 n 很大时，核矩阵的存储空间的开销急剧增加，算法的计算复杂度也急剧上升，使问题很可能无法求解。因此，在核空间中，对大数据集有必要对分类器的训练数据进行减少[5]。文献[6]提出了贪婪核主成分分析(GKPCA)，从训练数据中选择子集，用子集代替全部训练数据，其中，子集是在核空间中具有最小表示误差的样本。文献[7-8]应用GKPCA分别对减少分类器的训练数据和人体运动数据去噪等问题进行了详细的研究。文献[9]提出用核主成分分析(KPCA)[10]从样本对降维的贡献大小选择样本子集。由于GKPCA和KPCA是无监督方法，不能考虑已存在的类别信息(监督信息)，因此，提取的子集不一定适合于训练分类器模型。为了更好地利用已有的类别信息，本文提出了用NSVM方法提

收稿日期：2013－03－06；修回日期：2013－06－25
基金项目：四川省教育厅重点项目(11ZA124)；人工智能四川省重点实验室开放基金(2011RYJ02)
作者简介：王晓(1961－)，男，副教授，主要从事智能信息处理和计算机应用方面的研究.

取子集代替全部训练数据，训练分类器，以减少训练分类器的数据。

1 NSVM减少分类器的训练数据

NSVM[11]是线性SVM通过核理论扩展而形成的，其基本思想是通过非线性变换将输入变量 $\boldsymbol{x}$ 转化到某个高维空间中，然后在变换空间求最优分类面。给定训练数据集 $\boldsymbol{T}_{XY}=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_n,y_n)\}$，其中，$\boldsymbol{x}_i\subseteq\boldsymbol{R}^q$，$y_i$ 为样本 $\boldsymbol{x}_i$ 的类别，并且 $y_i\in Y=\{1,2,\cdots,c\}$。对NSVM方法，其分类函数和目标函数都只涉及训练样本间的点积 $(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)$。

NSVM通过非线性映射 $\boldsymbol{\Phi}:\boldsymbol{R}^q\to\boldsymbol{H}$ 将输入空间映射到高维的特征空间 $\boldsymbol{H}$ 中，当在 $\boldsymbol{H}$ 中构造分类超平面时，NSVM方法仅使用空间中的点积运算 $\boldsymbol{\Phi}(\boldsymbol{x}_i)\cdot\boldsymbol{\Phi}(\boldsymbol{x}_j)$，而没有单独出现 $\boldsymbol{\Phi}(\boldsymbol{x}_i)$ 的形式。由泛函理论可知，满足Mercer定理的函数都可以作为核函数 $\boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)$，这些核函数对应着某一高维特征空间 $\boldsymbol{H}$ 中的点积运算 $\boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{\Phi}(\boldsymbol{x}_i)\cdot\boldsymbol{\Phi}(\boldsymbol{x}_j)$，而这种运算可以通过原数据空间中的函数实现，如径向基核函数、多项式核函数、感知器核函数和样条核函数等[11]。根据以上原理，对非线性分类，构造最优分类超平面问题转换成如下的二次规划问题：

$$\begin{cases}\min & \boldsymbol{J}(\boldsymbol{w},b,\boldsymbol{\xi})=\dfrac{1}{2}\|\boldsymbol{w}\|^2+C\sum\limits_{i=1}^{n}\xi_i\\ \text{s.t.} & y_i(\boldsymbol{w}\cdot\boldsymbol{\Phi}(\boldsymbol{x}_i)+b)\geqslant 1-\xi_i\quad \xi_i\geqslant 0,\ i=1,2,\cdots,n\end{cases}\tag{1}$$

式中，C 为正则化常量，其值越高对错误的惩罚越重；$\boldsymbol{\xi}$ 是一个用于放宽约束条件的松弛项。

式(1)的最优解为拉格朗日函数的鞍点，即有：

$$L(\boldsymbol{w},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\beta})=\frac{1}{2}\|\boldsymbol{w}\|+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i[y_i(\boldsymbol{w}\cdot\boldsymbol{\Phi}(\boldsymbol{x}_i)+b)+\xi_i-1]-\sum_{i=1}^{n}\beta_i\xi_i\tag{2}$$

