一种基于稀疏编码的多核学习图像分类方法

亓晓振,王 庆
(西北工业大学计算机学院,陕西西安 710072)

摘 要: 本文提出一种基于稀疏编码的多核学习图像分类方法.传统稀疏编码方法对图像进行分类时,损失了空间信息,本文采用对图像进行空间金字塔多划分方式为特征加入空间信息限制.在利用非线性SVM方法进行图像分类时,空间金字塔的各层分别形成一个核矩阵,本文使用多核学习方法求解各个核矩阵的权重,通过核矩阵的线性组合来获取能够对整个分类集区分能力最强的核矩阵.实验结果表明了本文所提出图像分类方法的有效性和鲁棒性.对Scene Categories场景数据集可以达到83.10%的分类准确率,这是当前该数据集上能达到的最高分类准确率.

关键词: 图像分类;多核学习;稀疏编码;空间金字塔
中图分类号: TP319.7 **文献标识码:** A **文章编号:** 0372-2112 (2012) 04-0773-07
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2012.04.025

An Image Classification Approach Based on Sparse Coding and Multiple Kernel Learning

QI Xiao-zhen, WANG Qing
(*School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi* 710072, *China*)

Abstract: A novel image classification method based on sparse coding and multiple kernel learning is proposed in the paper. Traditional methods of image classification used common sparse coding but lose the spatial information. We add this spatial information by dividing the image with the spatial pyramid. With the nonlinear SVM for image classification, each level of spatial pyramid has its own kernel, and we adopt machine learning for the optimal trade-off between different kernels. A much more discriminative kernel can be seen as the linear combination of base kernels corresponding to different pyramid levels. The experiments on the benchmark dataset show the effectiveness and robustness of our method. The precision on scene categories dataset can reach 83.10%, and it is the best result comparing to the state-of-the-art work.

Key words: image classification; multiple kernel learning (MKL); sparse coding; spatial pyramid

1 引言

图像分类是根据图像具有的某种属性将其划分到预先设定的不同类别中[10].对于人类而言,图像分类不是难事,但是使用计算机对图像进行分类存在一系列问题:(1)图像内包含着大量信息,这些信息具备复杂多样性和不可描述性;(2)图像的物理表达和人类熟知的概念性信息之间差异巨大.

近年来图像分类技术得到迅速发展.Nister[1]等提出了一种基于词典树(VT, Vocabulary Tree)的图像特征表示方法并用于目标识别和图像检索.这种方法通过k-means算法对图像SIFT特征聚类生成词典,每个聚类中心构成一个视词,进而将图像量化为视词直方图.刘硕研[14]等提出一种基于上下文语义信息的图像块视词生成方法,在一定程度上提高了视词的区分性,但是该方法与传统词典树方式一样存在着特征空间信息丢失的问题.鉴于此,Lazebnik[2]等提出了一种空间金字塔匹配方法用于自然图像的分类与识别.这种方法对图像在空间上进行不同级别的划分,将划分的每个图像块分别量化为视词直方图.对不同图像进行处理时,计算对应空间金字塔层次间的相似度,通过对不同层次间的相似度进行加权求和,得到总的图像相似度.Varma[3]等提出了一种衡量不同图像特征重要性的多核学习(MKL, Multiple Kernel Learning)方法.这种方法使用不同图像低层特征组合进行图像分类,通过机器学习方法得到不同特征的权值,利用对核矩阵加权求和得到总的图像分类

收稿日期:2010-10-19;修回日期:2011-12-08
基金项目:国家自然科学基金(No.60873085,No.61103060);博士后基金(No.20110491691)

核矩阵,最后利用其进行图像分类.

作为压缩感知技术的一个主要代表,稀疏编码是使用一些已训练的能够表示低层特征的基向量的线性组合来表示输入图像特征[4,12].Yang[5]等使用一种稀疏编码方法对图像特征进行表示,用稀疏向量来表示图像特征,取得了较好的分类效果.该方法使用了一种空间金字塔划分方法进行稀疏编码,它首先训练用于稀疏编码的基向量,再利用得到基向量和对图像进行空间划分方法得到图像的特征表示.在使用空间金字塔过程中,该方法直接将不同金字塔层次得到向量直接相连构成一个向量,用一个向量来表示一幅图像.

空间金字塔在本质上对图像在空间上进行划分,空间划分方式是多样的.通常在图像的两个坐标方向进行 2 的指数倍划分,本文采用此种方式.记 $l=0,1,\cdots,L$ 为金字塔层次,金字塔总层数为 $L+1$.当 $l=0$,表示金字塔结构的第 0 层,在水平方向把图像分成 $2^l=2^0=1$ 块,在竖直方向将图像划分成 $2^l=2^0=1$ 块,其最终将把图像划分成 $2^l\times 2^l=4^l=1$ 块.故从 0 层开始,图像划分块数分别为 $4^0,4^1,4^2$ 等块.对特征进行空间金字塔划分方式如图 1 所示,其中的 3 种符号分别代表 3 种视词,最下面一行表示的是不同划分下各个子区域中 3 种视词的统计直方图.这些直方图将在后面的特征表示中被进一步处理.

