

引用格式:赵海英,张炼. 基于多层字典学习的服饰纹样多标签标注算法[J]. 中国传媒大学学报(自然科学版),2024,31(05):64-72.
文章编号:1673-4793(2024)05-0064-09

基于多层字典学习的服饰纹样多标签标注算法

赵海英*,张炼*

(北京邮电大学人工智能学院,北京 100876)

摘要:中国传统服饰作为中华文化重要的符号,图像内容丰富,在图像分类领域有着巨大的潜在价值。针对图像分类中传统判别字典学习方法存在的原子局部特征利用不足且单层结构在多标签分类上的局限性,提出了一种结合全局约束和支持向量引导的多层字典学习算法,实现单层到多层的过渡来获取高层复杂语义信息。该算法通过引入全局约束增强非相邻原子间相似性的考量,更精准地捕捉图像特征的非线性关系;同时采用支持向量判别项优化编码向量对的权重分配,提高模型的泛化能力。实验结果显示,此算法在明清服饰纹样数据集上相较于局部约束和单层字典学习方法,分类精度分别提高了3.33%和1.57%,在扩展耶鲁B人脸数据集上也展示了较高的分类准确度。

关键词:图像分类;字典学习;全局相似性;支持向量引导

中图分类号:TP319.56 文献标识码:A

Clothing pattern multi-label annotation algorithm based on multi-layer dictionary learning

ZHAO Haiying*, ZHANG Lian*

(School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: As an important symbol of Chinese culture, traditional Chinese clothing has rich image content and great potential value in the field of image classification. Aiming at the insufficient utilization of atomic local features in traditional discriminant dictionary learning methods in image classification and the limitations of single-layer structure in multi-label classification, a multi-layer dictionary learning algorithm combining global constraints and support vector guidance was proposed to achieve the transition from single layer to multi-layer to obtain high-level complex semantic information. The algorithm introduced global constraints to enhance the consideration of similarity between non-adjacent atoms and more accurately captured the nonlinear relationship of image features; at the same time, the support vector discriminant term was used to optimize the weight distribution of encoding vector pairs to improve the generalization ability of the model. Experimental results show that compared with local constraints and single-layer dictionary learning methods, the classification accuracy of this algorithm on the Ming and Qing clothing pattern dataset is improved by 3.33% and 1.57% respectively, and it also shows high classification accuracy on the extended Yale B face dataset.

Keywords: image classification; dictionary learning; global similarity; support vector guidance

基金项目:国家重点研发计划项目(2021YFF0901700)

作者简介(*为通讯作者):赵海英(1972-),女,博士,副教授,主要从事文化计算与媒体数据挖掘研究。E-mail: zhaohaiying@bupt.edu.cn;张炼(1999-),男,硕士生,主要从事文化计算与图像标注研究。E-mail: zhanglian@bupt.edu.cn

1 引言

中华民族传统文化博大精深,蕴含丰富的价值。其中,传统民族服饰作为文化的重要载体,不仅在设计上独具特色,而且在纹样上更是别具匠心,呈现出丰富多彩的文化图案。因此,研究服饰纹样标注任务是中华民族传统服饰文化表征的基础步骤和底层关键,不仅有助于保护和传承传统文化,更能够深入挖掘其内涵与价值,为文化的传承与发展提供重要的支撑。传统图像分类技术主要依赖于图像低层次的视觉特征信息,如图像的颜色和纹理等,通过这些特征构建模型,但是在处理复杂任务时效果并不理想。近年来,字典学习算法在图像分类、特征提取和目标检测等计算机视觉领域受到广泛关注,并取得了显著进展。得益于这些成果的促进,基于字典学习的分类算法能够逐步适应图像特征信息的多样性,可以捕获到更高层次的视觉信息,极大提升了数据处理的效率。

近年来字典学习作为一种先进的表征学习方法,在图像分类领域受到了广泛关注。Yang等人^[1]提出Fisher判别字典学习(FDDL, fisher discrimination dictionary learning),将Fisher判别标准强加于编码系数以学习特定类的子词典,减小样本类内离散度,增大类间离散度,使类别之间的区分度变强。为进一步提升性能,Cai等人^[2]通过联合学习多类支持向量机分类器提出了支持向量引导字典学习(SVGDL, support vector guided dictionary learning),将权重参数化与对称性、一致性和平衡约束相结合,让权重分配更具适应性和灵活性,实现更高效的分类。受SVGDL启发,Song等人^[3]提出了基于宽松支持向量的字典学习(RSVDL, relaxed support vector dictionary learning),通过构建与标签信息密切相关的松弛系数子空间来提升字典学习的判别能力。由于单层字典对于复杂数据表征能力的不足,后续研究逐渐从单层过渡到了多层领域。其中,Song等人^[4]提出具有局部约束的多层判别字典学习(MDDL, multi-layer discriminative dictionary learning)用于图像分类,性能明显优于单层结构。Zhou等人^[5]提出深度语义字典学习(DSDL, deep semantic dictionary learning),应用自动编码器从类级语义生成语义字典,并利用该字典通过标签嵌入表示卷积神经网络提取的视觉特征。Wang等人^[6]提出深度共享字典学习(DSDL, deep and shared dictionary learning),结合深度学习的深层结构和共享结构让学习到的字典具有更好的稀疏表示。Guo等人^[7]提

出的判别深度字典学习(DDDL, discriminative deep dictionary learning),引入类内和类间诱导机制学习多层字典之间的判别信息。在另一领域,Shaban A等人^[8]提出一种从局部相似性到全局相似性的转换方法,并将其应用于图像分类。

