

## 基于 $l_p$ -范数约束的 LSSVR 多核学习算法

李琦, 李晓航, 邢丽萍, 邵诚

(大连理工大学 辽宁省工业装备先进控制系统重点实验室, 辽宁 大连 116024)

**摘要:** 针对核函数选择对最小二乘支持向量机回归模型泛化性的影响, 提出一种新的基于  $l_p$ -范数约束的最小二乘支持向量机多核学习算法. 该算法提供了两种求解方法, 均通过两重循环进行求解, 外循环用于更新核函数的权值, 内循环用于求解最小二乘支持向量机的拉格朗日乘数, 充分利用该多核学习算法, 有效提高了最小二乘支持向量机的泛化能力, 而且对惩罚参数的选择具有较强的鲁棒性. 基于单变量和多变量函数的仿真实验表明了所提出算法的有效性.

**关键词:** 最小二乘支持向量机;  $l_p$ -范数; 多核学习; 泛化性

**中图分类号:** TP181

**文献标志码:** A

## Multiple kernel learning LSSVR algorithm based on $l_p$ -norm constraint

LI Qi, LI Xiao-hang, XING Li-ping, SHAO Cheng

(The Liaoning Key Lab of Advanced Control Systems for Industry Equipments, Dalian University of Technology, Dalian 116024, China. Correspondent: LI Qi, E-mail: liqi\_dut@163.com)

**Abstract:** In order to improve generalization performance of learning least squares support vector machines regression(LSSVR), a novel multiple kernel learning least squares support vector machines regression algorithm based on  $l_p$ -Norm constraint is proposed. Two wrapper methods are provided to solve the proposed algorithm, and both the training method are two-step methods. The inner loop is used to update the combination function parameters while fixing the least squares support vector machine(LSSVM) parameters, the outside loop is used to update the parameters of LSSVM while fixing the combination function parameters, and these two steps are repeated until convergence. The simulation on the one-variable function and multivariable function shows that the proposed algorithm is useful and outperforms the traditional LSSVR algorithm for generalization performance.

**Keywords:** least squares support vector machine;  $l_p$ -norm constraint; multiple kernel learning; generalization performance

## 0 引言

最小二乘支持向量机回归(LSSVR)算法中, 核函数的选择扮演着重要的角色. 传统的最小二乘支持向量机中常采用单核模型, 近年来的理论和应用表明, 多核代替单核能增强决策函数的可解释性, 并能获得比单核模型更优的性能<sup>[1]</sup>. 根据 Mercer 原理<sup>[2]</sup>和线性加权原则, Lanckriet 等<sup>[3]</sup>提出了多核学习问题(MKL), 即实现对多个基核函数的线性组合. 通常在多核框架下, 将原始核函数的选择问题转化为确定多核组合系数的问题.

多数针对最小二乘支持向量机的多核学习算

法采用  $l_1$ -范数约束, 对于核函数权值的求解, Jian 等<sup>[4]</sup>提出了基于半定规划和基于二次约束型二次规划的求解方法. Chen 等<sup>[5]</sup>提出了基于半无限线性规划的多核学习求解方法. 上述方法中,  $l_1$ -范数的多核函数的权值具有稀疏性, 即权值系数可以为 0, 具有可解释性和可拓展性. 但是, 稀疏性并不总是有效的, 也具有局限性, 例如, 当核函数之间存在互补的核函数信息时, 稀疏解可能丢失有用的信息, 弱化模型的泛化性能.

针对多核学习问题, Klöf 等<sup>[6]</sup>提出了一种非稀疏化的模式, 即  $l_p$ -范数 MKL 学习, 进一步增强了多

收稿日期: 2014-06-03; 修回日期: 2014-12-10.

基金项目: 国家自然科学基金项目(61403058); 中国石油科技创新基金项目(2014D-5006-0601); 中央高校基本科研业务费项目(DUT14LAB15).

作者简介: 李琦(1979—), 男, 副教授, 博士, 从事工业过程软测量、监测算法的研究; 李晓航(1988—), 女, 硕士生, 从事机器学习算法的研究.

核机器学习的鲁棒性和通用性. 由于  $l_p$ -范数 MKL 学习的有效性, Yan 等<sup>[7]</sup>将 MKL 的思想应用于 Fisher 判别分析中, 提出了基于  $l_p$ -范数的非稀疏的多核 FDA, 将优化问题转换为半无限规划问题, 并给出了迭代的 Wrapper 算法求解. Sun 等<sup>[8]</sup>为了消除  $l_p$ -范数 MKL 学习的冗余性, 提出了选择性多核学习方法, 首先根据子核函数之间的 Manhattan 距离, 利用  $C$  均值聚类的方法将子核函数分为几类, 然后选择每类中差异性最大的子核函数, 将  $l_p(p > 1)$ -范数的 MKL 问题求解约简成  $l_\infty$ -范数约束的特殊类型.