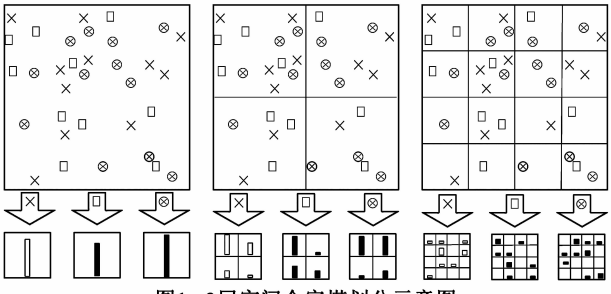

图1 3层空间金字塔划分示意图

2 本文方法

传统的词典树方法[1,2]对图像特征进行表示和分类时,本质上是对图像中每一个待量化特征,选取离其最近邻的一个视词作为它的特征表示,即在视词固定的情况下,使用只含有一个非零元素的向量来表示一个特征.这种方法对视词间分界线附近的特征量化误差较大,并且该量化误差是无法避免的.使用稀疏编码方式可以尽量减少量化误差,它可以通过解凸优化问题来求取待量化特征关于基向量的稀疏表示,使用多个向量的线性组合来表示一个特征.在理论上,使用稀疏编码方法较词典树方法在图像特征表示方面具备优势.使用 Yang[5] 的稀疏编码方法进行图像分类时也存在一些问题:单纯使用单一向量来表示一幅图像时,该单一向量维数会非常大,在计算时会造成不便;其次,使用各层次向量直接相连的方法,虽然依然可以对各层次向量进行加权,但是这些权值都是经验获得的.这种直接利用经验值进行加权的方法,不能保证它的核矩阵对图像集具有最强的区分能力.

本文提出了一种基于稀疏编码的多核学习图像分类方法.通过在空间上对图像进行金字塔划分,对每个金字塔层次分别进行稀疏编码,并将该层次特征转化一个向量表示.在本文中,使用多个稀疏向量来共同表示一幅图像.在采用核方法的 SVM 进行图像分类时,需要求取每幅图像的对应空间金字塔层次上的核矩阵,空间金字塔中每个层次分别对应一个核矩阵.使用机器学习方法求解各个核矩阵的权值,再利用这些权值对核矩阵进行加权求和得到区分能力最强的核矩阵.通过解决凸优化问题,我们可以保证加权线性求和后的核矩阵是最优的.通过在 2 个标准数据集上的实验,结果证明了本方法的有效性和鲁棒性.同时,本文在仅使用一种特征描述子的条件下,对 Scene Categories[6] 数据集的分类准确率可以达到 83.10%,这是目前该数据集上能达到的最高分类准确率.

2.1 多核学习

图像分类的目标是求解函数 $f(x)=w^t\phi_l(x)+b$,其中 $K_l(x_i,x_j)=\phi_l^t(x_i)\phi_l(x_j)$ 是第 l 空间金字塔的特征向量 x_i 和 x_j 映射到 ϕ 特征空间后向量的点积.SVM 分类方法是根据已有的训练数据集 $\{(x_i,y_i)\}$ 来学习得到最优的 w 和 b.由于待学习的分类器具备多个区分能力不同的核矩阵,故需要在求解传统 SVM 优化问题的基础上考虑加入多个核矩阵的线性组合系数 d_l,进而得到最终分类器的核矩阵 $\boldsymbol{K}=\sum_{l=0}^{L} d_lK_l$.如何求解最优的核矩阵线性组合,使得其对目标分类能力最强是我们要解决的一个问题.求解核矩阵 $\boldsymbol{K}$ 的最优值问题,可在 SVM 分类器学习框架下,利用训练集学习核矩阵系数 d,使得其对测试集能够达到最大的分类效率.定义如下代价函数:

$$T(d)=\min_{w,b}\frac{1}{2}w^tw+C1^t\xi+\sigma^td \tag{1}$$

$$\text{s.t. } y_i(w^t\phi(x_i)+b)\geqslant 1-\xi_i,\ \xi\geqslant 0,\ d\geqslant 0,\ Ad\geqslant p$$

式(1)与 l_1 C-SVM 相类似,C 为误分类惩罚常量,$Ad\geqslant p$ 是将已知先验信息作为限制条件.核矩阵可以改写为

$$\boldsymbol{K}(x_i,x_j)=\phi^t(x_i)\phi(x_j)=\sum_{l=0}^{L} d_l\phi_l^t(x_i)\phi_l(x_j).$$

求解最小 $\boldsymbol{T}$ 的策略是采用投影梯度下降法,即 $d^{n+1}=d^n-\varepsilon^n\nabla\boldsymbol{T}$,在求解过程中要始终满足 $d^{n+1}\geqslant 0$,$Ad^{n+1}\geqslant p$.利用投影梯度下降法求解上述优化问题时,最关键的一步是计算 $\nabla\boldsymbol{T}$.上述优化问题的对偶式是:

$$\boldsymbol{W}(d)=\max_{\alpha}1^{t}\alpha+\sigma^{t}d-\frac{1}{2}\sum_{l=0}^{L}d_l\alpha^{t}\boldsymbol{Y}\boldsymbol{K}_l\boldsymbol{Y}\alpha \tag{2}$$

$$\text{s.t. } 0\leqslant\alpha\leqslant C,1^{t}\boldsymbol{Y}\alpha=0$$

其中非零变量 α 与支撑向量相关,$\boldsymbol{Y}$ 是标记训练数据类别的对角矩阵,K_l 是 $\boldsymbol{K}$ 的第 l 列.

