

文章编号:1672-058X(2014)11-0056-06

# 一种改进的局部多核学习算法

丁 跃

(重庆大学 数学与统计学院,重庆 401331)

**摘 要:**局部多核学习算法(LMKL)是一种变系数的多核支持向量机算法,其利用选通函数局部的选取合适的合成核函数;但是其选通函数有严重的参数冗余的问题,为此提出了改进的局部多核学习算法(ILMKL),在其目标函数中加入正则项,区别于 LMKL 中选通函数的  $l_1$  范数形式,使用选通函数的  $l_p$  范数形式,增强核函数间的“互补”作用;采用该算法在模拟数据集和 UCI 数据集上实验,结果表明该算法取得较高的分类能力。

**关键词:**支持向量机;局部学习;多核学习;正则化

**中图分类号:**TP181

**文献标志码:**A

## 0 引 言

支持向量机<sup>[1]</sup>(Support Vector Machine)是建立在核函数基础上的机器学习算法,在模式识别、机器学习等领域有着广泛的应用.不同的核函数,同一核函数不同核参数,对于支持向量机泛化能力的影响是显著的<sup>[2]</sup>,在一些复杂的问题中,特别是数据具有异构性,样本分布不平坦,单一核函数已经难以满足问题的需要,为此 Lanckriet 提出来多核学习算法.多核学习算法通过将这些核函数组合起来学习,得到更高的泛化能力。

关于多核学习算法的主要研究成果有 Lanckriet 等<sup>[3]</sup>首先使用了半正定规划技术,完成了对于核矩阵的学习问题,但是其在大规模数据和高维数据的学习上面,学习效率较低 Sonnenburg 等<sup>[6]</sup>在多核函数锥组合的基础上,提出了一种通用而更有效的多核学习算法.该方法将多核学习问题改写为一种半无限线性规划(Semi-infinite linear program, SILP)形式,新的规划形式可以在标准的 SVM 应用问题中,利用成熟的线性规划方法进行求解. Rakotomamonjy 等提出的简单多核学习(SimpleMKL)<sup>[7]</sup>,引入新的混合范数将候选核函数的权系数包含在标准支持向量机的经验风险最小化中,目标函数转化为凸的、且光滑函数,在权系数的确定中,使用梯度下降算法,SimpleMKL,相比于其他算法获得更高的效率。

以上的多核学习算法对于所有输入样本,核函数的权系数是一样的,这无形中对样本进行了一种平均处理,局部多核学习算法(LMKL)利用选通函数,局部的选用了合适的核函数,但是,LMKL 有以下的不足:LMKL 算法所选用的选通函数有严重的参数冗余,其使用选通函数  $l_1$  范数的形式,提高了核函数间的稀疏性,但是削弱了核函数间的“互补”作用.为此,本文提出了改进的局部多核学习算法(ILMKL),针对选通函数的参数冗余的问题,在其目标函数中加入正则项,并且使用选通函数的  $l_p$  范数形式。

收稿日期:2014-02-14;修回日期:2014-04-26.

作者简介:丁跃(1988-),男,安徽六安人,硕士,从事模式识别、机器学习算法研究.

## 1 局部多核学习算法

假定训练集为  $\{(x_i, x_i)\}_{i=1 \dots N}, x_i \in X \subset \mathbf{R}^n, N$  为训练集的个数,  $y \in \{-1, 1\}, K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$  为核函数, 针对平稳组合的合成核的权系数是常数, 无法体现局部信息, Gönen and Alpaydin 提出的局部多核学习算法 (LMKL), 其合成核的构造:

$$K_\beta(x_i, x) = \sum_{m=1}^M \beta_m(x_i) \beta_m(x) K_m(x_i, x) \quad (1)$$

其核函数  $K_m$  的权系数不再是常数, 而是依赖于输入空间的样本数据, 此时对应的判别函数为

$$f(x) = \sum_{m=1}^M \beta_m(x) \langle w_m, \varphi_m(x) \rangle + b \quad (2)$$