本文在此基础上, 将最小二乘支持向量机与  $l_p$ -范数 MKL 学习相结合, 提出一种新的基于  $l_p$ -范数多核学习的最小二乘支持向量机算法, 并给出两种求解算法. 仿真实验表明, 所提出算法能够有效提高模型的泛化性能, 而且对惩罚参数的选择具有较强的鲁棒性, 具有较好的应用前景.

## 1 多核学习问题描述

假设训练集数据

$$D = \{(x_i, y_i)\}, i = 1, 2, \dots, n.$$

其中:  $x_i$  为输入数据集,  $y_i \in y \in R$  为输出数据集. 目标是找到一个映射空间  $f \in H$ , 对已知的数据和未知的数据均具有良好的泛化性能. 结构风险的最小化问题  $h^*$  可以表示为

$$h^* \in \arg \min_h R_{\text{emp}}(h) + \lambda \Omega(h). \quad (1)$$

其中:  $R_{\text{emp}}(h) = \frac{1}{n} \sum_{i=1}^n V(h(x_i), y_i)$  为经验风险,  $V: R \times y \rightarrow R$  为一个凸损失函数;  $\Omega: H \rightarrow R$  为一个调整项;  $\lambda$  为一个调节参数. 模型表达式为

$$f_\omega(x) = \langle \tilde{\omega}, \Phi(x) \rangle + b. \quad (2)$$

其中:  $\tilde{\omega}$  为权值向量;  $\Phi: x \rightarrow H$  为映射函数, 将  $x$  映射到高维的 Hilbert 空间  $H$  中;  $b$  为偏置量. 调整项  $\Omega(h) = \|\tilde{\omega}\|_2^2/2$  用来控制模型和算法的核的权值, 核函数以数据之间的内积的方式实现, 表达方式为

$$k(x, x') = \langle \Phi(x), \Phi(x') \rangle_H.$$

对于 MKL 多核学习, 假设有  $M$  个不同的映射函数  $\Phi_m: x \rightarrow H_m, m = 1, 2, \dots, M$ , 每个映射表示不同的核函数  $k_m$ . 线性组合多核学习的约束表示为

$$k_\theta = \sum_{m=1}^M \theta_m k_m, \theta_m \geq 0.$$

对比式 (2), 多核学习的模型表达式应该拓展为

$$f_\omega(x) = \sum_{m=1}^M \sqrt{\theta_m} \langle \tilde{\omega}, \Phi_m(x) \rangle_{H_m} + b = \langle \tilde{\omega}, \Phi_\theta(x) \rangle_{H_\theta} + b. \quad (3)$$

其中:  $\tilde{\omega} = (\tilde{\omega}_1^T, \dots, \tilde{\omega}_M^T)$  为参数向量,  $\Phi_\theta = \sqrt{\theta_1} \Phi_1 \times$

$\dots \times \sqrt{\theta_M} \Phi_M$  为相对应的映射空间模块.

对于多核学习, 主要目的是使训练数据集的结构风险 RSM 最小, 增加线性多核组合优化  $\sum_{m=1}^M \theta_m k_m$  和  $\theta$  调整项, 避免过拟合问题. 因此, 多核学习的结构风险 RSM 最小化问题描述为

$$\inf_{\omega, b, \theta: \theta \geq 0} \frac{1}{n} \sum_{i=1}^n V \left( \sum_{m=1}^M \sqrt{\theta_m} \langle \tilde{\omega}_m, \Phi_m(x_i) \rangle_{H_m} + b, y_i \right) + \frac{\lambda}{2} \sum_{m=1}^M \|\tilde{\omega}_m\|_{H_m}^2 + \tilde{\mu} \Omega(\theta), \quad (4)$$

其中  $\tilde{\mu} > 0$ . 式 (4) 的解是训练误差的上界,  $\Omega(\theta)$  表示  $\theta$  的任意范数, 当  $\Omega(\theta) = \|\theta\|_1$  是多核学习的 1-范数表示形式时, 核函数的权值具有稀疏解,  $V(\cdot)$  表示损失函数.

