

非负稀疏局部线性编码*

庄连生⁺, 高浩渊, 刘超, 俞能海

(中国科学技术大学 信息科学技术学院, 安徽 合肥 230027)

Nonnegative Sparse Locally Linear Coding

ZHUANG Lian-Sheng⁺, GAO Hao-Yuan, LIU Chao, YU Neng-Hai

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China)

+ Corresponding author: E-mail: lszhuang@ustc.edu.cn

Zhuang LS, Gao HY, Liu C, Yu NH. Nonnegative sparse locally linear coding. Journal of Software, 2011,22 (Suppl.2):89-95. <http://www.jos.org.cn/1000-9825/11029.htm>

Abstract: Feature quantization is an important component in Bag of word model. This paper proposes a novel method called nonnegative sparse locally linear coding (NSLLC) to improve the performance of locally linear coding. The core ideas of NSLLC is to use nonnegative sparse representation to select the nearest neighbors in the same subspace and then encode the local feature with respect to the local coordinate consisting of these nearest neighbors. Experimental results have shown NSLLC has outperformed state-of-the-art local feature coding methods and is in favor of image classification problem.

Key words: image classification; bag of word; nonnegative sparse representation; feature quantization

摘要: 针对视觉词袋模型中的特征量化问题,提出一种非负稀疏局部线性编码方法.它能够有效地改善局部特征编码性能,提高图像非线性特征的区分能力.其核心思想是,利用非负稀疏表示技术选择与待编码局部特征处于同一线性空间中的近邻点,然后以这些近邻点作为局部坐标系对当前局部特征进行线性编码.实验结果表明,该局部特征编码方法显著优于现有的特征编码方法,有效地提高了图像非线性特征的区分能力,更有利于图像分类任务.

关键词: 图像分类;词袋模型;非负稀疏表示;特征量化

图像分类问题是计算机视觉和多媒体领域的一个基本问题,在图像检索、视频监控等领域中有着广泛应用.图像分类问题的研究首先要解决图像表示问题.视觉词袋(bag of visual words)模型是目前使用最为广泛的一个图像表示模型.它通过提取图像局部特征(如 SIFT 特征^[1]),把局部特征量化为视觉词汇(visual words),最后统计各视觉词汇出现的频率直方图来表示整张图像.由于对平移、旋转等几何变换较为鲁棒,视觉词袋模型在场景分类^[2]、物体识别^[3]等具体图像分类任务中取得了巨大成功,受到学术界和工业界的广泛关注.

字典学习和特征量化是视觉词袋模型的两个关键步骤,直接决定了词袋模型的性能.字典学习的目标是寻找新的具有较强表达能力的特征空间.特征量化的目标则是寻找一种局部特征编码方法,实现局部特征从原始

* 基金项目: 国家自然科学基金(60933013, 61103134); 国家科技重大专项(2010ZX03004-003); 中央高校基本科研业务经费(WK2100230002, WK2101020003)

收稿时间: 2011-07-20; 定稿时间: 2011-12-01

特征空间到新特征空间的映射.长期以来,研究人员主要集中于字典学习问题的研究,忽视了特征量化问题的重要性,经常采用最简单的最近邻方法实现局部特征映射^[4].但是,近期研究结果表明,当字典达到一定规模之后,不同字典学习方法对最终分类结果的影响非常有限.在实际应用(如图像检索)中,字典规模一定都比较大,往往在几十万、数百万数量级之上,此时使用简单的随机采样也可以得到一个分类性能令人满意的字典.相反,特征量化方法则会对最终分类结果产生显著影响.因此,特征量化成为近年来一个新的研究热点^[4-9].针对特征量化问题,本文提出一种特征量化方法,用于改善局部特征编码性能,提高图像分类算法的分类性能.

