

## 结合组稀疏效应和多核学习的图像标注<sup>\*</sup>

袁莹, 邵健<sup>+</sup>, 吴飞, 庄越挺

(浙江大学 计算机科学与技术学院 人工智能研究所, 浙江 杭州 310027)

### Image Annotation by the Multiple Kernel Learning with Group Sparsity Effect

YUAN Ying, SHAO Jian<sup>+</sup>, WU Fei, ZHUANG Yue-Ting

(Institute of Artificial Intelligence, College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China)

+ Corresponding author: E-mail: jshao@zju.edu.cn

Yuan Y, Shao J, Wu F, Zhuang YT. Image annotation by the multiple kernel learning with group sparsity effect. *Journal of Software*, 2012, 23(9): 2500–2509 (in Chinese). <http://www.jos.org.cn/1000-9825/4154.htm>

**Abstract:** Since different kinds of heterogeneous features (such as color, shape and texture) in image have different intrinsic discriminative power for image understanding, this paper proposes a multiple kernel learning with group sparsity (MKLGS) to select groups of discriminative features for image annotation to effectively utilize those heterogeneous visual features. Given each image label, the MKLGS method embeds the nonlinear image data with discriminative features into a Hilbert space, and then utilizes the kernel function in the Hilbert space and group LASSO to select groups of discriminative features. Finally, a classification model can be trained for image annotation. In comparison to other image annotation algorithms, experiments show that the proposed MKLGS for image annotation achieves a better performance.

**Key words:** group LASSO; group sparsity; multiple kernel learning; feature selection; image annotation

**摘要:** 图像中存在的纹理、颜色和形状等异构视觉特征,在表示特定高层语义时所起作用的重要程度不同.为了在图像标注过程中更加有效地利用这些异构特征,提出了一种基于组稀疏(group sparsity)的多核学习方法(multiple kernel learning with group sparsity,简称 MKLGS),为不同图像语义选择不同的组群特征.MKLGS 先将包含多种异构特征的非线性图像数据映射到一个希尔伯特空间,然后利用希尔伯特空间中的核函数以及组 LASSO(group LASSO)对每个图像类别选择最具区别性特征的集合,最终训练得到分类模型对图像进行标注.通过与目前其他图像标注算法进行对比,实验结果表明,基于组稀疏的多核学习方法在图像标注中能取得很好的效果.

**关键词:** 组 LASSO;组稀疏;多核学习;特征选择;图像标注

中图法分类号: TP391 文献标识码: A

近年来,从统计信号处理中发展出的压缩感知(compressive sensing,简称 CS)受到越来越多的关注.该理论表明,如果信号的表示是足够稀疏的,那么就可以通过凸优化的方法来计算.美国斯坦福大学的 Tibshirani 和加州大学伯克利分校的 Breiman 几乎同时提出了对特征系数施以  $l_1$ -范数约束的 LASSO(least absolute shrinkage

\* 基金项目: 国家自然科学基金(61070068, 60833006); 国家重点基础研究发展计划(973)(2010CB327904); 核高基项目(2010ZX 01042-002-003)

收稿时间: 2011-03-07; 修改时间: 2011-09-22; 定稿时间: 2011-11-17

and selection operator)思想,促使被选择出来的特征尽可能稀疏,以保证结果稳定性和提高数据处理过程的可解释性(interpretable).由此出现了不少以 LASSO 为基础的特征选择算法,应用于人脸识别和图像标注等领域<sup>[1-4]</sup>.

但是,以 LASSO 为基础的特征选择方法并没有考虑到特征之间存在的组效应(grouping effect)特性(即某一(类)特征与其他(类)特征之间存在很强相关性).传统 LASSO 方法对所有特征逐一进行选择,忽略了特征之间相关性.为了克服这一不足,Yuan 等人<sup>[5]</sup>提出了组 LASSO(group LASSO)方法结合了 LASSO 中的  $l_1$ -范数约束以及岭回归中的  $l_2$ -范数约束,使得组间尽可能稀疏,而组内保持尽可能小的损失误差.由于这一方法考虑了特征之间的组效应,组 LASSO 的方法近来在图像分类、图像标注、回归分析等等<sup>[6-10]</sup>被成功使用.

众所周知,图像中存在纹理、颜色和形状等全局特征以及 SIFT 和 LBP 等局部特征,不同视觉特征在表示特定高层语义时所起重要程度不同,如 SIFT 等局部特征在表达刚体对象时更为有效.这些异构特征之间存在明显的组效应,我们可将同一种类视觉特征归属为一组(如 SIFT 特征归属为一组,而颜色直方图归属为另一组),使得图像异构特征在表达时能充分利用这种结构性组效应.因此,如何根据图像异构特征之间存在的组效应,应用组 LASSO 来构造适合图像处理的异构特征选择机制,是一个值得关注的问题.

此外,从现实世界中所得数据中提取的高维异构特征往往分布在复杂非欧式空间且线性不可分,为了将这种线性不可分的数据变得线性可分以便分析理解,基于支持向量机(SVM)理论的核方法迅速发展起来.核方法实质上是把数据映射到一个希尔伯特空间,在该空间中寻找数据之间存在的线性关系.这种映射由核函数隐式定义,利用数据之间的两两内积,避免了在特征空间中所需的复杂计算.此后,核方法迅速渗透到机器学习的各个领域,包括分类、回归分析、多标注问题等等.但是这些方法大多是基于单一特征空间进行单个核函数学习,并没有考虑到数据特征空间所具有的组效应特性,从而忽略了不同核函数之间的相互作用.为了解决单个核函数学习存在的局限性,近年来,一些学者开始研究多核学习的方法(multiple kernel learning)<sup>[11-13]</sup>.最简单的多核学习方法就是对多个核函数取平均,每个核函数都有相同的权重系数.但是直接将这一思路应用于图像理解,却无法反映图像中不同种类特征在表达对象时所起作用程度不同这一现实.

