计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2014.20121087
51(8):1821-1832,2014

弱标签环境下基于语义邻域学习的图像标注

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)
(tianfeng80@gmail.com)

Image Annotation by Semantic Neighborhood Learning from Weakly Labeled Dataset

Tian Feng[1,2] and Shen Xukun[1]
[1](*State Key Laboratory of Virtual Reality Technology and Systems* (*Beihang University*), *Beijing* 100191)
[2](*School of Computer and Information Technology*, *Northeast Petroleum University*, *Daqing*, *Heilongjiang* 163318)

Abstract With the advance of Web technology, image sharing has become much easier than ever before. Automatic image annotation, which can predict relevant labels for images, is becoming more and more important. Traditional image annotation methods usually require a large number of complete, accurate labeled data to obtain good annotation performance. However, since obtaining weak labeled training data is often much easier and costs less efforts than obtaining a large amount of fully labeled training data, image labels are often incomplete and inaccurate in real world environment. In addition, different labels usually have large frequency differences. To effectively harness these weakly labeled images, in this paper, an automatic image annotation approach based on semantic neighborhood learning (SNLWL) is proposed. The missing labels are replenished by minimizing the reweighted error functions on training data. Then, the semantic neighborhood is obtained by a progressive neighborhood construction approach. We incorporate label completeness, global similarity, conceptual similarity, and partly correlation into the stage. In addition, an effective label inference strategy is proposed by minimizing the neighborhood reconstruction error to handle the noise in the labels. Extensive experimental results on different benchmark datasets show that the proposed approach makes a marked improvement as compared with other methods.

Key words image annotation; automatic annotation; weak label; semantic neighborhood; neighborhood learning

摘 要 图像语义自动标注是实现图像语义检索与管理的关键,是具有挑战性的研究课题.传统的图像标注方法需要具有完整、准确标签的数据集才能取得较好的标注性能.然而,在现实应用中获得数据的标签往往是不准确、不完整的,并且标签分布不均衡.对于Web图像和社会化图像尤其如此.为了更好地利用这些弱标签样本,提出了一种基于语义邻域学习的图像自动标注方法(semantic neighborhood learning from weakly labeled image, SNLWL).首先在邻域标签损失误差最小化意义下,填充训练集样本标签.通过递进式的邻域选择过程,保证建立的语义一致邻域内样本具有全局相似性、部分相关性和语义一致性,并且语义标签分布平衡.在邻域标签重构误差最小化意义下进行标签预测,降低噪声标签

收稿日期:2012-11-27;**修回日期**:2013-05-02
基金项目:国家"八六三"高技术研究发展计划基金项目(2009AA012103);国家自然科学基金项目(61170132,60533070);黑龙江省教育厅科学技术研究项目(12511011,12521055);东北石油大学青年科学基金项目(2013NQ120)

对性能的影响. 多个数据集上的实验结果表明,与已知的具有较好标注效果的方法相比,此方法更适用于处理弱标签数据集,标准评测集上的测试也表明了此方法的有效性.

关键词 图像标注;自动标注;弱标签;语义邻域;邻域学习

中图法分类号 TP391

图像自动标注的任务是给图像添加反映图像内容语义的文本标签. 传统的图像标注方法,一个基本假设就是训练集中样本标签是准确的和完整的,但是在真实应用环境中,这个假设很难成立,因为给出图像的所有文本标签是一项繁琐、费力的工作,用户往往只是给出少部分标签,而不是完整的标签列表. 此外,一幅图像的标签也不是一定和图像相关的. 例如一幅标注有"car"的图像,只是表示该幅图像是在车内拍摄的,图像中并没有"car"的对象. 不论是带有关联文本信息的网页图像,还是图像共享社区中用户人工标注的图像,其原始标签描述常常是不准确和不完整的. 这种不完备情况称为弱标签性.

保证每幅图像都具有完整准确的标签是一个非常苛刻的条件,即使在 COREL5K, IAPR-TC 12 等评测集[1]上很多图像的标签也不完整,在数据规模和标签集合数目较大时该问题更加明显. 同时,在 COREL5K 数据集上统计表明,对于 20%的低频标签,采用 JEC(joint equal contribution)方法[1]只能达到 19.7%的 $F1$ 值,但是对于 20%的高频标签,其 $F1$ 值却可以达到 50.6%. 这是由于数据集的不平衡引发的小区块问题导致的模型决策边界偏移,而如翻译模型、跨媒体相关模型、多伯努利相关模型[1]等,标注的准确率依赖于概率分布的准确估计,当标签的频率较低时,低频标签上的概率估计的准确率将远小于高频标签,因此标签不平衡性将降低标注方法性能.

本文的"弱标签图像"包含 3 个层面的含义:1)图像的原始标签并不完整,即训练集中图像的准确标签只是真实标签的一部分;2)原始标签可能并不准确,即其中存在着一定程度的噪声;3)标签分布频率差异较大,存在语义标签分布的不平衡. 因此弱标签图像集即是指由语义标签分布不均衡的、不准确的、不完整的标签描述的图像构成的集合. 由于收集样本的所有完整、准确标签需要消耗大量资源,而收集弱标签样本要相对容易,所以弱标签图像大量存在于标准评测集和互联网图像集.

传统的图像标注可以分为基于生成式模型的方法和基于判别式模型的方法[2-3]. 生成式模型需要大量高质量的数据集学习图像视觉特征和语义概念之间的联合概率. 判别式模型将每一个标签看作一个语义类,为每一个标签建立一个分类器,其同样需要大量类别平衡的训练模型. 但是这些方法都不能有效地处理真实环境下的弱标签图像集.