式中，$\boldsymbol{\alpha}\geqslant 0$ 为拉格朗日乘子。

根据Karush-Kuhn-Tucker(KKT)定理[12]可得以下的定理：

定理 1 (KKT定理) 给定一个定义在凸域 $\Omega\subseteq\boldsymbol{R}^q$ 上的最优化问题，则有：

$$\begin{cases}\min\ f(\boldsymbol{w}) & \boldsymbol{w}\in\Omega\\ \text{s.t. } g_i(\boldsymbol{w})\leqslant 0 & i=1,2,\cdots,k\\ h_i(\boldsymbol{w})=0 & i=1,2,\cdots m\end{cases}$$

式中，$f\in C^1$ 是凸的，并且 g_i 和 h_i 是仿射函数。一般地，一个点 $\boldsymbol{w}^*$ 是最优点的充要条件是存在 $\boldsymbol{\alpha}^*$ 和 $\boldsymbol{\beta}^*$ 满足，则有：

$$\frac{\partial\boldsymbol{L}(\boldsymbol{w}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)}{\partial\boldsymbol{w}}=0$$

$$\frac{\partial\boldsymbol{L}(\boldsymbol{w}^*,\boldsymbol{\alpha}^*,\boldsymbol{\beta}^*)}{\partial\boldsymbol{\beta}}=0$$

$$\alpha_i^*g_i(\boldsymbol{w}^*)=0\quad i=1,2,\cdots,k$$

$$g_i(\boldsymbol{w}^*)\leqslant 0\quad i=1,2,\cdots,k$$

$$\alpha_i^*\geqslant 0\quad i=1,2,\cdots,k$$

因此，根据KKT定理，式(2)的最优解满足以下条件：

$$\begin{cases}\dfrac{\partial\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\beta})}{\partial\boldsymbol{w}}=\boldsymbol{w}-\sum\limits_{i=1}^{n}\alpha_iy_i\boldsymbol{\Phi}(\boldsymbol{x}_i)=0\\ \dfrac{\partial\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\beta})}{\partial b}=\sum\limits_{i=1}^{n}\alpha_iy_i=0\\ \dfrac{\partial\boldsymbol{L}(\boldsymbol{w},b,\boldsymbol{\alpha},\boldsymbol{\xi},\boldsymbol{\beta})}{\partial\xi_i}=C-\beta_i-\alpha_i=0\\ \alpha_i[y_i(\boldsymbol{w}\cdot\boldsymbol{\Phi}(\boldsymbol{x}_i)+b)-1+\xi_i]=0\quad\forall i\\ \beta_i\xi_i=0\quad\forall i\\ \alpha_i,\beta_i,\xi_i\geqslant 0\quad\forall i\end{cases}\tag{3}$$

将式(3)代入式(1)，可得：

$$\begin{cases}\max\boldsymbol{W}(\boldsymbol{\alpha})=\sum\limits_{i=1}^{n}a_i-\dfrac{1}{2}\times\sum\limits_{i,j=1}^{n}\alpha_i\alpha_jy_iy_j\boldsymbol{K}(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)\\ \text{s.t.}\quad\sum\limits_{i-1}^{n}\alpha_iy_i=0\quad 0\leqslant\alpha_i\leqslant C,\ i=1,2,\cdots,n\end{cases}\tag{4}$$

因此，在式(1)中构造最优分类超平面问题转换成式(4)的二次规划性问题，即式(4)是式(1)的对偶形式。在式(4)中，对应非零拉格朗日乘子 α_i 的样本称为支持向量，它们是构造最优分类超平面的样本，也是最难被分类的样本。这些支持向量包括了构造分类超平面的所有必要信息，即如果在数据集中去掉所有非支持向量的样本，那么用剩余的样本(即支持向量)仍可以构造出与原数据相同的分类超平面[12]。这可以从对偶问题中求出，因为去除非支持向量的行和列，对剩余的子矩阵仍有相同的最优问题。因此，最优解保持不变。

定理 2[12] 考虑一个压缩方案。对 $(\boldsymbol{x}_i,y_i)\in\{-1,1\}$ 上的任意概率分布 D、大小为 d 的压缩集定义，假设在 n 个随机样本 $\boldsymbol{T}_{XY}=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_i,y_i),\cdots,(\boldsymbol{x}_n,y_n)\}$ 上的误差以概率 $1-\delta$ 不大于，则有：

$$\underset{D}{\operatorname{err}}(f)\leqslant\frac{1}{n-d}\left(d\ln\frac{en}{d}+\ln\frac{n}{\delta}\right)$$