由此,本文提出了一种基于组稀疏和多核学习的图像异构特征选择方法(multiple kernel learning with group sparsity,简称 MKLGS).区别于传统的多核学习方法,MKLGS 考虑到图像原始高维异构特征之间存在组群结构这一特性,利用异构特征所存在的组稀疏(grouping sparsity)特点去选择某一语义所对应的重要特征,即选择特征相对应的核函数,从而达到自动选择核函数的目的.与传统组 LASSO 不同之处在于,MKLGS 考虑到了高维异构特征分布在复杂非线性空间这一特点,利用这一特性来提高具有组效应高维异构特征选择的性能.对于不同图像语义,MKLGS 能够选择最具区别性特征集,进而训练该语义的分类模型,实现图像语义理解.该算法将原始图像的高维异构特征按照组群结构一一映射到新的希尔伯特空间,利用希尔伯特空间中的核函数将线性不可分问题转化成线性可分,通过自动选择核函数的机制来选择某一语义所对应的重要特征,从而实现图像标注.

## 1 相关工作

现实世界中的样本往往和多个类别相关联,例如,一幅图片可以同时被标注为“sea”,“sky”,“beach”等标签,这就需将样本同时归到多个类,这就是多标注分类问题.

解决多标注分类问题一种常用的方法是对每个标注构造一个二元分类器,具有此标注的样本为正样例,其余样本为负样例,这种方法已经被广泛应用于多种分类问题中<sup>[14,15]</sup>.为了克服这一传统方法未能考虑不同标签之间语义关联性的不足, Ji 等人<sup>[16]</sup>假设子空间是多个标注共享的,利用线性转化将子空间引入到算法模型中,以此提高标注准确率.为使选择出来的特征尽可能稀疏,以保证结果稳定性和提高数据处理过程的可解释性(interpretable),出现了不少以 LASSO 为基础的特征选择算法. Ma 等人<sup>[2]</sup>利用稀疏表示(sparse representation)的分析原理,提出了一种能够处理损毁和遮挡的全新的、足够鲁棒的人脸识别方法. Liu 等人<sup>[17]</sup>提出了 NNCW 方法对测试样本进行非负稀疏表示.但是,这些方法没有考虑到数据特征具有组群结构特性.由此, Cao 等人<sup>[6]</sup>提出了一个异构特征机器模型(heterogeneous feature machines,简称 HFM)对不同的特征学习不同的度量函数.在组群层面上,通过  $l_1$ -范数构造一个稀疏逻辑回归模型,以此来进行异构特征的选取. Wu 等人<sup>[7]</sup>提出了一种基于组

LASSO 逻辑回归模型(group LASSO for logistic regression,简称 GLLR)对图像的异构特征进行选择,并考虑标注之间的相关性,使得基于语义理解的图像标注更为准确.Han 等人<sup>[18]</sup>提出了一种对异构特征进行组间和组内选择的机制,这一机制利用异构特征所存在的结构性组稀疏(structural grouping sparsity)特点,选择某一语义所对应的重要特征.为了将非线性高维特征空间中的原始数据变得线性可分,一些基于核的学习方法被用于图像多标注中<sup>[19,20]</sup>.

如前所述,本文提出一种基于组稀疏的多核学习图像标注算法(MKLGs),该算法利用数据特征的组群结构特性构造具有组群特性的核函数,将原始空间中非线性的多维数据特征向量按照组群结构映射到一个新的希尔伯特空间,利用数据间的内积表示构造分类器,从而实现了图像异构特征的选择以及标注(异构特征选择过程如图 1 所示).

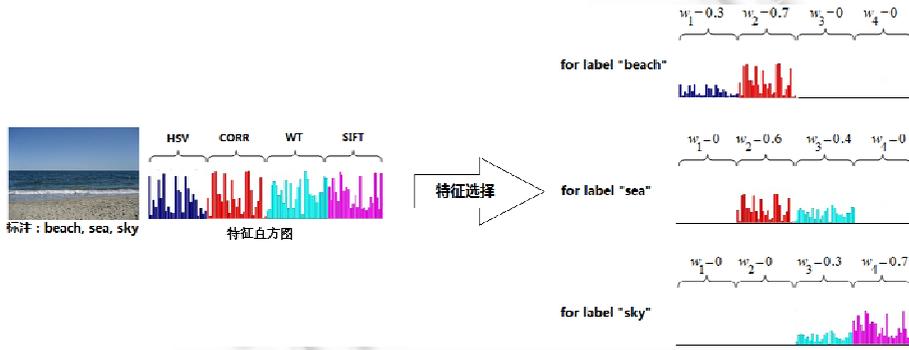


Fig.1 Flowchart of the heterogeneous feature selection  
图 1 异构特征选择机制示意图

在图 1 中,给定一幅标注为 beach,sea 和 sky 的图像,得到其 HSV,Corr,WT 和 SIFT 特征(HSV 为 HSV 空间中的颜色柱状图,Corr 为颜色相关图,WT 为图像的小波变换得到的纹理特征). $w=(w_1,w_2,w_3,w_4)$ 是组群特征的选择系数, $w_i$ 值为 0,表示第  $i$  种特征对此类图像的标注没有起到作用,在图像标注时可以忽略这种特征.可以看出,对于不同图像语义(label),不同视觉特征所起作用程度不同.

## 2 基于组稀疏的多核学习算法

### 2.1 符号表示

给定  $n$  幅带标注信息的图像集合  $\{(x_i,y_i) \in \mathbb{R}^p \times \{-1,1\}^C : i=1,2,\dots,n\}$ ,其中  $x_i=(x_{i1},\dots,x_{ip})^T \in \mathbb{R}^p$  表示第  $i$  幅图像的特征向量, $p$  为特征维数, $y_i=(y_{i1},\dots,y_{iC})^T \in \{-1,1\}^C$  表示第  $i$  幅图像所对应的标注信息,如果第  $i$  幅图像有标注  $j$ ,那么  $y_{ij}=1$ ;否则, $y_{ij}=-1$ ,令  $Y_i=(y_{i1},\dots,y_{iC})$ ,那么  $Y_i$  表示所有  $n$  幅图像的第  $i$  个标注集合.