文献[4]研究了当样本只有一个标签是正确的情况下,如何提高分类性能. 文献[5]结合主题模型和判别模型,降低像素标签缺失对图像区域标注的影响. 文献[6-7]考虑到不同标签包含的样本数量不同,利用 sigmoid 函数提高低频标签上的预测效果. 文献[8]采用清洗过滤方法,移除不相关标签. 文献[9]提出了一种从不完整类别数据中进行多标签学习的方法,针对遗漏标签在最小化标签排序误差中引入松弛项,并通过 Group Lasso 得到具有密集非 0 松弛项组的遗漏标签. 如果直接从弱标签数据集上学习,则在学习中要隐含假设,即样本标签是分布平衡的,包含的标签一定是准确的,样本不包含的标签一定不具有这个标签的语义,这很大程度上在学习样本中引入了噪声.

近年来,基于近邻模型的方法在图像标注任务处理中取得了很好的性能[1,6-7,10-12]. 本文期望通过邻域内样本的指导降低弱标签对标注性能的负面影响,由此,提出一种基于语义邻域学习的图像标注方法(semantic neighborhood learning for weakly labeled image, SNLWL). 首先,在正则化框架和标签损失误差最小化意义下,填充训练集样本标签,为每个样本构造标签分布均衡的语义平衡邻域(semantic balanced neighborhood, BN). 其次,在 BN 中进行标签信息嵌入测度学习,获得距离测度和图像语义的一致性,即保证图像和近邻样本处于同一语义子空间,以目标图像的邻域内样本为其局部坐标基,通过非负稀疏表示获得目标图像和近邻的部分相关性,得到与目标图像相关性最紧密的语义一致邻域(semantic consistent neighborhood, CN),CN 中样本具备和目标图像全局相似,部分相关和语义相似性. 最后,本文在 CN 内通过最小化标签重构误差进行标签预测,并通过双正则项降低噪声标签对性能的影响.

1　语义平衡邻域

目标图像的语义平衡邻域是指其邻域内标签分布频率差异较小,这样低频标签可以更有效地参与标注.由于图像标签不完整,即有遗漏标签的现象,所以本文首先填充训练集样本标签.令标签集合 $C=\{c_1,c_2,\cdots,c_q\}$,训练集为 $L=\{(\boldsymbol{x}_1,\boldsymbol{y}_1),\cdots,(\boldsymbol{x}_l,\boldsymbol{y}_l)\}\in\mathbb{R}^m\times\{0,1\}^q$,其中,$\boldsymbol{x}_i\in\mathbb{R}^m$ 表示第 i 幅图像的特征向量,$\boldsymbol{y}_i=(y_{i1},\cdots,y_{iq})\in\{0,1\}^q$ 表示第 i 幅图像所对应的标签向量,$\boldsymbol{Y}=[\boldsymbol{y}_1,\cdots,\boldsymbol{y}_l]^{\mathrm{T}}$ 为标签矩阵.如果第 i 幅图像有第 j 个标签,则 $y_{ij}=1$,否则为 0.令标签输出函数 f 对样本 x_i 的输出值为标签向量 $\boldsymbol{f}_i=[f_{i1},f_{i2},\cdots,f_{iq}]^{\mathrm{T}}$,$\boldsymbol{F}=[\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_l]$ 为输出的标签矩阵.学习的误差函数为:$E(f)=E_1(f)+\lambda E_2(f)$,其中,$\lambda\geqslant 0$ 为正则化参数,优化目标为:

$$\min_f\left\{\frac{1}{2}\sum_{j=1}^{q}\sum_{i=1}^{l}u_{ij}(y_{ij}-f_{ij})^2+\frac{1}{2}\lambda\sum_{i=1}^{l}\sum_{j=1}^{l}w_{ij}\left\|\frac{\boldsymbol{f}_i}{\sqrt{d_i}}-\frac{\boldsymbol{f}_j}{\sqrt{d_j}}\right\|^2\right\},$$

其中 $E_1(f)=\frac{1}{2}\sum_{i=1}^{q}\sum_{j=1}^{l}u_{ij}(y_{ij}-f_{ij})^2$ 为针对弱标签的加权误差函数.u_{ij} 表示样本 $\boldsymbol{x}_i$ 在标签 j 上的误差权重,$u_{ij}=\begin{cases}1, & y_{ij}=1;\\ \tau, & y_{ij}=0.\end{cases}$ 即如果 $\boldsymbol{x}_i$ 具有标签 j,则权值为 1,否则,设置为 $\tau(0\leqslant\tau\leqslant 1)$.最小化 $E_1(f)$ 确保 f 输出的标签与训练集原始标签近似一致.

以训练集样本为顶点,相似度为边建立一个近邻图,边权重设置为 $w_{ij}=\exp\left\{-\frac{\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2}{2\sigma^2}\right\}$,$\sigma$ 固定为样本平均距离.令 $d_i=\sum_{j=1}^{l}w_{ij}$,最小化误差项 $E_2(f)=\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}w_{ij}\left\|\frac{\boldsymbol{f}_i}{\sqrt{d_i}}-\frac{\boldsymbol{f}_j}{\sqrt{d_j}}\right\|^2$ 保证 f 在该图上输出的平滑性.