传统字典学习大多局限于单层编码框架,构建的特征空间不能充分捕获分类任务的本质。虽然目前一些多层编码技术通过增加表达维度提高了逼近能力,但仍面临编码判别性不足以及层间特征协同不足的挑战。此外,现有算法利用局部相似度约束来捕获数据流形结构的方法还有一定的优化空间。因此,本文提出基于联合全局约束和支持向量引导的多层字典学习算法(JGCSVG-MDL, join global constraint and support vector guide multi-layer dictionary learning)。该算法的优势有:1)引入全局约束,纳入非邻近原子相似性,深化特征表示的非线性关系,具有更好的类内相似性。2)对字典原子施加支持向量判别项,增大类间分散性,增强编码系数的判别能力。3)构建的多层字典结构能够捕捉数据的复杂结构和多尺度特征,可以更灵活地适应数据的多样性,学习到不同层次的特征,对数据中的噪声和异常值具有更好的鲁棒性。

2 相关工作

2.1 全局约束

传统字典学习常采用局部约束,虽有利于捕获数据流形的局域特性,但未能充分整合流形结构的全局关联。因此本文决定将其优化为全局约束机制,用其来弥补局部约束的局限性。数据的流形几何属性常通过邻域关联来表征,由原子的几何特性来继承训练样本中的结构信息。借鉴文献[9]采用高斯核方法,该方法通常用作流形学习方法中的局部相似性度量,构建局部相似性的公式如式(1)所示:

$$M_{ij} = \begin{cases} \exp\left(-\frac{\|d_i - d_j\|^2}{\sigma}\right) & \text{if } d_j \in \text{KNN}(d_i) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

其中, σ 代表控制局部性的参数,KNN(d_i)表示原子的K近邻, M_{ij} 反映原子 d_i, d_j 之间的局部相似性,其中 M 节点表示原子或特征,边表示节点之间的局部相似性。局部约束的缺点在于不相邻原子间的相似性均设为零,而引入全局约束能够很好的解决该问题。首先,在得到具有局部相似性矩阵 M 之后,随机矩阵

\mathbf{P} 可以通过归一化 \mathbf{M} 来定义^[10],公式如式(2):

$$\mathbf{P} = \mathbf{D}^{-1} \mathbf{M} \quad (2)$$

其中, \mathbf{D} 是对角矩阵, $D_{ii} = \sum_j M_{ij}$,根据归一化原理可以确保 \mathbf{P} 的每一行元素和固定为1。由于 \mathbf{M} 包含相似性特征,因此 \mathbf{P} 代表的是随机矩阵,其中 $P_{ij} = p(m_j|m_i)$ 表示为在均值为 m_i 且方差为 σ 的高斯分布中, m_j 属于其中的概率大小。为引入全局约束并评估非直接相邻原子间的相似性,可以通过在 \mathbf{M} 构建的图上实施随机游走策略来间接实现^[11]。具体公式如式(3):

$$\begin{aligned} p^{(2)}(m_k|m_i) &= \sum_{l=1}^c p(m_k, m_l|m_i) \\ &= \sum_{l=1}^c p(m_k|m_l) p(m_l|m_i) \end{aligned} \quad (3)$$

其中, $p^{(2)}(m_k|m_i)$ 表示 m_i 经过两步间接属于 m_k 的概率,即使 m_k 不在 m_i 的最近邻集合内,由此可得, P^2 所包含的元素即为2阶的间接相似度。推广更高阶的相似性可以通过迭代上述过程获得,利用前期迭代结果作为输入以递推地更新矩阵 \mathbf{P} 。第 t 次迭代公式如式(4):

$$p^{(t)}(m_k|m_i) = \sum_{l=1}^c p(m_k|m_l) p^{(t-1)}(m_l|m_i) \quad (4)$$

使用上述的方法不仅加深模型间关系的理解深度,而且通过迭代的方式逐步细化和丰富了相似性度量的层次结构,从而在复杂系统中捕捉更深层次的关联模式。紧接着,结合上述公式所得可以求解全局约束的 \mathbf{S} 矩阵,然后定义图的拉普拉斯矩阵 \mathbf{L} ,如式(5)、式(6)所示:

$$\mathbf{S} = \frac{1}{t} \sum_{m=0}^{t-1} \mathbf{P}^m \quad (5)$$

$$\mathbf{L} = \mathbf{T} - \mathbf{S} \quad (6)$$

其中, $\mathbf{T} = \text{diag}(t_1, \dots, t_K)$, $t_i = \sum_{j=1}^K S_{ij}$, t 表示迭代次数, K 表示字典原子个数。图的拉普拉斯矩阵 \mathbf{L} 可以继承训练样本的流形结构,流形结构描述了样本数据在特征空间中的分布和排列方式,可以保留学习字典的全局信息。根据文献[10]与全局约束相结合,通过最小化式(7)即可构建出全局相似性约束。

$$\frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K \|\hat{z}_i - \hat{z}_j\|_2^2 S_{ij} = \text{tr}(\mathbf{Z}^T \mathbf{L} \mathbf{Z}) \quad (7)$$

其中, \mathbf{Z} 表示稀疏向量矩阵。局部约束虽考虑到了样本数据的流形结构,但全局约束能更充分地描述该结构,能更好地发现非线性数据内在流形结构。

2.2 支持向量引导

为进一步提升编码向量的判别能力,本文在字典

学习过程中引入了多类支持向量机(SVM, support vector machine)正则化项。在SVM中,每个类别或类别之间的边界都可以由一个超平面来表示,每个超平面都有一个与之垂直的法向量。SVM是一种二分类模型,其核心思想是在特征空间中寻找一个最优超平面,同时最大化两类样本到超平面的间隔,使得两类样本之间尽可能被严格分开。在支持向量引导方法中,正则化项是关键,通过引入正则化项可以更好地学习分类边界,为每对稀疏编码向量分配不同的权重,同时还可以防止过拟合现象发生,从而进一步提高分类的准确性。其定义式如式(8):

$$\mathbf{L}(\mathbf{Z}) = 2 \sum_{c=1}^c f(\mathbf{Z}, \mathbf{y}^c, u_c, b_c) \quad (8)$$

