

稀疏表示的图像分类研究综述

周 近

(江苏第二师范学院,江苏 南京 210013)

摘要:良好的特征提取方法能减轻后续图像分类与识别的工作量。针对具体的分类问题提出了不同的特征提取方法,并在图像分类和识别任务上取得了较好的效果。然而,已有的基于传统方法的特征提取存在一些明显不足,即随着视觉任务规模的增大,直接利用这些传统方法进行特征分类,效果并不理想。提出的特征表达方法,在图像最基本特征基础上进行矢量量化、稀疏编码或其它表达以形成一幅图像最后的特征。着重介绍基于稀疏表示的特征分类算法并对其进行分析,最后探讨存在的问题和今后研究的方向。

关键词:稀疏表示;图像分类;稀疏编码;特征编码

中图分类号:TP391.41 **文献标识码:**A **文章编号:**1671-5322(2015)03-0047-05

图像的分类作为计算机视觉领域的重要组成部分,能够有效地对图像的内容进行分析,获取图像中的关键信息,并给出正确的判断,对现实的工作生活及社会的发展具有重要的意义^[1]。图像分类包括图像的预处理、图像特征的提取、特征的降维及特征的选择、分类器的设计等步骤。其中,视觉特征的提取是最基础的工作也是最关键的步骤,它能有效地减轻对后续环节的依赖性,同时也制约着整个系统的性能表现。因此,针对具体的视觉问题或一系列的视觉任务,如何有效地提取特征是计算机视觉领域的一个热点和难点。

特征提取^[2]是一个综合性的研究课题,涉及到图像处理、模式识别、机器学习、生物学等领域的知识。其中,特征表达是指在最基本的特征基础上进行统计(矢量量化)、编码或其它方法以形成一幅图像最后的特征,通常情况下会比原始的基本特征具有更好的性能;而特征学习是指从大量的图像样本数据中采用不同规模的网络结构以及各种学习规则来学习出特征,具有方法上的统一性,即针对不同类型的输入图像,可以采用同样的结构进行特征提取,不再需要人工设计特征。

近年来,稀疏编码理论已经成为信号分析处理、计算机视觉分析、模式识别与控制等国内外学

术界的重点研究方向。它作为一种降维方法,有效性高于一般算法,因此,稀疏编码常被应用于图像特征提取。根据国内外的相关研究,基于稀疏编码的图像特征提取框架可以分为特征提取(Feature Extraction)^[3]、词典学习(Dictionary Learning)^[4]、特征编码(Feature Coding)^[5]和特征汇总(Feature Pooling)^[5]这几个步骤。其中,特征编码是基于视觉词典 $D(D=[d_1,d_2,\cdots,d_m]\in R^{d\times m})$,式中 d 与特征的维度相同,是视觉单词的个数,在满足特定性质的条件下一般设定的性质包括致密性、稀疏性或统计独立性,将图片的 n 个高维的局部特(Local Feature)征转换为编码 $C=[c_1,c_2,\cdots,c_m]\in R^{m\times n}$ 。不同的编码技术具有不同的规则和性质,因而对同一个局部特征会产生不同的编码。近来,出现很多基于改进词袋模型(Bag of Words, BoW)^[6]的新方法,它们不仅能够更加精确地表示图像而且能够提高这些方法的分类能力。图像虽然在图像分类方面取得了显著的进展,但是在如何对一幅图像的局部特征进行编码方面仍然有很大的提升空间。近年来各种不同的编码方法可以分成 3 类:

1) 基于重构(Reconstruction)的编码方式:稀疏编码(Sparse Coding, SCSPM)^[7]、局部约束线

性编码 (Locality - constrained Linear Coding, LLC)^[8]、局部约束稀疏自编码器 (Locality - constrained Sparse Auto - Encoder, LSAE)^[9]、局部约束和空间规范的编码 (Locality - Constrained and Spatially Regularized Coding, LCSRC)^[10]、低秩稀疏编码 (Low - Rank Sparse Coding, LRSC)^[11];

2) 基于配置 (Assignment) 的编码方式: 硬分配编码 (Hard - assignment Coding, HC)^[6]、软分配编码 (Soft - assignment Coding, SaC)^[12]、局部软分配编码 (Localized Soft - assignment Coding, LSC)^[13];

3) 基于显著性 (Salient) 的编码方式: 显著编码 (Salient Coding, SC)^[14]、组显著编码 (Group Salient Coding, GSC)^[15]。