将 $E_2(f)$ 表示为矩阵形式,$E_2(\boldsymbol{F})=\mathrm{tr}(\boldsymbol{F}^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{V}^{\frac{1}{2}})\boldsymbol{F})$,其中 $\boldsymbol{V}=\mathrm{diag}([d_1,\cdots,d_l])$.令目标函数梯度 $\nabla_{\boldsymbol{F}}E(f)=\boldsymbol{M}\odot[\boldsymbol{F}-\boldsymbol{Y}]+\lambda\boldsymbol{S}\boldsymbol{F}$ 为 0,其中 $\boldsymbol{S}=\boldsymbol{I}-\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{D}^{\frac{1}{2}}$,则可得到方程组:

$$(\mathrm{diag}(\boldsymbol{u}_j)+\lambda\boldsymbol{S})\boldsymbol{f}_j=\boldsymbol{Z}_j,\qquad(1)$$

其中,$\boldsymbol{u}_j=[u_{1j},\cdots,u_{lj}]^{\mathrm{T}}$ 为误差向量,$\boldsymbol{f}_j$ 为标签输出矩阵 $\boldsymbol{F}$ 第 j 列,$\boldsymbol{Z}_j=\boldsymbol{u}_j\odot\boldsymbol{y}_j$,$\boldsymbol{y}_j$ 为原始标签矩阵第 j 列,$\odot$ 为矩阵按位相乘运算.通过最小二乘法可获得该方程组的近似最优解.算法 1 给出了样本标签填充的方法描述.

算法 1. 样本标签填充算法.

输入:训练集 L、原始标签矩阵 $\boldsymbol{Y}$、误差权重 τ、正则项参数 λ、邻域范围 δ;

输出:训练集标签矩阵 $\boldsymbol{Y}'$.

① 建立训练集样本 δ 近邻图,初始化相似度矩阵 $\boldsymbol{W}$,矩阵 $\boldsymbol{S}$,误差权重矩阵 $\boldsymbol{U}$;

② 求解方程组(1),得到 $\boldsymbol{F}_j$,输出矩阵 $\boldsymbol{F}=[\boldsymbol{f}_1,\cdots,\boldsymbol{f}_q]$;

③ 令 $\boldsymbol{Y}'=[\boldsymbol{y}'_1,\cdots,\boldsymbol{y}'_l]^{\mathrm{T}}$,其中 $\boldsymbol{y}'_i=\boldsymbol{f}_i$.

在标签填充过程中会引入一定程度的噪声,对于噪声的处理在后续的标签推论中进行处理.将具有标签 c_i 的图像表示为一个集合 $L_i\subseteq L(\forall i\in\{1,2,\cdots,q\})$.对于图像 $\boldsymbol{x}$,在 L_i 中得到其 ε 个视觉近邻,并形成集合 $L_{x,i}\subseteq L_i$,这样 $L_{x,i}$ 中包含与 $\boldsymbol{x}$ 视觉相似并具有标签 c_i 的图像,定义其为 $\boldsymbol{x}$ 的一个语义组,语义组的集合 $BN(\boldsymbol{x})=\{L_{x,1}\cup\cdots\cup L_{x,q}\}$ 即为 $\boldsymbol{x}$ 的语义平衡邻域.由于 $BN(\boldsymbol{x})$ 中每个标签至少出现了 ε 次,邻域中标签分布达到平衡.算法 2 给出了语义平衡邻域构造方法描述.

算法 2. 语义平衡邻域构造算法.

输入:图像 $\boldsymbol{x}$、填充后的训练集 $L=\{(\boldsymbol{x}_1,\boldsymbol{y}'_1),\cdots,(\boldsymbol{x}_l,\boldsymbol{y}'_l)\}$、邻域范围 ε;

输出:$\boldsymbol{x}$ 语义平衡邻域 $BN(\boldsymbol{x})$.

① 初始化 $\boldsymbol{x}$ 的 q 个语义组 $L_i\subseteq L$;

② For $i=1$ to q

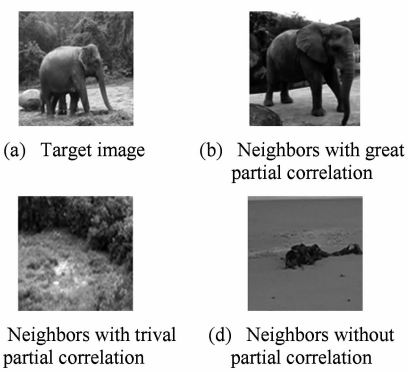
③　　获取 $\boldsymbol{x}$ 在 L_i 中 ε 近邻集合 $L_{x,i}$;

④　　$BN(\boldsymbol{x})=BN(\boldsymbol{x})\bigcup L_{x,i}$;

⑤ End For.

2　语义一致邻域

在语义平衡邻域的基础上构造语义一致邻域的目的是使得邻域内样本具有全局相似性的同时,具有部分相关性和概念相似性.图 1 给出了一个视觉



(a) Target image　(b) Neighbors with great partial correlation

(c) Neighbors with trival partial correlation　(d) Neighbors without partial correlation

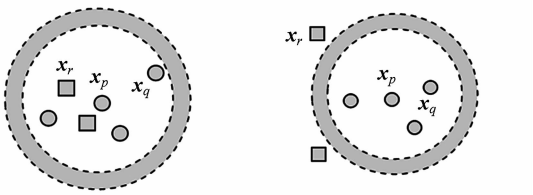
Fig. 1　Illustration of partial correlation.