特征量化也称为特征编码(feature encoding),主要研究如何把局部特征映射到由视觉词典所定义的特征空间中.早期词袋模型主要采取硬关联策略(hard assignment),利用最近邻算法选择词典中与待编码局部特征距离最近的视觉词汇来表示局部特征.在这种策略下,每个局部特征只能与1个视觉词汇相关联,造成信息损失,难以处理处于不同词汇中间区域(类边缘地带)的局部特征^[5].鉴于此,文献^[6]提出一种软关联(soft assignment)的策略,允许每个局部特征用多个视觉词汇来表示,有效降低了量化误差,提高了局部特征编码性能.但是,该方法对关联词汇的个数非常敏感.同时,该方法对局部特征量化之后,直接统计图像的词汇直方图作为图像的线性特征.该特征必须与非线性分类器相结合中才能取得较好的分类性能.但是,非线性分类器的计算复杂度一般比较高,泛化能力也较差,难以推广到大规模的应用中.

Yang等人^[7]把稀疏编码(sparse coding,简称SC)技术用于局部特征量化,结合非线性操作(max pooling)提取图像的非线性特征,利用线性SVM作为分类器,取得了很好的分类性能.由于采用线性分类器,该方法分类速度快,泛化能力也好,很适合大规模图像分类任务.通过对SC结果的研究^[9],Yu等人认为局部性比稀疏性更本质.局部非零一定导致整体稀疏,反之则不成立.基于该观点,他们提出一种局部约束线性编码方法(locality linear coding,简称LLC),提高了特征编码性能^[8].从本质上说,LLC编码方法只是局部线性嵌入(local linear embedding,简称LLE)算法^[10]在特征量化中的应用:首先利用 k 近邻算法寻找词典中与待编码的局部特征最接近的 k 个视觉词汇,然后利用这 k 个视觉词汇来重构当前局部特征,而重构系数则作为最终编码结果的非零系数项(其他不相关视觉词汇对应的系数为0).该算法隐含了一个基本假设:所寻找到的 k 个最近邻视觉词汇与当前局部特征位于同一个线性子空间(或仿射子空间),否则编码结果没有意义.从信号表示角度看,如果用不同子空间(类别)中的信号来重构当前信号,重构系数已经丧失了其应有的物理含义.但是, k 近邻算法显然难以保证所找到的 k 个视觉词汇处于同一子空间.如图1所示,圆圈和十字叉分别代表了来自两个不同子空间的基向量,五角星代表了当前待编码局部特征点(来自圆圈所在的子空间).圆形区域则是利用 k 近邻算法找到的近邻点范围.显然,我们只希望在五角星所在的子空间寻找近邻点,但是 k 近邻算法会同时找到两个子空间中的点.这会导致LLC方法的编码性能下降.

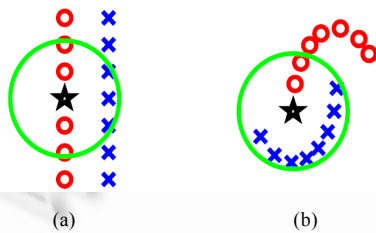


Fig.1 Diagram of k NN searching results
图1 k 近邻算法搜索结果示意图

圆形区域则是利用 k 近邻算法找到的近邻点范围.显然,我们只希望在五角星所在的子空间寻找近邻点,但是 k 近邻算法会同时找到两个子空间中的点.这会导致LLC方法的编码性能下降.

压缩感知理论研究表明^[11],稀疏表示的非零系数事实上揭示了信号所属的类别关系.受此结论启发,本文在LLC方法基础上提出一种非负稀疏局部编码方法(nonnegative sparse locally linear coding,简称NSLLC),以提高局部特征编码性能.其核心思想是:计算局部特征相对于视觉词典的稀疏表示系数,根据非零系数值的大小来选择近邻视觉词汇,然后利用近邻视觉词汇来重构局部特征点,重构系数作为编码结果的非零系数项.由于近邻样本是从非零稀疏表示系数对应的视觉词汇中选择的,这就确保当前局部特征和所选择的近邻视觉词汇来自于同一子空间(类别),保证随后的重构具有明确的物理含义.在标准稀疏编码技术中,稀疏表示系数可正、可负.这也意味着,基向量之间可以通过相互“抵消”来重构信号.尽管从数学上负系数不难理解,但对于图像分类应用而言,负系数却没有清晰、合理的物理解释.因此,本文采用非负稀疏编码技术(nonnegative sparse coding,简称NSC)来选择近邻点.与稀疏编码技术相比,非负稀疏编码技术更加符合生物学模型^[12,13],可以更好地揭示信号类别关系^[14].从本质上看,NSLLC方法是对LLC方法的改进.但与LLC方法相比,由于选择正确的近邻视觉词汇,