其中, c 代表类别; \mathbf{y}^c 表示类别 c 对应的类标签向量,保证了模型具有处理多类别分类问题的能力; u_c 是与SVM第 c 类超平面相关的法向量,定义了决策边界的方向; b_c 是相对应的偏差,决定了决策边界的位置; u_c 和 b_c 参数共同决定了超平面的具体形状和位置。 $\mathbf{y}^c = [y_1^c, y_2^c, \dots, y_n^c]$, n 表示为样本数量,如果 y_i^c 中样本 i 属于类别 c 则 $y_i^c = 1$,其他情况均为 $y_i^c = -1$ 。具体来说,判别项可以被定义为式(9):

$$f(\mathbf{Z}, \mathbf{y}^c, u_c, b_c) = \|u_c\|_2^2 + \theta \sum_{i=1}^n l(z_i, y_i^c, u_c, b_c) \quad (9)$$

其中, θ 代表惩罚参数, $\|u_c\|_2^2$ 有助于防止过拟合, $l(z_i, y_i^c, u_c, b_c)$ 代表铰链损失函数,描述了样本 z_i 距离类别 c 超平面的偏离程度。判别项通过最小化总损失来优化超平面参数 u_c 、 b_c ,使得样本尽可能被正确分类,并保持类间足够大的间隔,从而提升模型的泛化能力和对噪声的鲁棒性。

支持向量引导可以看作是Fisher判别的扩展,能够为每对稀疏编码向量分配不同的权重,有助于模型聚焦于最具判别力的信息,忽略无关或冗余的特征,提高分类器的精确性和泛化能力。在另一方面,支持向量引导方法通过引入正则化项,可以更好地学习分类边界,使得分类器能够更加适应复杂的数据分布,并且对于噪声和干扰的容忍度也会有相应的提高,具有更强的鲁棒性。

2.3 多层字典学习框架

在传统服饰纹样的分类任务中,纹样特征复杂且多样化,不同纹样间的组合以及高维度特征增加了分类难度。单层字典学习在提取数据流形结构和内在特征方面表现不足,且易受噪声影响,影响分类效果。此外,服饰图像的复杂性因拍摄角度和造型的变化而

增加,需要更强的非线性特征提取能力。相比之下,多层字典学习在近似非线性函数方面更具优势。

Tang 等人^[12]提出的 DDLCN (deep dictionary learning and code network)将字典学习融入深度学习,利用多层字典和编码替代传统卷积层,并在深层编码层中添加局部约束以提升字典库的聚合度。MDDL也是多层字典学习经典算法之一,在所提出的多层架构中,原始数据 $\mathbf{X} \in \mathbf{R}^{M \times N}$ 作为第一层的输入首先可以稀疏分解为 $\mathbf{X} \approx \mathbf{D}_1 \mathbf{Z}_1$,紧接着编码向量 $\mathbf{Z}_1 \in \mathbf{R}^{K_1 \times N}$ 作为第二层的输入被进一步稀疏分解为 $\mathbf{Z}_1 \approx \mathbf{D}_2 \mathbf{Z}_2$ 。按照这个逻辑继续往下可以得到 $\mathbf{Z}_{l-1} \approx \mathbf{D}_l \mathbf{Z}_l \in \mathbf{R}^{K_{l-1} \times N}$,其中 l 为层数, $l \geq 2$ 。理论上每层重建误差的概率分布是相同的,所以可以用 $\|\mathbf{X} - \mathbf{D}_1 \mathbf{Z}_1\|_F^2 + \|\mathbf{Z}_1 - \mathbf{D}_2 \mathbf{Z}_2\|_F^2 + \dots + \|\mathbf{Z}_{l-1} - \mathbf{D}_l \mathbf{Z}_l\|_F^2$ 表示整个重构误差项。

尽管在 MDDL 中没有引入非线性激活函数,但编码向量和字典学习仍具非线性特性^[13],因此两个字典不能合并为单个字典,并且最后学习到的编码向量 $\mathbf{Z}_l \in \mathbf{R}^{K_l \times N}$ 不等于直接从 $\mathbf{X} \approx \mathbf{DZ}$ 得到的 \mathbf{Z} 。两类编码向量在单层中可能是非线性可分的,但通过更深层次可以转化为线性可分,从而增强了对非线性函数的逼近能力,为图像分类任务奠定了良好的架构基础,本文的模型构建也是基于这种思想进行展开的。

3 JGCSVG-MDL 算法

3.1 系统整体架构

为了更好地发现和表达图像特征的非线性关系,并深度优化字典学习,本文提出了 JGCSVG-MDL 算法,在构建多层字典学习的基础上联合全局相似性约束和支持向量判别项,整体结构如图 1 所示。

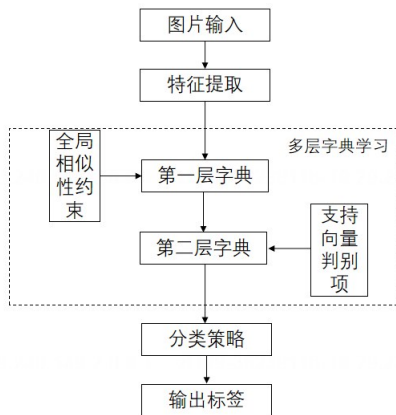


图 1 JGCSVG-MDL 算法整体结构

观察图 1 可知,该算法首先引入了全局相似性约束,在局部相似性的基础上同时考虑非相邻原子之间相似性,获得表示输入图像特征的非线性关系,在字典学习过程中更能体现数据的流形结构;其次引入了支持向量引导判别项,利用该判别项去表示所有编码向量对之间距离平方的加权和,并根据实际情况灵活为不同的编码向量对分配不同的权重,自动选择少数关键对来分配非零权重;最后通过构建多层字典学习结构将上一级的输出作为下一级的输入,实现单层到多层的过度,并利用双层结构的优势对输入数据更深层次地提取其内在属性,进一步增强图像分类的性能。算法的目标函数定义如下:

第一层公式由式(10)给出:

$$\min_{\mathbf{D}_1, \mathbf{Z}, \mathbf{L}} \|\mathbf{X} - \mathbf{D}_1 \mathbf{Z}\|_F^2 + \lambda_1 \text{tr}(\mathbf{Z}^T \mathbf{LZ}) + \lambda_2 \|\mathbf{Z}\|_F^2$$

$$s.t. \|d_k^1\|^2 = 1, k = 1, 2, \dots, K \quad (10)$$

第二层公式由式(11)给出:

$$\min_{\mathbf{D}_2, \mathbf{Z}, \mathbf{U}, \mathbf{b}} \|\mathbf{X} - \mathbf{D}_2 \mathbf{Z}\|_F^2 + 2\lambda_3 \sum_{c=1}^c f(\mathbf{Z}, \mathbf{y}^c, u_c, b_c) + \lambda_4 \|\mathbf{Z}\|_F^2$$

$$s.t. \|d_k^2\|^2 = 1, k = 1, 2, \dots, K \quad (11)$$

其中, \mathbf{X} 表示训练样本组成的矩阵, \mathbf{D}_1 表示第一层字典, \mathbf{D}_2 表示第二层字典, d_k^i 表示第 i 层的字典原子, k 表示原子个数, \mathbf{Z} 表示稀疏向量矩阵, \mathbf{L} 表示 \mathbf{D}_1 的图拉普拉斯矩阵, $\mathbf{U} = [u_1, u_2, \dots, u_c]$, $\mathbf{b} = [b_1, b_2, \dots, b_c]$, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 均是参数。

式(10)和式(11)中的第一项分别是字典 \mathbf{D}_1 和字典 \mathbf{D}_2 的重构误差项,最后一项是正则化项,这样设计的目的是可以保证最后求解稀疏向量具有一定的稳定性。式(10)中的第二项对 \mathbf{D}_1 施加了全局约束来追踪训练样本的流形结构。式(11)中的第二项对 \mathbf{D}_2 施加了支持向量判别约束来使得编码向量具备更强的辨别力。

3.2 模型求解

构建的两层字典之间存在联系,第一层字典所学习到能够追踪训练样本流形结构的字典 \mathbf{D}_1 和 \mathbf{Z} 作为第二层字典的输入,经过两层字典学习可以得到最终的 \mathbf{D}_2 和 \mathbf{Z} 。这里采用交替迭代法获取最优解,在求解过程中将问题分解为多个子问题,并交替迭代的求解每个子问题,在每次迭代时固定其他变量,只更新一个子问题所对应的变量。

在第一层首先固定 \mathbf{L} 和 \mathbf{Z} ,公式(10)关于 \mathbf{D}_1 优化问题能够被重写为式(12):

$$\min_{D_1} \|X - D_1 Z\|_F^2, \text{ s.t. } \|d_k^1\|^2 = 1, k = 1, 2, \dots, K \quad (12)$$

可以看到公式(12)成为具有二次约束的最小二乘问题,采用拉格朗日对偶法求解可得式(13):

$$D_1 = (XZ^T)(ZZ^T)^{-1} \quad (13)$$

得到 D_1 之后,使用公式(6)来构造图拉普拉斯矩阵 L 。接着固定得到的 D_1 和 L ,并对 Z 进行求解。通过对 Z 求解一阶偏导数,然后令其为0,得到最优 Z ,如式(14)所示:

$$Z = (D_1^T D_1 + \lambda_1 L + \lambda_2 I)^{-1} (D_1^T X) \quad (14)$$

第一层学习到的 D_1 和 Z 作为第二层 D_2 和 Z 的输入。公式求解思路相同,仍采用参数交替迭代方法。

更新 Z :公式(11)通过解决以下问题来固定其他变量并更新 Z ,如式(15):

$$\min_Z \|X - D_2 Z\|_2^2 + 2\lambda_3 \theta \sum_{c=1}^C l(z_i, y_i^c, u_c, b_c) + \lambda_4 \|Z\|_F^2 \quad (15)$$

在式(15)中, Z 的优化可以通过列进行,其公式可以写成式(16):

$$\min_{z_i} \|x_i - D_2 z_i\|_2^2 + 2\lambda_3 \theta \sum_{c=1}^C l(z_i, y_i^c, u_c, b_c) + \lambda_4 \|z_i\|_F^2 \quad (16)$$

其中,铰链损失函数 $l(z_i, y_i^c, u_c, b_c)$ 并非处处可微,这阻碍了基于梯度的优化方法的使用。为了促进优化效果,求解过程中采用可微分的二次铰链损失函数来近似原始函数。二次铰链损失函数定义为式(17):

$$l(z_i, y_i^c, u_c, b_c) = \begin{cases} \|y_i^c(u_c^T z_i + b_c) - 1\|_2^2, & y_i^c(u_c^T z_i + b_c) - 1 > 0 \\ 0, & y_i^c(u_c^T z_i + b_c) - 1 \leq 0 \text{ or } t = 1 \end{cases} \quad (17)$$

其中, t 表示迭代次数。当 $t = 1$ 时,式(16)可以表示为式(18):

$$\min_{z_i} \|x_i - D_2 z_i\|_2^2 + \lambda_4 \|z_i\|_F^2 \quad (18)$$

由式(18)可以求得式(19):

$$Z = (D_2^T D_2 + \lambda_4 I)^{-1} (D_2^T X) \quad (19)$$

当 $t \geq 2$ 时,式(16)可以表示为式(20):

$$\min_{z_i} \|x_i - D_2 z_i\|_2^2 + 2\lambda_3 \theta \sum_{c \in \phi} \|y_i^c(u_c^T z_i + b_c) - 1\|_2^2 + \lambda_4 \|z_i\|_F^2 \quad (20)$$