若是直接求解  $\beta_m(x_i) \beta_m(x)$ , 需要求解  $N \times M$  个参数, 并且目标函数不是凸函数, 为此, Gönen and Alpaydin 使用选通函数局部的选取合成核函数. 选通函数为

$$\beta_m(x) = \frac{\exp(\langle v_m, x \rangle + v_{m0})}{\sum_j \exp(\langle v_j, x \rangle + v_{j0})} \quad (3)$$

$v_m, v_{m0}$  为选通函数的参数, 选通函数建立了输入样本空间到核函数的权系数的函数关系, 为输入空间中的每一个样本局部的选取核函数, 其分母使用  $l_1$  范数的形式, 提高核函数的稀疏性。假设  $v = \begin{pmatrix} v_1, \dots, v_m \\ v_{10}, \dots, v_{m0} \end{pmatrix}$  与单核支持向量机类似, LMKL 算法对应的优化形式可以表示下面的优化问题:

$$\begin{aligned} \max_{\alpha, v} & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K_\beta(x_i, x_j) \\ \text{s.t.} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & C \geq \alpha_i \geq 0 \quad \forall i = 1, 2, \dots, n \end{aligned} \quad (4)$$

## 2 改进的局部多核学习算法 (ILMKL)

在式(1)的合成核构造中, 定义新的选通函数  $\eta_m(x)$  为

$$\eta_m(x) = \frac{\exp(v_m x + v_{m0})}{\left( \sum_j \exp(v_j x + v_{j0}) \right)^{\frac{1}{p}}} \quad (5)$$

其中  $v_m \geq 0$  为参数, 这里选用的是选通函数的  $l_p$  范数形式, 有助于增强各个核函数间的“互补”作用, 提高识别率。但是, 不论是式(3)还是式(5)中的选通函数都有参数“冗余”, 可以看出, 对于优化问题式(4), 若  $v$  是上述问题的解, 那么对于任意的常数  $v_0, v+v_0$  还是上述问题的解. 所以它们都有参数“冗余”, 为此, ILMKL 算法在目标函数中加入了正则项  $\|v\|_F$ , 这里使用的是 Frobenius 范数. 根据结构风险最小化原则. 其学习问题可以表示为

$$\begin{aligned} \min_{\{w_m\}, b, \xi, \{\beta(x)\}} & \frac{1}{2} \sum_{m=1}^M \|w_m\|_2^2 + C \|\xi\|_1 + u \|v\|_F \\ \text{s.t.} & y_i \left( \sum_{m=1}^M \eta_m(x) \langle w_m, \varphi_m(x) \rangle + b \right) \geq 1 - \xi_i \quad \forall i = 1, \dots, N \\ & \xi \geq 0 \end{aligned} \quad (6)$$

为继承传统多核学习中不断迭代标准支持向量机代码的优势, 将上述优化问题表示为

$$\min_{v \geq 0} T(v) = J(v) + u \|v\|_F \quad (7)$$

其中

$$J(v) = \begin{cases} \min_{w_m, b, \xi} & \frac{1}{2} \sum_{m=1}^M \|w_m\|_2^2 + C \|\xi\|_1 \\ \text{s.t.} & y_i \left( \sum_{m=1}^M \eta_m(x) \langle w_m, \varphi_m(x) \rangle + b \right) \geq 1 - \xi_i \quad \forall i = 1, \dots, N \\ & \xi \geq 0 \end{cases}$$

这里参数  $u$  是用来平衡参数“冗余”和  $J(v)$  的值,对于固定参数  $v$ ,将  $J(v)$  转化为其拉格朗日对偶形式:

$$W(v) = \begin{cases} \max_{\alpha} & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K_{\eta}(x_i, x_j) \\ \text{s.t.} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & C \geq \alpha_i \geq 0 \quad \forall i = 1, 2, \dots, N \end{cases} \quad (8)$$