压缩感知理论研究表明^[11],稀疏表示的非零系数事实上揭示了信号所属的类别关系.受此结论启发,本文在LLC方法基础上提出一种非负稀疏局部编码方法(nonnegative sparse locally linear coding,简称NSLLC),以提高局部特征编码性能.其核心思想是:计算局部特征相对于视觉词典的稀疏表示系数,根据非零系数值的大小来选择近邻视觉词汇,然后利用近邻视觉词汇来重构局部特征点,重构系数作为编码结果的非零系数项.由于近邻样本是从非零稀疏表示系数对应的视觉词汇中选择的,这就确保当前局部特征和所选择的近邻视觉词汇来自于同一子空间(类别),保证随后的重构具有明确的物理含义.在标准稀疏编码技术中,稀疏表示系数可正、可负.这也意味着,基向量之间可以通过相互“抵消”来重构信号.尽管从数学上负系数不难理解,但对于图像分类应用而言,负系数却没有清晰、合理的物理解释.因此,本文采用非负稀疏编码技术(nonnegative sparse coding,简称NSC)来选择近邻点.与稀疏编码技术相比,非负稀疏编码技术更加符合生物学模型^[12,13],可以更好地揭示信号类别关系^[14].从本质上看,NSLLC方法是对LLC方法的改进.但与LLC方法相比,由于选择正确的近邻视觉词汇,

NSLLC 方法具有更好的编码性能,对噪声更加鲁棒,更有利于图像分类任务.实验结果也验证了我们的结论.

本文第 1 节首先回顾非负稀疏编码相关模型,给出基于增广拉格朗日乘子法的求解算法.第 2 节介绍非负稀疏局部编码算法的实现框架.第 3 节是实验结果及分析.第 4 节是全文总结.

1 非负稀疏编码

1.1 非负稀疏表示模型

假设 $\mathbf{B}=[\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N] \in R^{d \times N}$ 是 d 维空间中的一组包含 N 个基向量的过完备基, $\mathbf{y}=(y_1, y_2, \dots, y_d)^T \in R^d$ 是 d 维空间中的向量.非负稀疏编码假设信号 \mathbf{y} 可以用 \mathbf{B} 中少数几个基向量的线性组合来表示,即

$$\mathbf{y} = \sum_{i=1}^M s_i \mathbf{b}_i = \mathbf{B}\mathbf{s} \quad (1)$$

其中, $\mathbf{s}=[s_1, s_2, \dots, s_M]^T$ 是线性组合系数构成的向量,每个分量都大于 0.通常,我们称 \mathbf{s} 为信号 \mathbf{y} 的非负稀疏表示 (nonnegative sparse representation, 简称 NSR).如果 \mathbf{s} 中非零元素个数不多于 k 个(且 $k \ll M$),则我们称信号 \mathbf{y} 是 k 稀疏的.实际应用中,如果 \mathbf{s} 最多只有 k 个比较显著的非零系数,而其他非零系数相对较小(接近于 0),那么我们也称信号 \mathbf{y} 是一个 k 稀疏信号.由于 \mathbf{B} 过完备,方程组(1)中未知数个数 M 远大于方程个数 d ,有无穷多解.非负稀疏编码寻找其中非零元素个数最少的 \mathbf{s} 解作为方程组(1)的最佳解 \mathbf{s}^* ,这可以通过求解如下最优化问题得到:

$$\mathbf{s}^* = \min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{s.t. } \mathbf{y} = \mathbf{B}\mathbf{s}, \mathbf{s} \geq 0 \quad (2)$$

其中, $\|\mathbf{s}\|_0$ 是向量 \mathbf{s} 的 0-范数,等于 \mathbf{s} 中非零元素个数.但是,式(2)是一个 NP 难的非凸优化问题,难以求解.根据压缩感知相关理论^[15,16],如果 \mathbf{s} 足够稀疏,我们通常可以用 1-范数代替 0-范数:

$$\min \|\mathbf{s}\|_1 \quad \text{s.t. } \mathbf{y} = \mathbf{B}\mathbf{s}, \mathbf{s} > 0 \quad (3)$$

显然,式(3)是凸优化问题,可以有很多方法来求解^[17].

