一种基于样本加权的多尺度核支持向量机方法

沈 健 蒋 芸 张亚男 胡学伟
(西北师范大学计算机科学与工程学院 兰州 730070)

摘 要 多核学习方法是机器学习领域中的一个新的热点。核方法通过将数据映射到高维空间来增加线性分类器的计算能力,是目前解决非线性模式分析与分类问题的一种有效途径。但是在一些复杂的情况下,单个核函数构成的核学习方法并不能完全满足如数据异构或者不规则、样本规模大、样本分布不平坦等实际应用中的需求问题,因此将多个核函数进行组合以期获得更好的结果,是一种必然的发展趋势。因此提出一种基于样本加权的多尺度核支持向量机方法,通过不同尺度核函数对样本的拟合能力进行加权,从而得到基于样本加权的多尺度核支持向量机决策函数。通过在多个数据集上的实验分析可以得出所提方法对于各个数据集都获得了很高的分类准确率。

关键词 多核学习,映射,非线性模式,数据异构

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.12.025

Novel Multi-scale Kernel SVM Method Based on Sample Weighting

SHEN Jian JIANG Yun ZHANG Ya-nan HU Xue-wei
(College of Computer Science and Engineering, Northwest Normal University, Lanzhou 730070, China)

Abstract Multi-kernel learning has been a new research focus in the current kernel machine learning field. Through mapping data into high dimensional space, kernel methods increase the computational power of linear classifier and they are an effective way to solve the problem of nonlinear model analysis and classification. In some complex situations, nevertheless, the kernel learning method of single kernel function can not completely satisfy the requirements of heterogeneous data or irregular data as well as samples of large size and non-flat distribution. Therefore, it is necessary to develop multiple kernel functions in order to get better results. In this paper, we proposed a new SVM method for multi-scale kernel learning based on sample weighting, which is assigned via fitting abilities of distinct scales kernel functions for samples. Through the experimental analysis on several data sets, we can get that the method proposed in this paper can attain better classification accuracy on each data set.

Keywords Multi-kernel learning, Map, Nonlinear model, Heterogeneous data

支持向量机(Support Vector Machine,SVM)[1]是一种基于统计学习的理论[2],其中,在VC理论和结构化风险最小原理的基础上实现的机器学习[3]方法能够较好地解决小样本、高维数、非线性、局部极小点等实际问题。由于支持向量机方法理论的发展和实际的应用,人们对核方法[4-6]的关注日益增多,核函数的采用使得线性的SVM很容易推广到非线性的SVM。其核心在于通过选择使用恰当的核函数替代内积,可以隐式地将训练数据非线性地映射到高维空间,并且不增加可调参数的个数,这种方法既避免了特征空间中复杂的内积计算,又避免了(学习机器)特征空间本身的设计。实际上,Mercer[7]早在1909年就研究了正负类型的函数与积分等式理论的联系;20世纪50年代,Aronszajn等人[8]研究并发展了再生核希尔伯特空间的理论;1964年Aizerman等[9]通过Mercer理论,将核函数解释为特征空间的内积,并将其引入到机器学习领域中,但是当时核方法并没有得到更多的发展和推广,直到1992年Boser等提出了SVM方法。SVM的广泛应用极大地促进了核方法的普及和发展,并且逐渐渗透到机器学习的许多领域,如回归估计[10]、模式分类[11]、概率密度估计[12]等,典型的包括核主成分分析[13]、核Fisher判别[14]、核判别分析[15]等。此后,核方法又得到了大量的改进和推广,并广泛应用于多个领域。

尽管上面提到的核方法在众多领域被证明有效并且实用,但是这些方法都是基于单特征空间的单核方法。由于不同的核函数具有不同的特性,并且对不同样本的拟合程度不同,核函数表现出的性能也有很大的差异:不同的核函数应用在相同样本上的性能会有很大差异,相同的核函数应用在不同样本上的性能也会有很大的差异[16]。此外,当样本特征含有异构信息[17-21]、样本规模很大[22,23]、数据不规则或者数据

到稿日期:2015-12-01 返修日期:2016-04-08 本文受国家自然科学基金资助项目(61163036),甘肃省高校研究生导师项目(1201-16),2012年度甘肃省高校基本科研业务费专项资金项目,西北师范大学第三期知识与创新工程科研骨干项目(nwnu-kjcxgc-03-67)资助。

沈 健(1991-),男,硕士生,主要研究方向为数据挖掘与粗糙集理论,E-mail:469635089@qq.com;**蒋 芸**(1970-),女,博士,教授,主要研究方向为数据挖掘与粗糙集理论、模式识别与机器学习;**张亚男**(1992-),女,硕士生,主要研究方向为软计算与数据挖掘,E-mail:1518015277@qq.com;**胡学伟**(1991-),男,硕士生,主要研究方向为数据挖掘与粗糙集理论。

分布不平坦[24]时，采用单个简单核函数将数据映射到高维空间的方式对所有样本进行处理是存在缺陷[16]的。针对以上问题，近年来出现了大量关于多核学习方法的研究。多核学习[25-27]模型是一类较单核具有更强灵活性和更高适应性的基于核学习的模型，不管是在理论还是实际应用中都已经证明使用多核学习方法比单核方法或者单核机器组合模型能获得更佳的学习性能[28,29]，如早期的基于 Boosting 的多核组合学习模型[30]、基于二次约束二次规划的学习方法[20]以及近来出现的简单多核学习方法[22]和多尺度核学习方法[31-33]。

