

基于局部空间变稀疏约束的多核学习方法

王庆超,付光远,汪洪桥,辜弘扬,王超

(火箭军工程大学信息工程系,陕西西安 710025)

摘要: 局部多核学习方法根据样本所在局部空间特性选择特定的核函数组合方式,具有较好的判别能力. 本文提出了一种基于局部空间变稀疏约束的多核学习方法,首先依据样本在特征空间的分布情况以软分组的方式将训练数据划分为若干数据子集. 以数据子集为单位,根据在相应的局部空间内的核函数相似程度,调整核组合的稀疏程度,使用交替优化的方法进行求解. 实验表明本文方法对于区分特征学习和对抗噪声方面具有的优势,因此也使得在图像场景分类问题上的准确率和稳定性得到明显提高.

关键词: 多核学习; 支持向量机; 局部学习; 变稀疏约束

中图分类号: TP181 **文献标识码:** A **文章编号:** 0372-2112 (2018)04-0930-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.04.022

Local Variable Sparsity Based Multiple Kernel Learning Algorithm

WANG Qing-chao, FU Guang-yuan, WANG Hong-qiao, GU Hong-yang, WANG Chao

(The Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China)

Abstract: Local multiple kernel learning method could learn a specific combination kernel function for various samples according to the local space characteristics, therefore it has better discriminant ability. In this paper, we propose a local variable sparsity based multiple kernel learning method. In our method, the samples are divided into a few groups with a soft grouping method and the sparsity of kernel weights in various local spaces is determined by the similarity of kernels. We use an alternative optimization method to solve this problem. The experiment on synthetic dataset indicates that our method has a strong advantage in discriminative feature learning and against noise. Finally we apply our method into image scene classification and the accuracy is improved obviously.

Key words: multiple kernel learning; support vector machine; local learning; variable sparsity constraint

1 引言

核方法通过将原始数据映射到高维空间,从而将线性不可分问题在高维空间内变得线性可分,同时避免了在高维空间上进行问题求解. 以支持向量机(Support Vector Machine, SVM)^[1]为代表的核方法在解决非线性模式分析问题上取得了很大成功. 然而基于单个特征空间的单核方法存在一定局限性. 首先,核函数的选择及核函数参数的确定对分类性能有显著影响,且至今没有完善的理论依据. 其次,当样本包含异构特征时,使用单个核函数描述并不合理. 多核方法(Multiple Kernel Learning, MKL)^[2]通过自动学习的方式选择多个基核的最佳组合,从而获得更好的分类性能. 多核方法已经广泛应用于分类^[3]、聚类^[4]、降维^[5]、回归^[6]等问

题. 在图像分类^[7]、姿态估计^[8]、视觉跟踪^[9]、人脸识别^[10]等方面有着优异的表现. 文献[2~11]综述了多核学习方法的研究进展. 近年来深度学习(Deep Learning, DL)^[12]的方法在诸多方面取得了巨大成功,然而其优势是建立在对大规模数据反复训练的基础上的,对于训练样本较少,计算实时性较高的问题,核方法相对于DL依然具有不可取代的优势.

传统多核方法通常在全部样本空间内采用统一的核组合形式. 然而在实际分类问题中不同样本的各种特征重要程度并不相同,因此部分研究者开始尝试核函数局部组合方式^[13,14]. 然而,为不同样本选择不同的核函数组合大大增加了模型的复杂程度,且难以求解. Gonen^[15]使用选通模型将局部核权值的求解转化为一个凸问题,通过梯度下降的方式求解,为局部多核学习

方法开辟了一个新思路.此后,更多关于核函数的局部组合方法不断涌现^[16-24].

然而样本级局部多核学习方法通常存在计算复杂度高且容易因过分强调单个样本特性而导致分类器不可靠. Yang^[17]提出了一种基于分组的局部多核方法,先对数据进行聚类操作,通过组内数据共享权值的方式,减少所需优化变量,且不会因个别样本的影响使分类器不可靠,这也使得图像场景分类的准确得到提高.

与此同时, Marius 等人^[25]研究了多核组合的稀疏性对分类器性能的影响. Marius 将 l_1 范数的多核学习方法扩展至任意 l_p 范数,并通过实验证明,针对不同数据集应当选择不同的范数约束,并且最佳的范数值与核函数的相似性有关.当数据集在多个核函数所对应的特征空间中相似程度较高时,意味着核函数间存在冗余,宜采用稀疏约束去冗余,反之则应采用非稀疏约束.

当前局部多核学习方法可以分为样本级方法^[15,21-24]和基于分组的方法^[16-20].样本级方法为每个样本指定不同的核组合方式,而基于分组的方法学习到的组合核作用于相应的样本组.样本级的方法通常计算量大并且没有考虑样本的相关性,而基于分组的方法则丢失了样本的多样性信息.

本文使用基于软分组的方式,通过样本对各个分组的隶属度不同,区分样本的多样性,并且相对与样本级方法优化变量数大大减少.同时本文通过区分局部空间核函数的相似性,调整核函数的稀疏约束项,从而使各个核函数的在各局部空间内优势得到充分发挥.值得注意的是,由于并不了解训练样本更多的先验知识,因此在训练之前需要通过聚类获取样本的分组情况.聚类结果实际上是样本之间相互关系的一种表示,从而达到依数据构建核的目的.

本文的创新点主要包括两个方面:(1)构建了一种新的局部多核组合模型,通过软分组的方式区分样本的多样性,实现了样本级的多核组合方式;(2)使用局部变稀疏的方法,根据局部空间核函数的相似性,调整核函数的稀疏性.