式中，d 为VC(Vapnik和Chervonenkis)维，在本文中是支持向量的个数。

根据定理2可知，最大间隔超平面是一个压缩方

案，只要给定了支持向量，就能重构相同的分类超平面。鉴于此，本文用支持向量代替原始训练数据集 $\boldsymbol{T}_{XY}$，训练分类器模型，以减少训练分类器的数据。假设支持向量集为 $\boldsymbol{T}_{XY}^{s}=\{(\boldsymbol{x}_1,y_1),\ \boldsymbol{T}_{XY}^{s}=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,\ (\boldsymbol{x}_{n_{sv}},y_{n_{sv}})\}\subset\boldsymbol{T}_{XY}$，其中 n_{sv} 是支持向量的个数。

标准的NSVM是求解两类问题，为了求解多类问题，本文通过一对多分解方法[13]把NSVM扩展到能求解多类问题，其分类函数为：

$$f(\boldsymbol{x})=\left[\sum_{i=1}^{n}y_i\alpha_i\boldsymbol{K}(\boldsymbol{x}_i\cdot\boldsymbol{x})+b\right] \tag{5}$$

2 实验结果及分析

2.1 实验数据

实验选取了两个数据集：UCI机器学习数据库的数据集IRIS[14]和一幅遥感图像数据。算法用MATLAB和C语言混合编程实现。遥感数据来至于3个波段(B_3、B_2和B_1)的北京一号小卫星的多光谱图像，其中B_3、B_2和B_1分别表示近红外、红和绿波段。成像时间为2007年9月14日11:30:29～11:34:59，图像大小为128×128像元。数据集IRIS的样本数为150个，样本类别数为3，属性数为4，训练数据和测试数据分别选取90和60个；遥感数据的样本数为16 384个，样本类别数为3，属性数为3，训练数据和测试数据各选取600个。北京一号小卫星按B_3、B_2和B_1组合的伪彩色图像如图1所示。

图1 北京一号小卫星按B_3、B_2和B_1组合的伪彩色图像

2.2 实验的流程

为了检验NSVM减少分类器的训练数据的有效性，本文与用GKPCA[6]和KPCA[9]提取的子集分别训练分类器，以进行对比验证。几种方法提取子集和训练分类器的流程图如图2所示。

在图2中，给定数据集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}\subset\boldsymbol{R}^q$，通过下式获得归一化数据，其目的是平衡不同范围的属性值对分类产生的影响，则有：

$$\tilde{x}_{jp}=(x_{jp}-\mu_p)/\sigma_p \qquad j=1,2,\cdots,n,\ \ p=1,2\cdots,q \tag{6}$$

式中，x_{jp} 表示第 j 个样本 $\boldsymbol{x}_j$ 的第 p 个特征值；$\mu_p=\frac{1}{n}\sum_{j=1}^{n}x_{jp}$ 和 $\sigma_p^2=\frac{1}{n}\sum_{j=1}^{n}(x_{jp}-\mu_P)^2$ 分别是第 p 个特征向量的均值和方差。

