软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2013,24(10):2405−2418 [doi: 10.3724/SP.J.1001.2013.04424]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种适合弱标签数据集的图像语义标注方法*

田 枫[1,2], 沈旭昆[1]

[1](虚拟现实技术与系统国家重点实验室(北京航空航天大学),北京 100191)

[2](东北石油大学 计算机与信息技术学院,黑龙江 大庆 163318)

通讯作者: 田枫, E-mail: tianfeng80@gmail.com

摘 要: 真实环境下数据集中广泛存在着标签噪声问题,数据集的弱标签性已严重阻碍了图像语义标注的实用化进程.针对弱标签数据集中的标签不准确、不完整和语义分布失衡现象,提出了一种适用于弱标签数据集的图像语义标注方法.首先,在视觉内容与标签语义的一致性约束、标签相关性约束和语义稀疏性约束下,通过直推式学习填充样本标签,构建样本的近似语义平衡邻域.鉴于邻域中存在噪声干扰,通过多标签语义嵌入的邻域最大边际学习获得距离测度和图像语义的一致性,使得近邻处于同一语义子空间.然后,以近邻为局部坐标基,通过邻域非负稀疏编码获得目标图像和近邻的部分相关性,并构建局部语义一致邻域.以邻域内的语义近邻为指导并结合语境相关信息,进行迭代式降噪与标签预测.实验结果表明了方法的有效性.

关键词: 图像语义标注;弱标签数据集;测度学习;非负稀疏编码;语义近邻

中图法分类号: TP391 **文献标识码**: A

中文引用格式: 田枫,沈旭昆.一种适合弱标签数据集的图像语义标注方法.软件学报,2013,24(10):2405−2418. http://www.jos.org.cn/1000-9825/4424.htm

英文引用格式: Tian F, Shen XK. Image semantic annotation method for weakly labeled dataset. Ruan Jian Xue Bao/Journal of Software, 2013,24(10):2405−2418 (in Chinese). http://www.jos.org.cn/1000-9825/4424.htm

Image Semantic Annotation Method for Weakly Labeled Dataset

TIAN Feng[1,2], SHEN Xu-Kun[1]

[1](State Key Laboratory of Virtual Reality Technology and Systems (BeiHang University), Beijing 100191, China)

[2](School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Corresponding author: TIAN Feng, E-mail: tianfeng80@gmail.com

Abstract: Automatic semantic annotation, which automatically annotates images with semantic labels has received much research interest. Although it has been studied for years, image annotation is still far from practical. The effectiveness of traditional image annotation techniques heavily relies on the availability of a sufficiently large set of correct, complete and balanced labeled samples, which typically come from users in an interactive manual process. However, in real world environment, image labels are often incomplete, noisy and imbalanced. This paper investigates the usefulness of weakly labeled information and proposes an image annotation method for weakly labeled dataset. First, the missing labels are automatically filled by a transductive method which incorporates label correlation and semantic sparsity, along with the consistency of visual and semantic similarity. Then approximate semantic balanced neighborhood is constructed. A distance metric learning method for large margin nearest neighbor embedded in multiple labels is supplied, making the retrieved neighbors by this metric appear in the same semantic subspace. Local semantic consistent neighborhood is obtained by local nonnegative sparse coding. Meanwhile, an iterative denoising method for label inference is proposed to simultaneously handle the noise and annotate images under the guidance of semantic nearest neighbors and contextual information. Experimental results demonstrate the effectiveness and capability of the proposed method.

Key words: image semantic annotation; weakly labeled dataset; metric learning; nonnegative sparse coding; semantic nearest neighbor

* 基金项目: 国家自然科学基金(61170132, 60533070); 国家高技术研究发展计划(863)(2009AA012103)
收稿时间: 2012-12-01; 定稿时间: 2013-05-03

给图像自动添加反映其内容的文本标签,是图像检索与管理的重要基础工作.但是传统的图像语义标注方法在现实应用中往往性能较低,用户感受不好.因为真实环境下的数据集往往来源于网络人工标注或者自动搜集,虽然其为标注任务提供了丰富的语义辅助,但是数据集所蕴含的噪声导致图像附带的标签通常是不准确和不完整的.Kennedy 与 Börkur 等人首先对人工手动标注的行为进行了研究,特别是对图像共享社区的人工标签进行了统计与分析[1,2].其研究结果表明,标签不完整和不准确的情况非常普遍.图像原始标签具有个性化强、噪声大和遗漏标签现象,相当一部分标签描述的是图像的上下文信息,如时间、地点和主观感受.例如,来自图像共享社区 Flickr 的图像,其附属标签仅有约 50%与图像内容相关,而且具有多于 4 个标签的图像低于 50%.此外,数据集中存在着标签分布欠平衡的现象.实验结果表明,通过特征均等贡献的贪婪搜索方法对 Corel 5K 数据集上 30%的低频标签进行标注,*F* 值为 23.1%;而对于 30%的高频标签,*F* 值却可以达到 47.5%.这些低质量的标签导致数据集具备弱标签性.本文提出的"弱标签数据集"具有下述特点:其一是图像语义标签不完整;其二是标签可能不准确,即存在一定的噪声干扰;其三是语义分布欠平衡.数据集的弱标签性在真实应用环境中普遍存在.图 1 给出了不同数据集中的弱标签图像的例子,粗体显示标签为遗漏标签,下划线显示标签为内容无关标签.

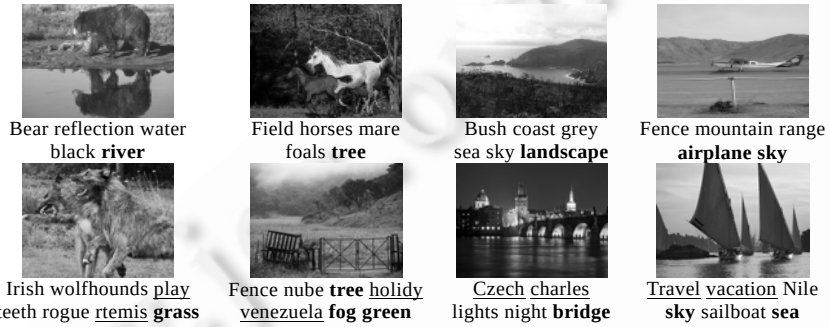


Fig.1 Illustration of weakly labeled dataset

图 1 弱标签数据集实例

传统的基于生成式模型的图像语义标注方法需要高质量数据集求取图像视觉特征和语义概念之间的联合分布,判别式方法将图像语义标注看作是多类问题.弱标签数据的标签遗漏、噪声干扰和欠平衡将导致模型的决策边界偏移,进而大幅度降低标注性能[3].早期的图像语义标注研究局限于受限环境下的小规模数据集上如何提高模型的标注精度,数据集的弱标签性并没有引起研究人员的关注[4].随着数据集规模的扩大,数据集的弱标签性逐渐引起了多媒体检索领域的重视.Jin 等人最早提出采用 WordNet 计算标签之间的语义关系,保留相关标签,滤除无关标签[5].Wang 等人通过马尔可夫随机游走改善标签与图像的相关性[6].Jin 等人在文献[5,6]的工作基础上,进一步采用融合 WordNet 和各种不同的语义度量的方式去除噪声标签,但是这种基于字典本体的方法对于具体语境的适用度较差[7].由于在现实环境中数据集中存在大量与图像内容不相关的标签,而采用标签相似度估计预测标签相关性的方法受噪声标签影响较大,Xu 等人提出了改进的潜在狄利克雷分布模型,通过交替迭代的方式进行标签相似度与标签相关性估计[8].