2 相关工作

传统多核学习方法由多个核函数的线性组合构成,形式如下:

$$K(x_i, x_j) = \sum_{m=1}^M \beta_m K_m(x_i, x_j) \quad (1)$$

其中, β_m 为第 m 个核的权重, M 为核函数的总数, 向量 β 需满足 $\beta_m \geq 0$, $\sum \beta_m = 1$ 的条件. 判别函数如式(2)所示:

$$f(x) = \sum_{m=1}^M \beta_m \langle \mathbf{w}_m, \phi_m(x) \rangle + b \quad (2)$$

相对于传统 MKL 方法为所有样本使用统一的核组合, LMKL 依据样本特性选择最优的核组合方式, 形式如下:

$$K(x_i, x_j) = \sum_{m=1}^M \gamma_m^x \gamma_m^{x_j} K_m(x_i, x_j) \quad (3)$$

其中, γ_m^x 对于样本 x_i 第 m 个核函数的权值, LMKL 对应的判别函数变为:

$$f(x) = \sum_{m=1}^M \gamma_m^x \langle \mathbf{w}_m, \phi_m(x) \rangle + b \quad (4)$$

不同于线性组合 MKL, LMKL 方法中的核权值 γ 是一个矩阵, 这使得目标函数非凸, 优化变得十分困难. 另一方面, 测试样本的核权值如何确定也是一个难题. Gonen^[15]通过使用式(5)所示的选通模型将对核权值 γ 的求解转化为对选通函数参数 \mathbf{v}_m, v_{m0} 的求解. 通过对 \mathbf{v}_m, v_{m0} 求导可以通过梯度下降的方式优化核权值 γ .

$$\gamma_m^x = \frac{\exp(\langle \mathbf{v}_m, x_i \rangle + v_{m0})}{\sum_{m'} \exp(\langle \mathbf{v}_{m'}, x_i \rangle + v_{m'0})} \quad (5)$$

相比为每个样本施以特定核权值的方式, Yang 等^[17]提出以组为单位进行优化, 组内样本共享权值, 形式如下:

$$K(x_i, x_j) = \sum_{m=1}^M \beta_m^{c(x_i)} \beta_m^{c(x_j)} K_m(x_i, x_j) \quad (6)$$

其中, $c(x_i)$ 为样本 x_i 所属分组, $\beta_m^{c(x_i)}$ 为在该分组内第 m 个核函数的权值. 由于核权值由样本所在分组决定, 因此也就不存在测试样本的核权值难以确定的问题, 求解方法可以有更多选择, 既可以使用类似 LMKL 的选通模型和梯度下降的方式求解^[17,20], 也可以在主问题上求得闭合解^[23].

3 基于局部空间变稀疏约束的多核学习方法

假设有 N 个训练样本 $D_L = \{\mathbf{x}_i, y_i\}_{i=1}^N$, 其中 \mathbf{x}_i 代表第 i 个训练样本, $y_i \in \{-1, +1\}$ 为 \mathbf{x}_i 的标记. 判别函数表示如下:

$$f(\mathbf{x}) = \sum_{g=1}^G (C_g(\mathbf{x}) \sum_{m=1}^M \beta_m^g \langle \mathbf{w}_m^g, \phi_m(\mathbf{x}) \rangle) + b \quad (7)$$

其中, β_m^g 代表第 m 个核函数在局部空间 g 内的权值, $C_g(\mathbf{x})$ 表示样本 \mathbf{x} 对于数据子集 g 的隶属度, \mathbf{w}_m^g 表示分类超平面的法向量, $\phi_m(\mathbf{x})$ 表示样本 \mathbf{x} 在第 m 个核函数所生成的特征空间的映射, b 为偏差. 模型应使得正负类间隔最大化, 预测误差最小化, 按照这一原则, 模型可以描述为求解以下优化问题:

$$\begin{aligned} \min_{\mathbf{w}, \beta, \xi, \beta} & \frac{1}{2} \sum_g \sum_m \frac{1}{\beta_m^g} \|\mathbf{w}_m^g\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s. t.} & y_i f(\mathbf{x}_i) \geq 1 - \xi_i \quad \forall i, \\ & \xi_i \geq 0 \quad \forall i \end{aligned} \quad (8)$$

$$\sum_{m=1}^M (\beta_m^g)^{p_g} = 1 \quad \forall g$$

$$\beta_m^g \geq 0 \quad \forall g, m$$

其中,优化目标的前一项代表分类间隔,后一项表示总体误差, C 为惩罚因子, ξ 为松弛变量, p_g 表示在局部空间 g 内对核函数权值的正则化约束项。

3.1 模型求解

分析式(8)可以发现,相比于传统 SVM,本文方法需要额外求解未知量 β . 并且目标函数对于 β 非凸,为求解该问题,先在 β 固定的情况下求式(8)的对偶问题. 使用拉格朗日乘法,构造目标函数 L_1 :

$$L_1 = \frac{1}{2} \sum_{g=1}^G \sum_m^M \frac{1}{\beta_m^g} \|\mathbf{w}_m^g\|^2 + \sum_{i=1}^N (C - \alpha_i - \gamma_i) \xi_i + \sum_{i=1}^N \alpha_i$$

$$- \sum_{i=1}^N \alpha_i \gamma_i \left(\sum_{g=1}^G \sum_{m=1}^M C_g(\mathbf{x}) \langle \mathbf{w}_m^g, \phi_m(\mathbf{x}) \rangle + b \right) \quad (9)$$