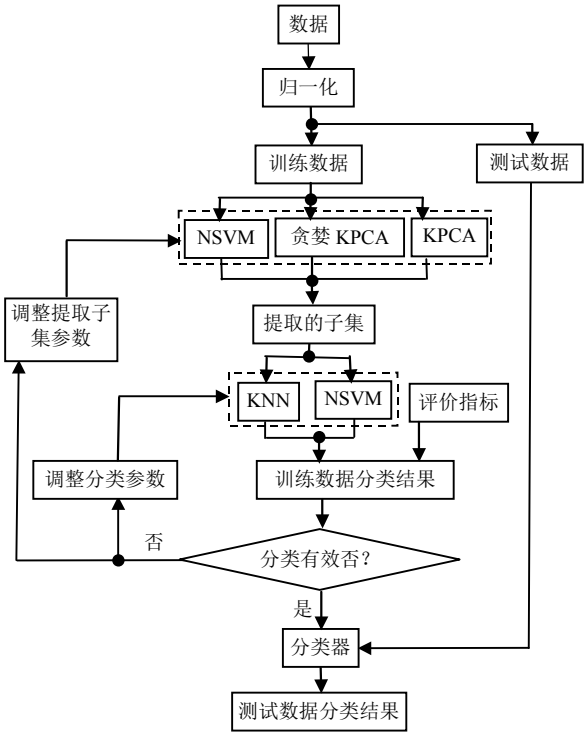


图2 提取子集和训练分类器的流程图

2.3 训练数据减少的结果

通过 5-重交叉验证法[15]可得，几种方法的参数和提取的子集数据的个数如表 1 所示。

表1 几种方法的参数和提取的子集数据个数

数据集	NSVM			GKPCA[6]		KPCA[9]	
	σ	C	子集数据	σ	子集数据	σ	子集数据
IRIS	1	10	30	3	22	1	30
遥感数据	10	100	56	40	55	10	60

核函数都选取径向基核函数 $\boldsymbol{K}(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/\sigma^2)$，$\sigma$ 为核参数；NSVM 方法还需要确定其正则化参数 C。几种方法分别从IRIS和遥感数据的训练数据中选取子集数据，其分布图如图3所示，*o*表示子集数据。用NSVM方法提取的子集分别如图3a和图3d所示，其子集主要分布在各类数据的边界；用GKPCA提取的子集分别如图3b和图3e所示，其子集在各类数据中较为分散；用KPCA提取的子集分别如图3c和图3f所示，其子集主要分布在各类数据的中心地带。

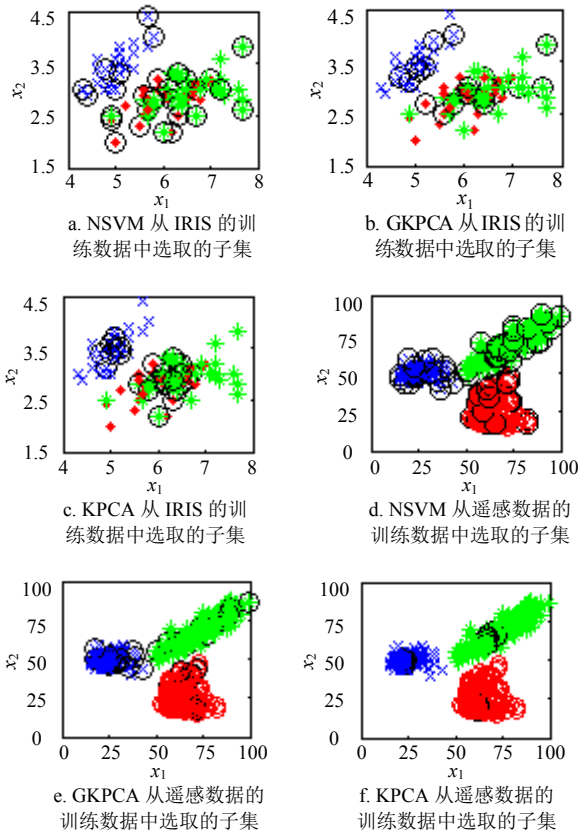
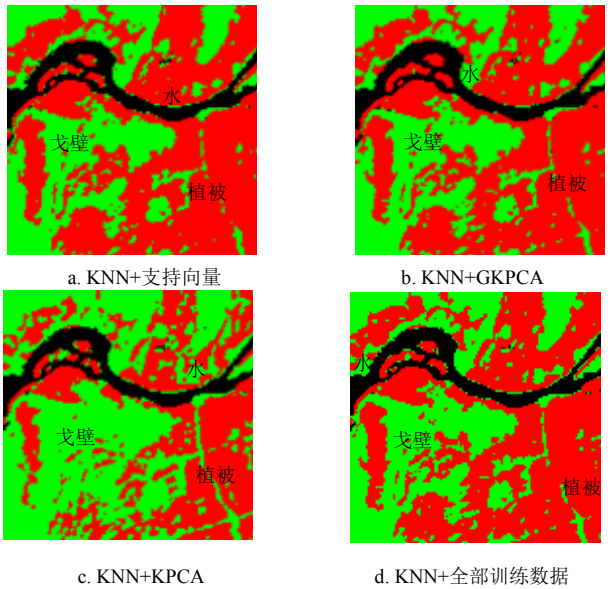


