

文章编号: 1004-4353(2018)02-0134-05

基于多尺度空间 LLC 的图像语义分类方法

宋正丹, 金小峰*
(延边大学 工学院, 吉林 延吉 133002)

摘要: 为了提高图像的空间分布和语义信息的有效利用,采用金字塔模型提出一种将多尺度空间、LLC 和图像语义分析相融合的图像语义分类方法. 首先对图像空间域金字塔划分的各个层次的局部块分别进行线性局部稀疏编码,并对不同层次上的量化编码进行串接生成共生矩;其次使用概率潜在语义模型对图像进行语义分析以获得最终的图像表示;最后采用线性多类别 SVM 对图像进行分类. 实验结果表明,本文提出的算法生成的图像特征具有较高的分类性能和效率.

关键词: LLC; 图像语义; 空间金字塔; 概率潜在语义模型; 多尺度空间

中图分类号: TP391.41 **文献标识码:** A

Semantic image classification method based on locality-constrained liner sparse coding in multi-scale space

SONG Zhengdan, JIN Xiaofeng*
(College of Engineering, Yanbian University, Yanji 133002, China)

Abstract: Adopting pyramid model, in order to make up for these problems, this paper proposes an image semantic classification method by fusing locality-constrained linear coding (LLC) based on multi-scale space and image semantic analysis. Firstly, locality-constrained linear coding is adopted to quantify local features of each block by using spatial pyramid approach, in order to produce the co-occurrence matrix by concatenating all local block. Secondly, the probabilistic latent semantic analysis (PLSA) is used to extract the latent semantic information to obtain the final image representation. Finally, the linear support vector machine (SVM) classifier is adopted in to improve the classification performance. Experimental results demonstrate that our approach has high classification performance and time efficiency.

Keywords: locality-constrained linear coding (LLC); image semantic; spatial pyramid; probabilistic latent semantic analysis (PLSA); multi-scale space

面对海量的图像数据,如何使计算机模拟人类对图像的认知方式,自动地把图像分类到不同的语义类别成为图像语义分析的一个关键问题. 目前,图像语义分类的典型方法大多是基于词袋模型(bag of word, BoW)^[1],但由于该方法丢弃了图像局部区域的空间位置关系,其表示能力大大降低. 为克服这一问题, S. Lazebnik 等提出了空间金字塔模型(spatial pyramid matching, SPM)^[2], 该模型在不同尺度下将图像划分成均等的子区域,并通过计算每个子区域的直方图,最终将所有子区域的直方图特征向量拼接成一个特征向量进行分类,以此获取图像更多的结构局部信息. 2009 年, Yang 等^[3]对 SPM 进行改进并提出了 ScSPM 算法,该算法克服了基于最近邻矢量量化(vector quantization, VQ)^[4]等传统方式在特征编码时重构误差大、优化时无法得到解析解和算法效率低

等问题,提高了图像分类效果.随后,Yang 等^[5]又通过局部限制的方法对稀疏编码进行了改进,使编码速度有了大幅提高.针对图像分类中出现的多义词和同义词问题,Li Feifei 等^[6]在 BoW 模型基础上使用概率潜在语义模型(probabilistic latent semantic analysis, PLSA)对图像进行了分类,通过概率生成模型进一步挖掘图像语义词汇之间的语义分析提高了图像语义特征的表达能力.上述研究方法都试图对图像特征编码模型进行改进,但由于采用了图像的底层特征,因此无法解决图像分类中普遍存在的语义鸿沟问题.为了进一步提取更具语义特征的图像信息,本文提出一种基于多尺度空间 LLC(locality-constrained linear coding, LLC)的图像语义分类方法,并通过实验验证该算法对图像分类的有效性.