其中, α_i 和 γ_i 为拉格朗日乘子。

通过对 \mathbf{w}, b, ξ 求偏导,并令导数为 0,可以得到三个等式:

$$\frac{\partial L_1}{\partial \mathbf{w}_m^g} = 0 \Rightarrow \mathbf{w}_m^g = \sum_i \alpha_i \gamma_i \beta_m^g C_g(\mathbf{x}_i) \phi_m(\mathbf{x}_i) \quad (10)$$

$$\frac{\partial L_1}{\partial b} = 0 \Rightarrow \sum_i \alpha_i \gamma_i = 0 \quad (11)$$

$$\frac{\partial L_1}{\partial \xi_i} = 0 \Rightarrow C = \alpha_i + \gamma_i \quad (12)$$

将式(10)(11)(12)代入式(9),并且根据 KKT 条件^[1]可以得到问题(8)的对偶形式及约束条件如下:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \gamma_i \gamma_j \left(\sum_{g=1}^G \sum_{m=1}^M \beta_m^g C_g(\mathbf{x}_i) C_g(\mathbf{x}_j) K_m(\mathbf{x}_i, \mathbf{x}_j) \right)$$

$$- \sum_{i=1}^N \alpha_i$$

s. t. $\sum_{i=1}^N \alpha_i \gamma_i = 0, 0 \leq \alpha_i \leq C \quad \forall i$ (13)

观察式(13)可知,求解问题(13)相当于解以 $K(\mathbf{x}_i, \mathbf{x}_j)$ 为核的支持向量机,其中核函数 $K(\mathbf{x}_i, \mathbf{x}_j)$ 变为:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_g \sum_m C_g(\mathbf{x}_i) C_g(\mathbf{x}_j) \beta_m^g K_m(\mathbf{x}_i, \mathbf{x}_j) \quad (14)$$

令

$$\gamma_m^g = \sum_{g=1}^G C_g(\mathbf{x}_i) \sqrt{\beta_m^g} \quad (15)$$

则合成核函数可表示为:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{m=1}^M \gamma_m^g \gamma_m^g K_m(\mathbf{x}_i, \mathbf{x}_j) \quad (16)$$

由式(16)可知,式(14)所示合成核相当于为每个样本赋予特定的核权值 γ_m^g . 因此在本文提出的框架下,通过软分组的形式实现了样本级的多核组合,并且计算过程较传统样本级方法大大简化. 同时软分组的过

程相当于对每个样本的核权值施以相应的空间约束,从而避免了个别噪声点的影响。

3.2 求解核函数权值

本节主要介绍核函数权值 β 的求解,首先考虑以下优化问题:

$$\min_{\beta} \sum_g \sum_m \frac{1}{\beta_m^g} \|\mathbf{w}_m^g\|^2 \quad (17)$$

s. t. $\sum_{m=1}^M (\beta_m^g)^{p_g} = 1 \quad \forall g$

$$\beta_m^g \geq 0 \quad \forall g, m$$

式(17)实质上是以 β 为变量的条件极值问题,可以使用拉格朗日乘法求解. 构造目标函数 L_2 :

$$L_2 = \sum_g \sum_m \frac{1}{\beta_m^g} \|\mathbf{w}_m^g\|^2 + \sum_{g=1}^G \theta_g \left(\sum_{m=1}^M (\beta_m^g)^{p_g} - 1 \right)$$

$$+ \sum_{g=1}^G \sum_{m=1}^M \gamma_m^g \beta_m^g \quad (18)$$

目标函数 L_2 对 β_m^g 求偏导并令导数为 0 可得:

$$\frac{\partial L_2}{\partial \beta_m^g} = -\frac{\|\mathbf{w}_m^g\|^2}{(\beta_m^g)^2} + p_g \theta_g (\beta_m^g)^{p_g-1} + \gamma_m^g = 0 \quad (19)$$

并且不等式约束项需满足 KKT 条件^[1]:

$$\gamma_m^g \beta_m^g = 0 \quad \forall m, g \quad (20)$$

由于 $\beta_m^g \neq 0$,由式(20)可知 $\gamma_m^g = 0$.

因此式(19)变为:

$$-\frac{\|\mathbf{w}_m^g\|^2}{(\beta_m^g)^2} + p_g \theta_g (\beta_m^g)^{p_g-1} = 0 \quad (21)$$

结合正则化约束项 $\sum (\beta_m^g)^{p_g} = 1$,可以非常容易解得 β_m^g :

$$\beta_m^g = \frac{(\|\mathbf{w}_m^g\|)^{\frac{2}{p_g+1}}}{\left(\sum_{m=1}^M (\|\mathbf{w}_m^g\|)^{\frac{2p_g}{p_g+1}} \right)^{\frac{1}{p_g}}} \quad (22)$$

$\|\mathbf{w}_m^g\|$ 为核函数 m 和局部空间 g 所确定的分类间隔, \mathbf{w}_m^g 的计算 3.1 节式(10)已经给出,根据式(10)可以很容易得知 $\|\mathbf{w}_m^g\|^2$ 的计算如下:

$$\|\mathbf{w}_m^g\|^2 = \sum_{i,j} \alpha_i \alpha_j \gamma_i \gamma_j (\beta_m^g)^2 C_g(\mathbf{x}_i) C_g(\mathbf{x}_j) K_m(\mathbf{x}_i, \mathbf{x}_j) \quad (23)$$

由此可以看出 β_m^g 的求解是建立在 α_i 已知的基础上,而 α_i 得求解则依赖于 β_m^g 所确定的合成核. 因此本文采用交替优化的方法,先初始化 $\beta_m^g = 1/m$,然后交替优化 β_m^g 和 α_i ,直至目标函数不再变化或达到最大迭代次数. 实际上这也是大多数多核学习方法所使用的策略,所幸 β_m^g 存在闭合解,可以直接计算,训练所需时间主要依赖于 SVM 的求解速度和迭代次数。

