基于字典学习的民族文化图案语义标注

赵海英^{1,2}, 陈洪^{1*}, 贾耕云², 郑桥², 王绍杰²

1. 北京邮电大学世纪学院移动媒体与文化计算北京市重点实验室, 北京 102101

2. 北京邮电大学网络技术研究院, 北京 100876

* 通信作者. E-mail: norman_chen@263.net

收稿日期: 2018-11-29; 接受日期: 2019-01-15; 网络出版日期: 2019-02-21

北京市科委基金 (批准号: D171100003717003) 资助项目

摘要 民族文化图案是中华民族宝贵财富, 对民族文化图案进行语义标注与分析, 是挖掘其文化价值, 进行再创作与应用的基础. 本文以此为研究对象, 在多分类字典学习的基础上, 提出了一种多标签字典学习标注算法 SCMIDL. 算法结合字典不相关性与系数相似性, 有效提高了多标签标注性能, 实现了民族文化图案的自动标注. 在收集并构建的三类民族文化图案数据集进行多标签语义标注实验, 实验结果验证了算法的有效性.

关键词 民族文化图案, 字典学习, 多标签, 图像标注, 相关性

1 引言

随着机器学习理论快速发展, 相关技术在越来越多的领域得到应用, 同时, 与其他学科的交叉研究和应用也受到越来越广泛的关注, 例如传统文化研究中的数字化分割、标注与智能设计等. 当前文化影像资料获取更加方便, 内容更加丰富, 这不但令人们能更方便地了解和学习各类中华优秀传统文化, 更为文化的深层分析提供了丰富的内容与素材.

民族文化图案是中华文化的宝贵财富, 孕育着丰富的文化内涵和文化价值. 如苗族服饰, 几乎每一个图案纹样都有来历或传说, 深含民族文化, 表达民族情感, 是苗族历史与生活的展示. 蝴蝶、龙、飞鸟、鱼、圆点花、浮萍花等图案都是苗族服饰纹样传承的内容, 色彩鲜艳、构图明朗、朴实大方. 图 1 是鲲鹏纹, 在苗绣中它有 3 种类型的大鱼和鸟结合的纹样. 第 1 种是龙首牛角, 鱼身鱼尾的大鱼形象, 腹下有两只爪形足, 如图 1(a); 第 2 种以鱼为身, 龙首牛角, 身上有鸟翅, 尾部为鱼尾, 腹下两只鸟足如图 1(b); 第 3 种以鸟为身, 龙首牛角, 或鸟首鱼身如图 1(c).

然而随着文化图像资料数量的迅速增长, 人工方法难以实现大量图像的标注与理解. 人们急需民族文化图案语义自动标注的方法或有效手段, 但相关的研究和应用仍然不足, 现有的一些方法也难以

引用格式: 赵海英, 陈洪, 贾耕云, 等. 基于字典学习的民族文化图案语义标注. 中国科学: 信息科学, 2019, 49: 172–187, doi: 10.1360/N112018-00203
Zhao H Y, Chen H, Jia G Y, et al. Semantic annotation of national cultural patterns based on dictionary learning (in Chinese). Sci Sin Inform, 2019, 49: 172–187, doi: 10.1360/N112018-00203

图 1 (网络版彩图) 苗族服装中的鲲鹏纹. (a) 鱼身; (b) 鸟翅; (c) 鸟身

Figure 1 (Color online) Roc pattern in miao nationality. (a) Body of the fish; (b) wings of the bird; (c) body of the bird

融入文化图案数字化研究的体系中. 为此, 需要引入先进的技术方法开展民族文化图案语义的自动化标注研究.

民族文化图案具有丰富的语义, 如图 1, 因此需要通过图像多标签自动标注技术实现语义解读. 字典学习算法是一种在稀疏表达框架下发展起来的算法, 在图像分类, 人脸识别等领域取得了很大成功, 但当前基于字典学习的多标签标注研究仍然较少. 因此, 本文以字典学习算法为基础, 提出了一种多标签字典学习标注算法 SCMIDL (similar coefficients & multi-label incoherent dictionary learning). 算法在两个方面做了改进. 首先通过加入字典不相关项, 以此来降低字典之间的相关性. 其次通过添加类内散度惩罚项, 来提升重建系数的判别能力. 本文还收集、构建了 3 种民族文化图案数据集并在数据集上进行了多标签语义标注实验, 验证了添加字典不相关项和类内散度惩罚项对实验效果的提升, 完善了民族文化数字化分析与研究体系.

2 相关工作

图像的多标签标注首先需要对图像进行特征提取, 而后特征被输入到多标签分类器中进行标注. 由于多标签标注与多分类关系密切, 其方法通常由多分类方法转化而来, 被分为两类^[1,2], 包括问题转化方法和算法转化方法. 问题转化方法是指把多标签分类问题转化为多分类问题, 这类算法首先对标签进行处理, 然后对一组标签训练多分类器^[3~6]. 算法转化的方法是指把多分类算法推广到多标签标注中, 比如 K 近邻^[7]、决策树^[8]、支持向量机^[9]和信息论方法^[10]等. 随着近年来深度学习的出现, 现今也有相关学者采用深度学习的方法进行多标签标注. 其中 Berger 等^[11]采用 CNN 与 GRU 模型, 在模型输出时根据阈值输出确定样本的标签. Kurata 等^[12]将每个样本的标签作为一个标签共现模式, 以此来对网络参数进行调整. Chen 等^[13]在多标签分类中的网络结构是 CNN 和 RNN 的结合. Zhang 等^[14]采用 CNN 网络和 deepwalk 结合的方法进行多标签分类, 并且取得了较好的分类效果. 但由于深度学习本身是由数据驱动, 需要大量的图片输入, 而由于民族文化图案数据集的规模较小, 无法达到深度学习的数据量要求.