根据强对偶原理,有 $\boldsymbol{T}(d)=\boldsymbol{W}(d)$.由于式(2)中采用的所有核矩阵都是严格可微的,记 $\boldsymbol{W}$ 取最大值时的 α 值为 α^{*},则 α^{*} 是唯一的,并且在 $\alpha=\alpha^{*}$ 时 $\boldsymbol{W}$ 是可微的[7].由文献[8]的引理 2 可得,如果 α^{*} 的值不依赖于 d 值,则对于不同的 d,$\boldsymbol{W}$ 是有差别的.故有下述等式成立,即

$$\frac{\partial\boldsymbol{T}}{\partial d_l}=\frac{\partial\boldsymbol{W}}{\partial d_l}=\sigma_l-\frac{1}{2}\alpha^{*}\,\boldsymbol{Y}\boldsymbol{K}_l\boldsymbol{Y}\alpha^{*} \tag{3}$$

多核学习的详细流程参见如下的算法 1.

算法 1

1: $n\leftarrow 0$

2:随机生成 $L+1$ 个数构成 d^0,并使得 $d^0=random(L+1)$,$d^0\in[0,1)$

3:Repeat

4: $\boldsymbol{K}=\sum_{l=0}^{L}d_lK_l$

5: 根据核矩阵 $\boldsymbol{K}$ 和训练数据,通过解决二次方程凸优化问题来求解 α^{*}

6: $d^{n+1}\leftarrow d^{n}-\varepsilon^{n}(\sigma_l-\frac{1}{2}\alpha^{*}\,\boldsymbol{Y}\boldsymbol{K}_l\boldsymbol{Y}\alpha^{*})$ (4)

7: 如果 d^{n+1}满足 $d^{n+1}\geqslant 0$ 和 $Ad^{n+1}\geqslant p$,则把 d^{n+1}设为可行解.

8: $n\leftarrow n+1$

9:Until 满足收敛条件或者超过最大迭代次数.

当求出 d 后,再根据训练数据进行学习,得到分类器的各项参数.对一个未知分类的输入特征,根据 $g(x)=\text{sign}(\sum_i\alpha_i y_iK(x,x_i)+b)$ 的结果可将其划分为相应的 ±1 类别.

2.2 稀疏编码

稀疏编码是一种特征表示方法,它的目标是通过求解少量未知的能代表低层特征信息基向量的系数,使用这些系数将基向量进行线性组合来表示输入特征.稀疏编码需要解决两个问题:(1)未知基向量的求解问题;(2)对输入向量,求解其关于基向量线性组合的系数问题.

为表述方便,本文使用 $\boldsymbol{X}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_M]^{\mathrm{T}}\in\mathbb{R}^{M\times D}$ 表示输入向量集合,其中 $\boldsymbol{x}_i\in\mathbb{R}^{1\times D}$,$i=1,2,\cdots,M$ 代表第 i 个输入向量,D 是输入向量的维数.相应地,使用 $\boldsymbol{B}=[\boldsymbol{b}_1,\boldsymbol{b}_2,\cdots,\boldsymbol{b}_N]^{\mathrm{T}}\in\mathbb{R}^{N\times D}$ 来表示基向量集合,其中 N 为基向量个数,$\boldsymbol{b}_j$,$j=1,2,\cdots,N$ 代表第 j 个基向量,使用 s_j,$j=1,2,\cdots,N$ 表示对应于基向量 $\boldsymbol{b}_j$ 的系数,所以对每一个输入向量都有 $\boldsymbol{x}_i\approx\sum_{j=1}^{N}\boldsymbol{b}_j s_j$.一般情况下,基向量个数大于基向量维数,即 $N>D$,故基向量集 $\boldsymbol{B}$ 是过完备的.

稀疏编码采用未标记数据来训练生成基向量集,再用基向量集的线性组合来表示输入向量.它是通过解决下面的优化问题来实现的.

$$\min_{\{\boldsymbol{b}_j\},\{s\}}\sum_{i=1}^{M}\left\|x_i-\sum_{j=1}^{N}\boldsymbol{b}_j s_j^{(i)}\right\|^2+\beta\sum_{i=1}^{M}\sum_{j=1}^{N}\phi(s_j^{(i)})$$

$$\text{s.t.}\quad\|\boldsymbol{b}_j\|^2\leqslant c,\forall j=1,2,\cdots,N \tag{5}$$

其中 $s_j^{(i)}$ 代表第 i 个输入向量中对应 $\boldsymbol{b}_j$ 基向量的系数.$\phi(\cdot)$是稀疏函数,β 是稀疏系数.本文选取 L_1 惩罚函数作为稀疏函数,即 $\phi(s_j)=\|s_j\|_1$.式(5)可以简记为,

$$\min_{\boldsymbol{B},S}\|\boldsymbol{X}-\boldsymbol{B}\boldsymbol{S}\|_F^2+\beta\sum_{i=1}^{M}\sum_{j=1}^{N}\phi(S_{i,j})$$

$$\text{s.t.}\quad\sum_{i=1}^{M}B_{i,j}^2\leqslant c,\forall j=1,2,\cdots,N \tag{6}$$