图 1　图像间的部分相关性

部分相关性的例子,图 1(b)和图 1(c)两幅图像的部分内容语义和图 1(a)相关,因此具有部分相关性,图 1(d)虽然和图 1(a)视觉相似,但是却不具有部分相关性.

压缩感知理论研究表明,稀疏表示的非零系数事实上揭示了信号所属的类别关系,受此结论启发,可以利用邻域图像之间的稀疏表示关系构建邻域.但是从信号表示角度看,如果用不同语义空间中的图像来重构当前图像,重构系数已经丧失了其应有的物理含义.虽然平衡邻域内样本具有和目标图像一定的视觉相似度,但是无法保证语义相似,因为图像的语义相似性受描述其的多个语义标签影响.

图 2 给出了一个图像邻域的示意,图 2(a)中 $\boldsymbol{x}_p$ 的语义相似近邻(用圆圈表示)和语义不相似近邻(用方块表示)均在其邻域内,这导致用 $\boldsymbol{x}_p$ 的近邻重构 $\boldsymbol{x}_p$ 缺乏物理含义.如果如图 2(b)所示,不相似近邻在 $\boldsymbol{x}_p$ 的语义邻域外,就可以用 $\boldsymbol{x}_p$ 邻域内样本重构 $\boldsymbol{x}_p$,进而获得样本的部分相关性.



(a) Unsatisfactory neighborhood (b) The ideal neighborhood

Fig. 2 Schematic illustration of neighborhood.

图 2 图像邻域示意

该问题属于测度学习的范畴,即为具有相同语义的样本学习一个距离度量,使得语义相似近邻距离较小,语义相异样本距离较大.然而,图像标注问题是一个多标签学习问题,经典的测度学习方法不能够直接应用.因此,需要在目标图像的语义平衡邻域内学习一个多标签语义嵌入的测度,利用该测度寻找目标图像的语义相似近邻.

令 $\{\boldsymbol{f}_a^1,\cdots,\boldsymbol{f}_a^n\}$ 与 $\{\boldsymbol{f}_b^1,\cdots,\boldsymbol{f}_b^n\}$ 表示图像 a 与 b,其中 $\boldsymbol{f}^i\in\mathbb{R}^{m_i}$,$i=1,\cdots,n$ 为视觉特征,每个特征维度为 m_i.由于常用的特征度量,如 L_1 距离、L_2 距离或者 χ^2 距离均可以表示为向量的点积,例如,令 $dist_{ab}^i(j)=|\boldsymbol{f}_a^i(j)-\boldsymbol{f}_b^i(j)|$,$\forall j\in\{1,\cdots,m_i\}$,则 L_1 距离可表示为 $L_1(\boldsymbol{f}_a^i,\boldsymbol{f}_b^i)=\sum_{j=1}^{m_i}u^i(j)dist_{ab}^i(j)$,$\boldsymbol{u}^i$ 是一个规一化的向量,表示不同维度的权重.这样,邻域内样本距离表示为

$$\tilde{d}(a,b)=\sum_{i=1}^{n}w(i)\sum_{j=1}^{m_i}u^i(j)dist_{ab}^i(j),\quad(2)$$

其中 $\boldsymbol{w}\in\mathbb{R}^n$ 是各种特征的组合系数向量.

由于 $\boldsymbol{x}_p$ 语义平衡邻域 $BN(\boldsymbol{x}_p)=\{L_{x_p,1}\cup\cdots\cup L_{x_p,q}\}$ 中样本可分为两部分,一部分是具有和 $\boldsymbol{x}_p$ 相同标签的语义组中的样本,表示为 $BN^{\mathrm{t}}(\boldsymbol{x}_p)=\{L_{x_p,r}\}_r(\forall r\ \mathrm{s.t.}\ y'_{pr}=1)$,其中样本和 $\boldsymbol{x}_p$ 视觉相似,又至少和 $\boldsymbol{x}_p$ 有一个相同标签,称为 $\boldsymbol{x}_p$ 的目标近邻,如图 2 中 $\boldsymbol{x}_q$.另一部分表示为 $BN^{\mathrm{f}}(\boldsymbol{x}_p)$,其中样本包含 2 种情况,1 种情况是和 $\boldsymbol{x}_p$ 不具有任何相同标签的样本;第 2 中样本是和 $\boldsymbol{x}_p$ 具有部分相同标签,但是其和 $\boldsymbol{x}_p$ 视觉相似度较低,导致其在 BN 构造中,排除在某个语义组的外面,但是因为图像具有多个标签,其在另外一个语义组中取得了相对高的 *rank* 值,这样其就进入了 $\boldsymbol{x}_p$ 的平衡邻域中,所以其和 $\boldsymbol{x}_p$ 就具有部分相同标签.$BN^{\mathrm{f}}(\boldsymbol{x}_p)$ 样本不论是何种情况,均称之为 $\boldsymbol{x}_p$ 的非目标近邻,如图 2 中 $\boldsymbol{x}_r$.学习目标为通过式(2)中 $\boldsymbol{w}$ 与 $\boldsymbol{u}^i$ 的适应,使得 $\boldsymbol{x}_p$ 与目标近邻 $\boldsymbol{x}_q$ 距离小于 $\boldsymbol{x}_p$ 与非目标近邻 $\boldsymbol{x}_r$ 的距离.误差函数定义为

$$\sum_{pq}\eta_{pq}\lambda_{pq}\tilde{d}(\boldsymbol{x}_p,\boldsymbol{x}_q)+\mu\sum_{pqr}\eta_{pq}(1-\lambda_{pr})[1+\tilde{d}(\boldsymbol{x}_p,\boldsymbol{x}_q)-\tilde{d}(\boldsymbol{x}_p,\boldsymbol{x}_r)]_+,$$