下面对常见的几类编码做详细的介绍。

1 几种特征编码方法

定义 $\mathbf{b}_i \in R^d$ 是字典中的视词 (基), 这里 d 代表局部特征的维度; 矩阵 $\text{MatrixB} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n]$ 代表字典或者说是视词的集合, n 代表视词的个数; $\mathbf{x}_i \in R^d$ 是一幅图像的第 i 个特征; $\mathbf{z}_i \in R^n$ 是 $\mathbf{x}_i$ 的编码, 其中 z_{ij} 是 $\mathbf{b}_j$ 的系数。

1.1 硬分配编码 (Hard - assignment Coding, HC)

对于一个局部特征 $\mathbf{x}_i$ 有且仅有一个非零系数, 根据预先设定好的距离寻找到离它最近的视词 (基) 来给它编码。该距离一般采用欧氏距离:

$$z_{ij} = \begin{cases} 1 & \text{if } j = \arg \min_{j=1, \dots, n} \|\mathbf{x}_i - \mathbf{b}_j\|_2^2 \\ 0 & \text{其他} \end{cases} \quad (1)$$

在视觉字典固定的情况下, 硬分配编码方法使用只含有 1 个非零元素的向量来表示 1 个特征, 不可避免地会产生很大的量化误差。而且在分类时若直接采用线性分类器, 结果往往很差; 而若采用非线性支持向量机, 计算复杂度就会很高。

硬分配编码方法会导致两个问题: 视觉单词的不确定性和视觉单词的合理性。视觉单词的不确定性是指图像局部特征到两个或几个视觉单词的距离都很近, 要从这两个或几个视觉单词中选出一个“正确的”, 硬分配编码方法只选择能最好表达局部特征的一个视觉单词, 而忽略其他候选视觉单词的相关性; 视觉单词的合理性是指图像局部特征到视觉词典中所有视觉单词的距离都很远, 但仍然要从这些不合理的候选者中选出一个去代表它。硬分配编码方法给出了对于每个局部

特征视觉词典中最合适的单词, 但是忽略了这个所谓“最合适”的视觉单词很可能并不真正代表局部特征。

1.2 软分配编码 (Soft - assignment Coding, SaC)

该方法的第 j 个编码系数代表了局部特征 $\mathbf{x}_i$ 和第 j 个视词之间的匹配程度, α 是平滑系数, 同时定义有 n 个视词用来计算 z_{ij} :

$$z_{ij} = \frac{\exp(-\alpha \|\mathbf{x}_i - \mathbf{b}_j\|_2^2)}{\sum_{k=1}^n \exp(-\alpha \|\mathbf{x}_i - \mathbf{b}_k\|_2^2)} \quad (2)$$

该方法通过稍微降低表示的稀疏性, 来达到高表示的性能的目的, 虽然计算速度会降低, 但是分类的准确性有较大的提升。

1.3 局部软分配编码 (Localized Soft - assignment Coding, LSC)

该方法基本思想是采用局部特征在字典中的 k 个近邻视词来更新 Soft - assignment Coding:

$$z_{ij} = \frac{\exp(-\alpha \|\mathbf{x}_i - \mathbf{b}_j\|_2^2)}{\sum_{k=1}^n \exp(-\alpha \|\mathbf{x}_i - \mathbf{b}_k\|_2^2)}, \mathbf{b}_k \in N_k(\mathbf{x}_i) \quad (3)$$

实验证明, 该方法在分类问题上取得了非常好的效果, 但是它并没有真正解决视词的合理性问题, 有可能这 k 个视词与局部特征之间的距离都很远 (都不相似), 而经过正则化之后变成了噪声, 从而影响了编码效果。

1.4 稀疏编码 (Sparse Coding, SCSPM)

该方法使用稀疏集合和字典中基向量的线性组合来表示局部特征 $\mathbf{x}_i$, 在求解系数向量 $\mathbf{z}_i$ 时, 加入了 l_1 范数的正则项:

$$\mathbf{z}_i = \arg \min \|\mathbf{x}_i - B\mathbf{z}_i\|_2^2 + \lambda \|\mathbf{z}_i\|_1 \quad (4)$$

该方法重建性能好, 稀疏特征更加线性可分。实验结果表明对较大的字典, 该方法表现出更好的性能, 与前面提到的向量化方法相比, 该方法造成的量化损失较小。稀疏性是由正则项控制的, 以保证学习到的表达能捕捉到局部特征的显著模式。但是为了保证稀疏性, 它非常有可能对相似的局部特征选择完全不同的视词, 导致编码之间的相关性较弱。

1.5 局部约束线性编码 (Locality - constrained Linear Coding, LLC)