针对图像共享社区的标签噪声问题,Li 等人提出了一种依据近邻投票思想进行相关性学习的方法,可以过滤出标签与图像是强相关的还是弱相关的.虽然其在测试数据集上只达到了 0.12 的平均准确率,但是作为一个数据驱动的方法,其通过视觉邻域样本降低单幅图像弱标签性的思路对于后续研究起到了指引作用[9].在此基础上,Li 等人对该方法进行了多视觉特征融合的拓展,证明了融合多种特征相关性得分可以提高标注效果[10].Zhu 等人依据标签和视觉的相关性假设,将图像标签相关矩阵分解,得到一个低秩的改善矩阵和一个稀疏的噪声矩阵[11],但其优化目标的假设前提过于严格.Chen 等人对每一个标签组织正反例样本并训练 SVM 分类器,估计标签与图像的初始相关分数,然后通过基于图的方法对数据集中的相关分数进行改善,可以提升数据集质量[12].然而,由于图像语义标注是一个多标签问题,文中正反例的构造方法对于稍大规模的概念集合而言并不适用.Fan 等人依据标签将图像聚类,以簇为单位估计标签相关性分数,然后通过图上的随机游走改善标签[13].

Bucak 等人将样本中标签不完整现象归结为“不完整类别下的多标签学习”问题,并通过基于排序的多标签学习机制提高了学习性能,但是并没有考虑标签噪声问题[14].文献[15,16]在判别测度学习过程中引入阈值函数,用以提高低频标签的标注效果.实验结果表明,该策略是有效的.Liu 等人首先依据 WordNet 本体对内容无关标签进行过滤,然后依据视觉相似图像语义相近的假设扩展标签,并通过基于知识的同义词、上位词对数据集标签进行扩展[17].在进一步的工作中他们指出,人工标注的数据集质量较低,标签的排序应该反映其与图像内容的相关性,可以通过核密度估计得到标签与视觉特征的初始相关度,并通过标签图上的随机游走获得相关分数,依据相关分数改善标签排序[18].在机器学习领域的近期研究工作中,Nguyen 等人提出,如果样本只有一个准确的类别标签,而其余标签的可信度未知,可以将样本分类描述为一个凸二次规划问题.UCI 数据集上的分类实验结果表明,该方法是有效的[19].但是对于图像语义标注任务而言,该工作的前提条件要求提供一个准确的类别标签,需要额外的信息量进行辅助.Sun 等人认为,多标签学习问题中,样本的已有标签是准确的,但是不能因为样本不具有某类标签就否定该样本的类属关系,而且分类边界须覆盖低密度区域,以解决数据集中正反例失衡现象[20].Pathipan 等人提出区域标注任务中的弱监督学习问题,并在多示例学习框架内求解.实验结果表明,其性能优于传统的基于多示例学习的区域标注方法[21].

上述相关工作从不同的角度处理低质量的数据集,但是大多数方法只着眼于数据集弱标签性的某一个方面,即标签遗漏、标签噪声和数据集欠平衡现象.针对上述弱标签数据集的 3 个特性,本文提出了一种利用语义近邻(semantic nearest neighbor,简称 SNN)的图像语义标注方法.图 2 给出了本文方法的处理流程.

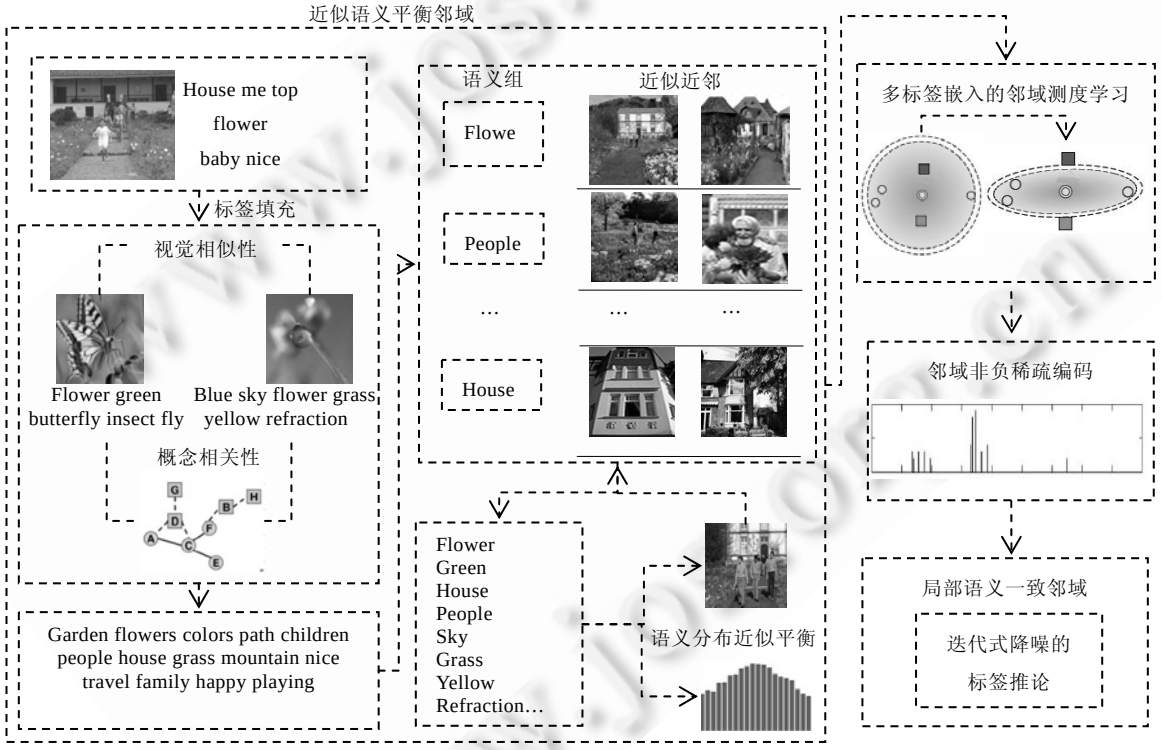


Fig.2 Schematic illustration of the proposed approach

图 2 处理流程示意图

本文方法的核心思想是:构造标签相对准确、语义丰富、分布均衡的语义邻域,以邻域内的语义近邻为指导,对目标图像进行标注.首先,在视觉内容与语义的一致性、语义相关性和稀疏性约束下填充样本标签,以语义组为基础得到近似近邻(approximate nearest neighbor,简称 ANN)和近似语义平衡邻域(approximate semantic

balanced neighborhood,简称 AN).鉴于 AN 中存在噪声干扰现象,进行多标签语义嵌入的邻域最大边际学习,使得近邻处于同一语义子空间;然后进行邻域非负稀疏编码,得到视觉相似并且语义一致的 SNN 和局部语义一致邻域(local semantic consistent neighborhood,简称 CN);最后,在邻域内进行标签推论.

1 近邻的语义平衡性

如果样本邻域内标签分布频率差异较小,其近邻所携带的概念总体上语义就会越丰富,低频标签就可以更有效地参与标注.标签的不完整性将加剧低频标签分布的不平衡[22],因此,首先对数据集进行标签填充.本节提出一种直推式的语义标签填充算法,该方法综合了视觉与标签的一致性约束、标签相关性约束和语义稀疏性约束,更适用于现实环境下的弱标签图像集.令训练集为 $L=\{(x_1,y_1),\ldots,(x_l,y_l)\}\in\mathbb{R}^m\times\{0,1\}^q$,标签集为 $C=\{c_1,c_2,\ldots,c_q\}$,其中,$x_i\in\mathbb{R}^m$ 为图像的视觉特征,特征向量矩阵为 $X\in\mathbb{R}^{l\times m}$. $\hat{y}_i=(\hat{y}_{i1},\ldots,\hat{y}_{iq})\in\{0,1\}^q$ 为对应的标签向量,标签向量的集合构成标签矩阵 $\hat{Y}\in\mathbb{R}^{l\times q}$,填充后的标签矩阵为 Y.$R\in\mathbb{R}^{q\times q}$ 为标签相关矩阵,$R_{i,j}$ 表示标签 c_i 与 c_j 之间的相关性,定义其为 $R_{i,j}=\dfrac{co_{i,j}}{o_i+o_j-co_{i,j}}$,其中,$o_i$ 表示训练集中 c_i 的频数,o_j 表示 c_j 的频数,$co_{i,j}$ 表示 c_i 与 c_j 的共现[2].学习误差函数为 $E(Y)=E_1(Y)+\lambda E_2(Y)+\tau E_3(Y)$.视觉特征相似的图像,其附属标签向量应该相近,因此令 $E_1(Y)=\|YY^T-XX^T\|^2$,即期望标签矩阵能够反映样本的视觉相似性.语义接近的标签,其共现相关性较高.例如,一幅描述“海洋”的图像,标签“沙滩”和“水”赋予该图像的可能性就较大,数据集中蕴含的这种语境相关信息应对标签填充起到指导作用,因此令 $E_2(Y)=\|Y^TY-R\|^2$.原始标签对标签填充具有指导作用,因此令 $E_3(Y)=\|Y-\hat{Y}\|^2$.