式 (4) 的非凸性不是固有的, 可以通过变换的方法解决. 应用替换的方法  $\omega_m \leftarrow \sqrt{\theta_m} \tilde{\omega}_m$ , 解除样本个数与调节参数之间的关系, 使用  $C \leftarrow \frac{1}{n\lambda}, \mu \leftarrow \frac{\tilde{\mu}}{\lambda}$  替换的方法, 在实际应用中有良好的效果. 多核学习的问题 (4) 可以表示为

$$\inf_{\omega, b, \theta: \theta \geq 0} C \sum_{i=1}^n V \left( \sum_{m=1}^M \langle \omega_m, \Phi_m(x_i) \rangle_{H_m} + b, y_i \right) + \frac{1}{2} \sum_{m=1}^M \frac{\|\omega_m\|_{H_m}^2}{\theta_m} + \mu \Omega(\theta). \quad (5)$$

## 2 $l_p$ -范数约束的多核学习描述与求解

### 2.1 $l_p$ -范数约束的多核学习描述

根据上述介绍的多核学习问题, 当式 (5) 的约束范数为  $P$  范数时, MKL 在  $l_p(p > 1)$  约束条件下的优化问题可以表示为

$$\inf_{\omega, b, \theta: \theta \geq 0} C \sum_{i=1}^n V \left( \sum_{m=1}^M \langle \omega_m, \Phi_m(x_i) \rangle_{H_m} + b, y_i \right) + \frac{1}{2} \sum_{m=1}^M \frac{\|\omega_m\|_{H_m}^2}{\theta_m} + \mu \|\theta\|_p^2. \quad (6)$$

其中:  $\mu > 0, \|\theta\|_p^2$  为  $\|\cdot\|^2$  属于  $R^M$  的  $P$ -范数.

MKL 学习的模型主要有两种表达形式: 一种是 Ivanov 正则优化<sup>[9]</sup>, 另一种是 Tikhonov 正则优化<sup>[10]</sup>, 式 (6) 为 MKL 的 Tikhonov 正则优化.  $l_p(p > 1)$  约束条件下多核学习的 Ivanov 正则优化问题表达式为

$$\inf_{\omega, b, \theta: \theta \geq 0} C \sum_{i=1}^n V \left( \sum_{m=1}^M \langle \omega_m, \Phi_m(x_i) \rangle_{H_m} + b, y_i \right) + \frac{1}{2} \sum_{m=1}^M \frac{\|\omega_m\|_{H_m}^2}{\theta_m}, \quad \text{s.t. } \|\theta\|_p \leq 1. \quad (7)$$

文献 [6] 证明, Ivanov 正则优化式 (7) 与 Tikhonov 正则优化式 (6) 的 MKL 优化问题是等价的.

### 2.2 $l_p$ -范数约束的多核学习求解

根据 Tikhonov 正则优化公式 (6), 假设参数  $(\omega, b)$  固定, 式 (6) 对  $\theta$  的偏导数等于零, 可以得到

$$-\frac{\|\omega_m\|_{H_m}^2}{2\theta_m^2} + \mu \frac{\partial(\|\theta\|_p^2/2)}{\partial\theta_m} = 0, \quad (8)$$

其中  $l_p$ -范数的偏导数为

$$\frac{\partial(\|\theta\|_p^2/2)}{\partial\theta_m} = \theta_m^{p-1} \|\theta\|_p^{2-p}. \quad (9)$$

可以将式 (8) 变换为

$$\exists \zeta \forall m = 1, 2, \dots, M : \theta_m = \zeta \|\omega_m\|_{H_m}^{\frac{2}{p+1}}. \quad (10)$$

$\|\theta\|_p \leq 1$  是式 (6) 的上界, 当  $\|\theta\|_p = 1$  时可以得到最优值, 代入到式 (10) 可以得到

$$\zeta = \left( \sum_{m=1}^M \|\omega_m\|_{H_m}^{2p/p+1} \right)^{1/p},$$

进而得到迭代的更新权值的方法, 即

$$\theta_m = \frac{\|\omega_m\|_{H_m}^{\frac{2}{p+1}}}{\left( \sum_{m'=1}^M \|\omega_{m'}\|_{H_{m'}}^{2p/p+1} \right)^{1/p}}. \quad (11)$$