图3　几种方法分别从IRIS和遥感数据的训练数据中选取的子集分布图

2.4　分类性能的评价

用NSVM、GKPCA[6]和KPCA[9]提取的子集分别训练K-最近邻(K-nearest neighbours，KNN)[15]和NSVM分类器。对K-近邻分类器，用欧氏距离作为样本之间距离的度量函数，参数$K=9$；NSVM分类时的参数与其提取子集时相同。几种方法的遥感图像分类图如图4所示。



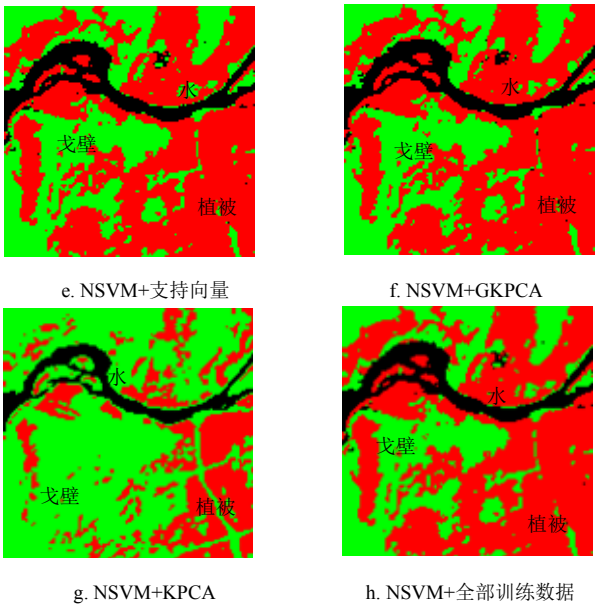


图4　几种方法的遥感图像的分类图

几种方法对IRIS数据和遥感数据提取的子集训练的分类器的分类结果分别如表2和表3所示。实验选取两种分类器，主要检验提取的子集对不同分类器的影响。测试数据的错分率反映出分类器对未知样本(即未参与训练分类器的样本)不能做出正确分类的能力，其值越大表示分类器的泛化能力越弱。

表2　几种方法对IRIS数据的分类结果

分类方法	训练分类器数据	训练数据错分率/(%)	测试数据错分率/(%)	训练分类器时间/($\times10^{-3}$s)
KNN	支持向量	2.2	1.7	3.8
	GKPCA提取子集	3.3	1.7	3.8
	KPCA提取子集	4.4	6.7	4.0
	全部训练数据	2.2	3.3	8.1
NSVM	支持向量	1.1	1.7	3.7
	GKPCA提取子集	4.4	3.7	2.3
	KPCA提取子集	8.9	10.0	3.5
	全部训练数据	1.1	5.0	7.9

表3　几种方法对遥感数据的分类结果

分类方法	训练分类器数据	训练数据错分率/(%)	测试数据错分率/(%)	训练分类器时间/($\times10^{-3}$s)
KNN	支持向量	3.8	8.3	5.5
	GKPCA提取子集	4.3	9.0	5.2
	KPCA提取子集	5.2	14.3	5.7
	全部训练数据	4.2	8.5	10.7
NSVM	支持向量	3.7	8.4	3.7
	GKPCA提取子集	4.1	8.7	3.9
	KPCA提取子集	5.9	18.4	4.4
	全部训练数据	4.6	8.8	7.2

对一大一小这两个数据集和两种分类器，用4种训练数据(全部训练数据、NSVM提取的子集、GKPCA提取的子集、KPCA提取的子集)分别训练分类器。仿真结果表明，用NSVM、GKPCA和全部数据提取的子集分别训练的分类器，它们的分类结果

基本相同，但全部数据训练分类器的泛化能力弱于NSVM和GKPCA提取的子集分别训练的分类器；用KPCA提取的子集训练的分类器的分类结果弱于NSVM和GKPCA，但用GKPCA提取的子集训练的分类器的泛化能力弱于NSVM。实际上，从图3的训练数据的子集分布图，也呈现出用NSVM和GKPCA提取的子集近似原始训练数据比用KPCA方法更合理。

3 结 论

针对核空间中大数据集的计算代价高问题，本文提出了用NSVM的支持向量代替全部训练数据训练分类器模型，以减少训练分类器的数据。为了检验本文方法的性能，与核空间中的其他两种提取子集方法GKPCA和KPCA进行了对比实验。结果表明，在两个数据集和两种分类器上，用KPCA提取的子集训练的分类器的分类性能弱于NSVM和GKPCA，但用GKPCA提取的子集训练的分类器的泛化能力弱于NSVM。因此，用NSVM方法提取训练数据的支持向量代替原始训练数据，训练分类器，不仅保证了分类器的泛化能力，而且也减少了训练分类器的数据，缩短了训练分类器的时间，也降低了分类算法的计算复杂度。

参 考 文 献

[1] SHAWE-TAYLO J, CRISTIANINI N. Kernel methods for pattern analysis[M]. Cambridge: Cambridge University Press, 2004: 15-25.