优化目标为 $\min\limits_{Y}\{\|YY^T-XX^T\|^2+\lambda\|Y^TY-R\|^2+\tau\|Y-\hat{Y}\|^2\}$,其中,$\lambda\geqslant 0,\tau\geqslant 0$ 为平衡参数.由于视觉特征表示特定语义时所起作用的程度不同,因此,定义权重向量 $w=(w_1,\ldots,w_m)\in\mathbb{R}^m$,令对角阵 $D=diag(w)$,则 x_i 与 x_j 相似度为 $x_i^TDx_j$.为了避免标签矩阵和权重向量过于稠密,分别引入约束 $\|Y\|_1$ 和 $\|w\|_1$.因此,优化目标改写为

$$E(Y,w)=\min_{Y,w}\{\|YY^T-XDX^T\|^2+\lambda\|Y^TY-R\|^2+\tau\|Y-\hat{Y}\|^2+\delta\|Y\|_1+\rho\|w\|_1\} \tag{1}$$

算法 1 给出了直推式标签填充算法.

算法 1. 直推式标签填充.

输入:特征矩阵 X;标签矩阵 $\hat{Y}$;误差权重 τ;正则项参数 λ,τ,δ,ρ;收敛阈值 ε;

输出:填充后的标签矩阵 Y.

(1) 初始化相关矩阵 $R=\hat{Y}^T\hat{Y}$,初始化 $w_1=1_d$, $Y_1=\hat{Y}$,$t=0$;

(2) 令 $t=t+1$,步长设置为 $\eta_t=1/t$,依据公式(4)求解 Y_{t+1} 和 w_{t+1};

(3) 如果 $\|E(Y_t,w_t)-E(Y_{t+1},w_{t+1})\|\leqslant\varepsilon\|E(Y_t,w_t)\|$,则转步骤(4);否则,转步骤(2);

(4) 输出补充后的标签矩阵 $Y=Y_t$.

将公式(1)表示成复合函数有利于求解,即令:

$$E(Y,w)=\min_{Y,w}\{F(Y,w)+\delta\|Y\|_1+\rho\|w\|_1\},F(Y,w)=\|YY^T-XDX^T\|^2+\lambda\|Y^TY-R\|^2+\tau\|Y-\hat{Y}\|^2 \tag{2}$$

采用次梯度下降求解公式(2),第 t 次迭代为

$$Y_{t+1}=\min_{Y}\frac{1}{2}\|Y-\hat{Y}_{t+1}\|_F^2+\delta\eta_t\|Y\|_1,w_{t+1}=\min_{w}\frac{1}{2}\|w-\hat{w}_{t+1}\|_F^2+\rho\eta_t\|w\|_1 \tag{3}$$

其中,$\hat{Y}_{t+1}=Y_t-\eta_t\nabla_Y F(Y_t,w_t),\hat{w}_{t+1}=w_t-\eta_t\nabla_w F(Y_t,w_t)$. 使用文献[23]中的结果,公式(3)的解为

$$Y_{t+1}=\max(0,\hat{Y}_{t+1}-\delta\eta_t 1_l 1_q),w_{t+1}=\max(0,\hat{w}_{t+1}-\rho\eta_t 1_d) \tag{4}$$

在样本标签的完整性得到提升的基础上,构建近似语义平衡邻域(AN).针对不平衡数据的传统处理手段主要包括数据层的重构和方法层的改进:数据重构是在数据层面上使得样本的数目趋于平衡,包括欠采样、过采样和混合采样策略等;算法层的改进是提出针对不平衡数据的算法.本文没有进行数据集上的全局处理,而只是围绕着样本的视觉邻域进行,即仅保证样本的局部邻域内的标签样本数量得到改善.给定图像 x,将具有标签 c_i

($\forall i \in \{1,2,\ldots,q\}$)的图像构成集合 G_i,x 在 G_i 中的视觉近邻定义为其近似近邻(ANN),近似近邻的集合构成一个语义组 $G_i^x, AN(x)=\{G_i^x\}_{i=1}^q$.图 3 给出了 AN 的实例,左侧为目标图像,右侧第 1 行为其 AN,第 2 行为文献[24]中 JEC 方法生成的邻域.可以看到,AN 中低频标签(landscape,bay,road,meadow)和高频标签(sky,house)均有出现.相对而言,ANN 携带的标签的信息量更大,语义信息也更加丰富.

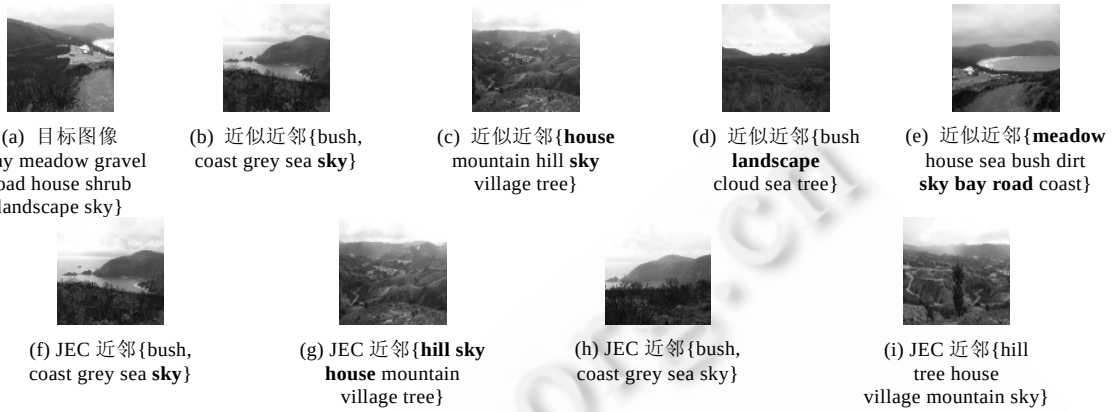


Fig.3 Examples of neighbors in AN and JEC neighborhood

图 3 局部语义平衡邻域实例

2 近邻的语义一致性

2.1 多标签语义嵌入的邻域最大边际学习

AN 中存在噪声干扰,如图 4 所示,I_1 为目标图像,I_2~I_6 为其 ANN,按照与 I_1 的相似度升序排列.其中,I_2,I_3 和 I_1 具有视觉部分相关性,线性重构的结果可以描述 I_1 的视觉内容.如果样本之间是视觉相似的,部分相关和概念相似的,则它们是语义一致的.本节的目标即在 AN 中寻找和 I_1 位于同一线性子空间的近邻,即 I_1 所处语义子空间的一组基.这组基与目标图像是语义一致的,定义其为 I_1 的语义近邻(SNN),SNN 的集合即为 I_1 的局部语义一致邻域(CN).