## 3 基于 $l_p$ -范数约束的 LSSVR 多核学习

### 3.1 算法描述

根据上述 MKL 问题的描述, 结合 LSSVR 的优化问题,  $l_p$ -范数约束的 MKL-LSSVR 问题求解表达式为

$$\begin{aligned} \min & \frac{1}{2} \sum_{m=1}^M \frac{\|\omega_m\|_2^2}{\theta_m} + C \sum_{i=1}^N \xi_i^2, \\ \text{s.t. } & y_i = \left( \sum_{m=1}^M \sqrt{\theta_m} \langle \omega_m, \Phi_m(x_i) \rangle_{H_m} + b \right) - \xi_i. \end{aligned} \quad (12)$$

其中:  $\omega_m \in R^{s_m}$ ,  $\xi_i \in R_+^N$ ,  $b \in R$ ,  $\theta \in R_+^M$ ,  $\|\theta\|_p \leq 1$ ,  $\theta_m \geq 0$ .

LSSVR 优化问题在对偶空间的拉格朗日表达式为

$$\begin{aligned} \max_a & \sum_{i=1}^n y_i a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j \sum_{m=1}^M \theta_m K(x_i, x_j) - \\ & \frac{1}{2C} \sum_{i=1}^n a_i a_j, \\ \text{s.t. } & \sum_{i=1}^n a_i = 0, \forall i = 1, 2, \dots, n. \end{aligned} \quad (13)$$

带  $l_p$ -范数约束的 MKL-LSSVR 在对偶空间的求解问题可以变换为

$$\begin{aligned} \min_{\theta} \max_a & a^T y - \frac{1}{2} a^T \sum_{m=1}^M \theta_m K(x_i, x_j) a - \frac{1}{2C} a^T a, \\ \text{s.t. } & \|\theta\|_p \leq 1, \theta_m \geq 0. \end{aligned} \quad (14)$$

对  $\|\theta\|_p$  进行 Taylor 二次展开, 有

$$\begin{aligned} \|\theta\|_p & \approx \\ & \frac{p(p-1)}{2} \sum_{m=1}^M \tilde{\theta}_m^{p-2} \theta_m^2 - \sum_{m=1}^M p(p-2) \tilde{\theta}_m^{p-1} \theta_m + \\ & \frac{p(p-3)}{2} + 1 = \nu(\theta), \end{aligned} \quad (15)$$

其中  $\tilde{\theta}_m$  为目前  $\theta_m$  的一个近似值. 将式 (15) 代入 (14), MKL-LSSVR 的求解可以变换为

$$\begin{aligned} \min_{\theta} \max_a & a^T y - \frac{1}{2} a^T \sum_{m=1}^M \theta_m K a - \frac{1}{2C} a^T a, \\ \text{s.t. } & \nu(\theta) \leq 1, \theta \geq 0. \end{aligned} \quad (16)$$

由式 (16) 可见, MKL-LSSVR 问题可以转换成一个二次约束的线性规划 QCLP (quadratically constrained linear program) 问题, 其中  $\nu(\theta) \leq 1$  是二次约束, 本文使用 MOSEK 优化工具箱对其进行求解.

### 3.2 $l_p$ -范数约束的 LSSVR 多核学习求解

针对  $l_p$ -范数约束 MKL-LSSVR 求解问题, 本文提出两种求解方法. 算法分为外循环和内循环两重循环, 内循环使用传统的 LSSVR 方法求解拉格朗日乘数  $a$  的值, 外循环使用迭代方法更新组合核函数  $\theta$  的权值.

对于组合核函数  $\theta$  的求取方法, 本文提出了两种方法. 算法 1 以 Klotz 的思想为根据, 利用式 (11) 进行更新组合核函数的权值, 根据 LSSVR 的求解得到  $\omega_m = \sum_{i=1}^N a_i \Phi(x_i)$ , 代入式 (11) 后更新权值, 具体过程如下.

**算法 1** 简单迭代求解  $l_p$ -范数约束的 MKL-LSSVR.

输入:  $K_1, K_2, \dots, K_M$ ,  $\theta_m^{(1)} = M^{-1/p}$ ;

输出: 线性组合核函数的权值  $\theta = (\theta_1^{(t)}, \dots, \theta_M^{(t)})^T$ .

若不满足停止条件则执行如下步骤.

Step 1: 根据标准 LSSVR 表达式求解  $a^{(t)}$ .

Step 2: 计算  $\|\omega_m\|^2 = \theta_m^2 a^T K_m a$ ,  $m = 1, 2, \dots, M$ .

Step 3: 根据式 (11) 更新核函数的权值.

上述算法的停止条件为: 在权值变化较小时停止循环, 即

$$|\max(\theta_m(t) - \theta_m(t-1))| \leq \xi,$$

其中  $\xi$  是一个非常小的正数. 算法 1 是多核学习在原空间的求解过程.