其中,$\mu>0$ 是前后的平衡项,η_{pq} 是一个指示变量,如果 $\boldsymbol{x}_p$ 是 $\boldsymbol{x}_q$ 的目标近邻,则 $\eta_{pq}=1$,否则为 0,$\lambda_{pq}=\frac{\|\boldsymbol{y}'_p\odot\boldsymbol{y}'_q\|_1}{\|\boldsymbol{y}'_q\|_1}\in[0,1]$,$\lambda_{pr}=\frac{\|\boldsymbol{y}'_p\odot\boldsymbol{y}'_r\|_1}{\|\boldsymbol{y}'_r\|_1}\in[0,1]$,$\odot$ 表示向量的按位相乘运算.$[z]_+=\max(0,z)$ 是 hinge 损失,间隔(margin)设定为 1,其最小化保证 $\boldsymbol{x}_q$ 比 $\boldsymbol{x}_r$ 离 $\boldsymbol{x}_p$ 近.λ_{pq} 和 λ_{pr} 依据样本的标签向量重叠情况分别对两个误差项进行缩放.优化目标为

$$\underset{w,u}{\operatorname{argmin}}\sum_{pq}\eta_{pq}\lambda_{pq}\tilde{d}(\boldsymbol{x}_p,\boldsymbol{x}_q)+\mu\sum_{pqr}\eta_{pq}(1-\lambda_{pr})\xi_{pqr},\quad(3)$$

$$\mathrm{s.t.}\ \tilde{d}(\boldsymbol{x}_p,\boldsymbol{x}_q)-\tilde{d}(\boldsymbol{x}_p,\boldsymbol{x}_r)\geqslant 1-\xi_{pqr},\xi_{pqr}\geqslant 0,$$

$$w(i)\geqslant 0,$$

$$\sum_{i=1}^{n}w(i)=1,u^i(j)\geqslant 0,$$

$$\sum_{j=1}^{m_i}u^i(j)=1.$$

式(3)中 ξ_{pqr} 为 hinge 损失.这是一个线性约束下的优化问题,对该问题的原始形式采用交替优化方式的随机梯度投影法[13]获得 $\boldsymbol{w}$ 与 $\boldsymbol{u}^i$ 的近似解.给定目标图像 $\boldsymbol{x}_i$,利用式(2)取得其 k 近邻构成 $\boldsymbol{x}_i$

的局部字典 $\boldsymbol{B}_i=[\boldsymbol{x}_{i_1},\boldsymbol{x}_{i_2},\cdots,\boldsymbol{x}_{i_k}]\in\mathbb{R}^{m\times k}$,其中 $i_p\in\{1,\cdots,l\}$,$p\in\{1,\cdots,k\}$.令 $\boldsymbol{\alpha}_i\in\mathbb{R}^k$ 表示 $\boldsymbol{x}_i$ 的重构系数,这里要保证 $\boldsymbol{\alpha}_i$ 的非负性,因为对于图像标注而言,负系数没有清晰、合理的物理解释,会导致基向量之间通过相互"抵消"来重构信号.所以引入如下约束:$\boldsymbol{x}_i=\boldsymbol{B}_i\boldsymbol{\alpha}_i+\boldsymbol{\zeta}$,其中 $\alpha_i(p)\geqslant 0$,$\sum_{p=1}^{k}\alpha_i(p)=1$.令非负变量 $\boldsymbol{\zeta}^+,\boldsymbol{\zeta}^-\in\mathbb{R}^m$,噪声项 $\boldsymbol{\zeta}=\boldsymbol{\zeta}^+-\boldsymbol{\zeta}^-$,$|\boldsymbol{\zeta}|=\boldsymbol{\zeta}^++\boldsymbol{\zeta}^-$,$\boldsymbol{B}_i'=\begin{bmatrix}\boldsymbol{B}_i & \boldsymbol{I}_m & -\boldsymbol{I}_m\\ \boldsymbol{E}_{1\times k} & \boldsymbol{0}_{1\times m} & \boldsymbol{0}_{1\times m}\end{bmatrix}\in\mathbb{R}^{(m+1)\times(k+2m)}$,$\boldsymbol{\alpha}_i'=[\boldsymbol{\alpha}_i\quad\boldsymbol{\zeta}^+\quad\boldsymbol{\zeta}^-]\in\mathbb{R}^{k+2m}$,$\boldsymbol{x}_i'=[\boldsymbol{x}_i\quad 1]^{\mathrm{T}}$.通过式(4)求解非负表示系数:

$$\min_{\boldsymbol{\alpha}_i'}\lambda\|\boldsymbol{\alpha}_i'\|_1+\frac{1}{2}\|\boldsymbol{x}_i'-\boldsymbol{B}_i'\boldsymbol{\alpha}_i'\|_2^2,\ \text{s.t.}\ \boldsymbol{\alpha}_i'\geqslant 0 \quad (4)$$

其中正则化参数 λ 为 $2\|\boldsymbol{B}'\boldsymbol{x}'\|_\infty$.