1 基于 LLC 的图像多尺度空间语义分析

Yang 等^[5]研究表明,多层的图像表示模型能够提高图像分类效果;因此,本文将图像特征提取分为3层.将提取的图像底层 dense SIFT 特征描述子视为第1层编码,将 BoW 和 ScSPM 等方法生成的 LLC 作为第2层编码,将融合 PLSA 生成的最终表示作为第3层编码.显然,第3层编码能够实现图像潜在的语义信息的挖掘,有利于提高图像分类的准确率.经过3层编码提取语义信息后,使用 SVM 分类器进行图像分类.图1是本文提出的图像语义分类方法的流程图.

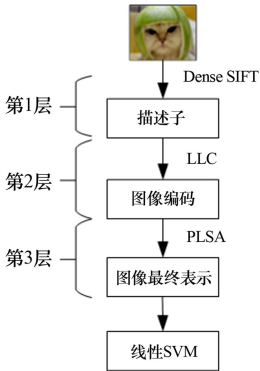


图1 图像分类流程

1.1 特征编码

为了降低传统的基于语义分类方法中矢量量化的重构误差,本文利用稀疏编码将矢量量化的约束条件 $\|s_i\|_0=1$ 放宽,并且利用 L_1 范式来正规

化稀疏编码(sparse coding, SC)^[5].假设 \mathbf{X} 是一个 K 维的图像局部特征提取描述子, $\mathbf{X}=[\mathbf{x}_1, \dots, \mathbf{x}_N] \in R^{K \times N}$, $\mathbf{B}=[\mathbf{b}_1, \dots, \mathbf{b}_M] \in R^{K \times M}$ 是字典中 M 个超完备基的集合,则优化模型为:

$$\min_{s_i} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{s}_i\|^2 + \lambda \|\mathbf{s}_i\|_{L_1}, \quad \|\mathbf{b}_k\| \leq 1 \text{ for } \forall k=1, \dots, M. \quad (1)$$

其中 $\mathbf{S}=[\mathbf{s}_1, \dots, \mathbf{s}_N]$ 是对 \mathbf{X} 的编码, λ 是正则化参数.

研究表明,稀疏编码得到的超完备基有助于提高图像的重构效果,同时编码的稀疏性有利于获取描述子的显著性^[7].另外,局部坐标编码(local coordinate coding, LCC)理论^[8]认为,在某些情况下局部性比稀疏性更有必要,同时局部性可以保证稀疏性.基于此,LLC方法使用局部限制项代替式(1)中的稀疏约束项,优化目标函数则变为:

$$\min_{c_i} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|^2 - \lambda \|d_i \odot \mathbf{c}_i\|^2, \quad \mathbf{A}^T \mathbf{c}_i = 1 \text{ for } \forall i=1, \dots, N. \quad (2)$$

其中 \mathbf{A} 是全部元素为1的向量; \mathbf{c}_i 表示系数; \odot 表示对应元素相乘; d_i 是局部调节系数,它对每一个基向量编码系数相似的描述子赋予不同的尺度.

$$d_i = \exp\left[\frac{\text{dist}(\mathbf{x}_i, \mathbf{B})}{\sigma}\right], \quad i=1, \dots, N, \quad (3)$$

式中 $\text{dist}(\mathbf{x}_i, \mathbf{B}) = [\text{dist}(\mathbf{x}_i, \mathbf{b}_1), \dots, \text{dist}(\mathbf{x}_i, \mathbf{b}_M)]^T$, $\text{dist}(\mathbf{x}_i, \mathbf{b}_j)$ 是 \mathbf{x}_i 和 \mathbf{b}_j 的欧氏距离, σ 是调节衰减的系数.