监督字典学习算法是一种广泛应用的多分类算法. 一类算法基于重建误差进行分类. Wright 等^[15]进行人脸识别时, 把每一类的训练样本作为字典进行分类. Yang 等^[16]和 Zhang 等^[17]则提出了字典应由学习得到而不是直接使用训练样本. Ramirez 等^[18]则在此基础上提出降低不同类字典间的相关性, 从而使得样本在不同字典上重建误差的差距变大. Mairal 等^[19]使用了 Softmax 损失函数, 希望最小化训练样本在本类字典上的重建误差, 还有一些算法假设每一类字典都是由一个共享字典和类特有字典组成^[20~22]. Gu 等^[23]则提出了一种 dictionary pair 的字典学习方式, 利用 F 范数

图 2 民族文化图案数字化研究框架
Figure 2 Framework of national cultural pattern digitalization

的正则化实现 group sparsity, 提高了运算速度. 另一类则基于重建系数进行分类, 例如基于 Fisher 判别准则学习重建系数^[24], 直接对字典和一个基于重建系数的线性分类器进行联合学习是最常用的方式^[25~27]. Jiang 等^[28]提出的 LCKSVD (label consistent KSVD) 算法引入了标签一致项, 它使得同类样本的重建系数趋向于非 0 值出现在相同的维度上, 因而其系数更相似.

上述工作多是将字典学习应用到传统的多分类任务中, 针对多标签标注的字典学习研究仍然较少, Cao 等^[29]提出了 SLED (semantic label embedding dictionary) 算法, 该算法为了利用多分类字典学习中的 Fisher 判别系数项, 构建了互斥标签集合, 每个子集内的标签都是互斥的, 即每个子集内都是一个多分类问题, 学习得到字典后, 算法又引入标签相关性来提高标注效果. Jing 等^[30]提出的多标签字典学习模型则以 LCKSVD^[28]算法为基础, 并基于多标签特性对特征进行线性变换.

3 民族文化图案数字化研究

民族文化图案形式复杂、种类多样, 但其数字化研究可概括为 3 个层面: 提取、解读和再利用, 如图 2 所示.

民族文化图案的提取主要指根据研究或应用目标, 从图案或图案组合中, 提取所需要的图案或图案元素, 这个过程应用到的技术有图像目标检测、图像自动分割、图案建模等. 传统的民族文化图案提取主要依靠人工观察、总结与绘制. Zhao 等^[31]通过大量观察研究, 对彝族装饰纹样进行了人工分类提取并进一步抽象, Wang^[32]提取了传统剪纸纹样的图元进行数字化存储. 当前也有很多工作对计算机辅助的民族图案提取进行了尝试, 它们借鉴了多种图像分割方法, 如 He 等^[33]结合图割算法对中国传统绘画中的图案进行提取, Xiang 等^[34]针对传统织物图像, 采用先去织物纹理后分割的方法提取图案形状. 此外还有多尺度分割方法^[35]等.

民族文化图案的解读是指获取图案的文化内涵, 称之为图案语义, 其中可以采取对民族文化图案进行多标签标注的方式来获取图案的语义信息. 人工方法主要根据领域知识进行分类, 如按照内容题材划分为剪纸图案^[36]、传统藏族服饰图案^[37]、蒙古族传统民族图案^[38]等; 当前自动标注研究主要针对某一类语义进行分类, 如对织物风格^[39]分类, 对藏族唐卡图案内容^[40]、传统中国画内容^[41]的分类, 以及对传统纺织纹理模式^[42]等分类.

民族文化图案的再利用是指对提取、解读后的图案或图案元素, 进行研究、加工、改造等, 并将其应用到当代社会的各个场景. 如将景颇族传统民族图案应用到园林设计^[43], 将维吾尔族的图案服饰应用于食品包装设计^[44]等. 同时, 基于民族风格特点的图案自动生成也受到广泛关注, 例如新疆地毯、

图 3 (网络版彩图) 民族文化图案数字化研究示例

Figure 3 (Color online) Research examples of national cultural pattern digitalization

民族织物等图案的自动生成^[45~48]等,这对民族文化图案的解构、诠释以及推广都有重要意义。

在民族文化图案数字化研究框架中,图案的自动解读是核心研究内容,是进一步研究的基础,例如对时空演化规律的研究,需要挖掘不同时间地域语义的一致性,当在不同场合的服饰设计中对传统民族图案进行创新应用时,需要准确把握图案的语义以避免误用。因此,通过解读民族文化图案,获取语义信息,才能把握其研究和应用的方向。

基于民族文化图案数字化研究框架,图 3 给出了一组民族文化图案数字化研究的示例。首先是第 1 层的特征提取,通过分析颜色、纹理、几何、结构等信息,获取如图所示的线条型植物图案基元;其次是第 2 层的民族文化图案解读,图中给出了对 3 幅图案的语义解读即标注结果;最下面一层是民族文化图案的再利用,借助数学方法,通过对大量图案解构和建模基础,智能设计出符合民族文化风格的图案。

4 SCMIDL 模型构建与学习

本节首先提出了多标签的不相关字典学习方法,然后在此基础上引入类内相似系数方法提高系数的判别能力,最后给出了 SCMIDL 的算法流程,如下图 4 所示。其中步骤 (a) 表示图像的特征提取,提取训练集和测试集中图像特征,并使用一个向量表示一个样本;步骤 (b) 表示将训练数据送入模型进行训练,得到每类的训练字典和每一个训练样本的稀疏表示,使用字典和稀疏表示可以近似重建原样本。步骤 (c) 将测试样本送入训练好的模型中,通过步骤 (d) 的标签排序,得到标注结果。

4.1 多标签不相关字典学习

令 $X_i \in \mathbb{R}^{d \times N_i}$ 表示含有标签 i 的训练样本集合, d 是样本特征的维度, N_i 是含有标签 i 的训练

图 4 (网络版彩图) 图像多标签标注流程示意图. (a) 图像特征提取; (b) 字典构建; (c) 和 (d) 为测试步骤

Figure 4 (Color online) Framework of multi-label image annotation. (a) Represents image feature extraction; (b) constructs the dictionary; (c) and (d) are the test stages

样本总数. 令 N 为训练集样本总数, 则有 $\sum_i N_i \geq N$, $y_i \in \mathbb{R}^L$ 表示任一样本的标注, L 为标签总数, $y_i(l) = 1$ 表示样本 i 含有标签 l , 否则 $y_i(l) = 0$.