求解式(6)优化问题,就可以得到输入向量集 $\boldsymbol{X}$ 的稀疏编码结果 $\boldsymbol{S}$,基向量集为 $\boldsymbol{B}$.式(6)的优化问题中 $\boldsymbol{B}$ 和 $\boldsymbol{S}$ 同时变化时,目标函数不一定是凸优化问题.然而,固定 $\boldsymbol{B}$,优化问题是关于 $\boldsymbol{S}$ 的凸函数;固定 $\boldsymbol{S}$,它也是关于 $\boldsymbol{B}$ 的凸函数.故可以通过固定一个变量求解另外一个变量的交替优化求解方法解决稀疏编码问题.本文采用特征符号搜索(feature sign)方法来求解稀疏矩阵 $\boldsymbol{S}$,使用拉格朗日对偶方法求解 $\boldsymbol{B}$[4].稀疏编码的详细流程参见算法 2,其中的收敛条件是两次迭代后式(6)中前项的量化损失和稀疏约束项的误差之和小于给定的阈值 ε.

算法 2

1: $\boldsymbol{X}\leftarrow$随机抽样

2:随机生成 N 个基向量,构成基向量集合 $\boldsymbol{B}$.

3:Repeat

4: 固定 $\boldsymbol{B}$ 不变,式(6)的优化问题可改写为

$$\min_{S}\|\boldsymbol{X}-\boldsymbol{B}\boldsymbol{S}\|_F^2+\beta\|\boldsymbol{S}\|_1$$

$$\text{s.t.}\quad\sum_{i=1}^{M}B_{i,j}^2\leqslant c,\forall j=1,2,\cdots,N \tag{7}$$

可以直接使用基于共轭梯度下降的特征符号搜索方法[4]求解系数矩阵 $\boldsymbol{S}$.

5: 固定 $\boldsymbol{S}$ 不变,式(6)的优化问题可重写为

$$\min_{B}\|\boldsymbol{X}-\boldsymbol{B}\boldsymbol{S}\|_F^2$$

$$\text{s.t.}\quad\sum_{i=1}^{M}B_{i,j}^2\leqslant c,\forall j=1,2,\cdots,N \tag{8}$$

可直接使用拉格朗日对偶方法[4]来求解基向量集 $\boldsymbol{B}$.

6:Until 满足收敛条件或者超过最大迭代次数.

3 实验结果与分析

本文实验采用的是稠密 SIFT 特征,提取 SIFT 特征的图像块为 16×16 像素,步长设为 8 像素.所有的图像都被预先转化为灰度图.空间金字塔总层数 $L+1=3$

层,划分的具体细节如图 1 所示.

在对提取的稠密 SIFT 特征进行稀疏编码时,使用了 1024 个基向量,稀疏系数设为 0.015.在多核学习过程中,由于没有加入额外先验性息,所以将 σ_l 的值取常数.在优化问题求解过程中,没有采用 $Ad^{n+1} \geqslant p$ 这一限制条件.在所有的实验中,误分类惩罚变量 $C = 1000$.在求解核函数时,采用非线性径向基核函数 $K_l(x_i, x_j) = \exp(-\gamma_l f(x_i, x_j))$,其中 γ_l 值取所有训练数据距离 $f(x_i, x_j)$的平均值.

本文实验中采用了 2 组最常用的标准数据集:Scene Categories[6] 和 Caltech101[9].Scene 数据集包含 15 个类别的场景,共有 4486 幅图像.Caltech101 数据集包含 102 个图像类别,每个类别图像数目自 31 至 800 个不等,共包含 9144 幅图像,图像内容涵盖动物、植物、飞机等.如图 2 所示.

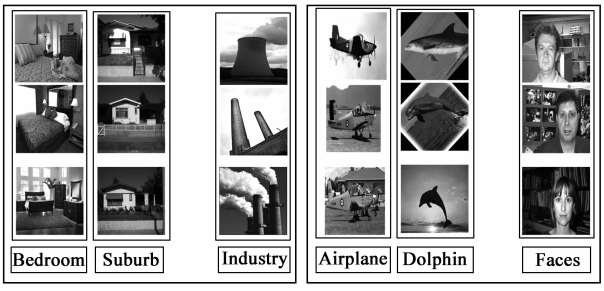

图2 实验中用到的2组图像分类集

3.1 多核学习优化

求解参数 d 的过程,是通过解式(2)二次凸优化问题 $\boldsymbol{W}(d)$的最大值来实现的,也可以通过求解 $\boldsymbol{W}'(d) = -\boldsymbol{W}(d)$的最小值来实现,则算法 1 中求解 α^* 的过程转换成了二次规划问题.本文采用 MATLAB 工具包 quadprog( )函数来解决该二次规划问题.α^* 求解出来后,则相应的 SVM 分类器的判别函数 $g(x)$也随之可解.

本文采用一对多的 SVM 分类器,分类器的判别函数与图像分类数目相同.在求解每一个判别函数时,都要首先使用多核学习的优化配置方法,需要解式(2)中的优化问题,求解核矩阵的系数.