一般而言,观测信号 \mathbf{y} 通常伴有随机噪声 \mathbf{e} ,此时需要对模型进行修正.如果噪声 \mathbf{e} 为高斯白噪声,则模型(3)被通常修正为

$$\min \|\mathbf{s}\|_1 \quad \text{s.t. } \|\mathbf{y} - \mathbf{B}\mathbf{s}\| \leq \epsilon, \mathbf{s} > 0 \quad (4)$$

其中, ϵ 表示所容忍的重构误差.式(4)通常转化为式(5)并利用著名的 Lasso 算法来求解:

$$\mathbf{s}^* = \min_{\mathbf{s}} \|\mathbf{y} - \mathbf{B}\mathbf{s}\|^2 + \lambda \|\mathbf{s}\|_1 \quad \text{s.t. } \mathbf{s} > 0 \quad (5)$$

其中,参数 λ 主要平衡重构误差和稀疏性.需要注意的是,根据压缩感知理论,式(5)中稀疏解存在的一个基本条件是,矩阵 \mathbf{B} 满足所谓的 RIP 性质.在实际应用中,我们很难验证 \mathbf{B} 是否满足 RIP 性质.为此,通常先利用一个高斯随机矩阵 Φ 来对信号 \mathbf{y} 和词典 \mathbf{B} 降维,即

$$\mathbf{s}^* = \min_{\mathbf{s}} \|\Phi\mathbf{y} - \Phi\mathbf{B}\mathbf{s}\|^2 + \lambda \|\mathbf{s}\|_1 = \min_{\mathbf{s}} \|\tilde{\mathbf{y}} - \tilde{\mathbf{B}}\mathbf{s}\|^2 + \lambda \|\mathbf{s}\|_1 \quad (6)$$

其中, $\tilde{\mathbf{y}} = \Phi\mathbf{y}$ 且 $\tilde{\mathbf{B}} = \Phi\mathbf{B}$.

此外,为了处理遮挡和损毁等情况,我们通常把 \mathbf{e} 建模为稀疏损毁噪声.此时,模型(3)通常被修正为

$$\min \|\mathbf{s}\|_1 + \|\mathbf{e}\|_1 \quad \text{s.t. } \mathbf{y} = \mathbf{B}\mathbf{s} + \mathbf{e}, \mathbf{s} > 0 \quad (7)$$

值得指出的是,由于增加了系数非负这一限制条件,文献[18]中的求解技巧并不能使用.也就是说,即使令 $\tilde{\mathbf{B}} = [\mathbf{B} \quad \mathbf{I}]$ 且 $\tilde{\mathbf{s}} = \begin{bmatrix} \mathbf{s} \\ \mathbf{e} \end{bmatrix}$, 式(7)也无法归结为式(3)来求解.如果要归结为问题(3)来求解,那么必然要求 $\tilde{\mathbf{s}} > 0$.但是,问题(7)只要求 $\mathbf{s} > 0$, \mathbf{e} 则可以为负.下面,我们利用增广拉格朗日乘子法来求解该问题.

1.2 增广拉格朗日求解算法

拉格朗日乘子法(augmented Lagrange multiplier, 简称 ALM)^[19]是求解高维凸优化问题的一种常用方法.为了求解式(7),我们首先定义如下增广拉格朗日函数:

$$L(\mathbf{s}, \mathbf{e}, \mathbf{h}, \mu) = \|\mathbf{s}\|_1 + \|\mathbf{e}\|_1 + \langle \mathbf{h}, \mathbf{y} - (\mathbf{B}\mathbf{s} + \mathbf{e}) \rangle + \frac{\mu}{2} \|\mathbf{y} - (\mathbf{B}\mathbf{s} + \mathbf{e})\|_F^2 \quad (8)$$

其中, μ 是一个正数, \mathbf{h} 是新引入的辅助变量, $\langle \bullet, \bullet \rangle$ 是向量内积算子. 与普通的拉格朗日函数相比, 增广拉格朗日函数多出一个关于约束的惩罚项. 增广拉格朗日乘子法每次迭代最小化增广拉格朗日函数, 得到一组新的 $(\mathbf{s}_k, \mathbf{e}_k)$, 通过 $(\mathbf{s}_k, \mathbf{e}_k, \mu_k)$ 更新乘子 \mathbf{h}_k , 然后继续求解下一个 $(\mathbf{s}_k, \mathbf{e}_k)$, 最终使得 $(\mathbf{s}_k, \mathbf{e}_k)$ 收敛到原问题最优解.