其中  $K_{\eta}(x_i, x) = \sum_{m=1}^M \eta_m(x_i) \eta_m(x) K_m(x_i, x)$ , 式(8)是一个标准的支持向量机问题,可以用现有的支持向量机算法求解,本文使用的是 smo 算法<sup>[13]</sup>.将  $W(v)$  代入到问题式(7)中,问题式(7)化为

$$\min_{v \geq 0} T(v) = W(v) + u \|v\|_F \quad (9)$$

假设  $\alpha^* = (\alpha_1^*, \dots, \alpha_N^*)$  是问题式(8)的最优解,那么向量  $w_m = \sum_{i=1}^N \alpha_i^* y_i \beta_m(x) \varphi_m(x_i)$ , 此时判别函数为

$$f(x) = \sum_{i=1}^N \sum_{m=1}^M \alpha_i y_i \eta_m(x) K_m(x, x_i) \eta_m(x_i) + b \quad (10)$$

在给定参数  $v$  下,由于  $W(v)$  的解  $\alpha^*$  存在并且是唯一的<sup>[1]</sup>,于是  $W(v)$  对参数  $v$  的梯度向量存在,且等于如下形式:

$$\begin{aligned} \frac{\partial W(v)}{\partial v_m} &= -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_{k=1}^M \eta_k(x_i) K_k(x_i, x_j) \eta_k(x_j) [x_i (\sigma_m^k - \eta_m(x_i)^p) + x_j (\sigma_m^k - \eta_m(x_j)^p)] \\ \frac{\partial W(v)}{\partial v_{m0}} &= -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_{k=1}^M \eta_k(x_i) K_k(x_i, x_j) \eta_k(x_j) [(\sigma_m^k - \eta_m(x_i)^p) + (\sigma_m^k - \eta_m(x_j)^p)] \end{aligned}$$

这里  $\sigma_m^k = 1$ , 当且仅当  $m=k$ 。进一步可以得到梯度向量  $\frac{\partial T(v)}{\partial v_m}$ , 本文采用梯度下降算法求解问题式(9)。

每求解一次目标函数值,需要求解一次标准支持向量机问题,通过不断迭代,直至收敛,解得最优参数  $v^*$ 。

## 2.1 ILMKL 算法描述

ILMKL 算法以  $T(v)$  前后两次迭代的值小于设定的值  $\varepsilon$  或者迭代次数达到最大值为停机准则,算法具体描述如下:

1) 设定候选核函数  $K_1, K_2, \dots, K_M$  和参数  $u$  和  $p$ , 参数  $v$  初始值  $v^0$ , 最大迭代次数  $G$ , 迭代停止条件  $\varepsilon$ , 迭代次数  $k=0$ ; 利用  $v^0$  计算选通函数  $\eta_m(x) m=1, 2, \dots, M$ , 进而得到初始合成核  $K^0$ ;

2) 求解以  $K^k$  为核函数的标准支持向量机, 计算目标函数值  $T^k(v)$ , 梯度向量:

$$D^k(v) = \begin{pmatrix} \frac{\partial T}{\partial v_1}, \frac{\partial T}{\partial v_2}, \dots, \frac{\partial T}{\partial v_M} \\ \frac{\partial T}{\partial v_{10}}, \frac{\partial T}{\partial v_{20}}, \dots, \frac{\partial T}{\partial v_{M0}} \end{pmatrix}^T$$

3) 线性搜索最优步长  $\mu^k \in [0, +\infty]$ ;

4) 迭代  $v^{k+1} = v^k - \mu^k * D^k(v)$ ;

5) 计算  $T^{k+1}(v)$ , 若  $|T^{k+1}(v) - T^k(v)| \leq \varepsilon$ , 或者  $k=G$ , 则算法停止, 否则,  $k=k+1$ , 转到步骤 2)。