一般可从图像中提取丰富的异构特征,假设从每幅图像提取了  $G$  组不同属性视觉特征,本文将  $p$  维异构特征向量按照属性(类别)分为  $G$  组,表达成  $G$  个不同的特征向量组,用  $m_g$  表示第  $g$  组(类)特征维数, $g \in \{1,2,\dots,G\}$ ,则有  $\sum_{g=1}^G m_g = p$ .于是,根据这样的结构性分组,第  $i$  幅图像的特征向量  $x_i$  可重新表达为  $x_i = (x_{i,1}^T, x_{i,2}^T, \dots, x_{i,G}^T)^T$ .

### 2.2 结构组稀疏特征选择

传统的  $l_1$ -范数(也叫 LASSO)对高维数据可以有效地实现特征的稀疏性选择,但是 LASSO 是在整个特征空间中对特征逐一进行选择,没有考虑特征之间存在的组群结构特性,也没有考虑特征组群之间的相关性.为了解决这个问题,研究者们在线性回归模型中提出了基于组 LASSO(group LASSO)的特征选择方法<sup>[5]</sup>.

在图像标注中,给定一个标注语义  $Y_i$ ,我们的目的是要找到一个线性分类器  $Y_i = \text{sign}(w^T x + b)$  来预测训练集以

外非标注图像的标注结果.这里的  $w$  与特征向量  $x$  有相同的结构,即  $w = (w_1, w_2, \dots, w_G) \in \mathbb{R}^{m_1 + \dots + m_G}$ . 由于给定一种图像语义,一般仅有若干视觉特征对这一语义具有区别性,也就是说,从图像中所提取的部分异构特征对某一特定语义而言是冗余的(over-complete),因此,特征选择的目的是使  $w$  在组群上表现为稀疏,即使得大部分  $w_i$  为 0.

考虑到  $w$  的组群结构特性,这一目的可通过最小化正则化项  $\left( \sum_{g=1}^G d_g \|w_g\|_2 \right)^2$  实现.

用  $\xi$  表示损失函数  $f(x) = 1 - y_i \left( \sum_g w_g^T x_{i,g} + b \right)$  的上限,那么图像标注中线性分类模型可以表示为

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \left( \sum_{g=1}^G d_g \|w_g\|_2 \right)^2 + \lambda \sum_{i=1}^n \xi_i \\ \text{w.r.t.} \quad & w \in \mathbb{R}^{m_1} \times \dots \times \mathbb{R}^{m_G}, \xi \in \mathbb{R}_+^n, b \in \mathbb{R} \\ \text{s.t.} \quad & y_i \left( \sum_g w_g^T x_{i,g} + b \right) \geq 1 - \xi_i, \forall i \in \{1, \dots, n\} \end{aligned} \right\} \quad (2.1)$$

其中,  $\lambda$  是可调节的参数,用来控制损失函数的权重;  $d_g$  是可调节的权重系数,通常令  $d_g = \sqrt{m_g}$ .

通过观察我们知道,问题(2.1)的目标函数不是凸函数,即问题(2.1)不是标准的凸优化问题,很难用数学方法来求解.但是我们可将其转化为拉格朗日对偶方法来求解.现有文献[21]证明,拉格朗日对偶是一个凸优化问题,可以用凸优化的方法来求解.于是,我们把问题(2.1)转换成拉格朗日对偶问题求解.

**定义 1 (拉格朗日对偶(the Lagrange dual)).** 给定一个优化问题,包括目标函数和约束函数(等式或不等式),引入拉格朗日乘子(也叫对偶变量),构造一个拉格朗日函数,对所有原始变量求导并使其等于 0 即可得到一个对偶函数,我们把这个对偶函数称为拉格朗日对偶.

为了将问题(2.1)转换成拉格朗日对偶问题,我们将该问题看做一个二阶锥规划(second-order cone program, 简称 SOCP)问题,表示如下<sup>[12]</sup>:

$$\left. \begin{aligned} \min \quad & \frac{1}{2} u^2 + \lambda \sum_i \xi_i \\ \text{w.r.t.} \quad & u \in \mathbb{R}, t \in \mathbb{R}^G, b \in \mathbb{R}, \xi \in \mathbb{R}_+^n, (w_g, t_g) \in \kappa_{m_g}, \forall g \\ \text{s.t.} \quad & y_i \left( \sum_g w_g^T x_{i,g} + b \right) \geq 1 - \xi_i, \forall i \\ & \sum_g d_g t_g \leq u \end{aligned} \right\} \quad (2.2)$$

其中,  $u = \sum_{g=1}^G d_g \|w_g\|_2$ ,  $\kappa_{m_g} = \{(w_g, t_g) \in \mathbb{R}^{m_g+1}, \|w_g\|_2 \leq t_g\}$ , 表示  $m_g$  的二阶锥.我们对问题(2.2)构造拉格朗日函数.

引入拉格朗日参数  $\alpha, \gamma$ , 对所有变量求导,去除多余的参数,可化简得到如下拉格朗日对偶问题.那么,我们的目标为求解问题(2.3).

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \gamma^2 - \alpha^T e \\ \text{w.r.t.} \quad & \gamma \in \mathbb{R}, \alpha \in \mathbb{R}^n \\ \text{s.t.} \quad & 0 \leq \alpha \leq \lambda, \alpha^T y = 0 \\ & \left\| \sum_i \alpha_i y_i x_{i,g} \right\| \leq d_g \gamma, \forall g \in \{1, \dots, G\} \end{aligned} \right\} \quad (2.3)$$

### 2.3 多核学习

由于原始数据特征空间中数据是线性不可分的,于是,我们通过一个非线性映射函数  $\phi: X \rightarrow \mathbb{R}^f$  将原始数据空

间中的数据映射到一个新的特征空间,使得新的特征空间中数据线性可分.考虑到原始空间数据的组群结构特性,我们假设映射函数 $\phi(x)$ 由 $G$ 个不同的子函数组成 $\phi(x)=(\phi_1(x), \dots, \phi_G(x))$ ,那么原始空间数据被映射到 $G$ 个不同的再生核希尔伯特空间(reproducing kernel Hilbert space,简称 RKHS).