在对式(2)进行优化过程中,通常将 d_i 减去 $\max(\text{dist}(\mathbf{x}_i, \mathbf{B}))$,使其规范于 $(0, 1]$,并约束 $\mathbf{A}^T \mathbf{c}_i = 1$ 以保证编码的平移不变性.需要注意的是,LLC在 l_0 范式意义上并不属于稀疏编码,只是保证解的稀疏性,在实际处理时,只是简单地用阈值法将一些小的无影响的值置零.因此,为更好地让式(2)得到闭合解,将 \mathbf{c}_i 按式(4)计算求得:

$$\begin{cases} \tilde{\mathbf{c}}_i = (\mathbf{C}_i + \lambda \text{diag}(\mathbf{d})) \backslash \mathbf{A}, \\ \mathbf{c}_i = \tilde{\mathbf{c}}_i / \mathbf{A}^T \tilde{\mathbf{c}}_i, \end{cases} \quad (4)$$

其中 $\mathbf{C}_i = (\mathbf{B} - \mathbf{A}\mathbf{x}_i^T)(\mathbf{B} - \mathbf{A}\mathbf{x}_i^T)^T$.

由式(4)即可保证 LLC 得到解析解,模型编码效率明显高于 ScSPM.图2描述了 SC 以及 LLC 两种编码映射方式的差异.相较于 SC 选择不同的原子作为线性组合表示,LLC 是将距离相近的描述子用视觉词典中多个相同的原子来表

示,即相似的描述子共享相似的编码,这样就可同时保证稀疏性和局部性。

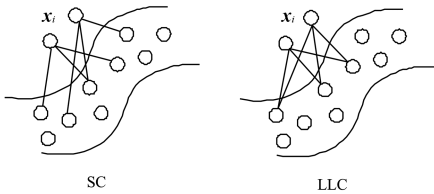


图 2 SC 和 LLC 的映射方式

1.2 图像语义提取

为了降低文本中同义词和多义词对文本语义理解的影响,Hofmann 提出了 PLSA 模型^[9],该模型较好地解决了“同义词”问题,并推广直图像词汇,因此本文采用 PLSA 建模并用期望最大化(expectation maximization, EM)算法进行参数估计。

假设有训练图像集 $D=[d_1,\cdots,d_N]$,其中 d_i 表示第 i 幅图像, $B=[b_1,\cdots,b_M]$ 表示生成的视觉词汇,PLSA 在生成词项-图像矩阵中引入了一个隐含变量(潜在主题) $Z=[z_1,\cdots,z_K]$,则对应的联合概率可由式(5) 计算得到:

$$P(d_i,z_k,b_j)=P(d_i)P(z_k|d_i)P(b_j|z_k). \tag{5}$$

由式(5) 可观察到变量的联合概率可以通过边缘化潜在主题 z_k 得到,即:

$$P(d_i,b_j)=P(d_i)\sum_{k=1}^K P(z_k|d_i)P(b_j|z_k). \tag{6}$$

其中: $P(d_i)$ 表示第 i 幅图像概率, $P(b|z)$ 为潜在语义在视觉词汇上的分布概率,它表示各个主题特征且对集外的图像仍适用; $P(z|d)$ 为图像中的潜在语义分布概率。

图 3(a) 为 PLSA 的模型图,图 3(b) 为图 3(a) 对应的矩阵分解图解。

利用 EM 算法对式(6) 参数进行估计,得

$$L=\sum_{i=1}^N\sum_{j=1}^M n(d_i,b_j)\log P(d_i,b_j), \tag{7}$$

其中 $n(d_i,b_j)$ 表示图片 d_i 中 b_j 的个数,EM 算法的 E 步骤和 M 步骤表示如下:

E 步骤 利用前面估计的模型参数,计算给定观察对 (d_i,b_j) 时,潜在主题 z_k 的条件概率分布为

$$P(z_k|d_i,b_j)=\frac{P(z_k|d_i)P(b_j|z_k)}{\sum_{l=1}^K P(z_l|d_i)P(b_j|z_l)}. \tag{8}$$