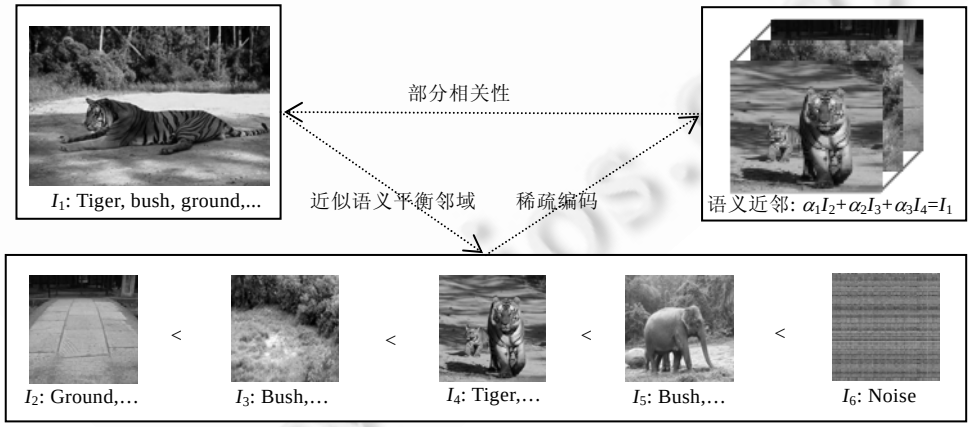


Fig.4 Illustration of partial correlation

图 4 图像间的部分相关性

图像的视觉特征重构有一个前提,即字典中样本与当前图像须处在同一个语义子空间,否则,重构编码就丧失了物理涵义.虽然 ANN 与目标图像视觉相似,但却无法保证语义相似,如图 5(a)所示.从信号重构理论角度来

分析,不能直接用 ANN 重构 x_i.要保证邻域内样本处在同一个语义子空间,重构系数才可以有效地描述其与目标图像的部分相关性,如图 5(b)所示.确保邻域内样本具有语义相似性的关键是指定一个语义度量,该问题是一个多标签测度学习问题.但是现有的测度学习方法不能直接适用于该应用环境,因为以相关成分分析、区分成分分析为代表的方法均需要数据集的辅助信息(side information),即同类样本和异类样本信息,而图像语义标注是一个多类问题,异类样本信息不易获取.因此,本节提出了一种多标签语义嵌入的邻域最大边际学习算法,以解决上述问题.首先构造每一个样本的 Side Information,为了简化符号表示,本节将 x_i 的 AN 用一个三元组表示:

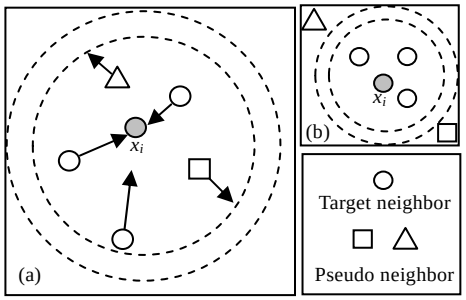


Fig.5 Schematic illustration of AN

图 5 邻域示意图

$$A_i=\{(x_i,A_i^+,A_i^-)\,|\,1<i<l\}.$$

即,将图像 x_i 的 AN 划分为两部分: A_i^+ 与 A_i^-, $A_i^+=\{x_j^+\}_j$.其中,样本和 x_i 至少有一个相同标签,同时具有相对的视觉相似性,为 x_i 的目标近邻(target neighbor),如图 5(a)所示.AN 中剩余的样本构成 $A_i^-=\{x_k^-\}_k$,其样本来源有两种:第 1 种是经过第 1 节样本填充后仍然与 x_i 不具有任何相同标签的样本;第 2 种是与 x_i 具有部分相同标签的样本,但是视觉相似度较低,即 A_i^+ 中已有其代表, A_i^- 中的样本称为 x_i 的伪邻(pseudo neighbor),如图 5(b)所示.需要学习一个线性变换 $L:\mathbb{R}^m\to\mathbb{R}^m$,使得 x_i 与目标近邻的距离须大于其与伪邻的距离.优化目标为

$$\min\left\{\sum_{x_i\in L}\sum_{x_j\in A_i^+}\lambda_{ij}L(x_i,x_j)+\mu\sum_{x_i\in L}\sum_{x_j\in A_i^+}\sum_{x_k\in A_i^-}(1-\lambda_{ik})[1+L(x_i,x_j)-L(x_i,x_k)]_+\right\} \tag{5}$$

$\mu>0$ 为平衡项, $\lambda_{ij}=\dfrac{\|y_i\odot y_j\|_1}{\|y_j\|_1}\in[0,1],\lambda_{ik}=\dfrac{\|y_i\odot y_k\|_1}{\|y_k\|_1}\in[0,1]$,两个参数依据 ANN 的标签相似度进行尺度调节,$\odot$表示向量的按位相乘运算,$[z]_+=\max(0,z)$为 hinge 损失,margin 设定为 1.需要说明的是,公式(5)中距离的度量统一采用点积形式:

$$L(I_1,I_2)=\sum_{i=1}^{n}w(i)\sum_{j=1}^{m_i}u^i(j)\,|\,f_{I_1}^i(j)-f_{I_2}^i(j)\,| \tag{6}$$

其中,n 为特征数,f^i 为第 i 个特征,相应的维度为 m_i, $\sum m_i=m$. u 与 w 是权重向量.公式(6)的求解形式为

$$\left.\begin{aligned}&\arg\min_{w,u}\left\{\sum_{x_i\in L}\sum_{x_j\in A_i^+}\lambda_{ij}L(x_i,x_j)+\mu\sum_{x_i\in L}\sum_{x_j\in A_i^+}\sum_{x_k\in A_i^-}(1-\lambda_{ik})\varsigma_{ijk}\right\}\\&\text{s.t. }L(x_i,x_j^+)-d(x_i,x_k^-)\geqslant 1-\varsigma_{ijk},\varsigma_{ijk}\geqslant 0,w(i)\geqslant 0,|\,w\,|=1,u^i(j)\geqslant 0,|\,\mu\,|=1\end{aligned}\right\} \tag{7}$$

其中,ς_{ijk} 为松弛变量.采用随机梯度投影法可以得到 w 与 u 的近似解[25].

2.2 邻域非负稀疏编码

利用公式(6)取得目标图像的近邻,进行邻域非负稀疏编码.给定样本 x,其 k 个近邻向量构成 x 的一组局部坐标基 $B=[x_1,x_2,\ldots,x_k]\in\mathbb{R}^{m\times k}$.稀疏编码假设信号 x 可以由 B 中少数基向量的线性组合来表示,即

$$x=\sum_{i=1}^{k}b_i a(i)=Ba,$$

其中,a 为 x 的重构系数构成的向量.a 中的负分量将使得坐标基之间的信号抵消,在图像重构中不具有物理含义,因此,a 的每个分量都要为正.如果 a 中非零元素个数不多于 r 个($r<<k$),则信号 x 就是 r 稀疏的.可建立如下模型:

$$\begin{cases}x=Ba\\a(i)\geqslant 0\\\sum a(i)=1\end{cases}.$$

由于数据集中存在噪声,令噪声 $e=e^+-e^-$,$|e|=e^++e^-$, $e^+,e^-\in\mathbb{R}_+^m$,则 $x'=B'a+e$,其中,

$$B'=\begin{bmatrix} B & I_m & -I_m \\ E_{1\times k} & 0_{1\times m} & 0_{1\times m}\end{bmatrix}\in\mathbb{R}^{(m+1)\times(k+2m)},a'=[a\ e^+\ e^-]\in\mathbb{R}^{k+2m},x'=[x\ 1]^T.$$

建立如下非参数模型:

$$\begin{cases}\min\|a'\|_1\\ \text{s.t. } x'=B'a'\\ a'>0\end{cases} \tag{8}$$

公式(8)可通过 YALL1 软件包求解,其解是非负稀疏的.重构向量中的非零分量对应的样本为 x 的 SNN,SNN 全体构成 x 的 CN.算法 2 给出了局部语义一致邻域构造方法.

算法 2. 局部语义一致邻域构造.

输入:数据集 $D=L\cup U$;AN 邻域矩阵 A;邻域范围 k;

输出:邻域矩阵 C.