多尺度核学习方法是多核学习方法的一种特殊情形，将多个尺度的单核函数进行融合，以达到最优解。这种方法具有更高的灵活性，并且比合成核方法具有更完备的尺度。随着小波理论以及多尺度理论的成熟与完备，通过引入尺度核空间，多尺度核具有更好的理论背景。这类方法在各领域也得到了较好的应用，如 Kingsbury[31]将多个不同尺度的核进行分步分类；Zheng[23]、Yang[32]提出多尺度支持向量回归，分别用作不平坦函数估计以及时间序列预测。此外，进一步将多尺度核学习方法与 SVM 结合，使得多尺度核学习方法在热点检测和图像压缩领域均得到了广泛应用。近年来，结合多尺度核学习方法，基于 Hilbert 空间中的再生核进行函数重构得到重视并有学者对其进行相关的应用研究。此外，多尺度核学习方法又逐步推广到高斯过程的建模与处理[32]。这些应用和研究为核机器学习提供更丰富的设计思路和更广泛的应用前景。

上述方法在确定多尺度核函数合成系数时，有的采取的是平均的方式，人为地使不同尺度的核函数对决策函数的贡献率相同；有的通过一些寻优方法进行迭代，获得使目标函数最优的参数值，这类寻优方法的大量迭代步骤使得学习时间大大增加。基于上述方法的诸多问题，本文提出一种基于样本加权的多尺度核支持向量机方法，对于样本的分类准确率，使用不同尺度核函数自动求取不同核函数的加权系数，其后利用求得的多尺度核函数进一步求得决策函数，最终实现多尺度核学习方法。通过多组实验可知，本文提出的方法获得了很好的分类结果。

1 多尺度核学习方法和支持向量机方法

1.1 多尺度核学习方法

核学习方法通常是将原始输入数据映射到某一特征空间，而对于支持向量分类器，则将原始输入数据空间的非线性问题演化成特征空间中的线性问题。然而，不同的特征空间有各自的优势，因此希望在这些特征空间所组成的一个增广空间中找到一个融合它们各自优势的更好的解，通常使用 Φ 表示非线性映射。常见的支持向量分类问题的决策函数都会表示为：$f(X)=\mathrm{sgn}(w^t\Phi(X)+b)$，对于多特征空间（如 m 个特征空间）有：$w^t=(w_1^t,w_2^t,\cdots,w_m^t)$，$\Phi(X)=(\Phi_1(X),\Phi_2(X),\cdots,\Phi_m(X))$，其中 $\Phi_1,\Phi_2,\cdots,\Phi_m$ 是 m 个不同的非线性映射，分别定义 m 个不同的特征空间 $F_1,F_2,\cdots,F_m$。非线性映射 Φ 被隐含地定义为核函数：$K(X_i,X_j)=\Phi(X_i)\cdot\Phi(X_j)$，因此核函数的选择对于基于核函数的学习方法是极其重要的。多尺度核学习方法就是要找到一组具有多尺度表示能力的核函数，在广泛被使用的核函数中，高斯径向基核函数：

$$K(X_i,X_j)=\exp(-\frac{\|X_i-X_j\|^2}{2\sigma^2})$$

是最常用的，因为它具有通用普遍的近似能力，同时它也是具有多尺度表示能力的核函数。将高斯径向基核多尺度化：

$$\exp(-\frac{\|X_i-X_j\|^2}{2\sigma_1^2}),\exp(-\frac{\|X_i-X_j\|^2}{2\sigma_2^2}),\cdots,\exp(-\frac{\|X_i-X_j\|^2}{2\sigma_n^2})$$

其中，$\sigma_1<\sigma_2<\cdots<\sigma_n$。可以看出，当 σ 较小时，高斯径向基核函数对变化剧烈的样本具有更好的映射能力；当 σ 较大时，高斯径向基核函数能更好地映射变化平缓的样本。

为简单起见，考虑一个两尺度核 K_1 和 K_2 加权合成分类问题，得到合成的决策函数：

$$f(X)=f_1(X)+f_2(X) \tag{1}$$

其中，$f_1(X)=\sum_{i=1}^{N}\alpha_iK_1(X_i,X)+b_1$，$f_2(X)=\sum_{i=1}^{N}\beta_iK_2(X_i,X)+b_2$。

假设 K_1 是一个大尺度核函数，核相关系数 α_i 选择平滑区域对应的支持向量，而 K_2 是小尺度核函数，核相关系数 β_i 选择变化剧烈的区域对应的支持向量。定义两个输入空间到特征空间的非线性映射 Φ_1 和 Φ_2，使得对于任意输入 x_i 和 x_j，有：$\Phi_1(x_i)\cdot\Phi_1(x_j)=K_1(x_i,x_j)$，$\Phi_2(x_i)\cdot\Phi_2(x_j)=K_2(x_i,x_j)$，且可通过增加特征空间的维数，让上述的 2 个函数享有特征维上相互独立的子集，使得在 2 个函数的计算中至少有一个特征维元素为 0，从而有 $\Phi_1(x_i)\cdot\Phi_2(x_j)=0$。

设

$$w_1=\sum_{i=1}^{N}\alpha_iK_1(X_i,X),w_2=\sum_{i=1}^{N}\beta_iK_2(X_i,X) \tag{2}$$

得到合成的决策函数：

$$f(X)=(w_1+w_2)(\Phi_1(x)+\Phi_2(x))+b \tag{3}$$

其中，$b=b_1+b_2$，由于 Φ_1 和 Φ_2 正交，$w_1\cdot\Phi_2$ 和 $w_2\cdot\Phi_1$ 均为 0。令 $w=w_1+w_2$，$\Phi(x)=\Phi_1(x)+\Phi_2(x)$，则上式形式与单核学习决策函数完全相同，通过拉格朗日方法求解 α_i 和 β_i：

$$w_1=\sum_{i=1}^{N}\lambda_iy_i\Phi_1(X_i),w_2=\sum_{i=1}^{N}\lambda_iy_i\Phi_2(X_i) \tag{4}$$