算法 2 是多核学习的 LSSVR 在对偶空间的求解方法, 基于 QCLP 方法对组合核函数  $\theta$  的权值进行更新. 式 (16) 转换为

$$\begin{aligned} \min S; \\ \text{s.t. } a^T y - \frac{1}{2} a^T \sum_{m=1}^M \theta_m K a - \frac{1}{2C} a^T a \leq S, \\ \theta \geq 0, \nu(\theta) \leq 1. \end{aligned} \quad (17)$$

内循环拉格朗日乘数  $a^{(t)}$  在  $\theta^{(t)}$ 、 $S^{(t)}$  固定的情况下由标准的 LSSVR 求解得到; 外循环求解  $\theta^{(t)}$ 、 $S^{(t)}$  的更新时, 需要满足如下约束条件:

$$\begin{aligned} \{\theta^{(t+1)}, S^{(t+1)}\} = \underset{\theta, S}{\operatorname{argmin}} S, \\ a^T y - \frac{1}{2} a^T \sum_{m=1}^M \theta_m K_m a - \frac{1}{2C} a a^T \leq S, \\ \theta \geq 0, \nu(\theta) \leq 1. \end{aligned} \quad (18)$$

算法的停止条件为  $|1 - S^{(t)}/\delta^{(t)}| \leq \xi$ , 其中  $\xi$  是一个非常小的常数.

对于  $l_1$ -范数约束的 MKL 求解问题, LSSVR 的  $l_1$ -范数多核求解问题可以表示为

$$\begin{aligned} \min S; \\ \text{s.t. } S \geq a^T y - \frac{1}{2} a^T \sum_{m=1}^M \theta_m K a - \frac{1}{2C} a^T a, \\ \theta \geq 0, \sum_{m=1}^M \theta_m = 1, \\ \text{for all } a \text{ where } a^T \mathbf{1}_M = 0. \end{aligned} \quad (19)$$

式(19)变换成外循环后可以利用 SILP 方法<sup>[5]</sup>求解; 如果同时更新  $a^{(t)}$  和  $\theta^{(t)}$ , 则可以使用 QCQP 方法<sup>[4]</sup>进行求解. 从  $l_1$ -范数的验证来看, 上述表示对  $l_1$ -范数是可行的.

**算法2** 优化方法求解  $l_p$ -范数约束的 MKL-LSSVR.

输入:  $K_1, K_2, \dots, K_M, \theta_m^{(1)} = M^{-1/p}, \delta^{(1)} = -\infty$ .

输出: 线性组合核函数的权值  $\theta = (\theta_1^{(t)}, \dots, \theta_M^{(t)})^T$ .

for  $t = 1, 2, \dots$ , do

Step 1: 根据标准 LSSVR 表达式求解  $a^{(t)}$ .

Step 2: 计算

$$S^{(t)} = a^T y - \frac{1}{2} a^T \sum_{m=1}^M \theta_m K(x_i, x_j) a - \frac{1}{2C} a^T a,$$

Step 3: 如果满足  $|1 - S^{(t)}/\delta^{(t)}| \leq \xi$ , 则停止循环, 否则  $\delta^{(t)} = S^{(t)}$ .

Step 4: 根据式(18), 多核求解问题转换成一个二次约束的线性规划(QCLP)问题, 使用 MOSEK 优化工具箱更新  $\{\theta^{(t+1)}, S^{(t+1)}\}$ , 其中  $\nu(\theta)$  定义为式(15),  $\tilde{\theta} = \theta^{(t)}$ .

Step 5: end for.

## 4 仿真实验

### 4.1 核标准化

由于不同核函数核矩阵的大小差异性较大, 为了减少核矩阵对组合权值的影响, 需要将核矩阵进行标准化.

本文利用球形标准化将核矩阵标准化为

$$K(x, \bar{x}) \leftrightarrow \frac{K(x, \bar{x})}{\sqrt{K(x, x)K(\bar{x}, \bar{x})}}. \quad (20)$$

通过上述操作, 对于任何输入数据  $x$ , 均有  $\|x\| = K(x, x) = 1$ , 即所有输入数据都被重新变换在一个单位的球体内, 球形标准化的方法消除了输入数据的量纲信息.