3.3 确定正则化约束参数 p_g

p_g 为控制核函数权值在局部空间 g 内的稀疏性的

正则化约束参数. 当 $p_g = 1$ 时, 表现最优的核被选中, 其余核的权值较小甚至趋向于 0; 当 $p_g \rightarrow \infty$ 时, 所有核均以接近为 1 的权值被选中. 文献[25]的研究表明, 最佳的稀疏约束参数与在相应数据集特性有关. 当在该数据集上各个核函数相似性较高时宜采用稀疏组合, 当各核函数相似性较低时宜采用非稀疏组合. 本文借鉴文献[25]的结论使用 KTA (Kernel Target Alignment)^[26] 作为核函数相似性的度量标准, 依据各数据子集内各个核函数间的 KTA 值确定 p_g 值.

在局部空间 g 内的 KTA 值计算方法如下:

$$\hat{A}_g(\mathbf{K}_m, \mathbf{K}_{m'}) = \frac{\langle \mathbf{K}_m^g, \mathbf{K}_{m'}^g \rangle_F}{\sqrt{\langle \mathbf{K}_m^g, \mathbf{K}_m^g \rangle_F \langle \mathbf{K}_{m'}^g, \mathbf{K}_{m'}^g \rangle_F}} \quad (24)$$

其中

$$\langle \mathbf{K}_m^g, \mathbf{K}_n^g \rangle_F = \sum_{i,j} C_g(x_i) C_g(x_j) \mathbf{K}_m(i,j) \mathbf{K}_n(i,j)$$

同时计算在全局空间上核函数之间的 KTA 值:

$$\hat{A}(\mathbf{K}_m, \mathbf{K}_n) = \frac{\langle \mathbf{K}_m, \mathbf{K}_n \rangle_F}{\sqrt{\langle \mathbf{K}_m, \mathbf{K}_m \rangle_F \langle \mathbf{K}_n, \mathbf{K}_n \rangle_F}} \quad (25)$$

其中

$$\langle \mathbf{K}_m, \mathbf{K}_n \rangle_F = \sum_{i,j} \mathbf{K}_m(i,j) \mathbf{K}_n(i,j)$$

根据在局部空间 g 内的核函数的相关程度调整局部核权值的正则化约束参数 p_g :

$$p_g = p \frac{\sum_{m,m'} \hat{A}_g(\mathbf{K}_m, \mathbf{K}_{m'})}{\sum_{m,m'} \hat{A}(\mathbf{K}_m, \mathbf{K}_{m'})} \quad (26)$$

3.4 算法流程

总结以上过程, 本文算法训练阶段的计算流程可描述为:

Step 1: 根据空间位置, 使用模糊 C 均值聚类算法将训练样本划分为若干子集;

Step 2: 计算各局部空间内的核函数的约束参数 p_g ;

Step 3: 初始化 $\beta_m^g = 1/m \quad \forall m, g$;

Step 4: 利用式(16)计算核矩阵;

Step 5: 通过解算标准 SVM 的方法计算 α ;

Step 6: 通过式(22)更新局部核权值 β ;

Step 7: 判断是否满足迭代停止条件, 若不满足, 重复 Step 4 至 Step 6, 直到满足停止条件.

3.5 时间复杂度分析

本文方法与样本级方法、分组方法一样通过交替优化的方式进行优化. 第一部分均为解以多核组合 SVM, 这一部分的计算量相同. 第二部分为核权值矩阵的优化, 样本级方法的时间复杂度为 $O(N^2 M^2 d)$, 其中 N 为样本数量, M 为核函数数量, d 为样本维度. 本文方法的时间复杂度为 $O(N^2 M G)$, G 为分组数量. 本文算法与样本级方法的时间复杂度高低跟样本数据的维度和

核函数数量有关. 分组方法的时间复杂度最低, 为 $O(N^2 M)$. 算法的具体运行时间还与迭代次数有关, 由于本文算法可以求得核权值的闭合解, 因此可以更快收敛至最优解, 实际计算效率会优于样本级方法, 介于分组方法和样本级方法之间.

4 实验及分析

本文使用 MATLAB 作为编程环境, 使用 SMO 算法^[27]解 SVM. 并将本文算法与三种具有代表性的多核学习方法: 变稀疏多核方法 lp -MKL^[25] ($p = 1, 2, 4$)、样本级局部多核方法 LMKL^[15]、组敏感法多核学习方法 GS-MKL^[20] 作对比. 本文通过构造一个包含两个特征的人工数据集验证本文方法在区分特征重要程度方面的能力以及噪声对分类准确率的影响, 而后通过两个图像数据集做了对比实验, 并对算法性能作细致分析. 分类准确率以总体准确率计算, 即正确分类样本数与所有待分类样本数的比值.

实验预处理阶段, 使用核模糊 C 均值算法^[28]进行聚类, 分别在各个特征上构建核, 以各特征空间内的平均距离作为样本间距离, 模糊因子 m 取 2, 随机选择距离初始中心, 重复多次聚类操作选择误差最小的一次进行实验.