[2] CAMPS-VALLS G, SHERVASHIDZE N, BORGWARDT M K. Spatio-spectral remote sensing image classification with graph kernels[J]. IEEE Geoscience and Remote Sensing Letters, 2010, 7(4): 741-745.

[3] 刘小芳. 基于核理论的遥感图像分类方法研究[D]. 成都: 电子科技大学, 2011.
LIU Xiao-fang. Research on remote sensing image classification methods based on kernel theory[D]. Chengdu: University of Electronic Science and Technology of China, 2011.

[4] 王介生, 高宪文, 张勇. 基于图像纹理特征和多级SVM的浮选过程状态识别方法[J]. 控制与决策, 2010, 25(10): 1523-1535.
WANG Jie-sheng, GAO Xian-wen, ZHANG Yong. Research on recognizing flotation states based on image texture features and multi-layer SVMs[J]. Control and Decision, 2010, 25(10): 1523-1535.

[5] FRANC V, HLAVÁČ V. Greedy algorithm for a training set reduction in the kernel methods[M]. Heidelberg: Springer-Verlag Berlin, 2003.

[6] FRANC V. Optimization algorithms for kernel methods[D]. Czech: Czech Technical University, 2005.

[7] TANGKUAMPIEN T, SUTER D. Human motion de-noising via greedy kernel principal component analysis filtering [C]//Proc of the IEEE 8th International Conference on Pattern Recognition. Piscataway, USA: IEEE, 2006.

[8] 刘小芳, 何彬彬, 李小文. 基于greedy GDA的训练数据减少和非线性特征提取方法[J]. 控制与决策, 2011, 26(10): 1511-1514.
LIU Xiao-fang, HE Bin-bin, LI Xiao-wen. Greedy GDA method for training data reduction and nonlinear feature fxtraction[J]. Control and Decision, 2011, 26(10): 1511-1514.

[9] 秦建玲, 李军. 基于核的主成分分析的特征提取方法与样本筛选[C]//第十二届工业工程与工程管理国际学术会议. 北京: 中国机械工程学会, 2005: 1577-1580.
QIN Jian-ling, LI Jun. Theory of feature extraction and samples filtration based on KPCA[C]//Proc of the 12th International Conference on Industrial Engineering and Engieering Management. Beijing: Industrial Engineering Institute of China, 2005: 1577-1580.

[10] SCHÖLKOPF B, SMOLA A, MULLER K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998(10): 1299-1319.

[11] BANKI M H, SHIRAZI A A B. New kernel function for hyperspectral image classification[C]// The 2nd International Conference on Computer and Automation Engineering. Piscataway: IEEE Computer Society, 2010.

[12] CRISTIANINI N, SHAWE-TAYLOR J. 支持向量机导论[M]. 李国正, 王猛, 曾华军, 译. 北京: 电子工业出版社, 2004.
CRISTIANINI N, SHAWE-TAYLOR J. An introduction to support vector machines and other kernel-based learning methods[M]. Translated by LI Guo-zheng, WANG Meng, ZENG Hua-jun. Beijing: Publishing House of Electronic Industry, 2004.

[13] HSU C W, LIN C J. A comparison of methods for multiclass support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415-440.

[14] ASUNCION A, NEWMAN D J. UCI machine learning repository[DB/OL]. [2013-02-20]. http://www.ics.uci.edu/mlearn/ML-Repository.html.

[15] DUDA R O, HART P E, STORK D G. 模式分类[M]. 第2版 李宏东, 姚天翔, 译. 北京: 机械工业出版社, 2006.
DUDA R O, HART P E, STORK D G. Pattern classification[M]. 2nd ed. Translated by LI Hong-dong, YAO Tian-xiang. Beijing: China Machine Press, 2006.

编 辑 黄 莘