基于重建误差的字典学习标注算法基于每一个标签 i 的样本集 X_i 分别学习一个字典 D_i , 对任一样本 $x \in X_i$, 满足 $x \approx D_i a$, $D \in \mathbb{R}^{d \times p}$, $a \in \mathbb{R}^p$, 且 $\|a\|_0$ 尽可能小. 比较常见的实现方法为优化下列目标函数. 虽然该方法要求所有类的重建误差最小, 但由于一类训练样本不影响其他类字典的训练, 因此求解过程可以并行进行:

$$\min_{D, A_i} \sum_{i=1}^q \|X_i - D_i A_i\|_F^2 + 2\lambda \|A_i\|_1, \quad (1)$$

其中 $A_i \in \mathbb{R}^{N_i \times p}$ 为重建系数. 第 1 项控制重建误差, 使字典原子的线性组合 $D_i A_i$ 与样本集 X_i 相似; 第 2 项控制 A_i 中非零项的数量, 用 L_1 范数代替 L_0 范数实现解的稀疏性.

上述方法对每一类字典分别训练, 各类字典描述的特征空间仍可能存在较大相似性, 这使得一个不含特定标签的样本, 仍可能被该标签对应的字典较好地表示, 因此, 为了提高字典的区分度, 引入了字典不相关项^[18], 降低所学得的字典的相关性:

$$\min_{D, A} \sum_{i=1}^q \left\{ \|X_i - D_i A_i\|_F^2 + 2\lambda_1 \|A_i\|_1 + \lambda_2 \sum_{j \neq i} \|D_j^T D_i\|_F^2 \right\}, \quad (2)$$

然而, 直接不加区分地降低所有字典间的相关性的做法无法直接应用到多标签标注中. 如果两个标签相关性较高, 则没有必要显著降低这两个标签对应字典的相关性; 反之, 如果两个标签几乎没有相关性, 则等同于多分类问题, 因此在式 (2) 的基础上, 引入标签相关性, 目标函数如下:

$$\min_{D, A} \sum_{i=1}^q \left\{ \|X_i - D_i A_i\|_F^2 + 2\lambda_1 \|A_i\|_1 + \lambda_2 \sum_{j \neq i} R_{ij} \|D_j^T D_i\|_F^2 \right\}, \quad (3)$$

R_{ij} 为标签相关系数, 可以由多种方法计算得到. 本文使用余弦相似度, 通过计算两个向量的夹角余弦

值来评估它们的相似度:

$$R_{ij} = 1 - \frac{Y_i \cdot Y_j}{|Y_i| \cdot |Y_j|}, \quad (4)$$

其中 Y_i 表示标签 i 在整个数据集上的分布, $Y_i(j) = 1$ 表示第 j 个样本含有标签 i , 反之 $Y_i(j) = 0$. 标签 i, j 相关性越高, R_{ij} 越小, 因此可以使用 R_{ij} 作为标签相关系数, 以此来衡量两个标签之间的相关性.

4.2 类内相似系数

在监督字典学习中, 样本的重建系数 A 可以看作样本特征的进一步提取, 因而也含有有效的监督信息, 可以参与标注提高性能.

在多分类字典学习中, 一种最常用的方法是在训练时引入 Fisher 判别准则加强系数的分类能力^[24]. 4.1 小节提出的模型中, 训练样本在各自标签对应的字典上重建时, 由于字典不统一, 在不同字典上的重建系数关系难以确定, 它与字典中原子的排列顺序有关. 在这种情况下, 重建系数的类间差异性被抑制. 另一方面, 对多标签标注而言, 含有不同标签的数据集交集可能不为空, 使重建系数的差异性更难以确定, 但上述特性不影响重建系数的类内相似性.

因此, 为了进一步提高标注性能, 本文在式 (3) 基础上, 引入重建系数的类内散度惩罚项, 提高含有同标签样本的重建系数的相似度:

$$\min_{D, A} \sum_{i=1}^q \left\{ \|X_i - D_i A_i\|_F^2 + 2\lambda_1 \|A_i\|_1 + \lambda_2 \left(\sum_{j \neq i} R_{ij} \|D_j^T D_i\|_F^2 \right) + \lambda_3 \text{tr}(S_w(A_i)) \right\}. \quad (5)$$

第 4 项为重建系数的类内散度项, 在优化时通常考虑其等价形式:

$$\text{tr}(S_w(A_i)) = \sum_{x_k \in X_i} \text{tr}((a_k - m_i)(a_k - m_i)^T) = \|A_i - M_i\|_F^2 = \|A_i B_i\|_F^2, \quad (6)$$

$$B_i = I - \frac{1}{N_i} \mathbf{1}\mathbf{1}^T. \quad (7)$$

式 (6) 中, a_k 为矩阵 A_i 的第 k 列, m_i 是矩阵 A_i 所有列的均值, M_i 的每一列都是矩阵 A_i 所有列的均值, 列数与 A_i 保持一致. N_i 是含有标签 i 的训练样本总数. 式 (7) 中, $\mathbf{1}\mathbf{1}^T$ 为全为 1 的列向量乘以行向量所得的矩阵. I 为单位矩阵.

4.3 模型学习与图像标注

字典学习的目标函数单独对字典 D 和重建系数 A 是凸的, 但是对两者非凸, 因此其一般采用迭代求解的方法, 每次固定一个变量求解另一个, 最终使目标函数收敛.

SCMIDL 算法采用迭代求解方法, 每一次循环固定字典 D 和重建系数 A 中的一个求解另一个. 但是不直接求解目标函数 (5), 根据文献 [18], 虽然在目标函数中降低不同标签字典的相关性, 但是字典重建原样本的性能对标注效果起决定性作用. 为了避免得到一组虽然相关性很低, 但重建性能很差的字典影响标注效果, 求解时首先忽略字典相关性惩罚项, 得到一组重建性能较好的字典, 然后再降低各类字典的相关性. 具体的求解过程如下所示.