和上面说的 SCSPM 方法不同, 该方法考虑更多的是局部性而不是稀疏性, 这就造成了相关系数小的基向量远离 $\mathbf{x}_i$, 用一个局部约束项来代替

SCSPM 中的稀疏约束项,编码 $\mathbf{z}_i$ 的计算就变成了解决以下正则化的最小二乘问题:

$$\mathbf{z}_i = \arg \min_{\substack{\mathbf{z}_i \\ \|\mathbf{z}_i\|_1 = 1}} \|\mathbf{x}_i - B\mathbf{z}_i\|_2^2 + \lambda \|\mathbf{d}_i \odot \mathbf{z}_i\|_2^2 \quad (5)$$

其中,

$$\mathbf{d}_i = \exp(\text{dist}(\mathbf{x}_i, B)/\delta, \text{dist}(\mathbf{x}_i, B)) = (\text{dist}(\mathbf{x}_i, \mathbf{b}_1), \text{dist}(\mathbf{x}_i, \mathbf{b}_2), \dots, \text{dist}(\mathbf{x}_i, \mathbf{b}_n))^T \quad (6)$$

$\text{dist}(\mathbf{x}_i, \mathbf{b}_j)$ 定义为 $\mathbf{x}_i$ 和 $\mathbf{b}_j$ 之间的 l_2 距离, δ 是一个正的参数,用于调整局部性适配器的权重衰退速度。上式中第 1 项限制了重建损失,第 2 项保证了相似的局部特征可以获得相似的编码。该方法提出后,在图像分类中取得了非常好的效果,它的优势在于:具有良好的重建性能,具有局部的光滑性,编码具有解析解(不需要迭代)。

1.6 拉普拉斯稀疏编码 (Laplacian Sparse Coding, LScSPM)

这是第一种对稀疏编码一致性进行改进的方法,它考虑了数据库中相似的局部特征应该具有相似的稀疏编码。该算法通过在 LASSO 问题中加入图嵌入的正则项来实现,字典的学习和稀疏编码的求解可以通过交替迭代得到:

$$(B, Z) = \arg \min_{B, Z} \|X - BZ\|_2^2 + \lambda \sum_i \|\mathbf{z}_i\|_1 + \beta \text{tr}(ZLZ^T) \quad (7)$$

s. t. $\|\mathbf{b}_j\|_2^2 \leq 1, \forall j = 1, \dots, n$ 这里 L 是拉普拉斯矩阵,它在编码时考虑了局部特征之间的关系, i 为样本数目。由于数据库中的特征非常多,当构建拉普拉斯矩阵和学习稀疏编码同时进行计算上是不可行的,所以产生了一些具有启发性的措施来改善计算复杂度。

1.7 显著编码 (Salient Coding, SC)

这是一种替代的编码方式,通过设置距离 $\mathbf{x}_i$ 最近的几个字典中基 $\mathbf{b}_j$ 的显著性程度来体现局部性特点。这里, $\varphi(\cdot)$ 是一个单调递减的函数, $\{\hat{\mathbf{b}}_m\}_{m=1, \dots, k}$ 是 $\mathbf{x}_i$ 的 k 个最近邻组成的集合:

$$z_{i,j} = 1 - \varphi\left(\frac{\|\mathbf{x}_i - \mathbf{b}_j\|_2^2}{\frac{1}{k-1} \sum_{m \neq j}^k \|\mathbf{x}_i - \hat{\mathbf{b}}_m\|_2^2}\right) \quad (8)$$

1.8 组显著编码 (Group Saliency Coding, GSC)

定义 $s_i^{(k)}$ 是 $\mathbf{f}_i$ (特征向量) 根据 group code 的大小得到的编码结果, $\phi^{(k)}(\mathbf{f}_i)$ 是经过改进的显著性程度的描述函数

$$s_i(j) = \begin{cases} \phi(\mathbf{f}_i) & \text{if } j = \arg \min_j \|\mathbf{f}_i - \mathbf{c}_j\|_2 \\ 0 & \text{其他} \end{cases} \quad (9)$$

$$\phi(\mathbf{f}_i) = \frac{\sum_{t=2}^{K_s} (\|\mathbf{f}_i - \tilde{\mathbf{c}}_t\|_2 - \|\mathbf{f}_i - \tilde{\mathbf{c}}_1\|_2)}{\sum_{t=2}^K \|\mathbf{f}_i - \tilde{\mathbf{c}}_t\|_2} \quad (10)$$

$g(\mathbf{f}_i, k)$ 是由距离 $\mathbf{f}_i$ 最近的 k 个视词组成的 codewords, k 是 group - code 尺寸的最大值。