比较式(1)和式(3)，得出 $\alpha_i=\lambda_iy_i$，同时 $\beta_i=\lambda_iy_i$，因此，对于任意 i，都有 $\alpha_i=\beta_i$，这个结果并不理想，这个结果意味着每一个支持向量都是由等量的 K_1 和 K_2 组成的，这是存在缺陷的，因此本文提出一种基样本加权的多尺度核支持向量机方法，引入加权系数 $\frac{acc_i}{\sum_{j=1}^{M}acc_j}$ 来解决上述 α_i 和 β_i 相等的问题，即：

$$f(X)=\sum_{j=1}^{M}\frac{acc_j}{\sum_{i=1}^{M}acc_i}f_j(X) \tag{5}$$

其中，M 是不同尺度核函数的个数，acc_i 是每个尺度核函数对样本分类的准确率，$\sum_{i=1}^{M}acc_j$ 是 M 个不同尺度核函数对样本分类准确率的和。

1.2 基于多尺度核学习的支持向量机方法

SVM 是于 1992 年基于统计学习理论提出的一种全新的机器学习方法。假设训练样本集合：$D=\{(X_i,y_i),X_i\in R^n,y_i\in\{-1,+1\},i=1,2,\cdots,m\}$，其中，每一个 X_i 对应一个数据向量；y_i 是 X_i 对应的类标签，训练样本可以被超平面 $W\cdot X+b=0$ 无误地分开，并且离超平面最近的向量离超平面的

距离是最远的(最大分离超平面)，则这个超平面是最优超平面[34]。SVM 的主要思想就是通过某种方式将低维非线性输入向量 X_i 映射到高维线性特征空间 Z 中，并在空间 Z 中构造最优超平面。但是如何构造最优超平面？又如何解决高维空间出现的维数灾难问题？文献[35]构造了最优超平面，即如下最优化问题：

$$\min \tau(W,\xi)=\frac{1}{2}\|W\|^2+C\sum_{i=1}^{m}\xi_i \quad \text{s.t. } y_i(W^TX_i+b)\geqslant 1-\xi_i,\text{and } \xi_i\geqslant 0, i=1,2,\cdots,m \tag{6}$$

其中，W 是超平面的法向量；b 是超平面的偏倚变量；C 是惩罚参数；ξ_i 是松弛变量。为了解决上述优化问题，引入 Lagrange 乘子 α_i 和 β_i：

$$L(W,\xi_i,b,\alpha_i,\beta_i)=\frac{1}{2}\|W\|^2+C\sum_{i=1}^{m}\beta_i\xi_i-\sum_{i=1}^{m}\alpha_i[y_i(W^T\cdot\Phi(X_i)+b)-1+\xi_i]+\sum_{i=1}^{m}\beta_{\xi_i} \tag{7}$$

该优化问题的解由上述拉格朗日函数的鞍点给出，并且在鞍点处满足对 W,b 和 ξ_i 的偏导为 0，则：

$$\max W(\alpha)=\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\alpha_j y_i y_j\langle X_i\cdot X_j\rangle \quad \text{s.t. } 0\leqslant\alpha_i\leqslant C,\text{and } \sum_{i=1}^{m}\alpha_i y_i=0, i=1,2,\cdots,m \tag{8}$$

解得决策函数为：

$$f(x)=sgn\{[W\cdot x]+b\}=sgn\{\sum_{i=1}^{m}\alpha_i y_i\langle x\cdot X_i\rangle+b\} \tag{9}$$

由式(7)易知，只有一部分 α_i 大于 0，称其对应的训练样本为支持向量。

在解决线性不可分、高维空间带来的维数灾难问题时，*SVM* 引入了核函数的概念，将核函数带入上述解得的决策函数，得到：

$$f(x)=sgn\{\sum_{i=1}^{m}a_i y_i\langle\Phi(X_i)\cdot\Phi(x)\rangle+b\} \tag{10}$$

即

$$f(x)=sgn\{\sum_{i=1}^{m}a_i y_i K(X_i,x)+b\} \tag{11}$$

将多尺度核学习方法(见式(1))带入解得的决策函数中，得到：

$$f(x)=sgn\{\sum_{j=1}^{M}\sum_{i=1}^{m}a_{j,i}y_{j,i}K_j(X_i,x)+b_j\} \tag{12}$$

其中，M 是选定的不同尺度核的个数。

2 基于样本加权的多尺度核支持向量机方法

为了获得更高的分类准确率以及提高学习的效率，本文提出一种基于样本加权的多尺度核支持向量机方法，通过一种固定的样本学习方法获得式(5) 中每一个尺度核函数的权重系数$\frac{acc_j}{\sum_{i=1}^{M}acc_i}$，由于 $acc_j(j=1,2,3,\cdots,M)$表示的是第 j 个尺度函数对于样本分类的准确率和精度，反映的是该尺度函数对于样本映射到高维空间的能力和性能，决定了样本是否线性可分，即对于样本使用非线性尺度函数 Φ_j 映射到高维线性特征空间，如果 Φ_j 对于样本具有良好的映射到高维空间的能力，那么样本在高维空间的线性可分性就较高，因此能得到较高的分类准确率和精度，即代表非线性尺度函数 Φ_j 对样本的分离性较好，对样本的适应能力也较强(即证明样本用非线性尺度函数 Φ_j 映射到高维特征空间的数量比较多)，因此对 Φ_j 使用较大的权重，对于样本分类获得较低分类准确率和精度的非线性尺度函数如 Φ_i，可以推导出非线性尺度函数 Φ_i 对于整个样本不能很好地映射到高维特征空间，但是样本中的部分样本(如不能被非线性尺度函数 Φ_j 很好地映射到高维特征空间的样本)是可以很好地映射到高维特征空间的，但是比重可能较低，因此对于 Φ_i 使用较小的权重。如式(5)所示，使用$\frac{acc_j}{\sum_{i=1}^{M}acc_i}$来对各决策函数进行加权，同时又可以对如 Φ_j 这样的对大部分样本都能进行较好映射的非线性尺度函数进行较大权重的加权，还可以对如 Φ_i 这样的对小部分样本进行较好映射的非线性尺度函数进行较小权重的加权，依据非线性尺度函数对样本进行非线性映射到高维空间的性能对非线性尺度函数进行加权；同时，根据本文的实验可以得到本文提出的基于样本加权的多尺度支持向量机方法是有可行性的，也能获得较高的分类准确率和精度，通过这种加权得到的非线性尺度函数的权重是合理的，也是切实可行的。