用核函数 $k(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \sum_{g=1}^G \phi_g(x_i)^T \phi_g(x_j) = \sum_{g=1}^G k_g(x_i, x_j)$ 表示 $\mathbb{R}^f$ 中数据的两两内积,使得非线性映射隐式地进行.通过这种映射,就得到 $G$ 个不同的核矩阵 $(K_1, \dots, K_G)$ .

通过核函数的引入,我们把问题(2.3)转换成如下多核学习问题<sup>[12]</sup>:

$$\left. \begin{aligned} \min \quad & \frac{1}{2} \gamma^2 - e^T \alpha \\ \text{w.r.t.} \quad & \gamma \in \mathbb{R}, \alpha \in \mathbb{R}^n \\ \text{s.t.} \quad & 0 \leq \alpha \leq \lambda, \alpha^T y = 0 \\ & (\alpha^T D(y) K_g D(y) \alpha)^{1/2} \leq \gamma d_g, \forall g \end{aligned} \right\} \quad (2.4)$$

这里, $D(y)$ 表示 $y$ 的对角矩阵; $K_g$ 表示点集 $\{x_i\}$ 对应第 $g$ 个核的核矩阵,核 $K_g$ 的稀疏性表现了问题(2.1)中 $w$ 的组群稀疏特性.核矩阵 $K$ 可以表示为 $K = \sum_g \eta_g K_g$ ,其中,非零 $\eta_g$ 对应的 $K_g$ ,我们称它为支持核.那么,核矩阵 $K$ 是支持核的一个稀疏组合.通过优化求解多核学习问题(2.4),我们可以选出核空间中最具差别性的非线性特征子集.

把问题(2.4)中的约束函数加入目标函数中,重排问题(2.4),我们得到如下最优化问题:

$$\left. \begin{aligned} \min \max \quad & \left\{ \frac{1}{2d_g^2} \alpha^T D(y) K_g D(y) \alpha - \alpha^T e \right\} \\ \text{w.r.t.} \quad & \alpha \in \mathbb{R}^n \\ \text{s.t.} \quad & 0 \leq \alpha \leq \lambda, \alpha^T y = 0 \end{aligned} \right\} \quad (2.5)$$

令 $J_g(\alpha) = \frac{1}{2d_g^2} \alpha^T D(y) K_g D(y) \alpha - \alpha^T e, J(\alpha) = \max_g J_g(\alpha)$ ,那么问题(2.5)可表达为在线性约束下最小化凸函数 $J(\alpha)$ .令 $\Gamma(\alpha)$ 表示主动核的集合,即使得 $J_g(\alpha) = J(\alpha)$ 的 $g$ 的集合.令 $F_g(\alpha) \in \mathbb{R}^n$ 表示 $J_g$ 的梯度,那么,

$$F_g = \frac{\partial J_g}{\partial \alpha} = \frac{1}{d_g^2} D(y) K_g D(y) \alpha - e.$$

为了求解问题(2.5),需要引入优化条件.在这之前,我们先要给出以下定义.

**定义 2(次微分(subdifferential)).** 给定任意函数 $J(\alpha)$ ,定义 $J$ 在 $\alpha$ 处的次微分 $\partial J(\alpha)$ 为

$$\partial J(\alpha) = \{g \in \mathbb{R}^n, \forall \alpha', J(\alpha') \geq J(\alpha) + g^T (\alpha' - \alpha)\}.$$

我们对问题(2.5)构造拉格朗日函数

$$L(\alpha) = J(\alpha) - \delta^T \alpha + \xi^T (\alpha - \lambda e) + b \alpha^T y.$$

这里, $b \in \mathbb{R}^n, \xi, \delta \in \mathbb{R}_+^n$ .对应 $\alpha$ 的 $L(\alpha, \delta, \xi, b)$ 全局最小化通过下式得到:

$$0 \in \partial L(\alpha) \Leftrightarrow \delta - \xi - b y \in \partial J(\alpha).$$

那么, $\alpha, (b, \delta, \xi)$ 是优化解当且仅当

$$\left. \begin{aligned} \delta - \xi - b y & \in \partial J(\alpha) \\ \forall i, \delta_i \alpha_i & = 0, \xi_i (\lambda - \alpha_i) = 0 \\ \alpha^T y & = 0, 0 \leq \alpha \leq \lambda \end{aligned} \right\} \quad (2.6)$$

公式(2.6)中的优化条件很难用数学方法直接求解,因此我们引入一个新的定义近似求解这个优化问题.

**定义 3( $\varepsilon$ -次微分( $\varepsilon$ -subdifferential)).** 给定任意函数 $J(\alpha)$ ,定义 $J$ 在 $\alpha$ 处的 $\varepsilon$ -次微分 $\partial_\varepsilon J(\alpha)$ 为

$$\partial_\varepsilon J(\alpha) = \{g \in \mathbb{R}^n, \forall \alpha', J(\alpha') \geq J(\alpha) - \varepsilon + g^T (\alpha' - \alpha)\}.$$

令 $(\varepsilon_1, \varepsilon_2)$ 表示优化参数,优化条件(2.6)可以转换成如下线性规划(LP)存在问题.即,找到 $\eta$ 满足

$$\left. \begin{aligned} \eta \geq 0, \eta_g = 0 \text{ if } g \notin \Gamma_{\varepsilon_1}(\alpha), \sum_g d_g^2 \eta_g = 1 \\ \max_{i \in I_M \cup I_{0-} \cup I_{\lambda+}} \{(K(\eta)D(y)\alpha)_i - y_i\} \leq \min_{i \in I_M \cup I_{0+} \cup I_{\lambda-}} \{(K(\eta)D(y)\alpha)_i - y_i\} + 2\varepsilon_2 \end{aligned} \right\} \quad (2.7)$$