(1) 字典初始化: $D_i^0, i = 1, \dots, q$, 初始化 $t = 0, \rho = 0$.

(2) 基于目标函数,

(i) 固定字典为 D_i^t , 求解重建系数 A_i^t :

$$A_i^t = \arg \min_{A_i} \|X_i - D_i^t A_i\|_F^2 + 2\lambda_1 \|A_i\|_1 + \lambda_3 \|A_i - M_i\|_F^2.$$

(ii) 固定重建系数为 A_i^t , 更新字典:

$$D_i^{t+1} = \arg \min_{D_i} \|X_i - D_i^t A_i^t\|_F^2 + 2\lambda_1 \|A_i^t\|_1 + \rho \sum_{j \neq i} R_{ij} \|D_j^T D_i\|_F^2,$$

(iii) $t = t + 1$.

(3) 判断收敛条件:

(i) 若 $\rho = 0$, 且目标函数未收敛, 则返回步骤 (2);

(ii) 若 $\rho = 0$, 且目标函数收敛, 则令 $\rho = \lambda_2$, 返回步骤 (2);

(iii) 若 $\rho = \lambda_2$, 且目标函数收敛或 t 达到最大循环次数 $t_{\max}$, 则输出当前字典 D_i^t 与类平均重建系数 m_i , $i = 1, \dots, q$, 训练阶段完成.

对任一给定的测试样本 y , 算法首先求解它在每个字典上的重建系数:

$$\hat{a}_i = \arg \min_{a_i} \|y - D_i a_i\|_2^2 + \lambda_1 \|a_i\|_1 + \gamma \|a_i - m_i\|_2^2, \quad (8)$$

其中 m_i 表示训练样本集 X_i 在字典 D_i 上重建系数的均值. 由于在训练阶段引入了类内散度惩罚项, 因此在重建测试样本时, 也采用同样的处理. 令测试样本在 D_i 上的重建系数与训练样本集 X_i 重建系数的均值相似, 从而使重建系数更有利于标注.

求得测试样本在每个字典上的重建系数后, 计算下述指标:

$$e_i = \|x - D_i \hat{a}_i\|_2^2 + \lambda_1 \|\hat{a}_i\|_1 + \gamma \|\hat{a}_i - m_i\|_2^2. \quad (9)$$

它包含 3 项: 第 1 项为测试样本在 D_i 上的重建误差, 其越小则 y 越有可能包含标签 i ; 第 2 项用以衡量重建系数的稀疏度, 若某测试样本在两个字典上的重建误差 (即第 1 项) 几乎相同, 但是其中一个的重建系数远比另一个稀疏, 则认为样本更有可能含有更稀疏的系数对应的标签; 第 3 项计算了重建系数与各训练样本集重建系数均值的距离, 距离越小, 表明测试样本越有可能含有标签 i .

综合以上 3 项并按从小到大排序, 即可得到对应标签的排序, 排序越靠前, 样本含有该标签可能性越大. 为了得到准确标签, 还可以根据数据集标签的数量, 选取排序的前 k 个作为测试样本的标签.

5 民族文化图案语义标注实验与分析

5.1 民族文化图案数据集构建与特征提取

本文对明清宫廷服饰图案、传统民族服饰图案和传统地毯图案进行了数字化采集、分割, 构建了 3 类民族文化图案数据集. 各数据集的属性如表 1 所示, 各属性含义见参考文献 [49]. 其中 LCard 度量的是每个样本平均的标签个数, LDen 是 LCard 除以标签个数获得的数据集的标签密度. DL 统计的是数据集中不同标签组合的数量. PDL 是 DL 除以数据集样本个数, 表示不同标签集的比例.

所有数据集包含所有样本特征矩阵和其相对应的标签矩阵, 在 3 个数据集中, 分别随机取固定数量的样本作为训练集, 剩下的部分作为测试集进行 3 组实验.

表 1 民族文化图案数据集属性

Table 1 Attributions of national cultural pattern data sets

Data set	Number of images	Number of labels	LCard	LDen	DL	PDL	Number of trianing samples
Ming and Qing court dress pattern	899	5	1.3582	0.2716	12	0.0133	600
Traditional national dress pattern	782	6	2.7724	0.4621	18	0.0230	400
Traditional carpet pattern	536	3	1.5690	0.5230	7	0.0131	200

明清两朝离现代较近, 传统服饰的设计、织造、纹绣等工艺均达到传统封建社会的高峰, 因此留存的服饰、织物等文物数量较多、种类齐全. 本文对收集的明清宫廷服饰进行分割预处理, 并选取了蝴蝶纹、花朵纹、龙纹、凤鸟纹、云纹等典型纹样进行标注.

传统民族服饰图案展现着不同的民族文化, 它是民族文化在平面图案中的体现, 不同民族有着特色鲜明、极具区分力的服饰图案. 本文收集了 3 个传统民族的服饰图案, 对其进行了分割预处理, 然后对苗族、蒙古族、维吾尔族、植物纹、动物纹、几何纹等典型纹样图案进行标注.

传统地毯是我国优秀文化遗产的重要组成部分, 具有装饰作用的同时也蕴含着丰富的文化内涵. 本文通过搜集一些基本纹样作为地毯主体图案, 然后通过添加边框、角隅再生成地毯图案, 且边角颜色会根据主题图案颜色自动做出相应调整, 以适应地毯整体色调, 最终构建了本文实验所用数据集, 其标注词为动物纹、植物纹、几何纹.

对 3 个数据集采用统一的特征提取方法, 特征包含两部分.