(1) 求解公式(7),得到 w 与 u 的近似解;

(2) 对于样本 x_i,以公式(6)得到其 k 个近邻,构成局部坐标基 $B_i=[x_{i_1},x_{i_2},...,x_{i_k}]\in\mathbb{R}^{m\times k}$;

(3) 初始化 $B_i'=\begin{bmatrix} B_i & I_m & -I_m \\ E_{1\times k} & 0_{1\times m} & 0_{1\times m}\end{bmatrix}\in\mathbb{R}^{(m+1)\times(k+2m)}$,$x_i'=[x_i\ 1]^{\mathrm{T}}$,求解模型(8),得到近邻权重 $C_{ij}=a'(j)$.

图 6 给出了语义近邻(SNN)与 JEC 方法得到近邻的实例,SNN 与目标图像具备更好的视觉相似性和语义相似性.

Fig.6 Examples of SNN and neighbors obtained by JEC method

图 6 语义近邻实例

3 迭代式降噪的标签推论

CN 邻域矩阵中的权重越大,对应样本相关性越强,标签向量应越相近,因此,目标图像的标签向量可以由对应的 SNN 重构得到,而重构系数即对应的权重.令标签矩阵 $Y=[Y_l\ Y_u]^T$,Y_l 为训练集矩阵,Y_u 为测试集矩阵.目标函数为

$$\min\sum_i\left\|y_i-\sum_j C_{ij}y_j\right\|^2 \tag{9}$$

考虑到标签之间的相关性可以丰富标注结果,因此,标签向量中的非零元素应体现出原始语境中所蕴含

的相关性.在公式(9)的基础上,引入第 1 节中定义了标签相关性矩阵 R,令 y_i^c 表示矩阵 Y 的第 i 列,则公式(9)可改写为 $\min\left\{\sum_i\left\|y_i-\sum_j C_{ij}y_j\right\|^2+\sum_i\sum_j R_{ij}(y_i^c-y_j^c)\right\}$,即 $\min_Y\{tr[(I-C)Y]^T[(I-C)Y]+tr(YRY^T)\}$. 令其对 Y 的梯度为 0,得到 $GY+YC=0$,其中,$G=\frac{1}{2}(H+H^T)$,$H=(I-C)^T(I-C)$,将矩阵 G 分块表示为 $G=\begin{pmatrix}G_{ll} & G_{lu}\\ G_{ul} & G_{uu}\end{pmatrix}$,则可得:

$$G_{uu}Y_u+Y_uR=-G_{ul}Y_l \tag{10}$$

公式(10)是控制系统分析中常用的 Sylvester 方程,由于该类方程是广义的 Lyapunov 方程,利用 Matlab 的 lyap 函数可以求解.因为原始数据集具有噪声干扰,虽然邻域的构造过程可降低噪声对标注性能的影响,但是原始数据集的噪声残留和计算过程的累积误差还是会对最终的标注结果产生负面影响.因此,令预测标签矩阵 $T=[T_l\ T_u]^T$,T_u 为测试集向量(初始元素均为 0),Y_l^* 表示训练集理想标签.引入约束 $\|Y_l^*-T_l\|^2$,保证理想标签与训练集预测结果的近似一致;引入约束 $\|Y_l^*-Y_l\|_1$,保证原始标签与理想标签的近似一致.优化目标改写为

$$\min_T\left\{\sum_{i,j}C_{ij}\|y_i-y_j\|^2+\sum_i\sum_j R_{ij}(y_i^c-y_j^c)+\lambda_1\|Y_l^*-T_l\|^2+\lambda_2\|Y_l^*-Y_l\|_1\right\}.$$

算法 3 给出了求解过程.

算法 3. 迭代式降噪标签推论.

输入:$D=L\cup U$,标签矩阵 Y,邻域矩阵 C,标签关系矩阵 R,参数 λ_1,λ_2;

输出:Y_u.

(1) 令 $Y_l^*=Y_l$,求解 $\min_T\{\|T-CT\|^2+\|TRT^T\|^2+\lambda_1\|T_l-Y_l^*\|^2\}$,获得 T_l;

(2) 固定 T_l,求解 1 范数最小化问题 $\min_{Y_l^*}\left\{\|T_l-Y_l^*\|^2+\frac{\lambda_2}{\lambda_1}\|Y_l^*-Y_l\|_1\right\}$,可利用 YALL1 软件包求解,得到 Y_l^*;

迭代步骤(1)与步骤(2),直至 Y_l 稳定;

(3) 用 Y_l^* 替代公式(10)中的 Y_l,求解公式(10)得到 Y_u.

4 实验与分析

4.1 实验设置

为了验证本文所提出的基于 SNN 的图像语义标注方法的有效性,实验中采用了如下 4 个数据集进行实验:Corel 5K 数据集,包含 5 000 幅图像和 260 个标签;IAPR-TC 12 数据集,包含 19 627 幅图像和 291 个标签;ESP-GAME 数据集,数据来自于网络图像标签生成游戏,包含 20 770 幅图像和 268 个标签;Flickr 数据集,数据来自于图像共享社区,取出现频数在 50 次以上的 457 个标签构成基准标签,最终数据集包含 12 148 幅图像和 457 个标签[24].抽取 3 组视觉特征,包括由 44 维颜色相关图、14 维颜色纹理矩和 6 维颜色矩构成的 64 维全局描述符,384 维的 GIST 特征和降维处理后的 30 维 DenseSurf 特征描述符.使用查准率、查全率、F 值验证平均标注性能.令 $\#(s)$ 表示标注结果中包含标签 w 的样本数,$\#(c)$ 为正确标注的样本数,$\#(t)$ 为测试集中包含标签 c 的样本总数,则有:

$$P(w)=\frac{\#(c)}{\#(s)},R(w)=\frac{\#(c)}{\#(t)},Fmeasure(w)=\frac{2\times P(w)\times R(w)}{P(w)+R(w)}.$$

对测试集所有标签的上述度量求取均值作为评价指标.此外,我们还统计了召回标签数($N+$),其为召回率大于 0 的标签个数,用来评价方法对标签集合的覆盖度.

4.2 参数确定

算法 1 中有 4 个参数需要确定.参数 λ 用以平衡标签填充前后的标签一致性误差和标签相关性误差,固定其

余 3 个参数为τ=1,δ=1,ρ=10,λ在{0.01,0.1,1,10,50,100}取值内发生变化.如图 7 所示(图 7 中的部分子图横轴作了适当的对数处理,以便于显示).