其中,  $K(\eta) = \sum_{g \in \Gamma_{\varepsilon_1}(\alpha)} \eta_g K_g$ ,  $I_M = \{i, 0 < \alpha_i < \lambda\}$ ,  $I_{0-} = \{i, \alpha_i = 0\}$ ,  $I_{\lambda}(\alpha) = \{i, \alpha_i = \lambda\}$ ,  $I_{0+} = I_0 \cap \{i, y_i = 1\}$ ,  $I_{0-} = I_0 \cap \{i, y_i = -1\}$ ,  $I_{\lambda+} = I_{\lambda} \cap \{i, y_i = 1\}$ ,  $I_{\lambda-} = I_{\lambda} \cap \{i, y_i = -1\}$ . 给定一个 $\alpha$ 以及优化参数 $(\varepsilon_1, \varepsilon_2)$ ,我们可以通过优化条件(2.7)来迭代求解 $\eta$ .

### 2.4 增强稀疏性

通过近似优化条件(2.7)得到优化解 $(\alpha, \eta)$ 后,我们考虑 $\eta$ 是否可以更加稀疏.当我们在对优化条件(2.7)迭代求解的过程中,对所有非零的 $\eta_g$ 进行循环,检查每个这样的 $\eta_g$ 可否被移除.也就是说,对所有满足 $g \in \Gamma_{\varepsilon_1}(\alpha)$ 的 $g$ ,令 $\eta_g = 0$ ,然后看 $\eta$ 是否仍然满足优化条件(2.7).如果满足,那么第 $g$ 个核可以被移除.通过这种迭代方式将权重系数 $\eta$ 变的更加稀疏,以此来突显数据的组群结构特性.

### 2.5 算法流程

至此,我们得到了特征选择的权重系数 $\eta$ 以及测试样本与训练样本的核矩阵 $K_1, \dots, K_G$ ,可计算得到多核矩阵 $K(\eta) = \sum_{g \in \Gamma_{\varepsilon_1}(\alpha)} \eta_g K_g$ ,通过计算 $\text{sign}(K(\eta)D(y)\alpha + b)$ 即可得到测试样本的标注信息.这里, $D(y)$ 表示训练样本标注的对角矩阵, $b$ 表示偏差.算法1给出了基于组稀疏的多核学习图像标注算法流程.

#### 算法 1.

1. 输入  $X \in \mathbb{R}^{(n_{tr} + n_{te}) \times p}$ ,  $Y \in \mathbb{R}^{n_{tr} \times c}$  ( $n_{tr}$  表示训练样本数目,  $n_{te}$  表示测试样本数目);
2. 计算  $X$  的核矩阵  $K_1, \dots, K_G$ ;
3. 初始化参数 $\alpha$ 和 $(\varepsilon_1, \varepsilon_2)$ ;
4. 对每个标注  $c$  迭代求解满足优化条件(2.7)的 $\eta$ 值;
5. 令 $\eta_g = 0$ ,其中,  $g \in \Gamma_{\varepsilon_1}(\alpha)$ .若此时 $\eta$ 仍满足(2.7),则 $\eta_g = 0$ ;否则,保留原来的 $\eta$ ;
6. 根据步骤2中测试样本与训练样本的核矩阵计算多核矩阵  $K(\eta) = \sum_{g \in \Gamma_{\varepsilon_1}(\alpha)} \eta_g K_g$ ;
7. 计算  $\text{sign}(K(\eta)D(y_{tr})\alpha + b)$ 得到测试样本的标注.

## 3 实验分析

本节给出 MKLGS 算法与其他算法在图像标注上的性能对比结果.

### 3.1 实验数据及评价标准

在实验中使用的数据集为:Kodak<sup>[22]</sup>,MSRA-MM 2.0<sup>[23]</sup>,NUS-WIDE<sup>[24]</sup>.表1给出了实验用到的3个数据集的详细信息.

Table 1 Datasets description

表 1 实验数据集详细信息

数据集	大小	特征维数	标注单词数目	特征类别数(组)
Kodak	3 590	409	20	4
MSRA-MM 2.0	30 000	409	100	4
NUS-WIDE	20 682	409	81	4

对每一幅图像提取异构特征,按照特征属性不同将这些异构特征分为不同组群.3个数据集特征描述如下:

- Kodak<sup>[22]</sup>:409 维特征向量,包括 4 种特征,分别为 128 维小波纹理特征、73 维边缘方向直方图、144 维颜色相关图以及 64 维 HSV 颜色直方图,本文将这 409 维特征分为 4 组;
- MSRA-MM 2.0<sup>[23]</sup>:411 维特征向量,包括 4 种特征,分别为 128 维小波纹理特征、75 维边缘方向直方图、

144 维颜色相关图以及 64 维 HSV 颜色直方图,本文将这 411 维特征分为 4 组;

- NUS-WIDE<sup>[24]</sup>:409 维特征向量,包括 4 种特征,分别为 128 维小波纹理特征、73 维边缘方向直方图、144 维颜色相关图以及 64 维颜色直方图,本文将这 411 维特征分为 4 组.

为了评估图像标注算法性能,本文采用了 ROC 曲线下面积(area under curve,简称 AUC)<sup>[25]</sup>,包括 MicroAUC 和 MacroAUC 以及准确率(precision)这 3 种评价指标.对每一个标注建立分类器,可得到一个分类准确率(分类准确的数据除以分类的数据),取多个标注的准确率平均值得到准确率(precision)的值.AUC 值越高代表标注性能越好,Precision 值越高表示标注准确率越高.