第 1 部分反映图像局部形状特征. 首先对所有图像, 统一图像大小, 并把图像转化为灰度图, 在每个图像中按照固定步长选择固定大小的局部区域, 在所有区域中提取 SIFT 特征. 在 SIFT 特征的计算过程中, 在固定大小的局部区域计算描述子, SIFT 描述子计算以关键点为中心的 16×16 像素区域内的梯度分布, 计算时每 $4 \times 4 = 16$ 个像素作为一个 patch, 计算每个 patch 内的像素梯度在 8 个角度区间内的幅值, 从而每个图像生成一组局部描述子. 然而, 局部特征对一幅图像表现为矩阵的形式, 无法直接作为常用分类器的输入. 而且, 即使将矩阵展开为向量, 过高的维度不但无法带来性能的明显提升, 还会引起维度灾难. 因此需要对局部特征进一步编码, 特征编码不但解决了上述问题, 而且也带了性能的显著提升. 对特征编码时, 首先对图像集的所有局部描述子聚类, 生成多个聚类中心, 再使用 LLC 算法^[50] 编码每一个局部描述子, 最后对一幅图像所有描述子的编码结果进行最大池化得到图像特征.

第 2 部分为颜色特征, 颜色特征可分为全局颜色特征和局部颜色特征两种, 颜色直方图是最常用的表达颜色特征的方法, 其通过统计分布在各个颜色区间内的像素数量, 表现了图像整体的颜色分布. 由于民族文化图案的语义多与图案颜色的位置分布无关, 因此可采用全局颜色直方图特征, 即在原彩色图像的 RGB 3 个通道上计算颜色直方图, 合并到一起作为颜色特征.

5.2 实验与分析

本文采用 4 种方法 one-error, coverage, ranking-loss 和 average-precision 来评价 SCMIDL 算法的标注效果.

One-error 衡量模型输出的置信度最高的标签不属于该测试样本的比例, 当多标签标注退化为传统多分类时, 该指标等价于错误率. Coverage 是指对测试样本在每个标签上的输出置信度, 按从大到小的顺序遍历, 直到遍历到所有该测试样本的所有标签时所需要的平均归一化步数. Ranking-loss 通

表 2 明清宫廷服饰图案数据集实验结果

Table 2 Experimental results of Ming and Qing court dress pattern data sets

	One-error	Coverage	Ranking-loss	Average-precision
ML-KNN	0.3946	1.1873	0.2152	0.7419
Rank-SVM	0.3679	1.1037	0.1909	0.7653
MLNB	0.5552	1.5418	0.3088	0.6533
LLSF-BR	0.4013	1.2642	0.2305	0.7340
LLSF-CC	0.7458	1.8428	0.3946	0.5427
LLSF	0.5686	1.6923	0.3464	0.6243
SCMIDL	0.3478	0.9565	0.1577	0.7911

表 3 传统民族服饰图案数据集实验结果

Table 3 Experimental results of traditional national dress pattern data sets

	One-error	Coverage	Ranking-loss	Average-precision
ML-KNN	0.0700	1.7853	0.0586	0.9313
Rank-SVM	0.0497	1.7382	0.0550	0.9431
MLNB	0.0497	1.8508	0.0699	0.9321
LLSF-BR	0.0445	1.7304	0.0528	0.9459
LLSF-CC	0.0602	1.7592	0.0620	0.9385
LLSF	0.0524	1.8455	0.0667	0.9357
SCMIDL	0.0445	1.6780	0.0450	0.9524

过两两对比的方法获得, 若一对标签中, 测试样本包含的标签排序反而低于不包含的标签, 则此为一对误排标签, 所有这种误排标签对占有所有标签对的比例即为该指标. Average-precision 衡量一个属于测试样本的标签排序比特定标签排序靠前的平均比例.

为了验证算法有效性, 本文与几种经典的多标签标注算法进行了对比实验, 包括 ML-KNN (multi-label KNN)^[7], Rank-SVM^[9], MLNB (multi-label naive Bayes)^[51], 同时, 使用了基于特征选择算法 LLSF (learning label specific features)^[52] 的 BR (binary relevance) 和 CC (classifier chains) 算法. MLNB 算法首次将特征选择技术运用于多标签标注中, 其中包含两个阶段的特征选择策略, 第 1 阶段使用 PCA 消除不相关和冗余特征, 第 2 阶段利用遗传算法将核实的特征子集进行分类. 同时使用基于朴素 Bayes 的多标签学习算法. LLSF 算法是一种线性分类算法, 它对每一个标签训练一个回归系数, 引入 L1 范数正则化使算法具有特征选择的作用, 同时根据标签相关性降低回归系数的相关性. LLSF-BR 算法是根据 LLSF 算法对数据特征进行选择, 然后再使用 BR 算法进行多标签标注, LLSF-CC 则是再使用 CC 算法进行多标签标注. 本文实验参数设置为 $\lambda_1 = 0.01$, $\lambda_2 = 0.01$, $\lambda_3 = 0.001$, $\gamma = 0.05$, 明清宫廷服饰图案数据集与传统地毯图案数据集字典原子数为 25, 传统民族服饰图案数据集字典原子数为 30. 三个数据集上的实验结果如表 2~4 所示. 部分图像的标注结果如图 5 所示.

算法对目标函数 (5) 使用迭代求解的策略, 在明清宫廷服饰图案数据集上, 其目标函数能量变化如图 6 所示, 可以看出, 目标函数能够在 600 次循环内收敛.

以 4 个关键指标 Average-precision, Coverage, One-error 和 Ranking-loss 在 3 个数据集上的实验结果, 验证算法对训练阶段 3 个超参数, λ_1 , λ_2 和 λ_3 , 以及测试阶段超参数 γ 的敏感度, 4 个指标之间的相关性. 为了能够清晰反映指标的变化趋势与指标之间的关系, 对所有指标的变化值分别进行归一

表 4 传统地毯图案数据集实验结果

Table 4 Experimental results of traditional carpet pattern data sets

	One-error	Coverage	Ranking-loss	Average-precision
ML-KNN	0.1380	0.8661	0.1887	0.9049
Rank-SVM	0.0675	0.7589	0.1196	0.9445
MLNB	0.2791	1.0327	0.3037	0.8341
LLSF-BR	0.0859	0.7827	0.1365	0.9351
LLSF-CC	0.0951	0.8393	0.1672	0.9225
LLSF	0.0767	0.7173	0.1043	0.9496
SCMIDL	0.0670	0.7366	0.0950	0.9530

图 5 (网络版彩图) 标注结果示例. (a) 传统地毯图案; (b) 传统民族服饰图案; (c) 明清宫廷服饰图案

Figure 5 (Color online) Examples of image annotation. (a) Traditional carpet pattern; (b) traditional national dress pattern; (c) Ming and Qing court dress pattern

化处理, 最小值为 0, 最大值为 1, 如图 7 所示.