令 S_τ 函数为收缩算子, 其定义如式(9)所示:

$$S_\tau(x) = \begin{cases} x - \tau, & \text{如果 } x > \tau \\ x + \tau, & \text{如果 } x < -\tau \\ 0, & \text{其他} \end{cases} \quad (9)$$

非负稀疏编码求解算法的主要流程如算法 1 所示. 在算法 1 中, 为加快计算速度, 减少迭代次数, 我们在算法中采用了非精确的 ALM 算法框架来求解. 有关 IALM 算法可以参考^[19].

算法 1. 基于 IALM 的非负稀疏编码算法流程.

输入: 视觉词典 \mathbf{B} .

初始化: $\mathbf{e}_0 = \mathbf{h}_0 = \mathbf{I}$, $\rho = 1.1$, $\mu_0 = 0.1$, $k = 0$.

While 不收敛 **do**

(1) 固定 $\mathbf{e}_k, \mathbf{h}_k, \mu_k$, 更新 \mathbf{s} :

$$\mathbf{s}_{k+1} = \arg \min_{\mathbf{s}: \mathbf{s} \geq 0} L(\mathbf{s}, \mathbf{e}_k, \mathbf{h}_k, \mu_k) = S_{(\mu_k \eta)^{-1}} \left(\mathbf{s}_k - \frac{1}{\eta} \mathbf{B}^T (\mathbf{B}\mathbf{s}_k - (\mathbf{y}_k + \mathbf{h}_k / \mu_k - \mathbf{e}_k)) \right)$$

其中, 参数 $\eta = \|\mathbf{B}\|_2^2$.

(2) 固定 $\mathbf{s}_{k+1}, \mathbf{h}_k, \mu_k$, 更新 \mathbf{e} :

$$\mathbf{e}_{k+1} = \arg \min_{\mathbf{e}} L(\mathbf{s}_{k+1}, \mathbf{e}, \mathbf{h}_k, \mu_k) = S_{\mu_k^{-1}} (\mathbf{y} - \mathbf{B}\mathbf{s}_{k+1} + \mathbf{h}_k / \mu_k)$$

(3) 更新变量 \mathbf{h} : $\mathbf{h}_{k+1} = \mathbf{h}_k + \mu_k (\mathbf{y} - \mathbf{B}\mathbf{s}_{k+1} - \mathbf{e}_{k+1})$

(4) 更新参数 μ : $\mu_{k+1} = \rho \mu_k$

(5) 更新迭代步骤: $k = k + 1$

End

输出: 最优的 \mathbf{s}^* .

2 非负稀疏局部线性编码

假设 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in R^{d \times N}$ 是一个包含 N 个 d 维局部特征的矩阵, $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M] \in R^{d \times M}$ 是一个包含 M 个视觉词汇的词典, 非负稀疏局部编码把所有局部特征 \mathbf{Y} 映射到由 \mathbf{B} 定义的编码空间. NSLLC 方法核心思想是: 利用非负稀疏表示技术在局部特征点所在子空间中选邻近点, 然后利用这些邻近点组成局部坐标系对局部特征点进行线性编码.