其中, $\phi = \{c | 1 \leq c \leq C, y_i^c(u_c^T z_i + b_c) - 1 > 0\}$, 求解可得式(21):

$$Z = \left(D_2^T D_2 + \lambda_4 I + 2\lambda_3 \theta \sum_{c \in \phi} u_c u_c^T \right)^{-1} \left(D_2^T X + 2\lambda_3 \theta \sum_{c \in \phi} u_c y_i^c - u_c b_c \right) \quad (21)$$

更新 D_2 :和第一层类似,式(11)可以转化为具有二次约束的最小二乘问题,这里采用拉格朗日对偶法进行求解,求解以后可得式(22):

$$D_2 = (XZ^T)(ZZ^T)^{-1} \quad (22)$$

更新 U, b :当其他变量固定时,式(11)关于 U, b 可归结为以下式(23)所示问题:

$$\min_{U, b} \sum_{c=1}^C \|u_c\|_2^2 + \theta \sum_{i=1}^C l(z_i, y_i^c, u_c, b_c) \quad (23)$$

其中,式(23)是一个多类线性 SVM 问题,根据文献[14]中提出的 SVM 求解器对其进行求解。算法整体计算流程如表1所示。

表1 算法JGCSVG-MDL计算流程

算法:JGCSVG-MDL算法整体计算流程

输入:训练数据矩阵 X ; 参数 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$;

输出:字典 D_2 , 稀疏矩阵 Z

1. 第一层字典学习:

初始化:通过K-SVD算法结合OMP算法得到原始的字典 D_1 和稀疏矩阵 Z

循环不收敛:

固定 Z, L , 求解(13)更新 D_1

固定 Z, D_1 , 用(6)及其相关公式更新 L

固定 D_1, L , 用(14)更新 Z

直到满足收敛条件,结束循环

2. 第二层字典学习:

初始化:第一层学习到的字典 D_1 和稀疏矩阵 Z 作为第二层 D_2 和 Z 的输入以及变量 U 和 b

循环不收敛:

固定其他变量,用(19)和(21)更新 Z

固定其他变量,用(22)更新 D_2

固定 Z, D_2 , 通过求解(23)更新 U, b

直到满足收敛条件,结束循环

3. 分类测试样本

使用多类线性SVM对编码向量进行分类并排序

3.3 分类方案

当字典学习完成后,这意味着图像可以被表示为字典中少数几个原子的线性组合,即具有稀疏系数的向量。因此需要进行下一阶段的分类任务,将已学习到的字典应用到新的、未标记的测试样本上,以确定它们所属的类别。首先给定一个新的测试样本 x_{new} , 为了使用学习到的字典 D 进行稀疏编码,这里定义了一个编码矩阵 P , 其计算公式为式(24):

$$P = (D^T D + I)^{-1} D^T \quad (24)$$

这是一个正则化的最小二乘解算子,其中 $D^T D$ 是字典的内积矩阵, I 是单位矩阵,两者相加并取逆是为了引入一定的正则化以增强计算稳定性。通过这个算子,可以得到测试样本 x_{new} 在字典 D 下的编码向量 z 如式(25)所示:

$$z = P x_{new} \quad (25)$$

对于编码向量 \mathbf{z} , 学习到的 SVM 分类器为类别 c 生成的决策函数值(得分)可表示为式(26):

$$s_c = \mathbf{u}_c^T \mathbf{z} + b_c \quad (26)$$

最后一步则是根据类别 c 的得分 s_c 来确定测试样本 x_{new} 的类别归属, 通常分类器会选择得分最高的类别作为预测结果。因此, 最后通过 $\text{identity}(x_{new}) = \arg \max_c (s_c)$ 进行分类, 选取得分最大的类别作为最终分类结果。

4 实验结果与分析

本节分别在两个具有显著差异的数据集上进行仿真实验, 数据集分别是专注于传统历史文化领域的明清宫廷服饰多标签数据集以及广泛应用于人脸识别研究的扩展耶鲁 B 人脸数据集。通过在这两个数据集上进行实验, 旨在深入探究算法模型在处理不同类型和风格图像时的性能表现。

4.1 实验参数设置

首先通过 K-SVD 算法进行稀疏矩阵 \mathbf{X} 与字典 \mathbf{D}_1 的初始化, 其中最大迭代次数为 20, 稀疏度为字典原子数的四分之一, 通过类别数确定子字典数量。在明清宫廷服饰多标签数据集中, 将子字典原子的个数设置为 30, $\lambda_1 = \lambda_3 = 0.01, \lambda_2 = \lambda_4 = 0.001$ 。在扩展耶鲁 B 人脸数据集中, 将子字典原子的个数设置为 38, $\lambda_1 = \lambda_3 = 0.01, \lambda_2 = \lambda_4 = 0.001$ 。

4.2 指标分析

对于第一个明清宫廷服饰多标签数据集, 实验采用四个基于样本的排名指标: 1-错误率 (One-Error), 该指标表示在预测标签排序中, 排名第一的标签不属于样本真实标签集中的比例; 覆盖率 (Coverage), 该指标表示排序后的标签序列平均向下移动以覆盖整个相关标签集所需要的探索深度; 排序损失 (Ranking Loss), 该指标表示不相关的标签排名高于相关标签排名的比例。平均精度 (AP, average precision), 该指标用于衡量整体样本标签预测的平均准确程度。One-Error、Coverage、Ranking Loss 三个指标结果越小就代表模型的性能越好, 指标 AP 数值越大则代表模型的准确度越高, 性能越好。