M 步骤 利用新的期望值 $P(z|d,b)$ 更新参数 $P(b|z)$ 和 $P(z|d)$:

$$P(b_j|z_k)=\frac{\sum_{i=1}^N n(d_i,b_j)P(z_k|d_i,b_j)}{\sum_{m=1}^M\sum_{i=1}^N n(d_i,b_m)P(z_k|d_i,b_m)}, \tag{9}$$

$$P(z_k|d_i)=\frac{\sum_{j=1}^M n(d_i,b_j)P(z_k|d_i,b_j)}{\sum_{j=1}^M n(d_i,b_j)}. \tag{10}$$

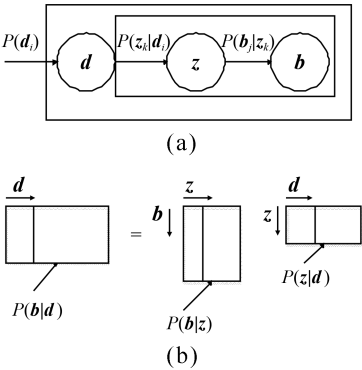


图 3 PLSA 模型

对于参数 $P(b|z)$ 和 $P(z|d)$,如果已知其中的一个分布,则另一个分布可以使用 folding-in 算法计算得到. folding-in 算法在迭代过程中保持已知参数不变,通过不断更新位置参数得到最大的 L 似然函数.经过上述方法最终得到 $P(z|d)$,通过它即可表示图像的最终特征^[10].

2 图像多尺度空间语义分析模型

结合图像的空间分布以及语义信息能够准确地对图像进行分类识别^[11],据此本文进一步分析图像空间分布,计算相应的位置是否包含同种语义信息,以提高图像分类性能.图 4 为本文构造的图像多尺度空间语义学习模型,其具体的算法描述如下:

- Step1 通过图像金字塔多尺度分割得到局部块;
- Step2 首先对每个局部块进行编码表示,然后对每个局部块所得的编码进行加权拼接,得到图像的底层特征表示;
- Step3 利用 PLSA 模型对已得到的特征表示进行分析,得到图像的最终特征表达;
- Step4 采用线性 SVM 进行训练.

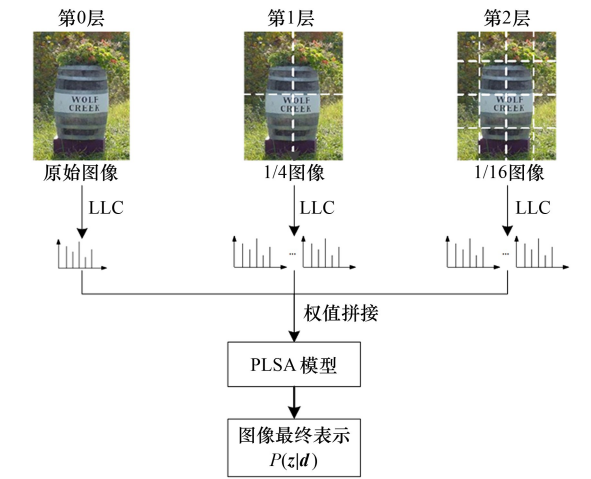


图 4 图像多尺度语义学习模型

金字塔分割是对图像第 l 层($l=0, \dots, L-1$)划分出 2^l 个局部块. 本文以一个 $L=3$ 层的空间金字塔模型为例进行说明: 第 0 层表示的是原图, 第 1 层是对图像进行 2×2 分块划分, 第 2 层是对图像进行 4×4 分块划分; 本文采用文献[2]中的方法设置加权拼接中的权值, 即将第 0 层的系数设置为 $1/2^{L-1}$, 第 l 层设置为 $1/2^{L-l-1}$, 因此 3 层金字塔结构的权值系数为 $[0.25, 0.5, 1.0]$.

测试阶段和训练过程类似, 只是在后期计算 PLSA 模型中的 $P(z|d)$ 时需要使用训练阶段所得到的 $P(b|z)$. 再进一步经 folding-in 算法计算, 最终得到 $P(z|d)$, 这个结果对训练集外的图像也仍有效. 图 5 为基于多尺度空间 LLC 的语义图像分类模型.