从本质上说, NSLLC 是对 LLC 方法的一个改进, 两者优化目标函数相同. 记 $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \in R^{M \times N}$ 为 \mathbf{Y} 相对于 \mathbf{B} 中的编码矩阵, \mathbf{s}_i 是局部特征 \mathbf{y}_i 的编码. 和 LLC 一样, NSLLC 的目标函数为

$$\min_{\mathbf{S}} \sum_{i=1}^N \|\mathbf{y}_i - \mathbf{B}\mathbf{s}_i\|^2 + \lambda \|\mathbf{d}_i \cdot \mathbf{s}_i\|^2 \quad \text{s.t.} \quad \mathbf{1}^T \mathbf{s}_i = 1, \quad \forall i \quad (10)$$

其中, \mathbf{d}_i 为局部特征 \mathbf{y}_i 到视觉字典 \mathbf{B} 的距离, 其定义为

$$\mathbf{d}_i = \exp \left(\frac{\text{dist}(\mathbf{y}_i, \mathbf{B})}{\sigma} \right) \quad (11)$$

且有 $\text{dist}(\mathbf{y}_i, \mathbf{B}) = [\text{dist}(\mathbf{y}_i, \mathbf{b}_1), \dots, \text{dist}(\mathbf{y}_i, \mathbf{b}_M)]^T$. $\text{dist}(\mathbf{y}_i, \mathbf{b}_j)$ 是局部特征 \mathbf{y}_i 到视觉词汇 \mathbf{b}_j 的欧式距离, 参数 σ 控制局部性衰减速度, 各参数物理含义可进一步参考文献[8]. $\|\mathbf{d}_i \cdot \mathbf{s}_i\|^2$ 是向量 \mathbf{d}_i 和 \mathbf{s}_i 对应元素的点积, 体现了编码的局部性. 该惩罚项要求编码 \mathbf{s}_i 中非零系数对应的视觉词汇总是与 \mathbf{y}_i 比较相似(欧式距离近).

由于直接求解式(10)比较困难,NSLLC 首先计算各局部特征点 \mathbf{y}_i 相对于基 \mathbf{B} 的非负稀疏表示系数,根据非零系数由大到小从中选择 k 邻近视觉词汇 $\{\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \dots, \mathbf{b}_{i_k}\}$, 这些邻近点构成了 \mathbf{y}_i 的局部坐标基 $\mathbf{B}_i = \{\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \dots, \mathbf{b}_{i_k}\}$, 最后通过求解如下小规模的最小二乘问题即可得到 \mathbf{y}_i 的局部线性编码:

$$\min_{\tilde{\mathbf{s}}_i} \sum_{i=1}^N \|\mathbf{y}_i - \mathbf{B}_i \tilde{\mathbf{s}}_i\|^2 \quad \text{s.t.} \quad \mathbf{1}^T \tilde{\mathbf{s}}_i = 1, \forall i. \quad (12)$$

其中, $\tilde{\mathbf{s}}_i$ 表示局部特征点 \mathbf{y}_i 相对于邻近视觉词汇 $\{\mathbf{b}_{i_1}, \mathbf{b}_{i_2}, \dots, \mathbf{b}_{i_k}\}$ 的编码,而非邻近视觉词汇对应的系数则为 0.显然,式(12)是一个带约束条件的最小二乘问题,具有闭解:

$$\tilde{\mathbf{s}}_i = ((\mathbf{B}_i - \mathbf{1}\mathbf{y}_i^T)(\mathbf{B}_i - \mathbf{1}\mathbf{y}_i^T)^T + \mu\mathbf{I}) \setminus \mathbf{1} \quad (13)$$

$$\tilde{\mathbf{s}}_i = \tilde{\mathbf{s}}_i / \mathbf{1}\tilde{\mathbf{s}}_i \quad (14)$$

其中, $\mu\mathbf{I}$ 是一个随机扰动项.当邻近点的个数大于特征维数时,局部协方差矩阵 $(\mathbf{B}_i - \mathbf{1}\mathbf{y}_i^T)(\mathbf{B}_i - \mathbf{1}\mathbf{y}_i^T)^T$ 非满秩,需要添加随机扰动项来保证方程有唯一解.需要说明的是,由于实际数据噪声的存在,最终计算得到的稀疏表示系数的稀疏性并非很好,我们通常把较小的系数直接置 0.

3 实验与分析

本文在目前常用 Caltech 101 数据集上验证算法的有效性,并与 ScSPM 方法^[7]、LLC 方法^[8]进行比较.实验在 PC 机上完成,机器配置为 Intel 2.4G 处理器,4GB 内存,实验平台是 Windows7(32 位)+Matlab 2010.所有都是用 Matlab 实现.受计算能力所限,本文实验最终只选择了 Caltech 101 数据集中的前 7 个类(accordion,airplanes, anchor,ant,Background_google,barrel,bass)进行测试,图像数据合计 1 507 张,如图 2 所示.