对于第二个扩展耶鲁 B 人脸数据集, 实验采用了三个指标: 准确度 (Accuracy), 该指标作为首要评估标准, 直接反映了模型在图像分类任务的精准程度, 其数值越

高就表明模型的分类性能越好; 训练时间 (Training Time), 该指标表示从模型初始化至学习过程结束所需的时间, 它不仅体现了算法的复杂度, 还间接反映了资源消耗与优化效率; 测试时间 (Testing Time), 该指标是衡量模型对未知样本进行预测的速度, 是评估系统实时响应能力和应用场景适应性的关键。训练时间和测试时间越短则代表模型的复杂度低以及识别速度快。

4.3 明清服饰多标签数据集

实验使用的是传统明清时期服饰纹样图像相关的数据集, 由于该领域的数据集较为稀缺, 因此在进行实验之前对少量的数据集进行了数据增强操作, 采用图像翻转、旋转、缩放等方式扩充数据集, 将传统服饰图像数据集从 910 张扩充到了 3640 张, 部分传统服饰纹样展示图如图 2 所示, 数据样本数量以及对应的纹样类型如表 2 所示。

表 2 传统服饰图像数据集构成

纹样类型	数量	纹样类型	数量
花纹	529	鸟纹+花纹	241
云纹	541	鸟纹+云纹	209
龙纹	536	云纹+龙纹	163
鸟纹	509	鸟纹+云纹+花纹	92
蝴蝶纹	518	云纹+鸟纹+龙纹	73
花纹+蝴蝶纹	177	鸟纹+花纹+蝴蝶纹	52
			合计: 3640

每次实验以近似 4:1 的比例随机选取 2800 张图片作为训练集, 剩下的均作为测试集。实验过程中采用尺度不变特征变换 (SIFT, scale-invariant feature transform) 方式对图像进行特征提取, 并与颜色特征相结合, 最终得到 2484 维特征向量。SIFT 算法因其尺度不变性和旋转不变性, 在图像识别领域得到了广泛应用, 本实验选择 SIFT 算法是因为它能够有效捕捉图像中的局部特征, 并且对噪声和部分遮挡具有较好的鲁棒性。此外, 为了保证实验结果的可靠性, 重复进行十次实验, 对十次的实验结果进行均值计算, 均值作为最终的实验结果。



图 2 部分传统服饰纹样展示图

实验过程中涉及到多种经典的多标签标注算法,其中包括ML-KNN (multi-label KNN)^[15]算法、LLSF (learning label specific feature)^[16]的BR(binary relevance)和CC (classifier chains) 算法、MDDL-2^[4]、SVGDL、CADL(class-aware analysis dictionary learning)^[17]、SADL(structure analysis dictionary learning)^[18]。另外,为了能够更好地对比全局相似性约束和局部约束的效果差异,以及单层字典学习和多层字典学习的性能差异,实验中引入了采用局部约束且为单层字典结构的LCDL-SV(locality constraint dictionary learning with support vector) 算法^[19],以及拥有全局约束但只有单层结构的JGCSVGDL 算法,该算法是在LCDL-SV和JGCSVG-MDL算法基础上进行了改进,最终保留全局约束和单层字典的特征。

实验结果如表3所示,表3整理了JGCSVG-MDL算法和其他多种模型在传统服饰图案数据集上的实验结果。观察表3结果可以发现,JGCSVG-MDL算法相比于其他模型,所有指标几乎都取得了最佳性能,在One-Error、Coverage、Ranking Loss、Average Precision四个指标下结果分别为0.3207,1.1134,0.2117和0.7908。一方面,相比于采用局部约束的LCDL-SV算法,采用全局约束的JGCSVGDL和JGCSVG-MDL算法在Average Precision指标上分别实现了1.73%和3.33%的提升,其他三个指标的数值也有一定的降低,算法模型的性能表现更佳。由此可证明在传统文化数据集上,字典学习引入全局相似性约束比局部相似性约束更加有效。在另一方面,相比于JGCSVGDL 单层字典学习算法,JGCSVG-MDL 多层字典学习在One-Error、Coverage、Ranking Loss三个指标上分别实现了2.76%、4.25%、2.35%的降低,而在Average Precision指标上实现了1.57%的提升,图像分类准确度得到了显著提升。因此,通过实验结果证明了多层字典学习算法相比于单层字典学习算法表现更出色,效果更佳。

此外,对比实验不仅局限于字典学习算法之间的比较,同时还引入了深度学习的对比。众所周知,深度学习的关键在于拥有大量标注良好的训练数据,类似包含上百万张图像的ImageNet这种大规模的数据集。为了能够更好地适应本实验这种小规模的传统服装图像数据集,本文决定在AlexNet的基础上进行修改和训练,用于提取图像特征,同时还对AlexNet模型进行调整,减少了整体网络结构,修改后的AlexNet结构只由一层卷积层和两层全连接层组成,这样的简化可以明显降低模型的复杂度,减少过拟合的风险,

本实验中将其命名为AlexNetLite。实验中将反向传播学习率设置为0.001, batch size 设置为64,并使用交叉熵损失函数计算损失。

从表3的结果观察到,AlexNetLite方法的One-Error、Coverage、Ranking Loss、Average Precision指标分别为0.3513,1.1087,0.1917,0.7691。可以看出,AlexNetLite模型相比于其他多种模型都显示出其优势,它在覆盖率 and 排名损失标准下取得了最好的结果,但是在平均精度上与JGCSVG-MDL算法相比还是有一定的差距,JGCSVG-MDL算法在平均精度方面仍然保持着显著的优势。通过分析可知,深度学习模型比较擅长于对类似于ImageNet这种大规模单标签图像数据的识别,然而,本实验所涉及的数据集特征与此有所不同,主要体现在两个方面:首先,数据集中包含了大量的多标签图像,这增加了模型训练的复杂性;其次,数据集的总体规模相对较小,这可能限制了深度学习模型性能的充分发挥。因此,实验结果表明,尽管深度学习在大规模单标签数据集上具有优势,但在面对小规模多标签数据集时,其性能可能受到一定影响。这一发现进一步验证了JGCSVG-MDL算法在适应不同数据规模和标签复杂性方面的优越性。JGCSVG-MDL算法能够更好地处理多标签图像数据,即使在数据量有限的情况下,也能保持较高的识别准确率和鲁棒性。