本文选用的核函数为 Gaussian Kernel, 即

$$\begin{aligned} K_{\text{GAU}}(X_i, X_j) = \exp(-\|X_i - X_j\|_2^2 / \delta^2), \\ \delta \in R^+. \end{aligned} \quad (21)$$

### 4.2 评价标准

利用均方根误差 RMSE 和最大绝对值误差 MAE 评估模型, 有

$$\text{RMSE} = \sqrt{\sum_{i=1}^N (y_i - f_i(x)) / N}, \quad (22)$$

$$\text{MAE} = \max(|y_i - f_i(x)|), \quad i = 1, 2, \dots, N. \quad (23)$$

其中:  $y_i$  为实测数据,  $f_i(x)$  为最小二乘支持向量回归算法预测结果,  $N$  为预测样本个数.

针对参数选择, 选用不同宽度的 RBF 核函数, 核函数宽度<sup>[11]</sup>范围为  $[2^{-2}, 2^{-1.5}, \dots, 2^{1.5}, 2^2]$ , 间隔为 0.5,  $C$  取值为  $[2^{-1}, \dots, 2^{10}]$ ,  $P$  取值为  $[1, 4/3, 2, 4, \infty]$  (与 Klotz<sup>[6]</sup>保持一致), 探讨不同  $P$  值对 LSSVR 算法泛化性的影响.

### 4.3 仿真结果分析

为了验证本文提出的  $l_p$ -范数 MKL-LSSVR 算法的有效性, 考虑算法 1 和算法 2 两者的算法目标是一致的, 只是求解方式不一样, 此处利用算法 1 对单变量函数和多变量函数进行数值仿真.

1) 单变量函数是简单的混合高斯分布函数<sup>[12]</sup>, 也是典型的非平坦函数, 有

$$\begin{aligned} r(x) = \\ \frac{\exp\left(-\frac{(x-2)^2}{2 \times 0.3^2}\right)}{\sqrt{2\pi} \times 0.3} + \frac{\exp\left(-\frac{(x-7)^2}{2 \times 1.2^2}\right)}{\sqrt{2\pi} \times 1.2}. \end{aligned} \quad (24)$$

加入噪声  $y_i = r(x_i) + n_i$ , 输入数据  $x_i \in [0, 10]$ , 均匀采样生成 200 个样本, 输出加入高斯噪声  $n_i \in N(0, 0.05^2)$ , 作为训练集, 然后随机在  $x_i \in [0, 10]$  生成 200 组测试样本集.

2) 多变量定义<sup>[13]</sup>为

$$y(t+1) = \frac{y(t)y(t-1)(y(t)+2.5)}{1+y^2(t)+y^2(t-1)} + \mu(t) + e(t+1). \quad (25)$$

非线性系统输入为

$$\mu(t) = \sin(2\pi t/25) + \sin(2\pi t/10),$$

均匀分布在区间  $[-2, 2]$  内, 构造 200 组数据, 其中前 100 组数据为训练集, 后 100 组数据为测试集.  $e(t)$  是高斯白噪声序列, 定义为  $e(t) \in N(0, 0.1)$ . 非线性系统可以表示为

$$y(t) = f(x(t)) + e(t),$$

$$x(t) = [y(t-1), y(t-2), \mu(t-1)].$$

单变量函数和多变量函数均独立实验 10 次, 实验结果取每次实验的平均值. 对于单变量函数, LSSVR 在参数  $C = 2^{10}, \delta = 0.4$  时最优, 测试结果对比如表 1 所示. 由表 1 可见, 多核学习的效果优于传统 LSSVR 方法, 当  $P = \text{Inf}$  时泛化性能最优.

表 1 单变量函数在不同  $P$  范数的多核学习 LSSVR 与传统 LSSVR 的测试结果对比

	LSSVR	1-norm	4/3-norm	2-norm	4-norm	Inf-norm
RMSE	0.0164	0.0087	0.0069	0.0054	0.0043	0.0035
MAXE	0.0663	0.0791	0.0598	0.0444	0.0314	0.0203

对于多变量函数, LSSVR 在参数  $C = 2^{10}, \delta = 0.1$  时最优, 对比结果如表 2 所示. 由表 2 可见,  $l_p$ -范数的多核学习效果远优于传统的 LSSVR 算法, 且不同范数的效果不同, 在  $P = 4/3$  时效果最好, 当  $P$  增加时, 泛化性能变差. 误差随  $P$  值的不同有波动, 合适的  $P$  值和不同的  $C$  取值对误差变化均会产生影响.