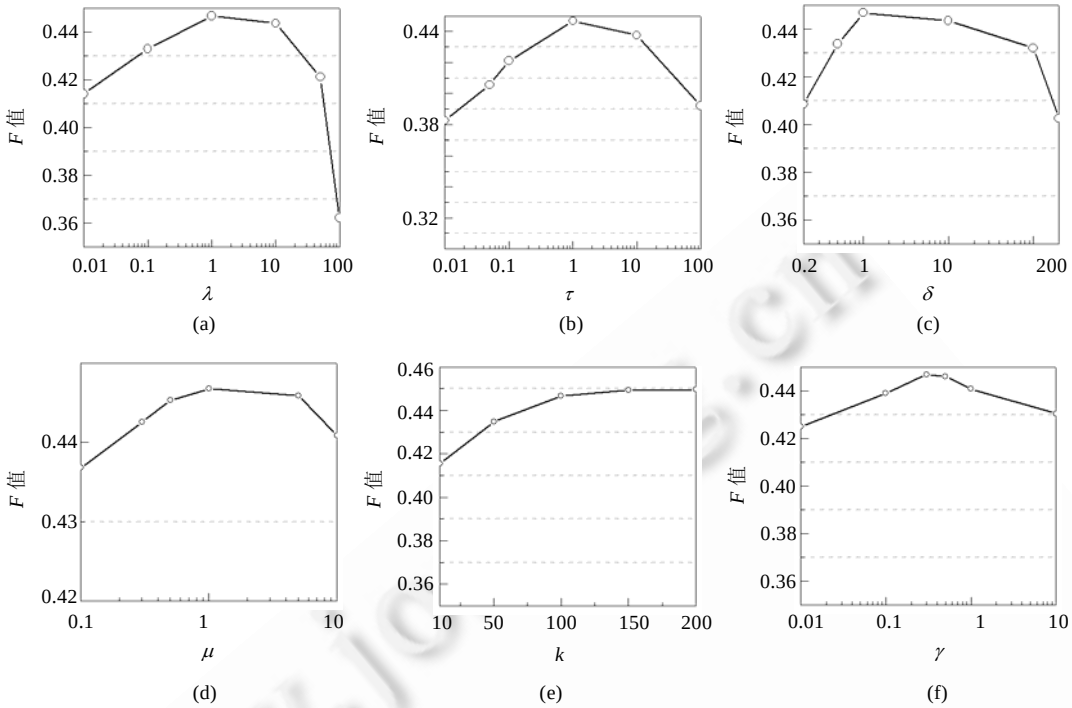


Fig.7 Performance of SNN with different parameter value

图 7 SNN 性能随参数取值的变化

由图 7(a)可以观察到:随着λ取值的递增,系统性能升高,说明标签相关性在标签填充过程中起到了指导作用;但是,当λ大于 1 时,系统性能开始下降,这是因为标签相关性的保持力度过高,权重过大,弱化了公式(1)中其他误差项的约束力,进而导致整体性能下降;而λ过小,将导致目标函数在训练集上过拟合,因此设置λ=1.参数τ用来约束原始标签对标签填充的指导性,令λ=1,δ=1,ρ=10,τ在{0.01,0.05,0.1,1,10,100}取值内变化,如图 7(b)所示,其最佳范围在也在 1 左右,较低的权重取值弱化了原始标签的指导作用,较高的取值导致了训练集上的过拟合,设定其为 1.参数δ用来约束标签的稀疏程度,δ在{0.2,0.5,1,10,100,200}取值内发生变化,如图 7(c)所示,随着δ的升高,系统性能有所增强,说明稀疏性约束对标签填充过程有效果,但是过高的δ取值同样导致了标签填充算法对原始标签的过度拟合,进而影响了标注性能,因此设定其值为 1.参数ρ用来约束视觉特征权重的稀疏程度,由于异构视觉特征表示特定高层语义时所起作用的重要程度不同,权重的适应性调节有助于增强特征的判别力,但是过高的取值会导致过拟合,其值设定为 10.参数ε为 AN 的邻域范围,在 Corel 5K,IAPR-TC 12,ESP-GAME,Flickr 数据集上分别设置为 4,3,3,2.第 2 节算法 2 中参数μ用于平衡伪邻与目标近邻和目标图像的距离,在{0.1,0.3,0.5,1,5,10}取值内发生变化,如图 7(d)所示,设置μ=1.参数 k 用于确定样本局部字典的范围,在{10,50,100, 150,200}取值内发生变化,如图 7(e)所示,过小的取值导致系统性能的下降,设置为 100.第 3 节算法 3 中参数λ_1 用于约束理想标签与训练集预测结果的近似一致,验证其在{50,100,200,300}取值下系统的性能变化,设置为 100.参数$\gamma=\lambda_2/\lambda_1$ 用于平衡预测结果和理想标签的误差,验证其在{0.01,0.1,0.3,0.5,1,10}取值情况下的系统性能,结果如图 7(f)所示,设置为 0.3,其值过大或者过小均会导致训练集上的过拟合,进而影响标注性能.

4.3 实验结果

实验中测试各种方法在 3 种环境下的性能:第 1 种是针对标签不完整情况下的标注性能,第 2 种是在标签

遗漏和不准确环境下的标注性能,第 3 种是在标准评测集下的标注性能.首先,我们测试第 1 种情况,为了模拟真实环境下用户提供弱标签样本的行为,在 Corel 5K 数据集上,依据一定比例随机删除样本的一部分标签,定义该比例为遗漏标签率(missing label rate),该比例越高,则标签越不完整.为了进一步评价算法利用弱标签数据集进行标注的性能,本文进行了第 2 种环境下的实验设置:依据一定比例随机删除样本的一部分标签,然后从词汇表中随机添加对应比例的标签.这里,定义该比例为弱标签率(weak label rate),该比例越高,则噪声越强.表 1 给出了标签不完整和不准确情况下的训练集样本示例(斜体显示标签为噪声).

Table 1 Illustration of incomplete and imprecise labels in training set

表 1 标签不完整和不准确情况下的训练集样本示例

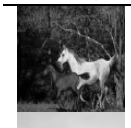
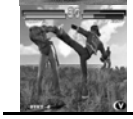
目标图像	基准标签	不完整标签	不完整和不准确标签
	Field horses mare foals	Horses	Horses field *building baby*
	Fence mountain range	Mountain	Mountain range *player*
	Fight grass game	Fight	Fight *bird tile*

图 8(a)为随遗漏率发生变化的实验结果,图 8(b)为随弱标签率发生变化的实验结果.其中,*r* 表示进行标签填充,*a* 表示构建 AN,*c* 表示构建 CN,*d* 表示降噪推论.从实验结果可以看到,随着遗漏率和弱标签率的增加,标注性能逐步降低,而且算法受弱标签率的影响更大,说明标签的不完整性和不准确性对标注结果有较大影响.可以观察到,SNN(*r*+*a*+*m*+*d*)与 SNN(*a*+*m*+*d*)的 *F* 值对比较为明显,两种情况下平均性能分别提升了 21.2%和 29.4%,所以算法 1 对于标签遗漏和噪声具有较强的鲁棒性.从实验结果还可以看到,构建 AN 能够提升标注结果性能,其与不构造 AN 的 SNN(*m*+*d*)相比,性能分别提升了 16.2%和 38.1%.CN 的构造使得近邻处于同一个语义子空间,与直接在样本邻域上进行稀疏表示的 SNN(*d*)相比,平均标注性能提升了 18.7%和 23.4%.

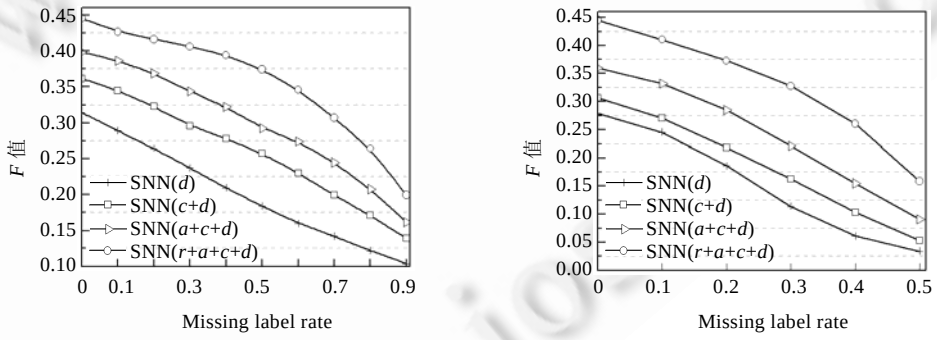


Fig.8 *F* value with various missing label rate and weak label rate

图 8 随遗漏率和弱标签率变化的 *F* 值

我们将 SNN 与其他标注方法进行了对比.对比方法包括 JEC[24],Tagprop[15,16],Tagprop(s)[15,16],GS[26]和 SML[27].表 2 记录了实验结果的 *F* 值.从表 2 的实验结果可知,各种方法在遗漏率和弱标签率上升的情况下,性能均逐步下降.因为标签的完整性与准确性对于 GS 建立视觉特征群组、SML 建立语义判别模型至关重要,所以两种方法受遗漏率和弱标签率影响较大.采用加权近邻模型和阈值函数的 Tagprop(s)提高了低频标签的标注效

果,性能优于 GS 与 SML.采用特征均等贡献的 JEC 方法在弱标签环境下的表现要好于 SML,其贪婪搜索的方式表现出了一定的适应能力.SNN 取得了显著优于其他方法的平均性能.