在进行算法性能评估的过程中,本文还深入探讨了字典原子个数对算法性能的影响,并针对LCDL-SV、JGCSVGDL以及JGCSVGDL-MDL三种算法设计了相应的对比实验。在实验中,通过改变字典中原子的个数,从K=30开始,以30的间隔差逐步增加至K=210,以致评估不同原子数量对算法性能的具体影响,具体实验结果如图3所示。

在图3所示的实验结果汇总中,通过细致的观察与分析,可以发现JGCSVGDL与JGCSVG-MDL算法在不同字典大小下的平均精度(AP)均高于LCDL-SV算法,展现出了引入全局相似性约束具有显著的性能优势。随着字典中原子个数的增加,三种算法的平均精度(AP)均呈现出上升趋势,这是因为更多的字典原子可以捕获到数据中更加细致的特征信息,从而进一步提升图像分类任务的性能。通过观察可知,LCDL-SV、JGCSVGDL和JGCSVG-MDL三种算法的平均精度(AP)分别在原子数量K=120,K=180,K=180的时候达到峰值。这不仅证实了全局相似性约束策略在提升算法性能方面的有效性,同时也凸显了多层字典学习框架在处理复杂数据集时的独特优势。

但需要注意的是,虽然增加字典原子个数可以提高算法的性能,但同时也可能带来计算复杂度的增加和过拟合的风险。因此,在实际应用中需要找到一个平衡点,以确保算法既有效又具有较好的泛化能力。

表3 传统服饰多标签数据集实验结果

方法	One-Error	Coverage	Ranking Loss	Average Precision
ML-KNN	0.4015	1.2173	0.2207	0.7341
LLSF	0.3822	1.2599	0.2241	0.7369
LLSF-BR	0.4013	1.2621	0.2307	0.7243
LLSF-CC	0.3889	1.3027	0.2385	0.7378
MDDL-2	0.3765	1.2359	0.2171	0.7431
AlexNetLite	0.3513	1.1087	0.1917	0.7691
CADL	0.3665	1.2113	0.2292	0.7497
SADL	0.3701	1.1917	0.2217	0.7556
SVGDL	0.3663	1.2278	0.2234	0.7492
LCDL-SV	0.3537	1.2523	0.2187	0.7653
JGCSVGDL	0.3298	1.1628	0.2168	0.7786
JGCSVG-MDL	0.3207	1.1134	0.2117	0.7908

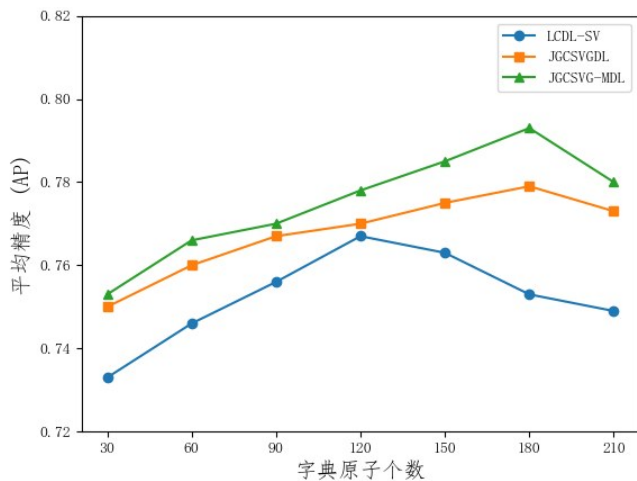


图3 不同原子数下明清宫廷服饰多标签数据集上的平均精度

4.4 扩展耶鲁B人脸数据集

扩展耶鲁B人脸数据集包含丰富的光照条件和人脸表情,其中包含38名受试者的2414张正面图像,每个受试者大约有64张图像。在预处理阶段对所有图像进行了标准化裁剪至统一尺寸192×168像素。接着,为了解决高维度特征空间带来的计算复杂度问题以及潜在的过拟合风险,本实验采用了主成分分析法(PCA, principal component analysis)进行特征降维,从原始的32256维(192*168=32256)降至50维,同时也保留了对于区分不同个体脸部特征至关重要的信息,最后对每一类别以4:1的比例分为训练集和测试

集。为保证实验结果的可靠性,重复进行十次实验,对十次的实验结果进行均值计算,均值作为最终的实验结果。实验过程中涉及到几种经典的图像分类算法,包括CADL、SADL、FDDL、SVGDL、AlexNetLite、LCDL-SV、JGCSVGDL和JGCSVG-MDL算法。