### 3.2 实验步骤

首先,我们将原始空间中的数据集(包括训练集和测试集)通过函数  $\phi(x)=(\phi_1(x),\dots,\phi_G(x))$  映射到新的希尔伯特空间,那么对应每一个组群有不同的映射函数.实验中选择最常用的高斯核函数作为映射函数,即

$$k_g(x_i, x_j) = e^{-\frac{(x_i - x_j)^2}{2\sigma}}.$$

对每个特征组群设定 12 个不同的  $\sigma$  值,即 12 个不同的核函数,那么对应 4 种特征向量,可以建立 48 个核矩阵.给定每一个标注  $c$ ,通过在训练集上运用基于组群稀疏的多核学习算法,可得到核矩阵  $K = \sum_g \eta_g K_g$  的权重系数  $\eta_g$ .根据  $\eta_g$  的稀疏性可以对图像特征进行选择,除去使  $\eta_g$  为 0 的第  $g$  类特征.此时,利用训练数据与测试数据的核矩阵及特征选择结果,计算  $y = \text{sign}(w^T x + b)$  的值.如果  $y=1$ ,那么给该测试图像标注  $c$ ;否则,该图像没有标注  $c$ .

我们随机分别选取了 100、200 和 300 个带标注的图像作为训练集(保证每个标注至少出现一次),在剩余的数据中分别选取 400、800 和 1 200 个未标注图像作为测试集,每组实验重复 10 次,然后算出 10 次结果的平均值和标准方差.

算法中需要初始化的参数有正则化参数  $\lambda$  和松弛变量  $(\epsilon_1, \epsilon_2)$ ,这些参数通过在 MacroAUC 评价指标上 5 次交叉验证得到其最优值作为最终参数值.

### 3.3 实验结果与分析

基于组群稀疏的多核学习算法利用了希尔伯特空间中核函数和数据特征的组群特性,本节讨论该算法与其他主流算法在图像多标注上的性能.

传统的多标注算法没有考虑标注之间的相关性,忽略了数据本身包含的特性.考虑到标注之间的相关性,一些算法利用了典型相关分析(CCA)或者共享空间来提高标注的准确率<sup>[16,17,26]</sup>.从文献[26]中可以看到,典型相关分析最大化数据集之间的协方差,因此可以通过最大化图像标注间的协方差来利用标注间固有的相关性,从而提高标注准确率.Ji 等人<sup>[16]</sup>提出了一种在多标注算法中提取共享子空间的模型.由于多标注总是共享相同的输入空间,而不同的标注对应的语义之间也有一定的相关性,因此在这个模型中,设定一个子空间是多个标注共享的,利用线性转化将子空间引入到算法模型中,利用不同标注对应语义之间的相关性,提高标注准确率.文献[27]将降维与多标注分类算法融合在一起,将降维的投影矩阵与分类算法(如 SVM)融合在一个模型中,通过该模型同时求解降维投影矩阵与分类回归系数.

本文比较下面 5 种多标注算法,包括基于组稀疏的多核学习算法(MKLGS)、典型相关分析以及 SVM(CCA+SVM)、典型相关分析以及岭回归(CCA+RR)、基于联合降维的多标注分类(joint dimensionality reduction for multi-label classification,简称 JDRMC)、基于共享结构的多标注分类(shared subspace for multi-label classification,简称 SSMC).在 MacroAUC 和 MicroAUC 评价指标下的实验结果见表 2、表 3.图 2 展示了在准确率评价指标下,Kodak 和 NUS-WIDE 数据集上的实验结果(左图为 Kodak 数据集上,训练数据分别为 100,200 和 300 时的精确率实验结果;右图为 NUS-WIDE 数据集上,训练数据分别为 100,200 和 300 时的精确率实验结果).

从实验结果中可以看到,在 Kodak 和 MSRA-MM 2.0 数据库上,将降维与多标注分类算法融合的 JDRMC 算法效果优于传统的先降维后分类算法,如 CCA+SVM 以及 CCA+RR,是由于传统的算法并没有考虑到降维与分

类之间的相关性,而 JDRMC 算法很好地将两者融合在一个模型中.

从实验结果中还可以看到,基于共享结构的多标注分类算法(SSMC)因为考虑到了多标注对应语义之间的相关性,其标注效果优于其他的多标注算法,如 CCA+SVM,CCA+RR,JDRMC.

而本文提出的基于组稀疏的多核学习算法(MKLGs)利用了希尔伯特空间核函数以及特征组群特性,其标注效果明显优于其他 4 种多标注算法.

**Table 2** Performance of the 5 compared methods in terms of MacroAUC

**表 2** 5 种多标注算法的 MacroAUC 比较

A. 训练数据和测试数据分别为 100 和 400 时的实验结果

	Kodak	MSRA-MM 2.0	NUS-WIDE
MKLGs	<b>0.6987±0.0012</b>	<b>0.6560±0.0034</b>	<b>0.6833±0.0013</b>
CCA+SVM	0.6693±0.0135	0.5775±0.0067	0.6364±0.0117
CCA+RR	0.6675±0.0128	0.5770±0.0183	0.6327±0.0112
JDRMC	0.6759±0.0076	0.5818±0.0121	0.6280±0.0057
SSMC	0.6894±0.0021	0.6367±0.0038	0.6716±0.0045

B. 训练数据和测试数据分别为 200 和 800 时的实验结果

	Kodak	MSRA-MM 2.0	NUS-WIDE
MKLGs	<b>0.7789±0.0022</b>	<b>0.6892±0.0038</b>	<b>0.7072±0.0018</b>
CCA+SVM	0.7016±0.0033	0.5821±0.0045	0.6584±0.0011
CCA+RR	0.7000±0.0124	0.5820±0.0087	0.6555±0.0101
JDRMC	0.7075±0.0102	0.6117±0.0127	0.6523±0.0092
SSMC	0.7710±0.0084	0.6727±0.0179	0.6956±0.0131