4.1 人工数据集

为验证本文算法在区分不同样本的核重要程度的能力, 本文构造了一个两特征数据集. 数据集包含正负两类各 200 个样本, 每个样本由 A、B 两种特征组成. 数据集由四个不同的高斯分布产生, 协方差矩阵均为 $[0.5, 0, 0, 0.5]$, 均值如下:

	特征 A	特征 B
正类	$\mu_{11} = (0, 0)$	$\mu_{12} = (0, 3)$
负类	$\mu_{21} = (3, 3)$	$\mu_{22} = (3, 0)$

分别从两类样本中随机选择 100 个样本在其 A 类特征中加入噪声, 然后再将另外 100 个样本在 B 类特征作同样操作. 分别从两个视角观察数据, 原始无噪声样本如图 1 所示, 加入 0.5 水平后的噪声后的样本如图 2 所示, 红色圆点代表正类样本, 绿色三角点代表负类样本.

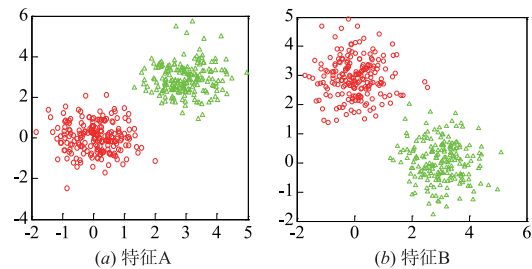


图1 人工数据集 (无噪声)

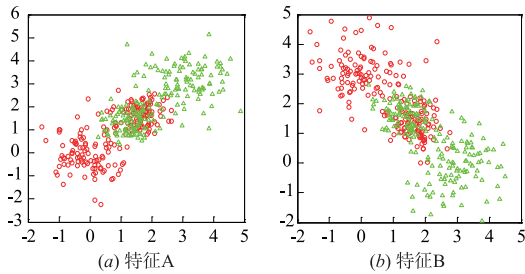


图2 人工数据集 (噪声水平为0.5)

如图2所示,加入噪声的数据在任一特征上均是不可分的,但是对于不同样本选择适合的特征,将使问题变得可分.在视觉场景分类问题中,通常需要提取多种特征,有时甚至多达上百种^[29],而通常大部分特征并不适用于所有样本,对于部分样本来说也可能只在个别特征空间中是可分的.因此对于包含异构特征的分类问题来说,区分特征学习十分重要.这个数据集的优点在于,两种特征对于样本的重要程度是已知的.可以通过观察算法在不同样本的核权值分布情况和分类准确率,评价算法在局部空间内的区分特征学习能力.

本文分别使用两个径向基核描述 A、B 两个特征,实验中各算法的参数定义如下:核参数 $\delta = 0.5$,惩罚因子 C 取 100,本文算法中核权值稀疏化参数 $p = 1$,分组参数 G 取 200.为克服随机性,选择 20 次实验的平均值作为实验结果.

图3所示为分类准确率随噪声变化的情况.从图中可以看出,随着噪声水平的增大,分类准确率不断降低.其中 l_p -MKL 下降最为明显,当噪声水平大于 0.75 时,已经失去了基本的分类判别能力(准确率接近于 0.5).这是由于 l_p -MKL 无论 p 取多少,均为所有样本使用统一的核加权方式,不能区分两种特征对于不同样本的重要程度.本文算法的分类准确率明显高于其它三种算法,并且随着噪声的增大优势更加明显.为了进一步分析本文方法对于局部多核学习的效果,图4描绘了两种局部多核学习算法在训练结束时所有支持向量的特征 A 所占权值.其中图4(a)为 LMKL 的计算结果,图4

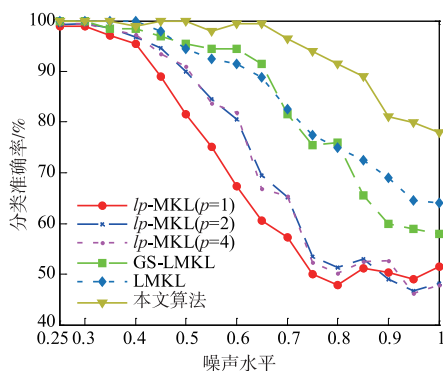


图3 分类准确率随噪声变化情况

(b)为本文算法的结果.为便于比较,将特征 A 包含噪声的样本放置于图的左侧,特征 B 包含噪声的样本放置于右侧.因此表现好的算法图右侧部分的特征 A 所占权值应该明显大于左侧部分.观察图5可以发现,本文算法计算结果更加符合两种特征的重要程度,因此获得了更高的分类准确率.

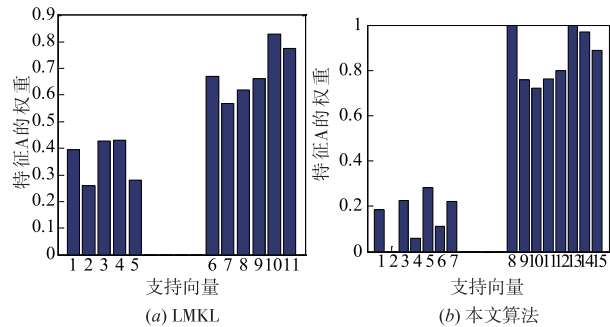


图4 特征A的权值

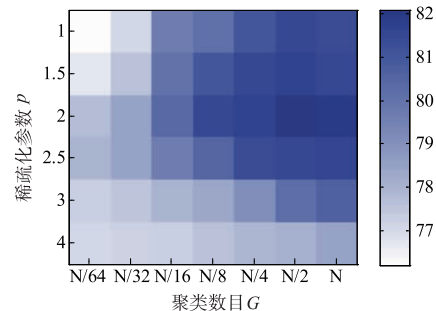


图5 分类准确率随参数变化图

4.2 图像数据集

本文通过两个图像数据集 Scene-15 和 Caltech-101 验证所提算法的有效性. Scene-15 数据集采集自 15 种不同的自然场景,每种类别包括 200 至 400 幅图像,共 4485 幅. Caltech-101 数据集共包含 9145 幅图像,分别属于 102 个类别,每个类别包含 31 到 800 不等数量的图像.