求解这个 ℓ_1 最小化问题,可以得到目标图像在其语义邻域中的非负稀疏线性表示,表示系数非 0 的近邻具备和 $\boldsymbol{x}_i$ 的区域视觉相似性,即部分相关性,由于这些近邻来自于语义平衡邻域,并且在一个语义子空间中,所以具备和目标图像的全局相似性、部分相关性和概念一致性.以这些近邻构成目标图像 $\boldsymbol{x}_i$ 的语义一致邻域 CN.算法 3 给出了语义一致邻域构造算法描述.

算法 3. 语义一致邻域构造算法.

输入:训练集、语义平衡邻域矩阵 $\boldsymbol{B}=[b_{ij}]$、邻域范围 k、正则项系数 μ;

输出:训练集样本的语义一致邻域矩阵 $\boldsymbol{C}=[c_{ij}]$.

① 依据语义平衡邻域矩阵 $\boldsymbol{B}=[b_{ij}]$,用随机梯度下降法求解式(3),获得权重向量 $\boldsymbol{w}$ 与 $\boldsymbol{u}^i$ 的近似解;

② 依据式(2)获取每一个样本 $\boldsymbol{x}_i$ 的 k 个近邻构成局部字典 $\boldsymbol{B}_i=[\boldsymbol{x}_{i_1},\boldsymbol{x}_{i_2},\cdots,\boldsymbol{x}_{i_k}]\in\mathbb{R}^{m\times k}$,其中 $i_p\in\{1,\cdots,n\}$,$p\in\{1,\cdots,k\}$;

③ 初始化 $\boldsymbol{B}_i'=\begin{bmatrix}\boldsymbol{B}_i & \boldsymbol{I}_m & -\boldsymbol{I}_m\\ \boldsymbol{E}_{1\times k} & \boldsymbol{0}_{1\times m} & \boldsymbol{0}_{1\times m}\end{bmatrix}\in\mathbb{R}^{(m+1)\times(k+2m)}$,$\boldsymbol{x}_i'=[\boldsymbol{x}_i\quad 1]^{\mathrm{T}}$,求解式(4);

④ 邻域矩阵元素 $c_{ij}=\begin{cases}\boldsymbol{\alpha}_i'(p), & j=i_p;\\ 0, & \text{else.}\end{cases}$

3　标签推论

依据 CN 中近邻向目标图像传播标签,图像的标签向量可以由其近邻的标签向量重构,且重构系数即为其视觉特征的重构系数,即 CN 邻域矩阵中权重 c_{ij} 反映了两个样本具有相同标签的可能性,用 c_{ij} 表示样本 $\boldsymbol{x}_i$ 和 $\boldsymbol{x}_j$ 的标签向量的重构系数.令矩阵 $\boldsymbol{f}=\begin{bmatrix}\boldsymbol{f}_{\mathrm{L}}\\ \boldsymbol{f}_{\mathrm{U}}\end{bmatrix}$,$\boldsymbol{f}_{\mathrm{L}}$ 为训练集样本标签向量矩阵,$\boldsymbol{f}_{\mathrm{U}}$ 为未知图像的标签向量矩阵(初始元素均为 0).标签重构误差函数为

$$E(f)=\sum_{i=1}^{n}\Big\|\boldsymbol{f}_i-\sum_{j\neq i}c_{ij}\boldsymbol{f}_j\Big\|^2,\ \text{s.t.}\ \boldsymbol{f}_i=\boldsymbol{y}_i', \quad (5)$$

其中,$\boldsymbol{y}_i'$ 为图像 $\boldsymbol{x}_i$ 填充后的标签向量,最小化式(5),优化目标表示为矩阵形式为

$$\min_{\boldsymbol{f}}[(\boldsymbol{I}-\boldsymbol{C})\boldsymbol{f}]^{\mathrm{T}}[(\boldsymbol{I}-\boldsymbol{C})\boldsymbol{f}],\ \text{s.t.}\ \boldsymbol{f}_{\mathrm{L}}=\boldsymbol{Y}'.$$

令目标函数梯度等于 0,即 $\nabla_f E(\boldsymbol{f})=(\boldsymbol{H}+\boldsymbol{H}^{\mathrm{T}})\boldsymbol{f}=\boldsymbol{M}\boldsymbol{f}=0$,其中 $\boldsymbol{M}=\boldsymbol{H}+\boldsymbol{H}^{\mathrm{T}}$,$\boldsymbol{H}=(\boldsymbol{I}-\boldsymbol{C})^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{C})$ 为对称阵.$\boldsymbol{M}$ 按 l 行,l 列分块表示为 $\boldsymbol{M}=\begin{bmatrix}\boldsymbol{M}_{\mathrm{LL}} & \boldsymbol{M}_{\mathrm{LU}}\\ \boldsymbol{M}_{\mathrm{UL}} & \boldsymbol{M}_{\mathrm{UU}}\end{bmatrix}$,则 $\nabla_f E(\boldsymbol{f})=0$ 等价于 $\begin{cases}\boldsymbol{M}_{\mathrm{LL}}\boldsymbol{f}_{\mathrm{L}}+\boldsymbol{M}_{\mathrm{LU}}\boldsymbol{f}_{\mathrm{U}}=0,\\ \boldsymbol{M}_{\mathrm{UL}}\boldsymbol{f}_{\mathrm{L}}+\boldsymbol{M}_{\mathrm{UU}}\boldsymbol{f}_{\mathrm{U}}=0.\end{cases}$ 标签向量矩阵可通过如下方程组求解

$$\boldsymbol{M}_{\mathrm{UU}}f_{\mathrm{U}}=-\boldsymbol{M}_{\mathrm{UL}}\boldsymbol{Y}'. \quad (6)$$