系数的 F1 范数项同时起到正则化和稀疏性约束的作用, 能够有效提高标注效果, 但其参数 λ_1 过高也会影响训练样本重建效果, 进而影响精度, 如图 7(a) 所示. λ_2 以及 λ_3 不同取值下的 4 个指标如图 7(b) 和 (c) 所示. 同理, 过高的取值也会影响所学习字典的重建效果, 其中, λ_2 值太小也会影响字典重建的效果. γ 用以权衡标注时更看重重建误差还是系数相似度, 其不同取值下的指标如图 7(d) 所示, 可以看出, 重建误差仍然对标注效果影响很大, 引入重建系数后可以提高一定的标注效果, 但过

图 6 目标函数收敛性示意图
Figure 6 Convergence of the objective function

图 7 (网络版彩图) 超参数: (a) λ_1 ; (b) λ_2 ; (c) λ_3 ; (d) γ
Figure 7 (Color online) Hyper-parameters: (a) λ_1 ; (b) λ_2 ; (c) λ_3 ; (d) γ

分依赖系数会导致标注精度极大的降低. 而且, Average-precision 与 Coverage, One-error, Ranking-loss 基本呈负相关. 在第 2 和 3 个数据集中, 情况基本相同. 字典不相关项对实验效果整体有提升, 但提升效果在一定范围内较为稳定. 而其他参数变化幅度较大.

同时, 为了验证目标函数中添加字典不相关项以及类内散度惩罚项对实验结果的影响, 分别将部

表 5 数据集 1/2/3 不同目标函数实验结果
Table 5 Experimental results of different objective function on data set 1/2/3

	Data set	Original objective function	$\lambda_2 = 0, \lambda_3 = 0, \gamma = 0$	$\lambda_2 = 0$	$\lambda_3 = 0, \gamma = 0$
Average-precision	1	0.7865	0.7829	0.7862	0.7809
	2	0.9459	0.9411	0.9467	0.9424
	3	0.9305	0.9277	0.9290	0.9277
Coverage	1	1.0736	1.0803	1.0702	1.0870
	2	1.6571	1.6518	1.6492	1.6466
	3	0.8184	0.8244	0.8214	0.8244
One-error	1	0.3378	0.3445	0.3411	0.3478
	2	0.0340	0.0628	0.0340	0.0602
	3	0.1018	0.1080	0.1049	0.1080
Ranking-loss	1	0.1826	0.1856	0.1826	0.1873
	2	0.0509	0.0525	0.0499	0.0516
	3	0.1388	0.1435	0.1404	0.1435

分参数设置为 0, 完成不添加字典不相关项和类内散度惩罚项 ($\lambda_2 = 0, \lambda_3 = 0, \gamma = 0$)、不添加字典不相关项 ($\lambda_2 = 0$)、不添加类内散度惩罚项 ($\lambda_3 = 0, \gamma = 0$) 的实验, 其他默认参数为 $\lambda_1 = 0.01, \lambda_2 = 0.01, \lambda_3 = 0.001, \gamma = 0.05$, 实验结果如表 5 所示.

从实验结果可以看出, 在第 1 个数据集中, 相比较于传统的字典学习算法, 添加了类内惩罚项后标注效果有了较大的提升. 虽然在只单独添加字典不相关项后, 标注结果有所下降, 但同时结合类内散度惩罚项后, 标注结果整体上升, 并且结果优于单独添加类内散度惩罚项的实验结果. 相较于原本的样本重建系数, 通过减少本类样本之间的距离, 得到的样本重建系数在类间的关系更加符合本数据集标签空间潜在的类间关系. 在第 2 个数据集中, 分别添加字典不相关项和类内散度惩罚项后, 标注结构都有一定提升, 在同时添加两项后, 整体效果相比只添加类内散度惩罚项有所下降. 在第 3 个数据集中, 单独添加字典不相关项对实验结果没有影响, 单独添加类内散度惩罚项后, 实验结果有所提升, 两项都添加, 整体实验结果优于其他所有结果. 由此证实添加字典不相关项和类内散度惩罚项的有效性. 虽然添加字典不相关项的效果对于不同数据集和目标函数其他参数较为敏感, 但整体对实验结果是有益的.

6 小结

本文以民族文化图案为研究对象, 构建了民族文化图案数字化研究的框架, 指出民族文化图案的自动标注是实现文化图案数字化深入研究以及创新再利用的基础; 并以字典学习算法为基础, 提出了一种多标签字典学习标注算法 SCMIDL; 在收集构建的 3 类典型民族文化图案数据集上进行了仿真实验, 证实了算法的有效性. 同时, 在未来随着民族文化图案数据集规模的扩充, 可考虑采用新的方法 (如深度学习) 进行仿真实验, 寻求更佳的多标签标注效果.