下面是以 Scene 数据集为例来图形表示算法优化求解过程.利用算法 1 步骤,目标函数 $\boldsymbol{W}'$ 的不断优化求解过程如图 3 所示,共有 15 条曲线,每条曲线对应一个图像类别.

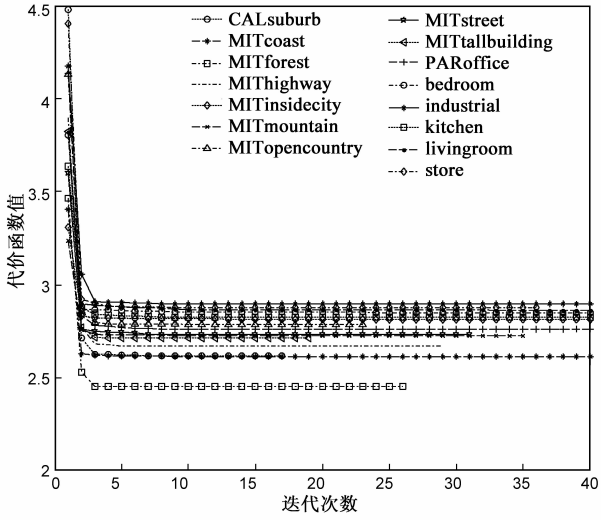

图3 Scene数据集上多核学习过程中代价函数变化曲线

利用上述的优化方法,在 2 个图像分类测试集中,可得到如图 4 所示的系数矩阵,其中 Scene 数据集的训练图像数目为 100,Caltech101 的训练图像数目为 30.

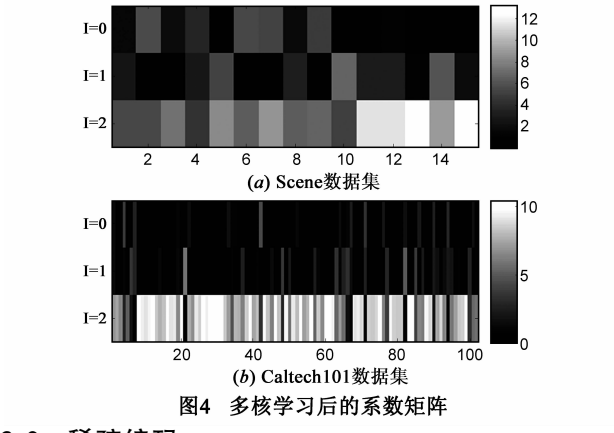

图4 多核学习后的系数矩阵

3.2 稀疏编码

由于本文选取的低层图像特征为 SIFT 特征,故基向量维数应该为 128.在初始阶段,对图像集进行特征抽样,形成 $\boldsymbol{X}$;再随机生成 1024 个 128 维基向量,构成 $\boldsymbol{B}$;固定 $\boldsymbol{B}$ 不变,使用 feature sign 方法求解与之对应的系数矩阵 $\boldsymbol{S}$.

由于对于每一个输入 x_i,都能用基向量的线性组合来表示出来,所以也可以将基向量视作一个 SIFT 特征.图像是由一系列特征构成,每个特征都可以通过稀疏编码转化为一个稀疏向量,故一幅图像经稀疏编码后变成了一个由基向量系数组成的稀疏矩阵.这点与传统使用词典树方法类似,词典树方法中图像特征经词典量化后的结果也是一个矩阵,它采用求和函数将量化后矩阵转化为视词直方图.当然,也可以采用同样量化策略来处理稀疏编码后的系数矩阵,选取的量化转换函数为 $\boldsymbol{Z} = F(\boldsymbol{S})$,其中 $\boldsymbol{Z} = [z_1, z_2, \cdots, z_N]$为稀疏系数矩阵 $\boldsymbol{S}$ 经转换后的向量,并且 $z_j = \max\{|s_{1j}|, |s_{2j}|, \cdots, |s_{Mj}|\}$.

3.3 图像分类正确率

Lazebnik[2] 和 Yang[5] 虽然都采用经验值作为权值的方法,但是他们对权值进行处理的过程是不同的.Lazebnik 在处理多层金字塔之间的核矩阵系数时,第 0

层系数置为 $1/2^L$,第 l 层的系数为 $1/2^{L-l+1}$,故 3 层金字塔结构的核矩阵系数为[0.5,0.25,0.25].与对核矩阵加权求和不同的是,Yang 直接对稀疏编码后的特征向量进行加权,然后依次连接组成一个向量,再利用该向量求取核矩阵,其3层金字塔划分经稀疏编码后各层的系数为[1,1,1].

实验结果如图 5、6、7 所示,图中分类结果是经 5 次重复实验后得到的统计结果,在每次实验过程中,都从整个数据集中随机选取一定数目的图像作为训练数据,其余图像作为测试数据.

3.3.1 Scene 场景分类数据集

本次实验中,使用从数据集的所有特征集合中随机选取 100000 个特征点来训练生成包含 1024 个基向量的集,其中交替优化的最大次数为 15.图 5 是 Scene 数据集的分类效果,其中第 0,1,2 层分别是仅使用当前空间金字塔层的核矩阵进行分类后的结果.