本文提出的这种方法充分考虑到了每一次需学习的样本的特性，以及每一个尺度核函数对于样本映射的能力，较其他多尺度核合成权重系数的方法可以更快、更高效地获得更高的学习准确率和精度。

假设现在有两类别样本 S 以及 M 个尺度核函数$k_j(x,z)$ $(j=1,2,3,\cdots,M)$，将 M 个尺度核函数代入式(11)，得到 M 个不同的关于样本 S 的支持向量机决策函数，其中第 j 个支持向量机决策函数为：

$$f_j(x)=\text{sgn}\{\sum_{i=1}^{m}a_i y_i k_j(X_i,x)+b\}, j=1,2,3,\cdots,M \tag{13}$$

通过 M 个决策函数可以获得每一个尺度核函数支持向量机的决策函数对于样本 S 的分类准确率：$acc_j(j=1,2,3,\cdots,M)$，可以使用不同尺度核函数支持向量机所获得的分类准确率描述尺度核函数对于样本映射到高维线性特征空间的能力、适应能力、分类准确率和预测精度，因此本文使用 $acc_j(j=1,2,3,\cdots,M)$描述多尺度核加权合成的权重系数，即：$\frac{acc_j}{\sum_{i=1}^{M}acc_i}$, $j=1,2,3,\cdots,M$。由于 acc_j 描述的是不同尺度核支持向量机的分类准确率，因此必然有 $acc_j\geqslant 0$，又由于 $\sum_{j=1}^{M}\frac{acc_j}{\sum_{i=1}^{M}acc_i}=1$，其中$\sum_{i=1}^{M}acc_i$ 表示所有分类准确率的和是常数，用$\frac{acc_j}{\sum_{i=1}^{M}acc_i}(j=1,2,3,\cdots,M)$描述每一个不同尺度核函数的决策函数权重系数，将其代入式(1)得式(5)，将式(13)代入式(5)得：

$$\begin{cases} f(X)=\sum_{i=1}^{M}\frac{acc_i}{\sum_{j=1}^{M}acc_j}f_i(X) \\ f_i(x)=\text{sgn}\{\sum_{i=1}^{m}a_i y_i k_j(X_i,x)+b\}, i=1,2,3,\cdots,M \end{cases} \tag{14}$$

解式(14)得决策函数：

$$f(x)=\text{sgn}\{\sum_{i=1}^{m}\sum_{j=1}^{M}a_i y_i\frac{acc_k}{\sum_{k=1}^{M}acc_k}k_j(X_i,X)+b\} \tag{15}$$

其中，acc_k 是 M 个尺度核函数支持向量机的决策函数对于样本分类的准确率。通过决策函数式(15)学习样本 S 可以获得更适合样本 S 的结果以及更高的分类准确率、预测精度。通

过基于样本加权的多尺度核学习方法可获得一种适合各类样本的加权多尺度核学习策略，因此所提方法相比其他通过迭代学习获得加权系数的方法具有更好的泛化性能和适用性（本文提出的方法由于与样本相关，因此对于任意样本都能适用）。

下面给出所提出的基于样本加权的多尺度核方法的详细步骤描述：

(1)为给定样本 S 选定 M 个将样本映射到特征空间的不同尺度核函数；

(2)将 M 个不同尺度核函数代入支持向量机，通过学习样本 S(将样本 S 分为训练数据集和测试数据集，通过训练可以获得 M 个选定的不同尺度核函数支持向量机模型，分别对测试数据集进行测试)，获得 M 个不同尺度核函数支持向量机模型对样本 S 的分类准确率 acc_j $(j=1,2,3,\cdots,M)$；

(3)将学习样本 S 获得的分类准确率归一化，得到 $\frac{acc_j}{\sum_{i=1}^{M} acc_i}$ $(j=1,2,3,\cdots,M)$，即得到 M 个不同尺度核函数的决策函数的权重系数；

(4)将获得的权重系数代入式(5)得到本文提出的基于样本加权的多尺度核学习方法；

(5)将得到的多尺度核学习方法带入支持向量机决策函数，得到基于样本加权的多尺度核支持向量机方法。

根据上述步骤即可得到本文提出的多尺度核学习方法，并对本文提出的方法建立支持向量机分类模型，以适应多种、异构数据集的分类任务。

对于多类别样本，采用有向无环图支持向量机 (Directed Acyclic Graph Support Vector Machine, DAG-SVM)[36] 的方法学习，该步骤与上述处理两类别样本 S 的步骤类似：

(1)为给定多类别样本 S 选定 M 个将样本 S 映射到特征空间的不同尺度核函数；

(2)将 M 个不同尺度核函数代入支持向量机，通过学习样本 S(将样本 S 分为训练数据集和测试数据集，通过训练可以获得 M 个选定的不同尺度核函数支持向量机模型，分别对测试数据集进行测试)，获得 M 个不同尺度核函数 DAG-SVM 模型对样本 S 的分类准确率 acc_j $(j=1,2,3,\cdots,M)$；

(3)将学习样本 S 获得的分类准确率归一化，得到 $\frac{acc_j}{\sum_{i=1}^{M} acc_i}$ $(j=1,2,3,\cdots,M)$，即得到 M 个不同尺度核函数的决策函数的权重系数；