参考文献

1 Gibaja E, Ventura S. Multi-label learning: a review of the state of the art and ongoing research. Wiley Interdiscip Rev Data Min Knowl Discov, 2014, 4: 411–444

- 2 Zhang M L, Zhou Z H. A review on multi-label learning algorithms. *IEEE Trans Knowl Data Eng*, 2014, 26: 1819–1837
- 3 Boutell M R, Luo J, Shen X, et al. Learning multi-label scene classification. *Pattern Recogn*, 2004, 37: 1757–1771
- 4 Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification. *Mach Learn*, 2011, 85: 333–359
- 5 Fürnkranz J, Hüllermeier E, Loza Mencía E, et al. Multilabel classification via calibrated label ranking. *Mach Learn*, 2008, 73: 133–153
- 6 Tsoumakas G, Katakis I, Vlahavas I. Random k-labelsets for multilabel classification. *IEEE Trans Knowl Data Eng*, 2011, 23: 1079–1089
- 7 Zhang M L, Zhou Z H. ML-KNN: a lazy learning approach to multi-label learning. *Pattern Recogn*, 2007, 40: 2038–2048
- 8 Clare A, King R D. Knowledge discovery in multi-label phenotype data. In: *Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery*. Berlin: Springer, 2001. 2168: 42–53
- 9 Elisseeff A, Weston J. A kernel method for multi-labelled classification. In: *Proceedings of International Conference on Neural Information Processing Systems: Natural and Synthetic*. Cambridge: MIT Press, 2001. 681–687
- 10 Ghamrawi N, McCallum A. Collective multi-label classification. In: *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*. New York: ACM, 2005. 195–200
- 11 Berger M J. Large Scale Multi-Label Text Classification With Semantic Word Vectors. Technical Report, Stanford University, 2015
- 12 Kurata G, Xiang B, Zhou B. Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016. 521–526
- 13 Chen G, Ye D, Xing Z, et al. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In: *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. New York: IEEE, 2017. 2377–2383
- 14 Zhang W, Yan J, Wang X, et al. Deep extreme multi-label learning. In: *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. New York: ACM, 2018. 100–107
- 15 Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell*, 2009, 31: 210–227
- 16 Yang M, Zhang L, Yang J, et al. Metaface learning for sparse representation based face recognition. In: *Proceedings of IEEE International Conference on Image Processing*. New York: IEEE, 2010. 1601–1604
- 17 Zhang Q, Li B. Discriminative K-SVD for dictionary learning in face recognition. In: *Proceedings of Computer Vision and Pattern Recognition*. New York: IEEE, 2010. 2691–2698
- 18 Ramirez I, Sprechmann P, Sapiro G. Classification and clustering via dictionary learning with structured incoherence and shared features. In: *Proceedings of Computer Vision and Pattern Recognition*. New York: IEEE, 2010. 3501–3508
- 19 Mairal J, Bach F, Ponce J, et al. Discriminative learned dictionaries for local image analysis. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2008. 1–8
- 20 Peng J. Learning inter-related visual dictionary for object recognition. In: *Proceedings of Computer Vision and Pattern Recognition*. New York: IEEE, 2012. 3490–3497
- 21 Kong S, Wang D. A dictionary learning approach for classification: separating the particularity and the commonality. In: *Proceedings of European Conference on Computer Vision*. Berlin: Springer-Verlag, 2012. 186–199
- 22 Gao S H, Tsang I W H, Ma Y. Learning category-specific dictionary and shared dictionary for fine-grained image categorization. *IEEE Trans Image Process*, 2014, 23: 623–634
- 23 Gu S, Zhang L, Zuo W, et al. Projective dictionary pair learning for pattern classification. In: *Proceedings of Advances in Neural Information Processing Systems*, 2014. 793–801
- 24 Yang M, Zhang L, Feng X, et al. Sparse representation based fisher discrimination dictionary learning for image classification. *Int J Comput Vis*, 2014, 109: 209–232
- 25 Lu J, Wang G, Zhou J. Simultaneous feature and dictionary learning for image set based face recognition. *IEEE Trans Image Process*, 2017, 26: 4042–4054
- 26 Pham D S, Venkatesh S. Joint learning and dictionary construction for pattern recognition. In: *Proceedings of IEEE*

- Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2008. 1–8
- 27 Mairal J, Ponce J, Sapiro G, et al. Supervised dictionary learning. In: *Proceedings of Advances in Neural Information Processing Systems*, 2009. 1033–1040
  - 28 Jiang Z L, Lin Z, Davis L S. Label consistent K-SVD: learning a discriminative dictionary for recognition. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35: 2651–2664
  - 29 Cao X C, Zhang H, Guo X J, et al. SLED: semantic label embedding dictionary representation for multilabel image annotation. *IEEE Trans Image Process*, 2015, 24: 2746–2759
  - 30 Jing X Y, Wu F, Li Z, et al. Multi-label dictionary learning for image annotation. *IEEE Trans Image Process*, 2016, 25: 2712–2725
  - 31 Zhao Q, Li Q M. Extraction and application of Yi nationality's decorative patterns in the tourist souvenirs. *Packag Eng*, 2017, 38: 240–244 [赵勤, 李萩敏. 旅游纪念品中彝族装饰纹样的提取与运用. *包装工程*, 2017, 38: 240–244]
  - 32 Wang Y. *The digital applications of folk art elements*. Taiyuan: Taiyuan University of Technology, 2014 [王毅. 民间艺术元素的数字化应用. 太原: 太原理工大学, 2014]
  - 33 He N, Lu K. An image segmentation method for chinese paintings by combining deformable models with graph cuts. In: *Proceedings of the 14th International Conference on Human-Computer Interaction: Design and Development Approaches*. Berlin: Springer, 2011. 571–579
  - 34 Xiang J, Zhang J, Pan R R, et al. Graphic contour extraction for printed fabric based on texture smoothing. *J Textile Res*, 2017, 38: 162–167 [向军, 张杰, 潘如如, 等. 采用纹理平滑的印花织物图案轮廓提取. *纺织学报*, 2017, 38: 162–167]
  - 35 Yang B W, Lu X Q, Chen C. Multiscale color segmentation algorithm for textile images. *J Zhejiang Univ (Eng Sci)*, 2005, 39: 530–533, 546 [杨彬蔚, 陆系群, 陈纯. 一种纺织印染图案的多尺度彩色分割算法. *浙江大学学报 (工学版)*, 2005, 39: 530–533, 546]
  - 36 Yue J F. On the subject classification and cultural connotation of paper-cut in Bozhou. *Home Darma*, 2017, 10: 255–256 [岳金凤. 亳州剪纸的题材分类与文化内涵探析. *戏剧之家*, 2017, 10: 255–256]
  - 37 Yang J L. A study on classification of traditional tibetan clothing patterns. *Big Stage*, 2015, 325: 230–231 [杨井兰. 藏族传统服饰图案的分类研究. *大舞台*, 2015, 325: 230–231]
  - 38 Liu C H. Classification and artistic characteristics of mongolian traditional patterns. *Big Stage*, 2015, 3: 81–82 [刘春恒. 蒙古族传统图案的分类及艺术特点. *大舞台*, 2015, 3: 81–82]
  - 39 Zhang H. Research and application of image classification of stylized fabrics. Dissertation for Master Degree. Zhejiang: Zhejiang University, 2012 [张辉. 风格化织物图像分类的研究与应用. 硕士学位论文. 浙江: 浙江大学, 2012]
  - 40 Qian J Q. A study on semantic-based thangka image annotation and retrieval. Dissertation for Ph.D. Degree. Lanzhou: Northwest Minzu University, 2010 [钱建军. 基于语义的唐卡图像标注与检索研究. 硕士学位论文. 兰州: 西北民族大学, 2010]
  - 41 Zhang D Q, Pham B, Li Y F. Modelling traditional Chinese paintings for content-based image classification and retrieval. In: *Proceedings of the 10th International Multimedia Modelling Conference*. Washington: IEEE Computer Society, 2004. 258
  - 42 Pawening R E, Dijaya R, Brian T, et al. Classification of textile image using support vector machine with textural feature. In: *Proceedings of International Conference on Information & Communication Technology and Systems*. New York: IEEE, 2016. 119–122
  - 43 An X L, Wei K Y. The discuss of the application of Jingpo ethnic traditional national pattern in landscape architecture. *South China Agr (Garden Flower Edition)*, 2011, 5: 1–6 [安星霖, 魏开云. 景颇族传统民族图案在园林中的应用探讨. *南方农业 (园林花卉版)*, 2011, 5: 1–6]
  - 44 Wang Y. A study on application of Uygur national patterns in food packaging design. Dissertation for Master Degree. Xi'an: Northwest University, 2015 [王宇. 维吾尔民族图案纹饰在食品包装设计中的应用研究. 硕士学位论文. 西安: 西北大学, 2015]
  - 45 Zhao H Y, Peng H, Zhang X L. A generation method of carpet pattern based on evolutionary computation. *J Graph*, 2015, 36: 41–46 [赵海英, 彭宏, 张小利. 基于进化计算的地毯图案生成方法. *图学学报*, 2015, 36: 41–46]

- 46 Zhao H Y, Pan Z G, Xu Z G. Automatic generation of Xinjiang ethnic fabric patterns based on configuration style. *J Graph*, 2013, 34: 17–21 [赵海英, 潘志庚, 徐正光. 基于构型风格的新疆民族织物图案自动生成. *图学学报*, 2013, 34: 17–21]
- 47 Zhao H Y, Yang Y F, Xu G M. Design method for xinjiang folk art pattern. *Comput Syst Appl*, 2011, 20: 94–99 [赵海英, 杨一帆, 徐光美. 新疆民族民间艺术图案生成方法. *计算机系统应用*, 2011, 20: 94–99]
- 48 Zhao H Y, Peng H, Yang Y F, et al. A generation method of carpet pattern based on topology configuration. *J Comput-Aided Design Comput Graph*, 2013, 25: 502–509 [赵海英, 彭宏, 杨一帆, 等. 基于拓扑构型的地毯图案生成方法. *计算机辅助设计与图形学学报*, 2013, 25: 502–509]
- 49 Zhang M L, Wu L. Lift: multi-label learning with label-specific features. *IEEE Trans Pattern Anal Mach Intell*, 2015, 37: 107–120
- 50 Wang J J, Yang J C, Yu K, et al. Locality-constrained linear coding for image classification. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, 2010. 3360–3367
- 51 Zhang M L, Peña J M, Robles V. Feature selection for multi-label naive Bayes classification. *Inf Sci*, 2009, 179: 3218–3229
- 52 Huang J, Li G R, Huang Q M, et al. Learning label specific features for multi-label classification. In: *Proceedings of IEEE International Conference on Data Mining*, Atlantic City, 2015. 181–190

Semantic annotation of national cultural patterns based on dictionary learning

Haiying ZHAO^{1,2}, Hong CHEN^{1*}, Gengyun JIA², Qiao ZHENG² & Shaojie WANG²

1. *Mobile Media and Cultural Computing Key Laboratory, Beijing Century College, Beijing University of Posts and Telecommunications, Beijing 102101, China;*

2. *Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*

* Corresponding author. E-mail: norman_chen@263.net

Abstract The national cultural pattern is a precious treasure of the Chinese nation. Semantic annotation and analysis of national cultural pattern is useful in cultural heritage and modern recreation. This article presents a multi-label dictionary learning algorithm called similar coefficient multi-label incoherent dictionary learning (SCMIDL) based on a multi-class dictionary learning algorithm. SCMIDL combines the incoherence of the dictionary and the similarity of coefficients, which significantly improves the performance of multi-label annotation. The superior annotation ability of the algorithm was confirmed on three kinds of national cultural pattern datasets constructed for evaluation purposes.

Keywords national cultural pattern, dictionary learning, multi label, image annotation, correlation

Haiying ZHAO was born in 1972. She received her Ph.D. degree from the University of Science and Technology, Beijing. She is now a distinguished professor of Century College, BUPT. Her research interests are multi-media data mining and cultural computing. She proposed a theoretical system of cultural computing, and has significantly advanced its applications.

Hong CHEN was born in 1976. He received his Ph.D. degree from China Agricultural University. He is now a distinguished professor of Century College, BUPT. His research interests are animation technology and interactive entertainment.

Gengyun JIA was born in 1992. He received his M.S. in information and communication engineering from Beijing University of Posts and Telecommunications in 2018. His research interests are cultural computing and pattern recognition.