参照文献[21]的设置,本文共选择 8 种特征描述图像,分别为 GIST, HOG2 \times 2, Dense SIFT, Sparse SIFT, SSIM, Line Histograms, Texton 以及 Gometric Map. 对于 GIST 特征使用 RBF 距离进行相似性度量,其余特征使用 χ^2 距离进行相似性度量.

本文使用“一对多”的策略处理多分类问题,即训练时对每个类别 i ,其它所有类别视为一类,这样共有 N 个分类器(这里 N 为类别数量).测试时,给定一幅未标记图像,先提取图像的视觉特征,再分别使用 N 个分类器进行识别,得到 N 个识别结果.最后选择置信度最大的类别作为测试图像最终识别结果.

(1) 参数选择与分析

由于本文算法增加了一些可调参数(p, G),本节通过实验考察不同参数设置对于算法性能的影响,以及

参数间的耦合性. 从 Caltech-101 数据集的每个类别中分别选择 20 个样本, 分别令 $G = \{N/64, N/32, N/16, N/8, N/4, N/2, N\}$ (N 为训练样本数), $p = \{1, 1.5, 2, 2.5, 3, 4\}$ 进行训练, 对其余样本进行预测. 惩罚因子 C 在 $[0.01, 0.1, 1, 10, 100]$ 使用网格法选取最优参数. 分类准确率如图 5 所示. 另外为便于观察, 图 6 描述了 $p = 2$ 时准确率随 G 的变化情况.

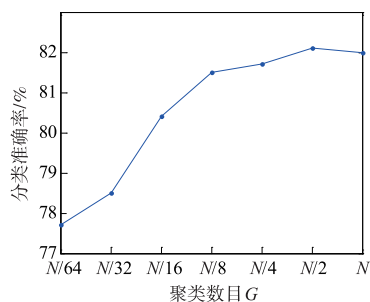


图6 分类准确率随聚类数变化图

p 为全局空间内的核权值稀疏参数, p 越大则会有越多的核在分类时发挥作用, 不同核的地位越平均. 从图 5 可以看出, 本次实验中 p 取 2 左右时效果最好, 因此在后面的对比实验中 $p = 2$. 值得注意的是 p 的取值是由不同核之间的冗余程度和重要性决定的, 不具有通用性. 本文算法可以根据核在不同局部空间内的冗余情况调整其稀疏化参数 p_g , 但是全局的稀疏参数 p 仍然需要人工确定.

聚类数目 G 表示对局部空间的区分程度, 当 $G = 1$ 时本文方法退化为普通多核方法, 不具有局部学习能力, 随着 G 的增大算法对局部空间的区分更加精细. 从图 5 和图 6 可以看出随分类数目 G 的增大, 分类准确率有不断增大的趋势. 但聚类数目的增加意味着更高计算量, 并且存在过拟合的风险, 从而影响分类准确率的提升, 因此后续对比实验中 G 取 $N/2$. 此外, 当 $p \geq 3$ 时, 这种准确率随聚类数目的变化并不明显. 分析可知, 当 p 较大时, 子核之间权重相差不大, 算法对核的选择能力减弱, 并逐步退化为平均加权的核组合方式, 因此分组多少对于性能的影响减弱.

(2) 性能对比分析

为便于比较, 本文使用与文献 [30, 31] 相同的设置, 即每次随机选取固定数目的样本训练, 重复多次取平均值作比较. 对于 Scene-15 数据集, 分别从每类数据中选择 10, 20, 30, 50, 100 作为训练集, 其余样本作为测试集. 对于 Caltech-101 数据集, 分别从每类数据中选择 10, 15, 20, 25, 30 作为训练集, 其余样本作为测试集. 在对比实验中除了选择前面三种算法进行对比外, 还选取了最近在图像分类方面表现优异的 CLMKL^[23]、BPMKL^[32] 和 MKL-SRC^[29] 进行对比, 同时将直接对多核

加权平均的 SVM 实验结果也列入表中进行比较.

重复实验 10 次的平均准确率以及方差如表 1 和表 2 所示. 由表 1 和表 2 可以看出, 本文方法在平均准确率上优于其他算法. 但是当训练样本数量较小时, 并不具有绝对优势. 这是由于本文方法是在对数据分组的基础上进行训练的. 当训练数据量较小时, 训练集内的多样性不明显, 分组的效果受到影响. 随着训练样本数目的增大, 本文方法的优势更加明显.

此外本文方法在 Scene 15 数据集的提高幅度要高于 Caltech 101 数据集, 这主要由于 Caltech 101 数据集中的目标较为简单, 样本的多样性并不突出; 而 Scene 15 数据集同一类别的样本差异性比较明显, 区分特征学习的需求更高. 这也体现了本文方法在区分特征重要性在样本间的差异方面具有较强的优势. 除此之外, 本文方法的分类准确率的方差较小, 体现了本文方法具有更好的稳定性.