Table 2 Comparison in terms of *F* value of SNN method and those reported in a selection of representative work

表 2 SNN 方法与其他代表性方法的 *F* 值比较

(a) 不同标签遗漏率下的 *F* 值

Missing label rate	SML	JEC	GS	TagProp	Tagprop(s)	SNN
0.1	0.215	0.271	0.283	0.301	0.353	0.427
0.2	0.164	0.235	0.250	0.276	0.336	0.416
0.3	0.131	0.206	0.224	0.259	0.323	0.406
0.4	0.104	0.172	0.192	0.246	0.307	0.394
0.5	0.078	0.146	0.161	0.232	0.288	0.374
0.6	0.060	0.125	0.135	0.211	0.262	0.346
0.7	0.039	0.107	0.122	0.186	0.233	0.307
0.8	0.029	0.092	0.108	0.167	0.196	0.264
0.9	0.021	0.084	0.093	0.147	0.175	0.199

(b) 不同弱标签率下的 *F* 值

Weak label rate	SML	JEC	GS	TagProp	Tagprop(s)	SNN
0.1	0.131	0.217	0.236	0.272	0.321	0.410
0.2	0.066	0.137	0.155	0.215	0.264	0.373
0.3	0.044	0.080	0.100	0.163	0.208	0.328
0.4	0.023	0.034	0.042	0.109	0.153	0.260
0.5	0.011	0.027	0.034	0.062	0.111	0.159

为测试本文方法对训练集规模的适用性,按照表 3 中训练集的 40%,60%和 100%这 3 种比例构造训练集合,每幅图像语义标注 5 个标签.表 4 记录了 50 次独立随机实验的准确率(*P*),平均标签召回率(*R*),*F* 值(*F*)和召回标签数(*N*+).

Table 3 Performance of various methods on benchmark dataset

表 3 标准评测集下的性能

(a) 40%训练集

标注方法	Corel 5K 数据集				IAPR-TC12 数据集				ESP-GAME 数据集				Flickr 数据集			
	P	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+
SML	0.15	0.18	0.164	114	0.13	0.16	0.143	147	0.11	0.13	0.119	126	0.10	0.12	0.109	216
JEC	0.17	0.22	0.192	128	0.22	0.24	0.230	185	0.16	0.21	0.182	170	0.15	0.19	0.168	279
Tagprop	0.21	0.28	0.240	133	0.39	0.21	0.273	212	0.41	0.19	0.260	195	0.36	0.14	0.202	315
Tagprop(s)	0.26	0.32	0.287	147	0.34	0.27	0.301	231	0.33	0.23	0.271	204	0.28	0.21	0.240	354
GS	0.19	0.24	0.212	134	0.26	0.24	0.250	203	0.24	0.18	0.206	183	0.23	0.15	0.182	301
SNN	0.34	0.36	0.350	165	0.44	0.32	0.371	264	0.43	0.26	0.324	238	0.41	0.26	0.318	372

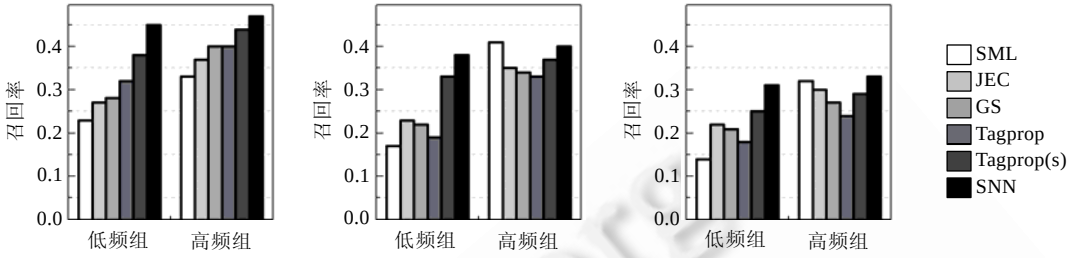
(b) 60%训练集

标注方法	Corel 5K 数据集				IAPR-TC12 数据集				ESP-GAME 数据集				Flickr 数据集			
	P	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+
SML	0.21	0.23	0.220	120	0.18	0.19	0.185	179	0.15	0.16	0.155	163	0.13	0.14	0.135	253
JEC	0.24	0.28	0.258	135	0.26	0.25	0.255	205	0.19	0.22	0.204	181	0.18	0.20	0.189	316
Tagprop	0.27	0.31	0.289	140	0.45	0.23	0.304	227	0.44	0.20	0.275	204	0.34	0.18	0.235	349
Tagprop(s)	0.30	0.34	0.319	152	0.43	0.31	0.360	257	0.35	0.25	0.292	224	0.30	0.23	0.260	385
GS	0.26	0.29	0.274	138	0.30	0.26	0.279	224	0.29	0.22	0.250	201	0.27	0.21	0.236	347
SNN	0.41	0.39	0.400	175	0.50	0.34	0.405	273	0.46	0.28	0.348	247	0.44	0.27	0.335	417

(c) 100%训练集

标注方法	Corel 5K 数据集				IAPR-TC12 数据集				ESP-GAME 数据集				Flickr 数据集			
	P	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+	*P*	*R*	*F*	*N*+
SML	0.25	0.28	0.264	132	0.27	0.29	0.280	226	0.21	0.23	0.220	201	0.17	0.20	0.184	318
JEC	0.26	0.32	0.287	145	0.28	0.29	0.285	231	0.23	0.26	0.244	224	0.21	0.21	0.210	355
Tagprop	0.30	0.36	0.327	158	0.48	0.26	0.337	245	0.48	0.21	0.292	227	0.43	0.18	0.254	382
Tagprop(s)	0.32	0.41	0.360	165	0.46	0.35	0.397	268	0.41	0.27	0.326	239	0.35	0.25	0.292	403
GS	0.31	0.34	0.324	149	0.33	0.28	0.302	243	0.37	0.24	0.291	225	0.31	0.22	0.250	375
SNN	0.45	0.46	0.455	197	0.53	0.39	0.449	284	0.53	0.32	0.399	253	0.50	0.31	0.383	431

从实验结果可知,本文提出的 SNN 方法在不同的训练规模下均取得了最好的标注效果,且其优势在 ESP-GAME 和 Flickr 两个网络数据集上更加明显.在 Flickr 数据集上,其 *F* 值比 Tagprop(s)高出 24%.在 ESP-GAME 数据集上,其 *F* 值比 Tagprop(s)高出 18.4%.进一步分析可知,SML 采用的语义建模方法在弱标签训练集上很不理想,而且利用特征相关性的 GS 方法较均等贡献的 JEC 方法性能略高.最后,依据频率将基准词汇表划分为规模均等的低频组与高频组,图 9 给出了低频组与高频组上各种方法的召回率对比.观察可知,SNN 和 Tagprop(s)两种方法的性能较为均衡,可以得出结论:构建 AN 并未大幅度降低高频标签的标注性能,而且 SNN 在低频和高频分组上的召回率均高于其他方法.



(a) Corel 5K (b) IAPR-TC12 (c) ESP-GAME

Fig.9 Mean recall for the partitions consists of the least frequent labels and frequent labels

图 9 低频标签组与高频标签组上的召回率

表 4 给出了不同方法对 3 个实例的标注结果,与基准匹配的词汇用粗体显示.可以发现,SNN 标注效果好于其他方法,一些标签虽然和基准不同,但也可以描述图像内容.