Fig.2 Image samples in experiments

图 2 实验图像样本

在本文实验中,我们提取 SIFT 特征作为图像底层特征,采用空间金字塔匹配模型(SPM)^[2]并结合“max pooling”操作来提取图像的非线性特征.此外,由于本文算法主要针对特征量化,因此直接采用 LLC 方法主页上所提供的视觉词典(包含 1 024 个视觉词汇).最后,本文采用线性 SVM 作为分类器,具体实验方案设置可以参考文献[7,8].为了保证比较的公平性,所有实验均跑 20 次,随机生成每次实验的训练集和测试集,最终取 20 次运算结果的平均作为最终实验结果,见表 1.其中,稀疏编码方法和 LLC 方法的代码分别从文献[20,21]下载.从表 1 中可以看出,本文所提特征编码方法更有效,更有利于图像分类任务.

Table 1 Classification performance on Caltech 101

表 1 在 Caltech 101 数据集上的分类性能

编码方法	训练样本数	近邻样本数	识别率	标准差
NSLLC	30	10	0.84	0.021
NSLLC+降维	30	10	0.85	0.026
ScSPM 方法 ^[20]	30	10	0.81	0.021
LLC 方法 ^[21]	30	10	0.82	0.029

(1) 非负限制条件的影响.本文采用非负稀疏编码技术来确保 k 近邻算法选择的近邻点与当前局部特征点

处于同一子空间中.与标准稀疏表示模型相比,非负稀疏表示模型更符合生物学模型,所得到的结果具有更好的稀疏性.图 3 中显示了同一样本用不同模型计算得到的稀疏表示系数的可视化结果.从图 3 中可以看出,非负稀疏表示模型可以得到更加稀疏的编码,更符合我们的期望.

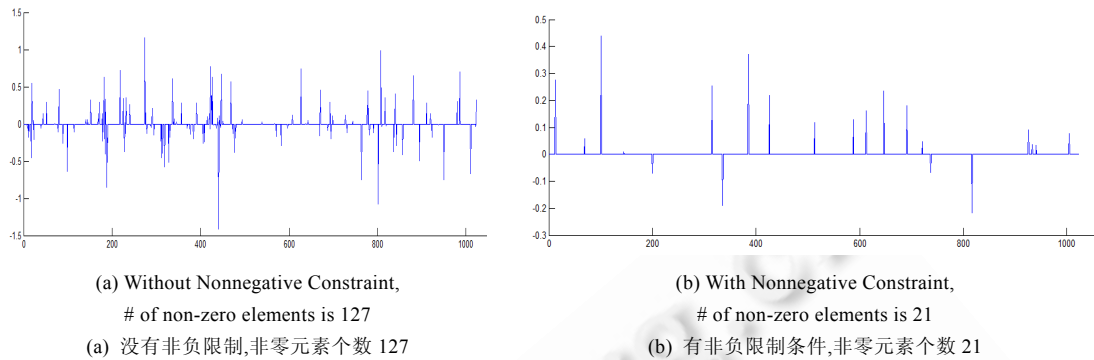


Fig.3 Influence of nonnegative constraints to the sparsity of coefficients

图 3 非负约束条件对稀疏表示系数求解结果的影响

(2) 特征降维的影响.在实际应用中,由于视觉词典 B 不一定满足所谓的 RIP 性质,此时稀疏表示系数解的精度和鲁棒性并没有理论上的保证.针对该问题,通常产生一个随机高斯矩阵,利用该矩阵对基以及所有观测样本进行降维,然后再求解稀疏表示系数.由于随机高斯矩阵满足 RIP 性质,根据压缩感知相关理论,该随机矩阵与另一个矩阵相乘,所得到的结果矩阵也满足所谓的 RIP 性质.为此,本文利用一个随机高斯矩阵(80×128)来对所有样本降维.图 4 显示了降维前后得到的稀疏表示系数的可视化结果.从图中可以看出,利用随机高斯矩阵降维后,稀疏表示系数具有更好的稀疏性.