(4)将获得的权重系数代入式(5)得到本文提出的基于样本加权的多尺度核学习方法；

(5)将得到的多尺度核学习方法带入 DAG-SVM 决策函数，得到基于样本加权的多尺度核 DAG-SVM 方法。

3 实验分析

为了对本文方法的分类准确率进行定量分析，采用 UCI 数据集中的 wine，iris，waveform，breast，pima 以及乳腺 X 光医学图像标准数据集 MIAS 进行分类实验，采用两类别数据集和三类别数据集对所提方法分别进行分析，对于两类数据集采用基于样本加权的多尺度核支持向量机的方法进行分类，对于三类数据集采用基于样本加权的多尺度核有向无环图支持向量机方法进行定量分析（基于样本加权的多尺度核有向无环图支持向量机方法同样适用于更多类别的数据集分类情况）。其中，wine 数据集共有 178 组数据，3 个类，13 个属性；iris 数据集共有 150 组数据，3 个类，4 个属性；waveform 数据集共有 480 组数据，3 个类，21 个属性；breast 数据集共有 699 组数据，2 个类，10 个属性；pima 数据集共有 768 组数据，2 个类，8 个属性；MIAS 数据集共有 322 组数据，3 个类，64 个属性。其中对 MIAS 乳腺 X 光医学图像标准数据集使用灰度共生矩阵[37] 法提取特征数据集，再使用本文方法对提取到的特征数据集进行分类实验。通过 10 次随机抽取，抽取 wine，iris，waveform，breast，pima 以及 MIAS 中 80% 的数据作为训练数据，20% 的数据作为测试数据，分别进行 10 组仿真实验，同时每组实验的训练样本数据和测试样本数据均包含了 3 类样本。本次实验环境是在 Windows 8.1 64bit OS，4GB RAM，CPU 主频 3.5GHz，JAVA 平台下进行的，其中选定 3 个不同尺度的核函数，核参数分别为 $\sigma_1=1$，$\sigma_2=0.5$，$\sigma_3=0.1$。本文提出了一种基于样本加权的多尺度核支持向量机方法，选择以上 3 个不同尺度的核参数，其中 $\sigma_1=1$ 表示大尺度核，$\sigma_2=0.5$ 表示常规尺度核，$\sigma_3=0.1$ 表示小尺度核，通过对这 3 种尺度的核进行基于样本的加权合成实验，证明本文提出的方法可以适用于更广的范围，包括常规核、大尺度核以及小尺度核。

同时本次实验采用参数寻优的方法对支持向量机的惩罚参数 C 进行寻优，具体操作如下：

(1)对每一组进行 10 次仿真实验，设定一个惩罚参数 C，C 从 1 增加到 10，每次增加 1，并获得设定的惩罚参数对应的 10 次仿真实验的平均分类准确率，获得准确率最高的区间（假设在 5 时获得最高分类准确率，这时准确率最高的区间就是[4,6]），并且获得多个最高准确率时选择给定惩罚参数小的区间；

(2)对于获得的准确率最高的区间，从区间的下界增加到上界，每次增加 0.2，重复步骤(1)的操作，选择给定的惩罚参数对应平均分类准确率最高的区间，并将区间的中点值设置为本次实验支持向量机的惩罚参数 C，本文对惩罚参数的精度精确到小数点后一位。

本次仿真实验的结果如表 1、表 2 所列，其中表 1 展示的是本文方法对选定的多类别数据集进行分类实验的结果，表 2 展示的是本文方法对选定的两类别数据集进行分类实验的结果。本文对选定的每个数据集进行 10 组仿真实验，表 1、表 2 中组号代表的是 10 次仿真实验的序号，分类准确率代表的是每次仿真实验的分类结果。

使用本文方法对 wine 数据集进行分类，分类准确率最高达到了 100%，并且平均分类准确率达到了 98.33%，同时对惩罚参数 C 采用本文的寻优方法，获得最佳的惩罚参数 C，表 1(a)展示的对 wine 数据集的分类准确率是在最佳惩罚参数 C 的条件下获得的，本文对 wine 数据集使用最佳惩罚参数 $C=1.0$。

使用本文提出的方法对 iris 数据集进行分类，分类准确率最高达到了 100%，并且平均分类准确率达到了 98.67%，对惩罚参数进行寻优，获得最佳的惩罚参数 C。表 1 展示的对 iris 数据集的分类准确率是在最佳惩罚参数 C 的条件下获得的，本文对 iris 数据集使用最佳惩罚参数 $C=1.8$。

使用基于样本加权的多尺度核支持向量机方法对 waveform 数据集进行分类，分类准确率最高达到了 100%，并且平均分类准确率达到了 97.38%，对惩罚参数进行寻优，获得最

佳的惩罚参数 C。表 1 展示的对 waveform 数据集的分类准确率是在最佳惩罚参数 C 的条件下获得的，本文对 waveform 数据集使用最佳惩罚参数 $C=1.0$。

使用本文提出的方法对 MIAS 数据集进行分类，分类准确率最高达到了 90.63%，并且平均分类准确率达到了 84.066%，对惩罚参数进行寻优，获得最佳的惩罚参数 C。表 1 展示的对 MIAS 数据集的分类准确率是在最佳惩罚参数 C 的条件下获得的，本文对 MIAS 数据集使用最佳惩罚参数 $C=8.0$。

表 1 本文方法在多类别数据上的分类结果(%)

组号	分类准确率			
	wine	iris	waveform	MIAS
1	99.12	100	95.83	84.38
2	95.59	100	95.83	84.38
3	98.24	93.33	93.75	87.5
4	100	100	97.92	71.88
5	98.24	93.33	97.92	78.13
6	97.35	100	97.92	90.63
7	98.24	100	100	84.38
8	98.24	100	97.92	87.5
9	98.24	100	97.92	84.38
10	100	100	95.83	87.5

使用基于样本加权的多尺度核支持向量机方法对 breast 数据集进行分类，分类准确率最高达到了 100%，并且平均分类准确率达到了 97.29%，对惩罚参数进行寻优，获得最佳的惩罚参数 C。表 2 展示的对 breast 数据集的分类准确率是在最佳惩罚参数 C 的条件下获得的，本文对 breast 数据集使用最佳惩罚参数 $C=1.6$。