C. 训练数据和测试数据分别为 300 和 1 200 时的实验结果

	Kodak	MSRA-MM 2.0	NUS-WIDE
MKLGs	<b>0.8098±0.0038</b>	<b>0.6965±0.0029</b>	<b>0.7087±0.0037</b>
CCA+SVM	0.7207±0.0058	0.5996±0.0033	0.6812±0.0061
CCA+RR	0.7007±0.0161	0.5876±0.0076	0.6701±0.0137
JDRMC	0.7153±0.0092	0.6164±0.0068	0.6862±0.0157
SSMC	0.8045±0.0108	0.6733±0.0196	0.7085±0.0085

**Table 3** Performance of the 5 compared methods in terms of MicroAUC

**表 3** 5 种多标注算法的 MicroAUC 比较

A. 训练数据和测试数据分别为 100 和 400 时的实验结果

	Kodak	MSRA-MM 2.0	NUS-WIDE
MKLGs	<b>0.6930±0.0032</b>	<b>0.6932±0.0043</b>	<b>0.6993±0.0031</b>
CCA+SVM	0.6589±0.0168	0.6092±0.0032	0.6716±0.0081
CCA+RR	0.6476±0.0106	0.6061±0.0076	0.6630±0.0051
JDRMC	0.6771±0.0057	0.6063±0.0077	0.6582±0.0162
SSMC	0.6809±0.0184	0.6651±0.0162	0.6861±0.0138

B. 训练数据和测试数据分别为 200 和 800 时的实验结果

	Kodak	MSRA-MM 2.0	NUS-WIDE
MKLGs	<b>0.7978±0.0021</b>	<b>0.7264±0.0039</b>	<b>0.7312±0.0052</b>
CCA+SVM	0.6999±0.0063	0.5788±0.0033	0.7074±0.0083
CCA+RR	0.6967±0.0176	0.5723±0.0183	0.6985±0.0111
JDRMC	0.7201±0.0182	0.6349±0.0126	0.6756±0.0132
SSMC	0.7885±0.0065	0.6976±0.0062	0.7266±0.0058

C. 训练数据和测试数据分别为 300 和 1 200 时的实验结果

	Kodak	MSRA-MM 2.0	NUS-WIDE
MKLGs	<b>0.8098±0.0056</b>	<b>0.7409±0.0028</b>	<b>0.7349±0.0076</b>
CCA+SVM	0.7077±0.0135	0.6332±0.0053	0.7130±0.0083
CCA+RR	0.7063±0.0087	0.6232±0.0064	0.7012±0.0163
JDRMC	0.7236±0.0161	0.6400±0.0187	0.6758±0.0035
SSMC	0.8012±0.0102	0.7025±0.0131	0.7289±0.0073

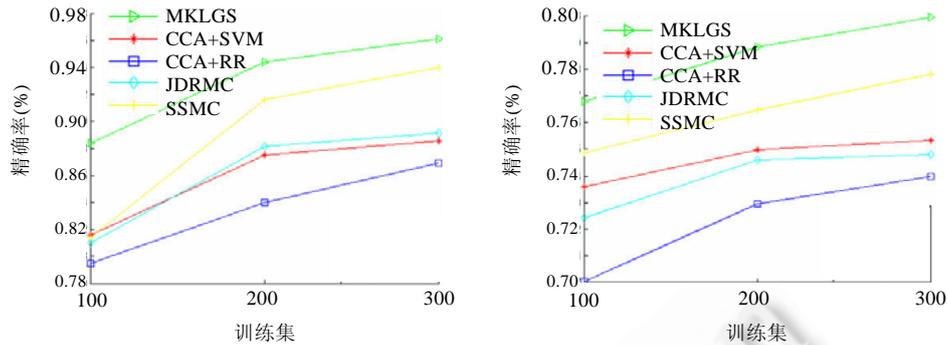


Fig.2 Performance of the 5 compared methods in terms of precision

图2 MKLGS,CCA+SVM,CCA+RR,JDRMC,SSMC 多标注算法的准确率比较

#### 4 结论

本文介绍了一种基于组稀疏的多核学习图像标注算法,该方法构造了一个希尔伯特空间,利用该空间中的核函数以及 group LASSO 正则化因子对图像特征进行选择,进而对图像进行有效地标注.本文算法充分利用了图像数据特征分布在非线性空间以及特征的组群性质,通过在 NUS-WIDE,MSRA-MM2.0 和 Kodak 等数据集上进行对比实验,结果表明了该算法的有效性.如何更加有效地对图像特征进行选择,使得图像标注更加准确,仍是今后的研究热点.如,文献[28]提出了一种同时在特征组间以及组内进行选择的机制,在图像标注实验上取得了较好的结果.同时,如何利用未标注图像数据以及如何扩展到大规模数据集,也是本文今后的研究重点.