采用广义最小残量方法 GMRES[14] 求解式(6).由于原始标签中存在错误标签或噪声标签,同时,标签填充过程也会引入噪声干扰,所以式(5)中训练集标签的指导性要弱化,即认为填充后的训练集标签并不是图像的理想标签,但是理想标签要和训练集标签在一定程度上一致.优化目标函数改写为

$$\min_{\boldsymbol{f}}\{\|\boldsymbol{f}-\boldsymbol{A}\boldsymbol{f}\|^2+\lambda_1\|\boldsymbol{f}_{\mathrm{L}}-\hat{\boldsymbol{f}}_{\mathrm{L}}\|^2+\lambda_2\|\hat{\boldsymbol{f}}_{\mathrm{L}}-\boldsymbol{Y}'\|_1\}. \quad (7)$$

其中,$\boldsymbol{f}_{\mathrm{L}}$ 表示预测标签,$\hat{\boldsymbol{f}}_{\mathrm{L}}$ 表示训练集的理想标签.第 1 项控制邻域标签的重构误差,第 2 项约束理想标签 $\hat{\boldsymbol{f}}_{\mathrm{L}}$ 和预测标签 $\boldsymbol{f}_{\mathrm{L}}$ 一致,第 3 项是个 1 范数,限制训练集标签 $\boldsymbol{Y}'$ 只有小部分标签和理想标签 $\hat{\boldsymbol{f}}_{\mathrm{L}}$ 不同.两个正则项的直观解释是,预测标签应该和理想标签相近,理想标签应该和训练集标签 $\boldsymbol{Y}'$ 相似.算法 4 给出了标签预测方法.

算法 4. 标签预测算法.

输入:标签矩阵 $\boldsymbol{Y}'$、邻域矩阵 $\boldsymbol{C}$、正则项参数 λ_1,λ_2;

输出:测试集标签矩阵 $\boldsymbol{f}_{\mathrm{U}}$.

① 以 $\boldsymbol{Y}'$ 作为理想标签 $\hat{\boldsymbol{f}}_{\mathrm{L}}$ 的初步估计,令 $\hat{\boldsymbol{f}}_{\mathrm{L}}=$

$\boldsymbol{Y}'$,采用式(6)求解$\min_{f}\{\|\boldsymbol{f}-\boldsymbol{C}\boldsymbol{f}\|^2+\lambda_1\|\boldsymbol{f}_L-\hat{\boldsymbol{f}}_L\|^2\}$,获得 $\boldsymbol{f}_L$;

② 固定 $\boldsymbol{f}_L$,求解$\min_{\hat{f}_L}\left\{\|\boldsymbol{f}_L-\hat{\boldsymbol{f}}_L\|^2+\frac{\lambda_2}{\lambda_1}\|\hat{\boldsymbol{f}}_L-\boldsymbol{Y}'\|_1\right\}$(求解该 1 范数最小化问题);

③ 用 $\hat{\boldsymbol{f}}_L$ 替代式(6)中的 $\boldsymbol{Y}'$并求解,得到未标注样本的标签. 迭代步骤②③,直至 $\boldsymbol{f}_L$ 稳定.

由步骤①②可知,随着迭代步数增加,每一次迭代得到的最小值都是单调下降的,而该最小值又是大于零的,因此当迭代步数趋于无穷时,该最小值的极限必为零. 因为该算法是确定性算法,而且输出是有界的,所以该算法是收敛和稳定的.

4 实验分析

4.1 实验数据及评价标准

实验中使用 COREL5K[1], IAPR-TC 12[1], ESP-GAME[1]和 FLICKR[15]数据集. FLICKR 数据集取的是出现频率大于 50 次的 457 个标签组成的测试子集. 表 1 给出了数据集的统计特性. 可以看到前 3 个图像集超过 74% 的标签频率低于均值, FLICKR 中 71%的标签频率低于均值,这充分说明了样本中标签的不均衡性. 同时,我们认为均值(或者中位数)和最大值之间的差异表明了很多图像并没有标注全部标签.

Table 1 Statistics of Datasets

表 1 数据集统计数据

Dataset	Number of Images	Number of Labels	Training Set	Test Set	Number of Images per Label			Ratio of Infrequent Labels/%
					mean	median	max	
COREL5K	5 000	260	4 500	500	58.6	22	1 004	75.0
IAPR-TC 12	19 627	291	17 665	1 962	347.7	153	4 999	74.6
ESP-GAME	20 770	268	18 689	2 081	326.7	172	4 553	75.0
FLICKR	12 148	457	10 933	1 215	149.0	350	1 484	70.9

提取 3 种视觉特征 COLOR64, GIST, Dense-Surf. COLOR64 包含 44 维颜色相关图, 14 维颜色纹理矩, 6 维颜色矩构成. 使用 8 个方向 3 个尺度上的 Gabor 滤波,图像分块采用 4×4 的分块方式,得到 384 维的 GIST 特征. 采用 Dense-Surf 对图像进行特征点检测与描述,得到 4 000 维的描述子,采用 PCA 对其进行降维,得到 30 维的特征,并用 *K*-MEANS 聚类中心构成码书. 采用标注准确率、召回率、*F*1 值和召回标签数等评价指标进行评测. 给定标签 c_i,若存在 n_i 幅图像具有该标签,而标注结果中包含该标签的图像个数为 n_i^s,其中 n_i^p 个是正确的,则 $P=\sum_{i=1}^{q}\frac{n_i^p}{n_i^s}$, $R=\sum_{i=1}^{q}\frac{n_i^p}{n_i}$, $F1=\frac{2\times P\times R}{P+R}$,召回标签数($N+$)为召回率大于 0 的标签个数. 平均标签准确率 P 衡量标注的准确性,召回率 R 衡量标注的完整性,$F1$ 衡量两者的均衡性能,$N+$ 衡量标注方法对标签集合的覆盖程度.