GaSC 的主要思想是考虑了 $\mathbf{f}_i$ 的 group - code 和其他 codewords 之间的相对位置。对于不同的 k , 总能找到固定的 $k+1$ 个近邻 codewords 来给每个特征编码, 这 k 个最近的基放在 group - codes 中, 其中, 第 k 个 codeword 作为 group - code 中最具代表性的元素用以计算显著性的程度。

1.9 低秩稀疏编码 (Low - Rank Sparse Coding, LRSC)

定义 X 是由 SIFT 描述子组成的矩阵, 每一列代表一个局部特征点, 通常是 128 维的, D 为字典, 在没有噪声的情况下, 局部特征 $\mathbf{x}_i$ 可以由字典中的基线性表示, 写成矩阵的形式就是: $X = DZ$

$$\min_Z \|X - DZ\|_F^2 + \lambda_1 \|Z\|_* + \lambda_2 \|Z\|_{1,1} \quad (11)$$

其中 $\|Z\|_*$ 是核范数, 用来刻画低秩, 它通过 l_1 范数来加入稀疏性, 即: $\|Z\|_{1,1} = \sum_{i=1}^n \|\mathbf{Z}_i\|_1$, 系数 λ_1 和用以权衡在特征编码中的稀疏性和低秩性。这样特征编码的问题就转化为一个求解稀疏低秩方程的优化问题。该方法的优点是它考虑了稀疏性, 也考虑了局部特征的相似性(从字典中选择近邻的基, 这和 LSC, LLC 相似), 同时还加入了空间一致性的信息(通过在局部区域加入低秩约束, 使得空间关系上紧邻的点具有相似的特征编码)。

1.10 局部约束稀疏自编码器 (Locality - constrained Sparse Auto - Encoder, LSAE)

在基于字典的编码方法中, 当给定字典 D 后, 往往需要求解一个优化算法来获得输入 x 的码字, 其计算复杂度随着字典规模增大而增大。另一方面, 基于自动编码器的编码模式只需要简单的内积运算和一个非线性变换就能获取码字, 但其在编码过程中丢失了近邻性。因此把基于字典编码中的近邻性引入自动编码器中, 使得自动

编码器在快速获取码字的同时能充分利用近邻性,使得相似的输入能够采用相似的基来编码,从而使得最后的码字具有相似的结构。该方法的优点在于:

(1)相对于视觉领域的字典编码来说,LSAE能够快速学习字典,并且只需要简单的前馈操作就能获得码字;

(2)相对于传统的(稀疏)自动编码器来说,LSAE在获取码字过程中充分利用了输入与基之间的近邻性,使得码字更为鲁棒;

(3)相对于 LLC 自动编码器来说,其学习效率更高。

存在的问题是:当字典规模非常大时,有些基仍很难学习到有用的特征。

2 编码方法讨论

现阶段的研究表明:给定一个字典,局部特征的编码方法将显著影响着分类性能。最早的方法是硬分配编码(HC),这种方法虽然简单,但是它对于字典的选择十分敏感,而且量化误差较大,因此出现了一个更加鲁棒的方案软分配编码方法(SC),它在提高分类性能的同时付出了一定的时间代价。为了改进软分配和硬分配这两种编码方法,通过稀疏学习的技术在图像局部特征编码中引入了稀疏性^[7],但是稀疏编码非常耗时而且常常导致编码不稳定,比如,具有相似描述子的局部特征会有不同编码。为了消除这种不稳定性,有人提出了另一种编码方案^[15]。这种体现局部性的编码方法让那些代表局部特征的视词(字典中的基)和特征描述子之间尽可能地相似,这就需要从原来的字典中选择和特征描述子最近邻的几

个视词作为特征字典,这样图像的描述子就可以用这些局部选择出来的基来编码。然而,在前面所提到的这些编码方案中,每个特征的编码都是相对独立的。在拉普拉斯稀疏编码(LScSPM)^[16]方法中,在引入稀疏限制的同时还考虑了局部特征之间的全局相似性,但是这种方法在特征集合较大的时候计算量巨大,而且没有结合相应特征的稀疏编码和空间布局之间的关系。空间一致性要求在图像中空间位置相近的点应该具有相似的稀疏编码,并可以由相似的字典中的基来表示。鉴于这种思想又出现了低秩稀疏编码(LRSC),兼顾了稀疏性、局部性和空间一致性信息。显著编码(SaC)的提出主要是为了解决局部约束线性编码(LLC)在编码过程中存在信息的丢失问题所进行的改进,组显著编码(GSaC)又对显著编码存在不能处理较大字典的问题进行了修改。