使用本文提出的基于样本加权的多尺度核支持向量机方法对 pima 数据集进行分类，分类准确率最高达到了 100%，并且平均分类准确率达到了 96.40%，同时对惩罚参数进行寻优，获得最佳的惩罚参数 C。表 2 展示的对 pima 数据集的分类准确率是在最佳惩罚参数 C 的条件下获得的，本文对 pima 数据集使用最佳惩罚参数 $C=1.0$。

表 2 本文方法在两类别数据上的分类结果(%)

组号	分类准确率	
	breast	pima
1	97.86	95.42
2	100	97.38
3	96.43	95.42
4	92.86	95.42
5	97.86	97.38
6	97.86	100
7	97.86	94.77
8	97.86	95.42
9	96.43	95.42
10	97.86	97.38

对于 wine，iris，waveform 3 个数据集，采用 DAG-SVM，1-vs-1 SVM[38]，1-vs-a SVM[39] 等方法进行对比实验，对每种方法使用惩罚参数寻优，并且使用分类准确率最高的一组实验对应的惩罚参数 C 作为所对应实验的惩罚参数，上述支持向量机方法使用径向基核函数，核参数 σ 取值为 1。以上方法对 wine，iris，waveform 3 个数据集的平均分类准确率如表 3 所列。

观察表 3 可以得出，本文方法在 wine，iris，waveform 3 个数据集上都取得了较好的分类准确率。在 wine 数据集上，本文方法较 DAG-SVM，1-vs-1 SVM，1-vs-a SVM 等方法分别提高了 1.886%，2.146%，2.146%；在 iris 数据集上，本文方法较 DAG-SVM，1-vs-1 SVM，1-vs-a SVM 等方法分别提高了 1.34%，2%，2%；在 waveform 数据集上，本文方法较 DAG-SVM，1-vs-1 SVM，1-vs-a SVM 等方法分别提高了 2.45%，2.51%，2.51%。

表 3 本文方法在分类准确率上与其他方法的比较(%)

方法	wine	iris	waveform
DAG-SVM	96.44	97.33	94.93
1-vs-1 SVM	96.18	96.67	94.87
1-vs-a SVM	96.18	96.67	94.87
本文方法	98.326	98.67	97.38

对于 pima 和 breast 两个数据集，采用 SVM、加权合成多核支持向量机(Weighted Summation Kernel Support Vector Machine，WSK-SVM)方法进行对比实验，对每种方法使用惩罚参数寻优，并且使用分类准确率最高的一组实验对应的惩罚参数 C 作为所对应实验的惩罚参数，对 SVM 使用径向基核函数，核参数取值为 1。以上方法对于 pima 和 breast 两个数据集的平均分类准确率如表 4 所列。

表 4 本文方法在分类准确率上与其他方法的比较(%)

方法	pima	breast
SVM	95.78	95.31
WSK-SVM	96.18	96.12
本文方法	96.40	97.29

观察表 4 可以得出，本文提出的方法在 pima 和 breast 两个数据集上也都取得了最佳的分类准确率，在 pima 数据集上本文方法较 SVM 方法和 WSK-SVM 方法分别提高了0.62%和0.32%；在 breast 数据集上本文方法较 SVM 方法和 WSK-SVM 方法分别提高了 1.98%和 1.17%。

对于 MIAS 数据集的特征集，采用 DAG-SVM，1-vs-1 SVM 和判别式受限波尔兹曼机(DRBM)[40] 等方法进行对比实验，使用惩罚参数对每个支持向量机方法寻优，并且使用分类准确率最高的一组实验对应的惩罚参数 C 作为所对应实验的惩罚参数，上述支持向量机方法使用径向基核函数，核参数取值为 1。以上方法对 MIAS 数据集的平均分类准确率如表 5 所列。

表 5 本文方法在分类准确率上与其他方法的比较(%)

方法	本文方法	DAG-SVM	1-vs-1 SVM	DRBM
分类精度	84.066	77.406	75.156	78.46

观察表 5 可以得出，本文方法在 MIAS 数据集上取得了较高的分类准确率，较 DAG-SVM 方法、1-vs-1 SVM 方法、DRBM 方法分别提高了 6.66%，8.91%，5.606%。

通过以上分析可知，基于样本加权的多尺度核支持向量机方法在多个数据集上的分类实验均取得了较高的分类准确率。无论是对于两类别问题还是对于多类别问题，无论是对于 UCI 标准数据集还是更贴近现实的 MIAS 乳腺 X 光医学图像数据集，所提方法都能取得较好的分类结果，因此可以得出：所提方法对于样本的适应性更高，灵活性更强，同时也具有更好的泛化性能。

结束语 本文提出了一种基于样本加权的多尺度核支持向量机方法，并将其应用于 UCI 数据集和 MIAS 乳腺 X 光医学图像标准数据集的分类中，获得了良好的分类效果。将该方法与 DAG-SVM 方法、1-vs-1 SVM 方法、1-vs-a SVM 方法、WSK-SVM 方法以及判别式受限波尔兹曼机进行比较，实

验结果表明，本文方法在 wine，iris，waveform，breast，pima 以及 MIAS 共 6 个数据集上的分类效果较其他几种方法有明显的提升。但是针对大数据分类问题，如何应用本文提出的方法是下一步研究的主要内容。

参 考 文 献

[1] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers[M] // Haussler D. eds., 5th Annual ACM Workshop on COLT, Pittsburgh, PA. ACM Press., 1992

[2] Vapnik V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995

[3] Wang L P. Support vector machine: theory and application [M]. Berlin: Springer-Verlag, 2005

[4] Schoelkopf B, Smola A, Muller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5): 1299-1319

[5] Scholkopf B, Mika S, Burges C J C, et al. Input space versus feature space in kernel-based methods[J]. IEEE Transactions on Neural Network, 1999, 10(5): 1000-1017