图 5 分别使用 5~30、60 和 100 幅作为训练数据,可见第 0 层对图像集的区分能力最弱,第 1 层和第 2 层对图像集的区分能力相近.第 1 层比第 0 层的分类准确率最大能提高 6.7%,这说明空间金字塔能有效保留图像特征的空间信息,使得第 1 层对图像集的区分能力增强.第 1 层和第 2 层对 Scene 数据集的区分能力都较强.本文综合第 0~2 层的方法比仅使用一层空间金字塔分类结果最大提高了 8.86%.

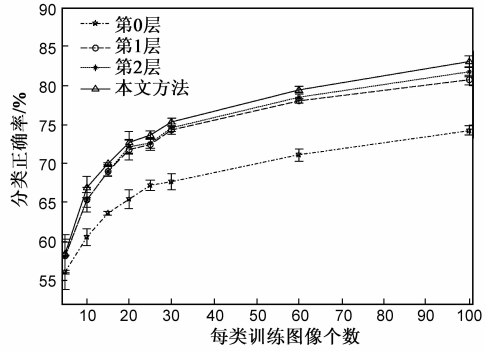

图5　本文方法在Scene场景图像集分类结果

表 1 是不同方法在 Scene 场景图像数据上的分类正确率,包括 5 次实验的平均正确率和方差.从表 1 可以看出,本文方法的分类准确率比 Lazebnik[2] 的基于词典树和空间金字塔划分方法有大幅提高(达 10.9%),比 Yang[5] 的基于稀疏编码方法亦提高了 2.8%,本文结果达到了目前该数据集上的最高准确率.

表 1　Scene 场景图像数据的分类正确率(%)

训练样本	5	10	15	20	25	30	60	100
Yang[5]	-	-	-	-	-	-	-	80.28
Lazebnik[2]	-	-	-	-	-	-	-	72.20
Cai[11]	-	-	-	-	-	-	-	73.40±1.0
[1,0,0]	56.10±2.2	60.60±1.1	63.67±0.2	65.44±1.2	67.21±0.6	67.71±1.0	71.13±0.7	74.24±0.6
[0,1,0]	58.20±1.8	65.40±1.6	69.02±0.6	71.81±1.3	72.42±0.8	74.28±0.8	77.84±0.2	80.70±0.6
[0,0,1]	58.10±2.2	65.40±0.9	69.03±0.6	72.13±0.9	72.65±0.6	74.60±0.5	78.45±0.8	81.77±0.7
[1,1,1]	59.10±2.2	65.70±1.5	69.06±0.6	71.35±1.2	72.76±1.2	73.60±0.9	76.48±1.1	80.98±1.2
[.5,.25,.25]	**60.90±1.2**	65.00±0.3	69.11±1.5	72.20±0.5	73.59±0.7	75.10±1.0	78.44±0.3	81.35±0.8
本文方法	58.50±2.4	**67.10±1.4**	**69.98±0.2**	**72.71±1.4**	**73.67±0.5**	**75.33±0.5**	**79.81±0.8**	**83.10±0.7**

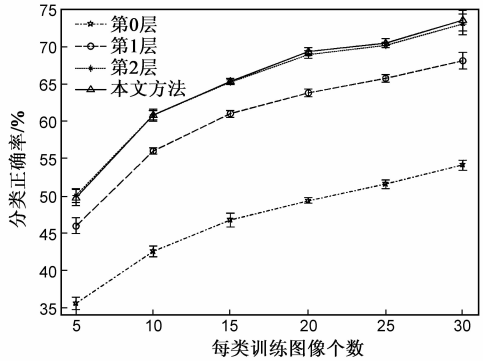

图6　本文方法在Caltech101目标图像集分类结果

3.3.2 Caltech101 目标分类数据集

针对 Caltech101 数据集,我们对原始稠密 SIFT 图像特征进行随机抽样,选取500000个特征用来训练基向量集,最大交替迭代次数为 15,基向量个数为 1024.

图 6 是本文算法对 Caltech101 数据集的分类结果,其中没加入空间信息的第 0 层对 Caltech101 数据集的图像类别区分能力依然很弱,第 1 层次之,第 2 层区分能力最强.本文算法的分类准确率与第 2 层的分类准确率相近.

表 2 是不同方法在 Caltech101 数据集上的分类正确率,其中,Gehler[17] 是使用 4 个核矩阵进行图像分类的结果.在 Gehler 的实验中最多选择了 39 个特征分别组成各自核矩阵,Caltech101 每个类别选取 30 幅作为训练图像,其分类准确率可达 77.7%.由于本实验是在 3 层空间金字塔上进行(3 个核矩阵),所以选择 Gehler 中与本文实验环境最相近的一组实验结果作对比.