3 总结

研究表明:有效的特征表达方法能够极大地改善视觉图像分类和识别的性能,而分类的准确率很大程度上依赖特征编码的具体方式。本文介绍了比较具有影响力的几种特征编码方法的原理、设计动机、优越性以及存在的问题。通过总结归纳发现,这些方法仍然是基于浅层模型的特征描述方法,它们的特征表达能力是有限的。所以最近深度学习在图像分类的任务中发挥了越来越重要的作用,主要因为它属于更深一层的模型,抽象效果更好,具有较强的特征表达能力,能够显著地提高分类的准确率,是一个非常前沿的研究方向且具有广泛的应用前景。

参考文献:

- [1] Lu D, Weng Q. A survey of image classification methods and techniques for improving classification performance[J]. International Journal of Remote Sensing, 2007,28(5):823-870.
- [2] Nixon M S, Aguado A S. Feature Extraction and Image Processing[M]. London: Elsevier,2008.
- [3] Yu N, Qiu T, Bi F, et al. Image Features Extraction and Fusion Based on Joint Sparse Representation[J]. IEEE Journal of Selected Topics in Signal Processing, 2011,5(5):1 074-1 081.
- [4] Yang M, Zhang L, Feng X, et al. Fisher Discrimination Dictionary Learning for sparse representation[C]//2011 IEEE International Conference on Computer Vision (ICCV), Piscataway:IEEE, 2011:543-550.
- [5] Gao S H, Tsang I W H, Chia L T. Sparse Representation With Kernels[J]. IEEE Transactions on Image Processing, 2013, 22(2):423-434.
- [6] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]//Computer Vision and Pattern Recognition, New York:IEEE, 2006:2 169-2 178.

- [7] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]. Computer Vision and Pattern Recognition, New York:IEEE, 2009;1 794 – 1 801.
- [8] Wang J, Yang J, Yu K, et al. Locality – constrained linear coding for image classification[C]//Computer Vision and Pattern Recognition (CVPR), New York:IEEE, 2010;3 360 – 3 367.
- [9] Luo W, Yang J, Xu W, et al. Locality – constrained Sparse Auto – Encoder for Image Classification[J]. IEEE Signal Processing Letters, 2015,22(8);1 070 – 1 073.
- [10] Shabou A, Borgne H L. Locality – constrained and spatially regularized coding for scene categorization[C]//Computer Vision and Pattern Recognition (CVPR), New York:IEEE, 2012;3 618 – 3 625.
- [11] Zhang T, Ghanem B, Liu S, et al. Low – Rank Sparse Coding for Image Classification[C]//Computer Vision (ICCV), Piscataway:IEEE, 2013;281 – 288.
- [12] Gemert J C V, Geusebroek J – M, Veenman C J, et al. Kernel Codebooks for Scene Categorization[C]//10th European Conference on Computer Vision, Heidelberg:springer Verlag, 2008;696 – 709.
- [13] Liu L, Wang L, Liu X. In defense of soft assignment coding[C]//Computer Vision (ICCV), Piscataway:IEEE, 2011: 2 486 – 2 493.
- [14] Huang Y, Huang K, Yu Y, et al. Salient coding for image classification[C]//Computer Vision and Pattern Recognition (CVPR), New York:IEEE, 2011;1 753 – 1 760.
- [15] Yu K, Zhang T, Gong Y. Nonlinear Learning using Local Coordinate Coding[C]//Advances in Neural Information Processing Systems 22 (NIPS 2009), Vancouver:Curran Associates, 2009;1 – 9.
- [16] Gao S, Tsang I W – H, Chia L – T, et al. Local features are not lonely – laplacian sparse coding for image classification [C]//Computer Vision and Pattern Recognition (CVPR), New York:IEEE, 2010;3 555 – 3 561.

Survey of Image Classification Based on Sparse Representation

ZHOU Jin

(Jiangsu Second Normal University, Nanjing Jiangsu 210013, China)

Abstract: Good feature extraction method can reduce the workload of subsequent image classification and recognition. Different feature extraction methods are proposed for the specific classification problem, and achieved good results in image classification and recognition tasks. However, there are some obvious shortcomings of the existing feature extraction based on the traditional method. With the increasing of the size of the visual task, direct use of these traditional methods for feature classification is not ideal. The feature expression method is proposed, which is based on the most basic features of the image, and the sparse encoding or other expressions are proposed to form a final image. Based on sparse representation and its analysis, this paper focused on the feature classification algorithm and finally discussed the existing problems and future research directions.

Keywords: sparse representation; image classification; sparse coding; feature coding

(责任编辑:李华云)