## 2.2 ILMKL 算法复杂度分析

ILMKL 算法的时间复杂度主要来源于以下几个方面: 所使用标准支持向量机的复杂度, 算法的迭代次数. 本文使用的是 SMO 算法<sup>[13]</sup>, 其算法复杂度为在  $O(n^{1.2})$  与  $O(n^{2.2})$  之间, 而其迭代次数受样本数据和

迭代步长的选择的影响,假设迭代次数为  $k$ ,所以算法的复杂度介于  $O(kn^{1.2})$  与  $O(kn^{2.2})$  之间。

### 3 实验分析

#### 3.1 模拟数据实验

为验证 ILMKL 算法的有效性,本文模拟 4 个二元正态分布,从每个正态分布中产生 100 个数据,4 个二元正态分布的均值分别为  $(-3,1), (-1,-2.5), (1,1), (3,-2.5)$ ,协方差矩阵分别为  $(0.8,0,0,2), (0.8,0,0,4), (0.8,0,0,2), (0.8,0,0,4)$ ,将第 1 和第 3 个二元正态分布作为一类,第 2 和第 4 个二元正态分布作为另一类.这里使用典型的传统多核学习算法 SimpleMKL、局部多核学习算法 LMSVM、和本文提出的 ILMKL 算法,分别对两类数据进行分类学习。选用 3 个线性核  $(K_L-K_L-K_L)$ ,ILMKL 算法选取参数  $u=(0,0.1,1)$ ,选取  $p=(1,2,3)$ ,由于知道数据的真实分布,所以贝叶斯分类器是最优得分类器<sup>[14]</sup>。分别绘制了个算法的分类界面与贝叶斯分类器比较。

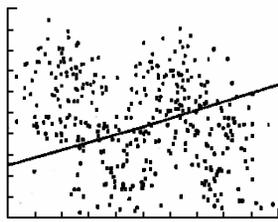


图 1 简单多核学习

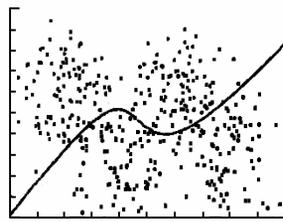


图 2 局部多核学习

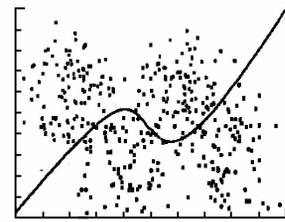


图 3 改进的局部多核学习

图 1、图 2、图 3 中,虚线表示贝叶斯分别类界面,实线表示各个算法的分类界面,黑点表示支持向量。将图 2 和图 3 与图 1 比较,明显地显示了局部多核学习的优势。由于选用的核函数是 3 个线性核,传统多核学习算法 SimpleMKL 做出的类界面只是单纯的直线,局部多核学习可以通过选通函数局部的结合 3 个线性核,达到非线性分类的能力.比较图 3 和图 2,ILMKL 分类界面与虚线拟合的更好,从储存支持向量的个数角度,ILMKL 算法和 LMKL 算法的支持向量的个数最少。

表 1 实验所采用的数据集

数据集	样本数	特征数	类别数
heart	270	13	2
Ionosphere	351	34	2
Liver	245	6	2
Sonar	208	60	2
Spambase	4601	57	2
Wbdc	198	33	2
Iris	150	13	2
Msplice	3175	240	2
Glass	214	9	2

#### 3.2 UCI 数据实验

将 ILMKL 算法、单核 SVM<sup>[1]</sup>、LMKL 算法<sup>[4]</sup>、SimpleMKL<sup>[7]</sup> 算法应用到 UCI 数据库上,验证 ILMKL 算法的有效性,实验选取 9 个 UCI 数据集,数据集的信息如表 1 所示。