3 实验结果与分析

为了验证本文算法的有效性和鲁棒性, 将本文方法与 ScSPM、PLSA 和 LLC 等方法进行对比实验. 实验数据采用 Caltech-101 图像集中的 5 类共 240 幅自然场景图像, 随机抽取每类中 30 张图片共 150 幅用作训练集, 剩余的 90 幅作为测试集. 另外为了减少运算量, 每幅图像均转换为灰度并在保持原比例前提下缩小至 300×300 像素以内. 提取 dense SIFT 特征时, patch 定义为 16×16 , $\text{stepsize}=8$.

分类器采用 Liblinear^[12], 该分类器支持多种 $l_1 l_2$ 规范的线性 SVM, 能够提高分类效率和速度, 其参数通过交叉验证的方法确定. 将 5 次随机生成的训练集和测试集获得的平均分类准确率和

标准差作为评价标准. 图 6 是语义主题数目对分类精度的影响. 从图 6 可以看出: 主题数目对分类精度影响很大, 主题数目太小或太大时均对图像的描述较弱; 当主题数目 $K=30$ 时, 精度达到最高.

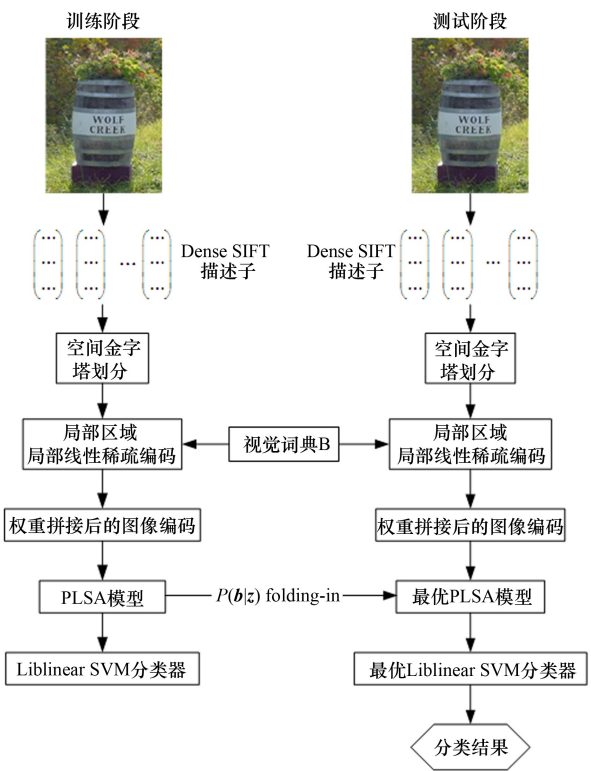


图 5 多尺度空间 LLC 的语义图像分类模型

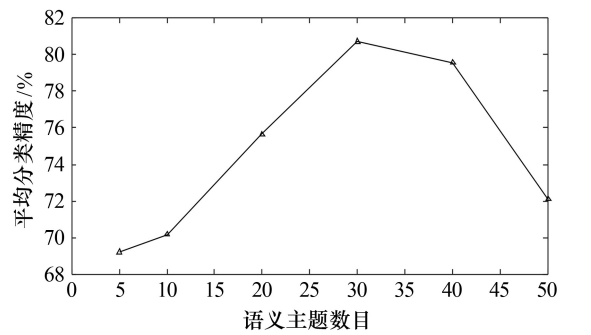


图 6 语义主题数目对分类精度的影响

为了进一步证明多尺度空间金字塔匹配对提升图像分类精度的影响, 将本文方法和 ScSPM + PLSA 分别在单尺度和多尺度空间进行对比实验. 实验中金字塔分层 $L=3$, 主题数目 $K=30$, 精度选择 5 次平均分类精度, 实验结果见表 1. 由表 1 可知: 图像的多尺度划分粒度越细, 两种分类方法的精度越高, 且本文方法的分类精度优于 ScSPM + PLSA.