表 2 Caltech101 目标图像分类集的分类正确率(%)

训练样本	5	10	15	20	25	30
Yang[5]	-	-	**67.00**	-	-	73.20
Lazebnik[2]	-	-	56.40	-	-	64.60
Yang[12]	**51.15**	59.77	65.43	67.74	70.16	73.44
Zhang[15]	46.60	55.80	59.10	62.00	-	66.20
Griffin[16]	44.20	54.50	59.00	63.30	65.80	67.60
Gehler[17]	37.80±0.4	48.60±0.7	54.50±0.9	57.90±0.8	60.70±0.8	63.80±1.0
[1,0,0]	35.58±0.8	42.61±0.7	46.80±0.9	49.40±0.4	51.59±0.6	54.16±0.7
[0,1,0]	45.99±1.1	56.02±0.4	61.06±0.5	63.81±0.5	65.84±0.5	68.19±1.1
[0,0,1]	50.11±1.0	60.85±0.6	65.31±0.3	68.97±0.5	70.26±0.3	73.06±1.4
[1,1,1]	48.90±1.2	59.77±0.9	64.50±0.7	67.82±0.5	70.40±0.7	72.79±1.4
[.5,.25,.25]	47.83±0.9	54.58±1.1	62.33±0.7	68.00±1.1	70.10±0.9	71.69±0.8
本文方法	49.81±1.1	**60.88±0.8**	65.42±0.4	**69.38±0.5**	**70.54±0.6**	**73.58±1.4**

3.3.3 混淆矩阵

图 5、6 给出的是图像分类中不同类别分类正确率的平均值,这种平均值掩盖了同一数据集内部各类别间的分类准确率.为了分析不同类别的分类正确率,我们在图 7 中给出了分类混淆矩阵(Confusion Matrix),横轴方向是分类器的类别,纵轴方向是测试图像经分类器分类判别后的图像类别.在这组实验中,我们使用 Scene 数据集,以每个类别 100 幅图像作为训练数据、其余图像作为测试图像.同时,图 7 中分类结果是由图 4(a)中系数矩阵对核矩阵进行加权后,再利用 SVM 分类得到的.可以看出,15 个分类中有 6 个类别的准确率超过了 90%,其中 CALsuburb 的准确率达到 100%.准确率最低的 3 个类别分别是:industrial (56.9%),livingroom (67.2%)和 MITopencountry (76.5%),这 3 个类别与其他类别最易混淆的类别是:store (16.6%), bedroom (13.8%)和 MITcoast(11.0%).

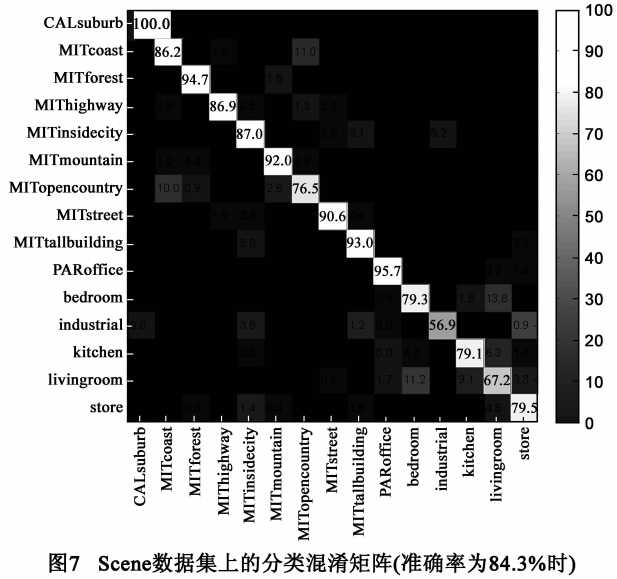

图7 Scene数据集上的分类混淆矩阵(准确率为84.3%时)

3.3.4 多核学习再讨论

本文实验中,空间金字塔各层分别对应各自的核矩阵,就每个核矩阵而言,它们对图像集的区分能力是不同的.空间金字塔层数影响着多核学习中核矩阵系数,通过观察图 5 和图 6 中的曲线,我们发现从单一空间金字塔层次考虑,第 0 层对图像的区分能力最弱,而在图 4 中,第 0 层对应核矩阵的系数可能会很大,因此本节对 MKL 进行进一步探讨,研究一下去除空间金字塔中对图像集区分能力最弱的一组核矩阵,观察其对图像分类的影响.如图 8 所示,它表示只考虑第 1 层和第 2 层情况下,利用多核学习得到的图像分类结果.

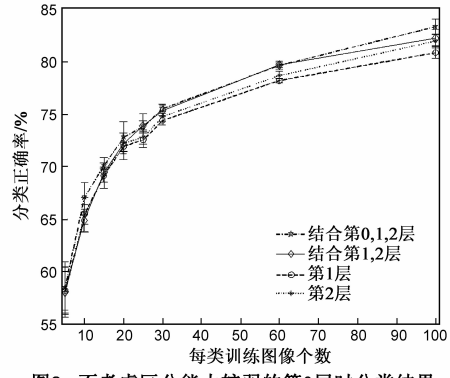

图8 不考虑区分能力较弱的第0层时分类结果

由图 8 可知,随着训练图像数目增多,不考虑第 0 层的分类方法将逐渐逼近加入第 0 层时分类准确率曲线.当选取每个类别 30 幅作为训练图像时,图像分类准确率为 75.20±0.7,与表 1 中的 75.33±0.5 分类较为接近;当选取 100 幅作为训练样本时,图像分类准确率为 82.04±0.7,这说明当考虑第 0 层时,虽然其自身对图像集的区分能力不强,但其对整体图像分类结果影响也不大.图 8 进一步表明,使用本文算法学习核矩阵系数,能有效抑制区分能力弱的核矩阵对整体分类能力的影响,使用 3 层金字塔划分和综合第 1 层和第 2 层对图像集的分类能力相近,这也反映了本文算法的有效性和鲁棒性.