最终实验结果如表4所示。通过观察与分析,可以发现LCDL-SV、JGCSVGDL和JGCSVG-MDL三种算法在该数据集上的准确率相比于其他字典学习算法都展现出了显著的优势,而且模型训练时间也比较理想,表现出了较高的效率,特别是JGCSVG-MDL算法没有出现因为引入多层字典学习而导致训练时间过长的现象。相比于采用局部约束的LCDL-SV算法,由于JGCSVGDL算法引入了全局约束,在原基础上还需考虑非相邻原子之间相似性,虽然这导致了训练时间的增加,但换来了分类准确率的提升。然而,JGCSVG-MDL算法由于充分研究与平衡全局约束和字典判别的结果,使得其训练时间相比于JGCSVGDL算法有了明显的缩短,但由于涉及到多层结构,学习特征的量会增大,所以训练时间会比LCDL-SV单层结构稍微长一点,但是综合精确度、训练时间、测试时间等多方面来评估的话,JGCSVG-MDL算法依然是性能最佳的。表4还展示了AlexNetLite算法的实验结果,该算法不仅达到了最高的准确率,而且训练时间也是最短的,这进一步印证了卷积神经网络(CNN, convolutional neural networks)在单标签图像分类任务中的卓越性能。然而,值得注意的是,尽管CNN在单标签任务上表现出色,但在处理多标签分类问题时可能会面临一些挑战。相比之下,基于字典学习的算法能够更好地适应这种复杂的分类场景,尤其是在多标签分类任务中,其灵活性和适应性显得尤为重要。通过本次实验,不仅验证了JGCSVG-MDL算法在特定数据集上的性能优势,也进一步探讨了不同算法在不同分类任务中的适用性和效率,为未来的研究方向提供了有价值的参考。

表4 扩展耶鲁B人脸数据集实验结果

方法	Accuracy(%)	Training Time(s)	Testing Time(s)
CADL	96.3	23.5	1.3e-3
SADL	95.1	27.2	1.3e-3
FDDL	92.8	352.6	3.4
SVGDL	95.7	29.8	1.3e-3
AlexNetLite	98.9	2.1	1.3e-3
LCDL-SV	96.5	23.7	1.3e-3
JGCSVGDL	97.1	87.3	1.3e-3
JGCSVG-MDL	98.2	37.2	1.3e-3

5 结束语

本文提出了一种基于联合全局约束和支持向量引导的多层字典学习算法。该算法首先引入了全局相似性约束,在局部相似性的基础上同时考虑非相邻原子之间相似性,能更好地发现和表达图像特征的非线性关系。其次,通过对字典原子施加支持向量引导,为不同编码向量分配不同的权重并自动选择少数关键对来分配非零权重,为图像分类提供更好的泛化能力。最后,引入多层字典学习方法鼓励来自同一类别的样本的编码向量共享更一致的稀疏模式,同时促进不同类别样本编码向量的可分离性,也进一步减少了字典学习过程中训练数据受到噪声的负面影响。通过在传统服饰图像数据集和扩展耶鲁B人脸两组数据集上进行对比实验,JGCSVG-MDL算法在性能方面都展现出了显著的优越性,并验证了相对于局部相似性约束和单层字典学习而言,全局相似性约束和多层字典学习架构更适合小规模传统文化多标签数据集。但是该算法还存在一定的改进空间,比如随着层数的不断增加会导致学习过程参数的选择变得复杂,因此在未来的实验中会进一步尝试设计更合理的算法结构和程序来优化这些问题。

参考文献(References):

- [1] Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation[C]// Proceedings of 2011 International Conference on Computer Vision, 2011: 543-550.
- [2] Cai S, Zuo W, Zhang L, et al. Support vector guided dictionary learning[C]// Proceedings of 13th European Conference on Computer Vision, 2014: 624-639.
- [3] Song J, Liu Z, Xie C, et al. Relaxed support vector based dictionary learning for image classification[J]. Multimedia Tools and Applications, 2024, 83(5): 12731-12755.
- [4] Song J, Xie X, Shi G, et al. Multi-layer discriminative dictionary learning with locality constraint for image classification[J]. Pattern Recognition, 2019, 91: 135-146.
- [5] Zhou F, Huang S, Xing Y. Deep semantic dictionary learning for multi-label image classification[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 3572-3580.
- [6] Wang H, Dong G, Chen J, et al. A novel dictionary learning named deep and shared dictionary learning for fault diagnosis[J]. Mechanical Systems and Signal Processing, 2023, 182: 109570.
- [7] Gou J, Yuan X, Yu B, et al. Intra-and inter-class induced discriminative deep dictionary learning for visual recognition[J]. IEEE Transactions on Multimedia, 2023, 25: 1575-1583.
- [8] Shaban A, Rabiee H R, Farajtabar M, et al. From local similarity to global coding: an application to image classification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2794-2801.
- [9] Belkin M, Niyogi P, Sindhvani V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples[J]. The Journal of Machine Learning Research, 2006, 7: 2399-2434.
- [10] Li Z, Lai Z, Xu Y, et al. A locality-constrained and label embedding dictionary learning algorithm for image classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 28(2): 278-293.
- [11] Szummer M, Jaakkola T. Partially labeled classification with markov random walks[C]// Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, 2001: 945-952.
- [12] Tang H, Liu H, Xiao W, et al. When dictionary learning meets deep learning: deep dictionary learning and coding network for image recognition with limited data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(5): 2129-2141.
- [13] Gregor K, LeCun Y. Learning fast approximations of sparse coding[C]// Proceedings of the 27th International Conference on Machine Learning, 2010: 399-406.
- [14] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]// Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1794-1801.
- [15] Zhang M L, Zhou Z H. ML-KNN: a lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048.
- [16] Huang J, Li G, Huang Q, et al. Learning label-specific features and class-dependent labels for multi-label classification[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3309-3323.
- [17] Wang J, Guo Y, Guo J, et al. Class-aware analysis dictionary learning for pattern classification[J]. IEEE Signal Process Letters, 2017, 24(12): 1822-1826.
- [18] Tang W, Panahi A, Krim H, et al. Analysis dictionary learning based classification: structure for robustness[J]. IEEE Transactions on Image Processing, 2019, 28(12): 6035-6046.
- [19] Yin H F, Wu X J, Chen S G. Locality constraint dictionary learning with support vector for pattern classification[J]. IEEE Access, 2019, 7: 175071-175082.