#### References:

- [1] Yang J, Yu K, Gong Y, Huang T. Linear spatial pyramid matching using sparse coding for image classification. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. Florida: IEEE Computer Society Press, 2009. 1794–1801. [doi: 10.1109/CVPRW.2009.5206757]
- [2] Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(2):210–227. [doi: 10.1109/TPAMI.2008.79]
- [3] Yang JC, Wright J, Huang T, Ma Y. Image super-resolution as sparse representation of raw image patches. In: Proc. of the Conf. on Computer Vision and Pattern Recognition. Anchorage: IEEE Computer Society Press, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587647]
- [4] Yang JC, Tang H, Ma Y, Huang T. Face hallucination via sparse coding. In: Proc. of the Int'l Conf. on Image Processing. San Diego: IEEE Signal Processing Society Press, 2008. 1264–1267. [doi: 10.1109/ICIP.2008.4711992]
- [5] Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society—Series B (Methodological), 2006,68(1):49–67. [doi: 10.1111/j.1467-9868.2005.00532.x]
- [6] Cao L, Luo J, Liang F, Huang T. Heterogeneous feature machines for visual recognition. In: Proc. of the 2009 IEEE Int'l Conf. on Computer Vision. 2009. 1095–1102. [doi: 10.1109%2FICCV.2009.5459401]
- [7] Wu F, Yuan Y, Zhuang YT. Heterogeneous feature selection by group LASSO with logistic regression. In: Proc. of the 2010 Int'l Conf. on Multimedia. Firenze: ACM Press, 2010. 983–986. [doi: 10.1145/1873951.1874129]
- [8] Meier L, Van de Geer S, Bühlmann P. The group LASSO for logistic regression. Journal of the Royal Statistical Society—Series B (Methodological), 2008,70(1):53–71. [doi: 10.1111/j.1467-9868.2007.00627.x]
- [9] Zhang S, Huang J, Huang Y, Yu Y, Li H, Metaxas D. Automatic image annotation using group sparsity. In: Proc. of the 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. San Francisco: IEEE Computer Society Press, 2010. 3312–3319. [doi: 10.1109/CVPR.2010.5540036]
- [10] Wu F, Han YH, Tian Q, Zhuang YT. Multi-Label boosting for image annotation by structural grouping sparsity. In: Proc. of the 2010 Int'l Conf. on Multimedia. Firenze: ACM Press, 2010. 15–24. [doi: 10.1145/1873951.1873957]
- [11] Lanckriet GRG, Cristianini N, Bartlett P, Ghaoui LE, Jordan MI. Learning the kernel matrix with semi definite programming. The Journal of Machine Learning Research, 2004,5:27–72.

- [12] Bach FR, Lanckriet GRG, Jordan MI. Multiple kernel learning, conic duality, and the SMO algorithm. In: Proc. of the 2004 Int'l Conf. on Machine Learning. Banff: ACM Press, 2004. 41–48. [doi: 10.1145/1015330.1015424]
- [13] Zien A, Ong CS. Multiclass multiple kernel learning. In: Proc. of the 2007 Int'l Conf. on Machine Learning. New York: ACM Press, 2007. 1191–1198. [doi: 10.1145/1273496.1273646]
- [14] McCallum AK. Multi-Label text classification with a mixture model trained by EM. In: Proc. of the AAAI'99 Workshop on Text Learning.
- [15] Ueda N, Saito K. Parametric mixture models for multi-labeled text. In: Proc. of the Advances in Neural Information Processing Systems, Vol.15. 2002. 721–728.
- [16] Ji S, Tang L, Yu S, Ye J. Extracting shared subspace for multi-label classification. In: Proc. of the Int'l Conf. on Knowledge Discovery and Data Mining. Las Vegas: ACM Press, 2008. 381–389. [doi: 10.1145/1401890.1401939]
- [17] Liu YN, Wu F, Zhang ZH, Zhuang YT, Yan SC. Sparse representation using nonnegative curds and whey. In: Proc. of the 2010 IEEE Conf. on Computer Vision and Pattern Recognition. San Francisco: IEEE Computer Society Press, 2010. 3578–3585. [doi: 10.1109/CVPR.2010.5539934]
- [18] Han YH, Wu F, Jia JZ, Zhuang YT, Yu B. Multi-Task sparse discriminant analysis (MtSDA) with overlapping categories. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. Atlanta: AAAI, 2010. 469–474.
- [19] Elisseeff A, Weston J. A kernel method for multi-labelled classification. In: Proc. of the Advances in Neural Information Processing Systems, Vol.14. 2002. 681–687.
- [20] Arenas-García J, Petersen KB, Hansen LK. Sparse kernel orthonormalized PLS for feature extraction in large data sets. In: Proc. of the Advances in Neural Information Processing Systems, Vol.19. 2007. 33–40.
- [21] Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004.
- [22] Loui A, Luo J, Chang S, Ellis D, Jiang W, Kennedy L, Lee K, Yanagawa A. Kodak's consumer video benchmark data set: Concept definition and annotation. In: Proc. of the Int'l Workshop on Multimedia Information Retrieval. Augsburg: ACM Press, 2007. 245–254. [doi: 10.1145/1290082.1290117]
- [23] Li H, Wang M, Hua X. MSRA-MM 2.0: A large-scale Web multimedia dataset. In: Proc. of the 2009 IEEE Int'l Conf. on Data Mining Workshops. Miami: IEEE, 2009. 164–169. [doi: 10.1109/ICDMW.2009.46]
- [24] Chua T, Tang J, Hong R, Li H, Luo Z, Zheng Y. Nus-Wide: A real-world Web image database from national university of singapore. In: Proc. of the Int'l Conf. on Image and Video Retrieval. Santorini: ACM Press, 2009. 1–9. [doi: 10.1145/1646396.1646452]
- [25] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters, 2006,27(8):861–874. [doi: 10.1016/j.patrec.2005.10.010]
- [26] Hardoon D, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. Neural Computation, 2004,16(12):2639–2664. [doi:10.1162/0899766042321814]
- [27] Ji S, Ye J. Linear dimensionality reduction for multi-label classification. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence (IJCAI). Morgan Kaufmann Publishers, 2009. 1077–1082.
- [28] Yuan Y, Wu F, Zhuang YT, Shao J. Image annotation by composite kernel learning with group structure. In: Proc. of the Int'l Conf. on Multimedia. Scottsdale: ACM Press, 2011. [doi: 10.1145/2072298.2072049]



袁莹(1986—),女,浙江诸暨人,博士生,主要研究领域为跨媒体检索,稀疏表达.



吴飞(1973—),男,博士,教授,CCF 高级会员,主要研究领域为跨媒体分析与检索,稀疏表达,机器学习理论.



邵健(1982—),男,博士,讲师,主要研究领域为跨媒体检索,海量数据管理.



庄越挺(1965—),男,博士,教授,博士生导师,主要研究领域为网络多媒体内容分析与检索,计算机动画,数字图书馆,智能CAD.