表 2 多变量函数在不同  $P$  范数的多核学习 LSSVR 与传统 LSSVR 的测试结果对比

	LSSVR	1-norm	4/3-norm	2-norm	4-norm	Inf-norm
RMSE	0.0190	0.0134	0.0131	0.0132	0.0134	0.0137
MAXE	0.0739	0.0344	0.0340	0.0350	0.0349	0.0354

本文共有 9 个不同宽度的核函数, 对于不同的范数, 多核学习的核函数权值各不相同. 图 1 为单变量函数在不同范数下的核函数的权值分布图, 图 2 为多变量函数的权值分布图.

对应不同的范数, 核函数的权值是不一样的. 1-范数约束的多核学习具有稀疏解, Inf-范数所有的权值均为 1, 即针对所有的核函数都是无权值的直接叠加. 当  $P$  值越接近 1 时, 多核学习的稀疏性越好; 当数据分布比较平坦时,  $P$  的取值趋近于 1; 当数据分布越复杂时,  $P$  的取值越大.

针对不同的  $C$  值和不同的  $P$  值, 算法的泛化性能有所变化, 图 3 和图 4 分别为单变量函数和多变量

函数在不同  $C$  值和不同  $P$  值下测试集的均方根误差变化图. 由图 3 和图 4 可见, 多核学习的最小二乘支持向量机泛化性能优于单一核学习的 LSSVR 情况, 与 LSSVR 相比较,  $l_p$ -范数的多核学习 LSSVR 对于惩罚参数  $C$  的取值具有较强的鲁棒性.