Table 4 Examples of images with labels generated by different methods (labels in bold font are matching)

表 4 不同标注方法的标注结果实例(粗体显示标签为匹配标签)

Image	Groundtruth	JEC	Tagprop(s)	GS	SNN
	Building water house city urban boat landscape canal	**House** dreamy green scenery roof **water** travel blue art church	Harbor **city water house** lake travel people **boat** window river	**City** lake **boat** scenery **water house** holiday black stone street	**Building** river **water** green **city urban** window **canals** lake **boat**
	Family home man woman child girl person parent daughter	Hot hair white **woman** face party **person girl** boy eye	Hair **man woman person** dress hand toy **girl** smile boy	**Man** white portrait hair movie **woman girl** lady dance eye	**Man woman** hair **home** happy smile **person child mother girl**
	Ocean tree water rock clouds sky grass railroad	Sea snow **water** square lake **sky** beach **clouds** dress people	**Ocean grass** island ball boat **water** sea **clouds sky** landscape	**Clouds sky** sea island **water ocean** road boat green sand	**Grass tree sky** landscape green **clouds water rock** road **railroad**

5 结束语

真实环境下数据集中存在标签噪声和语义分布欠平衡现象,导致现有的图像语义标注方法在实际应用中性能有所下降.本文提出了一种适合弱标签数据集的图像语义标注方法.该方法首先在损失误差最小化意义下填充标签,保证数据集填充前后的视觉内容与标签语义的一致性、标签相关性和语义稀疏性,得到近似近邻.通过多标签嵌入的距离测度学习,使得近邻处于同一语义子空间.在此基础上进行邻域非负稀疏编码,得到语义近邻.以语义近邻为指导,结合语境相关信息,通过迭代式降噪的标签推论方法进行标签预测.实验结果表明,本文

所提方法有效地减弱了数据集弱标签性的影响,标注性能提高显著,更适用于弱标签数据环境.如何更好地融合概念相关性和视觉相关性,改善弱标签数据集上的标注性能,是下一步的主要研究内容.

References:

[1] Kennedy LS, Chang SF, Kozintsev IV. To search or to label? Predicting the performance of search-based automatic image classifiers. In: Proc. of the ACM Int'l Multimedia Conf. and Exhibition. New York: ACM Press, 2006. 245−258. [doi: 10.1145/1178677.1178712]

[2] Sigurbjörnsson B, van Zwlo R. Flickr tag recommendation based on collective knowledge. In: Proc. of the Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2008. 327−336. [doi: 10.1145/1367497.1367542]

[3] Zhang DS, Islam MM, Lu GJ. A review on automatic image annotation techniques. Pattern Recognition, 2012,45(1):346−362. [doi: 10.1016/j.patcog.2011.05.013]

[4] Wang M, Ni BB, Hua XS, Chua TS. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. ACM Computing Surveys, 2012,44(4):1−24. [doi: 10.1145/2333112.2333120]

[5] Jin Y, Khan L, Wang L, Awad M. Image annotations by combining multiple evidence & Wordnet. In: Proc. of the ACM Int'l Conf. on Multimedia. New York: ACM Press, 2005. 706−715. [doi: 10.1145/1101149.1101305]

[6] Wang CH, Jing F, Zhang L, Zhang HJ. Image annotation refinement using random walk with restarts. In: Proc. of the ACM Int'l Conf. on Multimedia. New York: ACM Press, 2006. 647−650. [doi: 10.1145/1180639.1180774]

[7] Jin Y, Khan L, Prabhakaran B. Knowledge based image annotation refinement. Journal of Signal Processing Systems, 2010,58(3): 387−406. [doi: 10.1007/s11265-009-0391-y]

[8] Xu H, Wang JD, Hua XS, Li SP. Tag refinement by regularized LDA. In: Proc. of the ACM Int'l Conf. on Multimedia. New York: ACM Press, 2009. 573−576. [doi: 10.1145/1631272.1631359]

[9] Li XR, Snoek CGM, Worring M. Learning social tag relevance by neighbor voting. IEEE Trans. on Multimedia, 2009,11(7): 1310−1322. [doi: 10.1109/TMM.2009.2030598]

[10] Li XR, Snoek CGM, Worring M. Unsupervised multi-feature tag relevance learning for social image retrieval. In: Proc. of the Int'l Conf. on Image and Video Retrieval. New York: ACM Press, 2010. 10−17. [doi: 10.1145/1816041.1816044]

[11] Zhu GY, Yan SC, Ma Y. Image tag refinement towards low-rank, content-tag prior and error sparsity. In: Proc. of the Int'l Conf. on Multimedia. New York: ACM Press, 2010. 461−470. [doi: 10.1145/1873951.1874028]

[12] Chen L, Xu D, Tsang IW, Luo JB. Tag-Based Web photo retrieval improved by batch mode re-tagging. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2010. 3340−3446. [doi: 10.1109/CVPR.2010.5539988]

[13] Fan JP, Shen Y, Zhou N, Gao YL. Harvesting large-scale weakly-tagged image databases from the Web. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2010. 802−809. [doi: 10.1109/CVPR.2010.5540135]

[14] Bucak SS, Jin R, Jain AK. Multi-Label learning with incomplete class assignments. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2011. 2801−2808. [doi: 10.1109/CVPR.2011.5995734]

[15] Guillaumin M, Mensink T, Verbeek J, Schmid C. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: Proc. of the Int'l Conf. on Computer Vision. Piscataway: Institute of Electrical and Electronics Engineers Inc., 2009. 309−316. [doi: 10.1109/ICCV.2009.5459266]

[16] Verbeek J, Guillaumin M, Mensink T, Schmid C. Image annotation with TagProp on the MIRFLICKR set. In: Proc. of the ACM SIGMM Int'l Conf. on Multimedia Information Retrieval. New York: Association for Computing Machinery, 2010. 537−546. [doi: 10.1145/1743384.1743476]

[17] Liu D, Hua XC, Wang M, Zhang HJ. Image retagging. In: Proc. of the ACM Multimedia Conf. New York: ACM Press, 2010. 491−500. [doi: 10.1145/1873951.1874031]

[18] Liu D, Hua XS, Yang LJ, Wang M, Zhang HJ. Tag ranking. In: Proc. of the Int'l World Wide Web Conf. New York: Association for Computing Machinery, 2009. 351−360. [doi: 10.1145/1526709.1526757]

[19] Nguyen N, Caruana R. Classification with partial labels. In: Proc. of the ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2008. 551−559. [doi: 10.1145/1401890.1401958]

[20] Sun YY, Zhang Y, Zhou ZH. Multi-Label learning with weak label. In: Proc. of the National Conf. on Artificial Intelligence. American Association for Artificial Intelligence, 2010. 593−598. http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1948/2045

[21] Siva P, Russell C, Xiang T. In defence of negative mining for annotating weakly labelled data. In: Proc. of the European Conf. on Computer Vision. Heidelberg: Springer-Verlag, 2012. 594−608. [doi: 10.1007/978-3-642-33712-3_43]

[22] Heymann P, Koutrika G, Garcia-Molina H. Can social bookmarking improve Web search? In: Proc. of the Conf. on Web Search and Web Data Mining. New York: Association for Computing Machinery, 2008. 195−206. [doi: 10.1145/1341531.1341558]

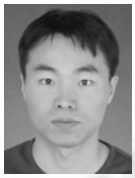
[23] Ioffe A. Composite optimization: Second order conditions, value functions and sensityvity. Analysis and Optimization of Systems, 1992,144(1):442−451.

[24] Makadia A, Pavlovic V, Kumar S. A new baseline for image annotation. In: Proc. of the European Conf. on Computer Vision. Berlin: Springer-Verlag, 2008. 316−329. [doi: 10.1007/978-3-540-88690-7_24]

[25] Shaler-Shwartz S, Singer Y, Srebro N. Pegasos: Primal estimated sub-gradient solver for SVM. Mathematical Programming, 2011, 127(1):3−30. [doi: 10.1007/s10107-010-0420-4]

[26] Zhang ST, Huang JZ, Huang YC, Yu Y, Li HS, Metaxas DN. Automatic image annotation using group sparsity. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2010. 3312−3319. [doi: 10.1109/CVPR.2010.5540036]

[27] Carneiro G, Chan AB, Moreno PJ, Vasconcelos N. Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(3):394−410. [doi: 10.1109/TPAMI.2007.61]



田枫(1980－),男,黑龙江安达人,博士生,讲师,主要研究领域为多媒体语义分析,模式识别,计算机视觉.
E-mail: tianfeng80@gmail.com

沈旭昆(1965－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为虚拟现实与可视化,计算机视觉,多媒体内容管理.
E-mail: xkshen@vrlab.buaa.edu.cn