4 结 论

特征量化问题是视觉词袋模型的一个研究热点,对图像分类算法的性能有非常重要的影响.本文根据稀疏表示相关理论,提出了一种非负稀疏局部线性编码,对局部线性编码进行改进.同时,本文给出了利用增广拉格朗日求解非负稀疏编码问题的相关算法,用于解决非负表示系数的获取.实验结果表明,本文方法能够有效地提高图像非线性特征提取的性能,更有利于图像分类任务.但是,受实验条件所限,本文只在少量数据上进行测试,下一步拟大规模测试,以验证算法的有效性.此外,本文算法虽然可以提高局部特征编码性能,但计算复杂度还有待降低,如何加快计算速度是后续的一个重要研究方向.

致谢 在此,我们感谢微软亚洲研究院马毅博士和林宙辰博士在稀疏编码以及求解算法方面给予的宝贵讨论和指导,感谢 HHME2011 会议承办单位为本文提供一个宝贵的交流机会,感谢匿名审稿人所给予的宝贵审稿意见.

References:

- [1] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos. In: Proc. of the ICCV 2003. 2003. 1470-1477.
- [2] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2006.
- [3] Yang L, Jin R, Sukthankar R, Jurie F. Unifying discriminative visual codebook generation with classifier training for object category recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2008.

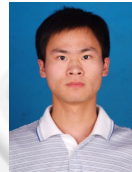
- [4] Coates A, Ng AY. The importance of encoding versus training with sparse coding and vector quantization. In: Proc. of the ICML. 2011.
- [5] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Anchorage, 2008.
- [6] Philbin J, Chum O, Isard M, Sivic J, Zisserman A. Object retrieval with large vocabularies and fast spatial matching. In: Proc. of the CVPR. Minneapolis, 2007.
- [7] Yang J, Yu K, Gong Y, Huang T. Linear spatial pyramid matching using sparse coding for image classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2009.
- [8] Wang J, Yang J, Yu K, Lü F, Huang T, Gong Y. Locality-Constrained linear coding for image classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2010. 3360–3367.
- [9] Yu K, Zhang T, Gong YH. Nonlinear learning using local coordinate coding. In: Proc. of the NIPS 2009. 2009.
- [10] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000,290(5500):2323–2326.
- [11] Elhamifar E, Vidal R. Sparse subspace clustering. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2009.
- [12] Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature, 1999,401:788–791.
- [13] Hoyer PO. Non-Negative sparse coding. In: Proc. of 2002 the 12th IEEE Workshop on Neural Networks for Signal Processing. 2002. 557–565.
- [14] He R, Zheng WS, Hu BG, Kong XW. Nonnegative sparse coding for discriminative semi-supervised learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2011.
- [15] Candes E, Tao T. Near-Optimal signal recovery from random projections: Universal encoding strategies. IEEE Trans. on Information Theory, 2006,52(12):5406–5425.
- [16] Donoho D. For most large underdetermined systems of linear equations the minimal l^1 -norm solution is also the sparsest solution. Communications on Pure and Applied Mathematics, 2006,59(6):797–829.
- [17] Yang A, Ganesh A, Sastry S, Ma Y. Fast l^1 -minimization algorithms and an application in robust face recognition: A review. In: Proc. of the ICIP 2010. 2010.
- [18] Wright J, Yang A, Ganesh A, Sastry S, Ma Y. Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(2):210–227.
- [19] Lin ZC, Chen MM, Ma Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrix. Technical Report, UILU-ENG-09-2215, UIUC, 2009.
- [20] <http://www.ifp.illinois.edu/~jyang29/ScSPM.htm>
- [21] <http://www.ifp.illinois.edu/~jyang29/LLC.htm>



庄连生(1978—),男,福建南靖人,博士,讲师,主要研究领域为计算机视觉,图像检索,机器学习.



高浩渊(1987—),男,硕士生,主要研究领域为计算机视觉,机器学习.



刘超(1987—),男,硕士生,主要研究领域为计算机视觉,信息检索.



俞能海(1964—),男,博士,教授,博士生导师,主要研究领域为多媒体信息检索.