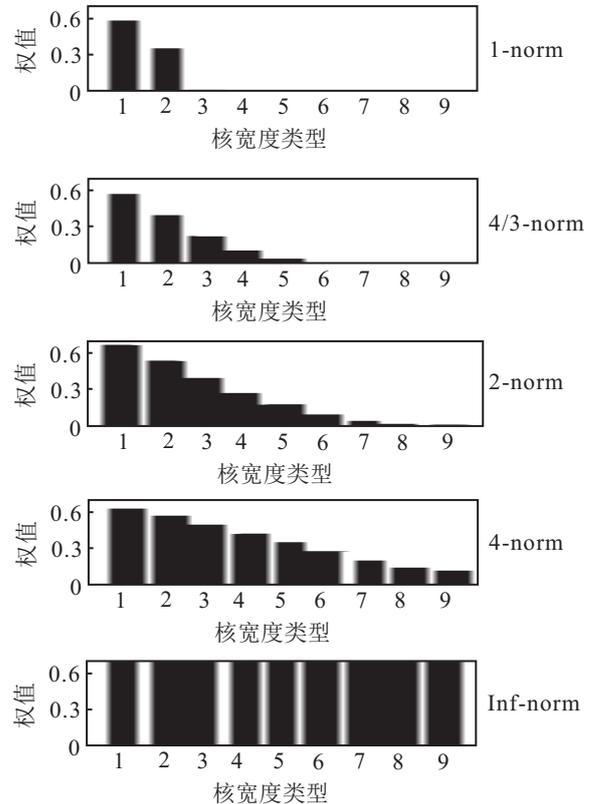


图 1 单变量函数在不同的范数情况下核函数的权值

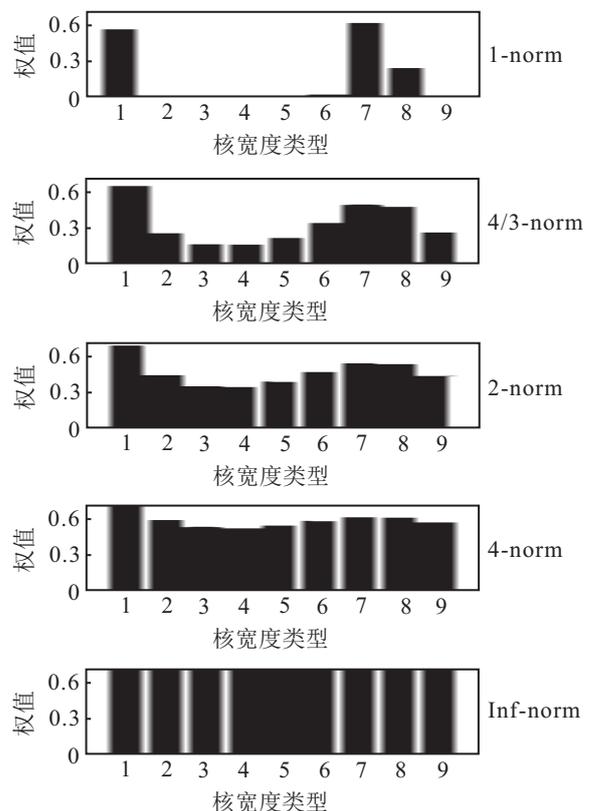


图 2 多变量函数在不同的范数情况下核函数的权值

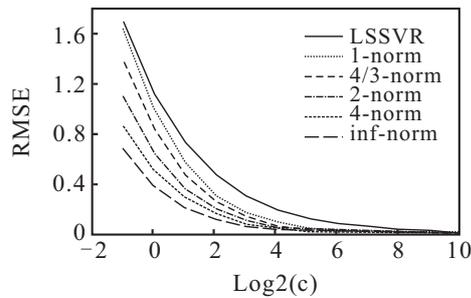


图3 单变量函数在不同 $C$ 值和不同 $P$ 值下的均方根误差

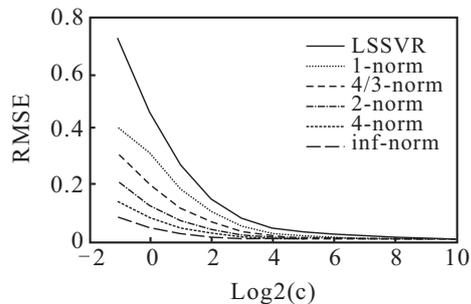


图4 多变量函数在不同 $C$ 值和不同 $P$ 值下的均方根误差

## 5 结论

本文针对最小二乘支持向量机的核函数构造问题,提出了一种基于 $l_p$ -范数的最小二乘支持向量机多核学习,并提出了两种求解算法.通过在单变量函数和多变量函数上的实验表明, $l_p$ -范数的多核学习具有鲁棒性和通用性. $l_p$ -范数的多核函数LSSVR相比于单核LSSVR具有更好的泛化性能和可解释性,且惩罚参数 $C$ 的取值鲁棒性更强.本文的研究工作主要集中在 $l_p$ -范数多核学习的核函数权值研究,如何选择合适的 $P$ 值、确定核函数的个数和核函数宽度的取值范围是下一步的研究方向.

## 参考文献(References)

- [1] 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.  
(Wang H Q, Sun F C, Cai Y N, et al. On multiple kernel learning methods[J]. Acta Automatica Sinica, 2010, 36(8): 1037-1050.)
- [2] Cristianino N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods[M]. Cambridge: Cambridge University Press, 2000: 45-78.
- [3] Lanckriet G, Cristianini N, Bartlett P, et al. Learning the kernel matrix with semi-definite programming[J]. The J of Machine Learning Research, 2004, 5(1): 27-72.
- [4] Jian L, Xia Z, Liang X, et al. Design of a multiple kernel learning algorithm for LS-SVM by convex programming[J]. Neural Networks, 2011, 24(5): 476-483.
- [5] Chen X, Guo N, Ma Y, et al. More efficient sparse multi-kernel based least square support vector machine[C]. Communications and Information Processing. Berlin: Springer-Heidelberg, 2012: 70-78.
- [6] Kloft M, Brefeld U, Sonnenburg S, et al. Lp-norm multiple kernel learning[J]. The J of Machine Learning Research, 2011, 12(3): 953-997.
- [7] Yan F, Kittler J, Mikolajczyk K, et al. Non-sparse multiple kernel fisher discriminant analysis[J]. The J of Machine Learning Research, 2012, 13(3): 607-642.
- [8] Sun T, Jiao L, Liu F, et al. Selective multiple kernel learning for classification with ensemble strategy[J]. Pattern Recognition, 2013, 46(11): 3081-3090.
- [9] Ivanov V K, Vasin V V, Tanana V P. Theory of linear ill-posed problems and its applications[M]. Walter de Gruyter, 2002: 124-199.
- [10] Tikhonov A N, Arsenin V I A, John F. Solutions of ill-posed problems[M]. Washington DC: Winston, 1977: 191-222.
- [11] Zhang X, Hu L, Zhang L. An efficient multiple kernel computation method for regression analysis of economic data[J]. Neurocomputing, 2013, 118(10): 58-64.
- [12] Ceperic V, Gielen G, Baric A. Sparse multikernel support vector regression machines trained by active learning[J]. Expert Systems with Applications, 2012, 39(12): 11029-11035.
- [13] Cai Y, Wang H, Ye X, et al. A multiple-kernel LSSVR method for separable nonlinear system identification[J]. J of Control Theory and Applications, 2013, 11(4): 651-655.

(责任编辑:郑晓蕾)