表 1 在 Caltech 101 数据集上的分类准确率

算法	每一类的训练样本数量				
	10	15	20	25	30
SVM	65.5 ± 0.8	68.7 ± 0.8	70.4 ± 0.7	71.8 ± 0.6	73.3 ± 0.5
l_1 -MKL	66.4 ± 1.1	70.5 ± 1.1	73.7 ± 1.0	75.2 ± 0.8	75.7 ± 0.8
l_2 -MKL	66.8 ± 1.0	72.1 ± 1.0	75.2 ± 1.0	78.1 ± 0.8	79.5 ± 0.7
l_4 -MKL	65.8 ± 0.9	69.6 ± 1.0	71.1 ± 0.9	72.9 ± 0.8	74.5 ± 0.8
LMKL	68.2 ± 1.5	75.3 ± 1.2	77.7 ± 1.1	79.5 ± 1.1	80.4 ± 1.0
GS-MKL	66.3 ± 1.3	75.3 ± 1.1	81.6 ± 1.1	83.6 ± 1.0	84.4 ± 0.9
CLMKL	67.0 ± 1.2	75.5 ± 1.0	80.8 ± 0.9	83.8 ± 0.9	84.7 ± 0.9
BPMKL	68.2 ± 1.1	74.8 ± 1.0	79.7 ± 1.0	81.8 ± 0.8	83.7 ± 0.8
MKL-SRC	68.9 ± 1.3	73.2 ± 1.3	78.8 ± 1.2	82.0 ± 1.1	82.2 ± 1.0
本文方法	68.1 ± 1.2	76.2 ± 1.1	82.1 ± 1.0	84.1 ± 1.0	85.0 ± 0.8

表 2 在 Scene 15 数据集上的分类准确率

算法	每一类的训练样本数量				
	10	20	30	50	100
SVM	58.2 ± 1.0	64.4 ± 0.9	68.3 ± 0.8	74.5 ± 0.7	80.9 ± 0.5
l_1 -MKL	59.7 ± 1.2	65.9 ± 1.1	71.0 ± 1.0	77.2 ± 0.8	82.9 ± 0.6
l_2 -MKL	60.1 ± 1.1	66.8 ± 1.0	71.2 ± 1.0	77.8 ± 0.7	83.5 ± 0.6
l_4 -MKL	58.2 ± 1.0	64.4 ± 0.9	68.3 ± 0.8	74.5 ± 0.7	80.9 ± 0.5
LMKL	62.5 ± 1.6	67.1 ± 1.3	72.7 ± 1.1	78.3 ± 0.9	84.3 ± 0.8
GS-MKL	61.3 ± 1.4	68.4 ± 1.2	74.3 ± 1.1	80.8 ± 1.0	86.6 ± 0.9
CLMKL	61.4 ± 1.5	68.2 ± 1.3	74.6 ± 1.0	79.9 ± 1.0	86.7 ± 1.0
BPMKL	62.6 ± 1.6	67.8 ± 1.1	75.0 ± 1.2	80.1 ± 1.1	85.5 ± 1.0
MKL-SRC	62.5 ± 1.7	67.5 ± 1.3	73.5 ± 1.3	78.2 ± 1.1	86.1 ± 1.0
本文方法	62.6 ± 1.3	69.1 ± 1.1	75.9 ± 0.9	82.0 ± 0.8	88.1 ± 0.7

5 结论

本文提出了一种基于局部空间变稀疏约束的多核学习方法. 使用软分组的方式为样本划分局部空间, 根据样本所在局部空间的核函数特性确定核函数组合的稀疏性, 通过交替优化的方式求解局部核函数权值和

支持向量系数. 实验表明本文方法具有更优的区分特征学习和抵抗噪声的能力, 因此也使得图像场景分类的准确率和稳定性得到明显提高. 此外本文方法也可以用于解决其他包含异构特征的分类回归等问题.

由于本文算法使得可调参数增加, 使得训练前的调参任务加重, 下一步考虑研究不同参数设置与数据集特性的关系.