[6] Muller K R, Mika S, Ratsch G, et al. An introduction to kernel based learning algorithms[J]. IEEE Transactions on Neural Networks, 2001, 12(2): 181-201

[7] Mercer J. Functions of positive and negative type and their connection with the theory of integral equations[J]. Philosophical Transactions of the Royal Society of London, Series A, 1909, 209: 415-446

[8] Aronszajn N. Theory of reproducing kernels[J]. Transactions of the American Mathematical Society, 1950, 68(3): 337-404

[9] Aizerman A, Braverman E M, Rozoner L I. Theoretical foundations of the potential function method in pattern recognition learning[J]. Automation and Remote Control, 1964, 25(5): 821-837

[10] Smola A J, Scholkopf B. A tutorial on support vector regression [J]. Statistics and Computing, 2004, 14(3): 199-222

[11] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Data Mining and Knowledge Discovery, 1998, 2 (2): 121-167

[12] Kerm P V. Adaptive kernel density estimation [J]. Stata Journal, 2003, 3(2): 148-156

[13] Schöolkopf B, Mika S, Smola A, et al. Kernel PCA pattern reconstruction via approximation preimages[C] // Proceedings of the International Conference of Artificial Neural Networks. Skovde, Sweden: IEEE, 1998: 147-152

[14] Mika S, Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels[C] // Proceedingsof the Conference on Neural Networks for Signal Processing. Washington D. C., USA: IEEE, 1999: 41-48

[15] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach[J]. Neural Computation, 2000, 12(10): 2385-2404

[16] Wang Hong-qiao, Sun Fu-chun, Cai Yan-ning, et al. On multiple kernel learning methods[J]. Acta Automatic Sinica, 2010, 36 (8): 1037-1050(in Chinese)
汪洪桥，孙富春，蔡艳宁，等. 多核学习方法[J]. 自动化学报，2010，36(8): 1037-1050

[17] Huang C, Chen Y, Chen W, et al. Gastroesophageal Reflux Disease Diagnosis Using Hierarchical Heterogeneous Descriptor Fusion Support Vector Machine[C] // IEEE Engineering in Medicine and Biology Society. 2015: 1-10

[18] Penga S, Hua Qing-hua, Chen Yin-li, et al. Interactive, Improved support vector machine algorithm for heterogeneous data [J]. Pattern Recognition, 2015, 48(6): 2072-2083

[19] Sonnenburg S, Ratsch G, Schafer C, et al. Large scale multiple kernel learning[J]. The Journal of Machine Learning Research, 2006, 7(7): 1531-1565

[20] Xiao Yu-lin, Zhong Shang-ping. An improved online multiple kernel classification algorithm based on double updating online learning [C] // 2014 International Conference on Cloud Computing and Internet of Things. 2014: 109-113

[21] Bach F R. Consistency of the group Lasso and multiple kernel learning[J]. The Journal of Machine Learning Research, 2008, 9 (6): 1179-1225

[22] Cortes C, Mohri M, Rostamizadeh A. Learning sequence kernels [C] // Proceedings of the International Conference on Machine Learning for Signal Processing. Washington D. C., USA: IEEE, 2008: 2-8

[23] Rakotomamonjy A, Bach F R, Canu S, et al. Simple MKL[J]. The Journal of Machine Learning Research, 2008, 9(11): 2491-2521

[24] Zheng D N, Wang J X, Zhao Y N. Nonflat function estimation with a multi-scale support vector regession[J]. Neurocomputing, 2006, 70(1-3): 420-429

[25] Wang Jing-yan, Bensmailc H, Gao Xin. Feature selection and multi-kernel learning for sparse representation on a manifold [J]. Neural Networks, 2014, 51(3): 9-16

[26] Xua Lin, Feng Yan-qiu, Liu Xiao-yun, et al. Robust GRAPPA reconstruction using sparse multi-kernel learning with least squares support vector regression [J]. Magnetic Resonance Imaging, 2014, 32(1): 91-101

[27] Wang Jing-yan, Huang Jian-hua, Sun Yi-jun, et al. Feature selection and multi-kernel learning for adaptive graph regularized nonnegative matrix factorization[J]. Expert Systems with Applications, 2015, 42(3): 1278-1286

[28] Lanckriet G R G, Cristianini N, Bartlett P, et al. Learning the kernel matrix with semi definite programming[J]. The Journal of Machine Learning Research, 2004, 5(1): 27-72

[29] Lee W J, Verzakov S, Duin R P. Kernel combination versus classifier combination[C] // Proceedings of the 7th International Workshop on Multiple Classifier Systems. Prague, Czech Republic: Springer, 2007: 22-31

[30] Bi J B, Zhang T, Bennett K P. Column-generation boosting methods for mixture of kernels[C] // Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, USA: ACM, 2004: 521-526

[31] Kingsbury N, Tay D B H, Palaniswami M. Multi-Scale Kernel Methods for Classification[C] // Proc. of the IEEE Workshop on Machine Learning for Signal Processing. Mystic, USA, 2005: 43-48

[32] Yang Zhen, Guo Jun, Xu Wei-ran. Multi-Scale Support Vector Machine for Regression Estimation[C] // Proc. of the 3rd International Symposiumon Neural Networks. Chengdu, China, 2006: 1030-1037

[33] Izmailov R, Bassu D, McIntosh A, et al. Application of multiscale singular vector decomposition to vessel classification in overhead satellite imagery[C] // Seventh International Conference on Digital Image Processing (ICDIP 2015). 2015