实验所选用的核函数是线性核函数  $K_L$ ,二项式核函数  $K_p(x,x')=(1+x^T x')^2$ ,高斯核函数  $K_G$ ,其中单核 SVM 分别使用上面的 3 个核函数计算,高斯核函数的参数  $\sigma$  采用以下方式估计:计算每一个样本与其他样本的最小欧式距离,将所有这些欧氏距离的平均作为核参数  $\sigma$ 。采用交叉验证确定单核 SVM、ILMKL、LMKL、SimpleMKL 算法的惩罚系数  $C$  值,其中  $C \in \{0.1,1,10\}$ ,选取 ILMKL 中参数  $\mu \in \{0.01,0.1,1,10\}$ , $p \in \{1,2,3,4,5,6,7,8\}$ ,表 2 记录了 10 重交叉验证的平均准确率,和多核支持向量机存储的平均支持向量数,所有实验结果,都是重复 10 次的平均结果。

表 2 实验结果

	$K_L K_p K_G$	SVM		SimpleMKL		LMKL		ILMKL( $p=1$ )		IMKL( $p>1$ )	
		Acc	SV	Acc	SV	Acc	SV	Acc	SV	Acc	SV
Heart	84.07	82.59	84.81	82.96	48.40	85.93	48.60	85.93	33.70	86.34	31.50
Ionosphere	87.44	94.00	88.86	94.29	54.57	88.90	49.73	95.14	44.55	95.14	43.75
Liver	70.15	71.20	57.98	73.03	71.63	71.45	73.27	73.32	69.92	75.65	64.38
Sonar	77.90	83.00	53.38	83.60	80.93	78.83	72.60	87.52	66.83	85.52	63.84
Spambase	93.31	93.44	93.26	94.02	22.61	93.57	24.94	93.44	26.03	94.28	19.02
Wbdc	76.26	76.79	76.29	76.29	97.70	76.32	77.22	76.32	72.22	80.32	79.85
Iris	100	99.33	88.00	99.33	9.33	99.33	18.44	99.33	8.59	100	8.44
Msplice	94.36	92.19	66.96	96.37	41.57	94.38	46.68	95.65	37.11	98.44	46.51
Glass	80.80	82.21	79.42	82.28	83.38	84.03	70.25	85.45	67.24	84.07	61.63

从分类精度的角度,LMKL,SimpleMKL 算法基本相同,ILMKL 算法分类精度最高,其中 ILMKL( $p>1$ )要比 ILMKL( $p=1$ )分类精度更高,这是因为核函数间不同的核映射有“互补”的作用,而使用 1 范数( $p=1$ )会提高核函数间的稀疏性,这样会削弱核函数间的“互补”作用,进而影响分类精度。与单核 SVM 比较,ILMKL 算法在这些数据集上都高于精度最高的单核 SVM,SimpleMKL 算法在除 Heart,Iris 数据集以外,其他数据集上都要高于精度最高的单核 SVM,LMKL 算法在除 Iris Ionosphere 数据集以外,其他数据集上都要高于精度最高的单核 SVM,这点也表明多核学习算法有更好的分类精度。从存储的支持向量上来看,ILMKL 算法存储的支持向量最少。

## 4 结 论

本文提出改进的局部多核学习算法,针对 LMKL 算法中选通函数的参数冗余的问题,在其目标函数中加入正则项,并且使用选通函数的  $p$  范数形式。对比实验结果表明:ILMKL 有更高的泛化能力和存储更少支持向量。未来的工作,将进一步提升该算法的泛化才能和效率,并将考虑将该方法推广到多分类或回归等其他领域。