参考文献

- [1] VAPNIK V. The Nature of Statistical Learning Theory [M]. New York: Springer-Verlag, 1995.
- [2] 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. (in Chinese)
- [3] Li X, Mao W, Jiang W. Multiple kernel learning based extreme learning machine for classification design[J]. Neural Computing and Applications, 2016, 27(1): 175 - 184.
- [4] Wang C D, Lai J H, Yu P S. Multi-view clustering based on belief propagation[J]. IEEE Transactions on Knowledge & Data Engineering, 2016, 28(4): 1007 - 1021.
- [5] Nazarpour A, Adibi P. Two-stage multiple kernel learning for supervised dimensionality reduction[J]. Pattern Recognition, 2015, 48(5): 1854 - 1862.
- [6] Sahoo D, Hoi S C H, Li B. Online multiple kernel regression[A]. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining[C]. USA: ACM, 2014. 293 - 302.
- [7] 亓晓振, 王庆. 一种基于稀疏编码的多核学习图像分类方法[J]. 电子学报, 2012, 40(4): 773 - 779.
QI Xiao-zhen, WANG Qing. An image classification approach based on sparse coding and multiple kernel learning [J]. Acta Electronica Sinica, 2012, 40(4): 773 - 779. (in Chinese)
- [8] 余家林, 孙季丰, 李万益. 基于多核稀疏编码的三维人体姿态估计[J]. 电子学报, 2016, 44(8): 1899 - 1908.
YU Jia-lin, SUN Ji-feng, LI Wan-yi. 3D human pose estimation based on multi-kernel sparse coding[J]. Acta Electronica Sinica, 2016, 44(8): 1899 - 1908. (in Chinese)
- [9] 苏娟, 王百合, 刘代志. 一种基于拓扑约束的多核跟踪算法[J]. 电子学报, 2015, 43(2): 353 - 357.
SU Juan, WANG Bai-he, LIU Dai-zhi. A multi-kernel tracking algorithm based on topology constraint [J]. Acta Electronica Sinica, 2015, 43(2): 353 - 357. (in Chinese)
- [10] 杨赛, 赵春霞, 刘凡. 多核学习融合局部和全局特征的人脸识别算法[J]. 电子学报, 2016, 44(10): 2344 - 2350.
YANG Sai, ZHAO Chun-xia, LIU Fan. Fusion of local and global features using multiple kernel learning for face recognition [J]. Acta Electronica Sinica, 2016, 44(10): 2344 - 2350. (in Chinese)
- [11] Mehment G, Ethem A. Multiple kernel learning algorithms [J]. Journal of Machine Learning Research, 2011, 12: 2211 - 2268.
- [12] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436 - 444.
- [13] Lin Y-Y, Liu T-L, Fuh C-S. Local ensemble kernel learning for object category recognition [A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. USA: IEEE, 2007. 1 - 8.
- [14] Lin Y-Y, Tsai J-F, Liu T-L. Efficient discriminative local learning for object recognition [A]. Proceedings of IEEE 12th International Conference on Computer Vision (ICCV) [C]. USA: IEEE, 2009. 598 - 605.
- [15] Gonen M, Alpaydin E. Localized algorithms for multiple kernel learning [J]. Pattern Recognition, 2013, 46(3): 795 - 807.
- [16] Han Y, Yang K, Liu G. Lp-norm localized multiple kernel learning via semi-definite programming [J]. IEEE Signal Processing Letters, 2012, 19(10): 688 - 691.
- [17] Yang J, Li Y, Tian Y, et al. Group-sensitive multiple kernel learning for object recognition [J]. IEEE Transactions on Image Processing, 2012, 21(5): 2838 - 2852.
- [18] 张瑞杰, 郭志刚, 李弼程, 等. 基于 E ~ 2LSH-MKL 的视觉语义概念检测 [J]. 自动化学报, 2012, 38(10): 1671 - 1678.
Zhang Ruijie, Guo Zhigang, Li Bicheng, et al. A visual semantic concept detection algorithm based on E2LSH-MKL [J]. Acta Automatica Sinica, 2012, 38(10): 1671 - 1678. (in Chinese)
- [19] Fu G Y, Wang Q C, Wang H Q, Bai D Y. Group based non-sparse localized multiple kernel learning algorithm for image classification [A]. Proceedings of the 4th IEEE International Conference on Cloud Computing and Intelligence Systems [C]. USA: IEEE, 2016. 191 - 195.
- [20] Fu G Y, Wang Q C, Bai D Y, Li L L. Group based localized multiple kernel learning algorithm with lp-norm [J]. International Journal of Innovative Computing, Information and Control, 2016, 6(12): 1835 - 1849.
- [21] Han Y, Yang K, Ma Y, et al. Localized multiple kernel learning via sample-wise alternating optimization [J]. IEEE Transactions on Cybernetics, 2013, 44(1): 127 - 148.
- [22] Liu X, Wang L, Zhang J, et al. Sample-adaptive multiple kernel learning [A]. Proceedings of the 28th AAAI Conference on Artificial Intelligence [C]. USA: AAAI Press, 2014. 1 - 16.

- [23] Lei Y, Binder A, Doganü, Kloft M. Theory and algorithms for the localized setting of learning kernels[J]. The Journal of Machine Learning Research, 2015, 7 (7): 173 – 195.
- [24] Antic B, Ommer B. Per-sample kernel adaptation for visual recognition and grouping[A]. IEEE International Conference on Computer Vision[C]. USA: IEEE, 2015. 1251 – 1259.
- [25] Marius K, Ulf B, Soren S, Alexander Z. Lp-norm multiple kernel learning[J]. The Journal of Machine Learning Research, 2011, 12(12): 953 – 997.
- [26] Cristianini N, Shawe-Taylor J, Elisseeff A, et al. On kernel-target alignment[A]. Proceedings of the International Conference on Neural Information Processing Systems [C]. USA: MIT Press, 2001. 367 – 373.
- [27] Platt J C. Sequential minimal optimization: a fast algorithm for training support vector machines [A]. Advances in Kernel Methods-Support Vector Learning[C]. USA: Microsoft, 1998. 212 – 223.
- [28] Chen L, Chen C L, Lu M. A multiple-kernel fuzzy C-means algorithm for image segmentation[J]. IEEE Transactions on Systems Man & Cybernetics, 2011, 41 (5): 1263 – 1274.
- [29] Shrivastava A, Patel V M, Chellappa R. Multiple kernel learning for sparse representation-based classification[J]. IEEE Transactions on Image Processing, 2014, 23 (7): 3013 – 3024.
- [30] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[A]. Proceedings of IEEE Conference on Computer Vision & Pattern Recognition[C]. USA: IEEE Computer Society, 2006. 2169 – 2178.
- [31] Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories[J]. Computer Vision & Image Understanding, 2004, 106(1): 178.
- [32] Ni B, Li T, Moulin P. Beta process multiple kernel learning[A]. Computer Vision and Pattern Recognition[C]. USA: IEEE, 2014. 963 – 970.

作者简介



王庆超 男, 1988 年生, 河北清河人, 博士研究生. 2014 年于第二炮兵工程大学获得硕士学位, 现为火箭军工程大学信息工程系博士研究生, 主要从事视觉目标识别及遥感图像分析方面的研究.
E-mail: wangqingchao@yeah.net



付光远 男, 1966 年生, 四川简阳人, 教授、博士生导师. 1993 年于西南交通大学获得硕士学位, 2004 年于第二炮兵工程学院获得博士学位. 主要从事图像匹配精确制导、C4I 系统建模与仿真方面的研究.
E-mail: dr-f@21cl.com