表 1 多尺度空间与单层空间对分类精度的影响

	分类精度/%	
	ScSPM+PLSA	本文方法
单层 $L=0$	73.56	76.70
金字塔分块 $L=0,1$	78.57	79.36
金字塔分块 $L=0,1,2$	79.34	80.69

文献[3]研究证明,ScSPM 的时间复杂度为 $O(M \times K)$, 其中 K 为非零元素个数. 文献[5]中利用近似方法使得 LLC 的时间复杂度相较于 ScSPM 降低到了 $O(M+K^2)$, 其中 $M \gg K$, K 为优化算法中最近邻字典原子个数. 表 2 为 ScSPM 与多尺度空间 LLC 的图像表示生成时间, 由表 2 可以看出, 多尺度空间 LLC 要明显低于 ScSPM, 证明了 LLC 算法的高效性.

表 2 ScSPM 与 LLC 的图像表示生成时间

方法	时间/s
ScSPM	624.46
多尺度空间 LLC	0.15

为了验证 PLSA 分析方法对提高分类精度的有效性, 本文实验模型采用最优主题数目 30 和较优金字塔层次 $L=3$. 表 3 为本文方法与 ScSPM、多尺度空间 LLC 和 ScSPM+PLSA 的分类精度. 由表 3 可知, 本文方法的分类精度高于多尺度空间 LLC、ScSPM 和 ScSPM+PLSA 3 种方法, 这说明加入语义信息的学习方法能够提高图像的分类准确率.

表 3 不同算法的分类精度比较

方法	准确率/%
ScSPM	76.12
多尺度空间 LLC	77.06
ScSPM+PLSA	79.34
本文方法	80.69

4 结论

本文提出的基于多尺度空间、LLC 和图像语义分析相融合的图像语义分类方法, 提高了图像分类速度和分类精度, 且优于 ScSPM、多尺度空间 LLC 和 ScSPM+PLSA 3 种方法, 因此本文方法在图像语义分析领域中具有一定的应用价值. 但本文方法中的 PLSA 主题数目以及词典规模都是采用经验值, 使得模型缺乏自主性, 在今后的

研究中我们将探讨基于自适应的主题数目及分层数目的选取方法.

参考文献:

[1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2):91-110.

[2] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]//IEEE Computer Society Conference on Computer Vision & Pattern Recognition. New York: IEEE Computer Society, 2006: 2169-2178.

[3] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami: IEEE Computer Society, 2009:1794-1801.

[4] Leborgne H, Shabou A. Locality-constrained and spatially regularized coding for scene categorization [C]//IEEE Conference on Computer Vision and Pattern Recognition. Rhode Island: IEEE, 2012: 3618-3625.

[5] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification [C]//IEEE Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010:3360-3367.

[6] Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories[J]. Computer Vision and Image Understanding, 2007,106(1):178.

[7] 焦李成,赵进,杨淑媛,等. 稀疏认知学习、计算与识别的研究进展[J]. 计算机学报,2016,39(4):835-852.

[8] Yu K, Zhang T, Gong Y. Nonlinear learning using local coordinate coding[C]//International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc, 2009:2223-2231.

[9] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis [J]. Machine Learning, 2001,42(1/2):177-196.

[10] 李志欣,施智平,李志清,等. 融合语义主题的图像自动标注[J]. 软件学报,2011,22(4):801-812.

[11] 赵仲秋,季海峰,高隽,等. 基于稀疏编码多尺度空间潜在语义分析的图像分类[J]. 计算机学报,2014,37(6):1251-1260.

[12] Yu H F, Hsieh C J, Chang K W, et al. Large linear classification when data cannot fit in memory [J]. Acm Transactions on Knowledge Discovery from Data, 2012,5(4):23.