4 结论

本文提出了一种基于空间金字塔对图像进行空间

划分和稀疏编码作为图像特征表示的图像分类方法,它能够利用多核学习方法求取不同核矩阵的权值,再利用核矩阵的线性组合来进行图像分类.通过解决式(1)的凸优化问题,可以寻求到最优的权值用于核矩阵的线性组合.为了说明金字塔层次中区分能力最弱的核矩阵对总体分类能力的影响,通过移除区分能力最弱的核矩阵后再进行多核学习图像分类,结果证明,区分能力最弱的核矩阵虽然其自身区分能力较低,但其对整体图像分类结果影响不大.本文算法能有效抑制区分能力弱的核矩阵对整体分类的影响,这也从另一个方面反映了本文算法的有效性和鲁棒性.

参考文献

[1] D Nister, et al. Scalable recognition with a vocabulary tree [A]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Press, 2006. 2161 – 2168.

[2] S Lazebnik, C Schmid, J Ponce. Beyond bag of features: spatial pyramid matching for recognizing natural scene categories [A]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Press, 2006. 2169 – 2178.

[3] M Varma, D Ray. Learning the discriminative power- invariance trade-off [A]. Proceeding of International Conference on Computer Vision [C]. USA: IEEE Press, 2007. 1 – 8.

[4] H Lee, A Battle, R Raina, A Y Ng. Efficient sparse coding algorithms [A]. Proceeding of Advances in Neural Information Processing System [C]. Canada: NIPS Press, 2007. 801 – 808.

[5] J Yang, K Yu, Y Gong, T Huang. Linear spatial pyramid matching using sparse coding for image classification [A]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Press, 2009. 1794 – 1801.

[6] F-F Li, A Oliva. Fifteen scene categories [OL]. http://www-cvr.ai.uiuc.edu/ ponce_grp/data/, 2006.10.

[7] J F Bonnans, A Shapiro. Perturbation Analysis of Optimization Problems [M]. Springer, 2000.

[8] O Chapelle, V Vapnik, O Bousquet, S Mukherjee. Choosing multiple parameters for support vector machines [J]. Machine Learning, 2002, 46(1-3): 131 – 159.

[9] F-F Li, M Andreetto, M A Ranzato. The Caltech-101 object categories [OL]. http://www. vision. caltech. edu/feifeili/Data sets.htm, 2003.9.

[10] 谢昭,高隽.基于高斯统计模型的场景分类及约束机制新方法[J].电子学报,2009,37(4):733 – 738.
Xie Zhao, Gao Jun. A novel method for scene categorization with constraint mechanism based on gaussian statistical model [J]. Acta Electronica Sinica, 2009, 37(4): 733 – 738. (in Chinese)

[11] H Cai, F Yan, K Mikolajczyk. Learning weights for codebook in image classification and retrieval [A]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Press. 2010, 2320 – 2337.

[12] 石光明,刘丹华,高大化,等.压缩感知理论及其研究进展[J].电子学报,2009,37(5):1071 – 1078.
Shi Guang-ming, Liu Dan-hua, Gao Da-hua, et al. Advances in theory and application of compressed sensing [J]. Acta Electronica Sinica, 2009, 37(5): 1071 – 1078. (in Chinese)

[13] J Wang, J Yang, K Yu, F Lv, T. Huang, Y Gong. Locality-constrained linear coding for image classification [A]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Press, 2010. 3360 – 3367.

[14] 刘硕研,须德,冯松鹤,刘镝,裘正定.一种基于上下文语义信息的图像块视觉单词生成算法[J].电子学报,2010,38(5):1156 – 1161.
Liu Shuo-yan, Xu De, Feng Song-he, Liu Di, Qiu Zheng-ding. A novel visual words definition algorithm of image patch based on contextual semantic information [J]. Acta Electronica Sinica, 2010, 38(5): 1156 – 1161. (in Chinese)

[15] H Zhang, A C Berg, et al. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [A]. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition [C]. USA: IEEE Press, 2006. 2126 – 2136.

[16] G Griffin, A Holub, P Perona. Caltech-256 object category dataset [R]. California Institute of Technology, USA, 2007.

[17] P Gehler, S Nowozin. On feature combination for multi-class object classification [A]. Proceeding of International Conference on Computer Vision [C]. USA: IEEE Press, 2009. 221 – 228.

作者简介

亓晓振　男.1985 年生,陕西咸阳人.2008 年和 2011 年于西北工业大学获工学学士和硕士学位.现在华为技术有限公司西安研究所从事研究与开发工作.
E-mail: tsi0120@163.com

王　庆(通讯作者)　男.1969 年生,陕西西安人.博士、教授、博士生导师、中国计算机学会高级会员、IEEE 会员、ACM 会员.1991 年和 2000 年分别在北京大学、西北工业大学获理学学士和工学博士学位.现为陕西省语音与图像信息处理重点实验室副主任,主要从事图像处理、计算机视觉和模式识别等方面的研究工作.
E-mail: qwang@nwpu.edu.cn