### 参考文献:

- [1] VAPNIK V. The Nature of Statistical Learning Theory[M]. NewYork:Springer-Verlag, 1995
- [2] LANCKRIET C, VAPNIK V,BOUSQUET O, Choosing multipleparameters for support vector machines[J]. Machine Learning, 2002,46(1): 131-159
- [3] LANCKRIET G, CRISTIANINI N, BARTKETT P, GHAOUI L,JORDAN M. Learning the kernel matrix with semi-definite programming[J].Machine LearningResearch,2004,5(1): 27-72
- [4] Gönen M, ALPAYDIN E.Localized multiple kernel learning[C]. In: Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland: ACM, 2008
- [5] BACH F, Lanckriet G G, Jordan M I. Multiple kernel learning, conic duality, and the SMO algorithm[C]. In: Proceedings of the 21st International Conference on Machine Learning Ban, Canada: ACM, 2004. 41-48
- [6] SONNENBURG S, RATSCH G, SCHAFER C. A general and efficient multiple kernel learning algorithm[J]. Advances in Neural Information Processing Systems. 2006,18(1):1 273-1 280
- [7] RAKOTOMAMONJY A, BACH F, CANU S, GRANDVALET Y.Simple MKL[J].Machine Learning Research, 2008, 9(11): 2491-2521
- [8] 汪洪桥,孙富春. 多核学习方法[J]. 自动化学报,2010,26(8):1037-1050

## A Kind of Improved Localized Multiple Kernel Learning Algorithm

**DING Yue**

(School of Mathematics and Statistics, Chongqing University, Chongqing 401331, China)

**Abstract:** Localized Multiple Kernel Learning (LMKL) Algorithm is a kind of multiple kernel support vector machine algorithm with varying coefficient and uses gating function to locally select suitable compound kernel function, however, its gating function has serious parameter redundancy, therefore, this paper proposes improved localized multiple kernel learning (ILMKL) algorithm, adds regularization term to objective function to discriminate norm form of the gating function in LMKL, uses lp norm form of gating function to enhance “complementary” role between kernel functions, uses this algorithm to conduct experiments on simulation data set and UCI data set and draws the conclusion that this algorithm gets higher classification capacity.

**Key words:** support vector machine; locally learning; multiple kernel learning; regularization

责任编辑:代小红

(上接第 49 页)

- [5] WU Jing, CHEN Tongwen. Design of Networked Control Systems With Packet Dropouts [J]. Automatic Control, IEEE Transactions on, 2007, 52(7): 1314-1319
- [6] WANG Xiaofeng, LEMMON M. Event-Triggering in Distributed Networked Control Systems [J]. Automatic Control, IEEE Transactions on, 2011, 56(3): 586-601
- [7] 徐雷鸣, 庞博, 赵耀. NS 与网络模拟 [M]. 北京: 人民邮电出版社, 2003
- [8] 于斌, 孙斌. NS2 与网络模拟 [M]. 北京: 人民邮电出版社, 2007
- [9] 柯志亨, 程荣祥, 邓德隽. NS2 仿真实验——多媒体和无线网络通信 [M]. 北京: 电子工业出版社, 2009
- [10] 高文宇, 王建新, 陈松乔. 网络仿真软件 NS2 中队列调度算法的扩展 [J]. 系统仿真学报, 2006(2): 521-525
- [11] 杨吉文, 张卫东. 基于 NS2 的主动队列管理仿真研究 [J]. 计算机工程, 2006(17): 189-191
- [12] 高鹏, 戴旭初. 基于 NS 的主动队列管理算法的仿真与分析 [J]. 计算机仿真, 2006(5): 98-100, 155

## Design of Preferential Queuing Management Algorithm Based on NS2

**PAN Feng-min, JIANG Ming, ZHOU Jia-wen, GE Yuan**

(Anhui Polytechnic University, Anhui Wuhu 241000, China)

**Abstract:** The signals of network control system are transmitted in internet by the grouping form, which causes the problems such as time delay, packet dropout and so on in the transmission, as a result, in order to study the performance of the network system, this paper uses NS2 to set up network transmission model of network control system, and designs preferential queuing management algorithm by taking its preferential business grouping as research object. This algorithm is improved on the basis of DropTail algorithm, inherits the advantages of original algorithm, and contains business preference authentication and packet dropout selection mechanism. The validity of PDropTail algorithm is tested by using the two algorithms to do experiment for the analysis and comparison from the perspective of time delay, packet dropout and throughput.

**Key words:** network control system; preferential level; queuing management algorithm; time delay; packet dropout; throughput

责任编辑:代小红