[34] Vapnik V N. 统计学习理论[M]. 许建华,张学工,译. 北京:电子工业出版社,2009
[35] Ding Shi-fei, Qi Bing-juan, Tan Hong-yan. An overview on theory and algorithm of support vector machines [J]. Journal of University of Electronic Science and Technology of China, 2011, 40(1): 2-10(in Chinese)
丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报, 2011, 40(1): 2-10
[36] Platt J C, Cristianini N, Shawe-Taylor J. Large Margin DAGs for multiclass classification[J]. Advances in Neural Information Processing Systems, 2000, 12(3): 547-553
[37] Manivannan K, Aggarwal P, Devabhaktuni V, et al. Particulate matter characterization by gray level co-occurrence matrix based support vector machines [J]. Journal of hazardous materials, 2012, 223-224(2): 94-103
[38] Hsu C W, Lin C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. IEEE Transactions on Neural Networks, 2002, 13(3): 415-425
[39] Kre U, et al. Pairwise classification and support vector machines [C]// Advances in Kernel Methods. MIT Press, 1999: 255-268
[40] Larochelle H, Bengio Y. Classification using Discriminative Restricted Boltzmann Machines[C]// Proceedings of the 25th International Conference on Machine Learning, 2008. Helsinki, Finland, 2008: 1-8

(上接第 129 页)

表 4 运用 5-折交叉验证运行 10 次对结论 1)的统计分析

数据集	均匀-高斯		均匀-指数	
	P-值 1	w	P-值 2	w
Forest	0.1428	1	5.9547E−08	2
Optical	0.3272	1	2.3619E−04	1
CT	0.1820	1	1.6185E−05	1
RenRu	0.2936	1	1.2497E−07	3

从表 3 和表 4 的 P-值 1 可以看出,用服从均匀分布和高斯分布的随机数初始化输入层权值和隐含层结点的偏置得到的测试精度没有本质的区别,这一结论成立的概率至少为 0.95。从表 3 和表 4 的 P-值 2 可以看出,用服从均匀分布和高斯分布的随机数初始化输入层权值和隐含层结点的偏置得到的测试精度有本质的区别,这一结论成立的概率至少为0.95。因此,本文得出的结论 1)具有可信性。

结束语 本文研究了在极限学习机中随机权分布对学习系统测试精度的影响。具体地,用实验方法分别研究了随机权服从均匀分布、高斯分布和指数分布对极限学习机性能的影响。在 18 个数据集上进行了大量的实验研究,通过对实验结果的分析发现了有趣的结论。例如,用服从[−1,+1]区间均匀分布的随机数初始化单隐含层前馈神经网络不一定是最优的选择。本文得出的结论对从事极限学习机网络研究的人员具有一定的借鉴作用。

参考文献

[1] Huang Guang-bin, Zhu Qin-yu, Siew C K. Extreme learning machine: A new learning scheme of feedforward neural networks [C]// Proceedings of International Joint Conference on Neural Networks(IJCNN2004). 2004: 985-990
[2] Huang Guang-bin, Wang Dian-hui, Lan Yuan. Extreme learning machines: a survey [J]. International Journal of Machine Learning and Cybernetics, 2011, 2(2): 107-122
[3] Huang Guang-bin, Zhu Qin-yu, Siew C K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70(1-3): 489-501
[4] Huang Guang-bin, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes [J]. IEEE Transactions on Neural Networks, 2006, 17(4): 879-892
[5] Huang Guang-bin, Chen Lei. Enhanced random search based incremental extreme learning machine [J]. Neurocomputing, 2008, 71(16-18): 3460-3468
[6] Huang Guang-bin, Chen Lei. Convex incremental extreme learning machine [J]. Neurocomputing, 2007, 70(16): 3056-3062
[7] Liang Nan-ning, Huang Guang-bin, Saratchandran P, et al. A fast and accurate on-line sequential learning algorithm for feedforward networks [J]. IEEE Transactions on Neural Networks, 2006, 7(6): 1411-1423
[8] Liu Qiu-ge, He Qing, Shi Zhong-zhi. Extreme support vector machine classifier [C]// Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. Springer-Verlag. 2008: 222-233
[9] Huang Guang-bin, Ding Xiao-jian, Zhou Hong-ming. Optimization method based extreme learning machine for classification [J]. Neurocomputing, 2010, 74(1-3): 155-163
[10] Huang Guang-bin, Zhou Hong-ming, Ding Xiao-jian, et al. Extreme Learning Machine for Regression and Multiclass Classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B, 2012, 42(2): 513-529
[11] Emilio S O, Juan G S, Martín J D, et al. BELM: Bayesian Extreme Learning Machine [J]. IEEE Transactions on Neural Networks, 2011, 22(3): 505-509
[12] Feng Guo-rui, Huang Guang-bin, Lin Qing-ping, et al. Error Minimized extreme learning machine with growth of hidden nodes and incremental learning [J]. IEEE Trans. Neural Networks, 2009, 20(8): 1352-1357
[13] Rong Hai-jun, Ong Y S, Tan A H, et al. A fast pruned-extreme learning machine for classification problem [J]. Neurocomputing, 2008, 72(1-3): 359-366
[14] Miche Y, Sorjamaa A, Bas P, et al. OP-ELM: Optimally pruned extreme lear-ning machine [J]. IEEE Transactions on Neural Networks, 2010, 21(1): 158-162
[15] Mohammed A A, Minhas R, Jonathan Q M, et al. Human face recognition based on multidimensional PCA and extreme learning machine [J]. Pattern Recognition, 2011, 44(10/11): 2588-2597
[16] Iosifidis A, Tefas A, Pitas I. Minimum Class Variance Extreme Learning Machine for Human Action Recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(11): 1968-1979
[17] Frank A, Asuncion A. UCI machine learning repository [OL]. http://archive.ics.uci.edu/ml
[18] Zhai Jun-hai, Li Ta, Zhai Meng-yao, et al. Experimental Research on Random Mapping Function in ELM Algorithm [J]. Computer Engineering, 2012, 38(20): 164-168(in Chinese)
翟俊海,李塔,翟梦尧,等. ELM 中随机映射作用的实验研究[J]. 计算机工程, 2012, 38(20): 164-168