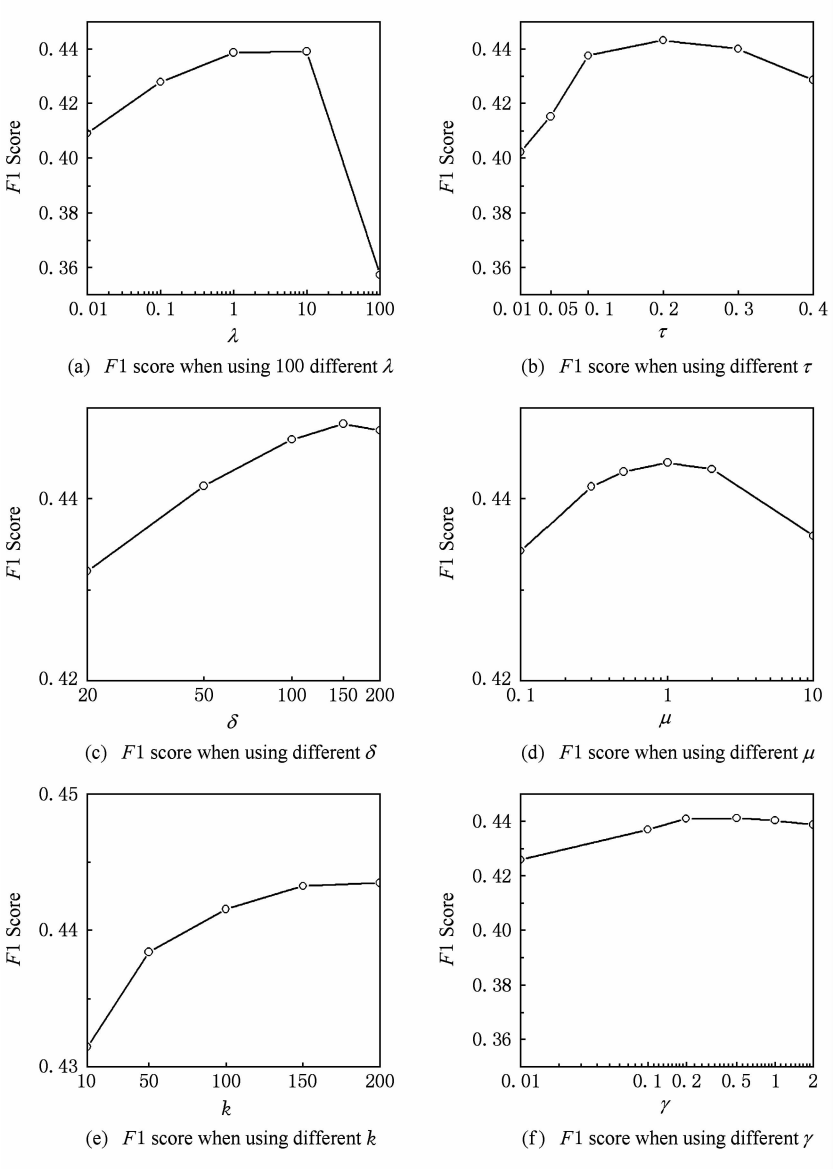
4.2 参数设置

通过验证集估计设置算法参数的最优值. 算法 1 中正则项参数 λ 是需要设置的正则化系数,用于平衡标签的加权误差和平滑误差. 检验参数 λ 在 $\{0.01, 0.1, 1, 10, 100\}$5 种情况下 SNLWL 的性能. 如图 3(a)所示,参数 λ 的最佳配置出现在 1～10 附近,此时 SNLWL 方法中的加权误差和平滑误差重要性接近. 参数 λ 过大,算法中的平滑性误差占有更大的权重,会降低标签填充的效果,同时导致原始标签的指导性也降低,从而性能下降. 同理,当参数 λ 过小,算法中加权误差权重加大,会忽略函数平滑性,导致函数在训练集上过拟合,因此性能也受到了影响.

误差权重 τ 用于设置潜在遗漏标签的误差权重. 检验参数 τ 在 $\{0.01, 0.05, 0.1, 0.2, 0.3, 0.4\}$6 种情况下 SNLWL 性能. 如图 3(b)所示, τ 的最佳配置出现在 0.2 附近. 当参数 τ 过大,未标注标签误差占有了更大的权重,会降低原始标签的指导性,性能下降速度加快. 当参数 τ 过小,会影响到 SNLWL 填充标签的能力,因此性能也受到了影响.

邻域范围 δ 用于设置标签填充的邻域范围,检验参数 δ 在 $\{20, 50, 100, 150, 200\}$ 5 种情况下 SNLWL 性能,如图 3(c)所示,过大的邻域范围并未对性能提升起到更多的贡献,实验中设置 $\delta=100$.

算法 2 中邻域范围 ε 用于从各语义组中选择目标图像的视觉近邻,在 COREL5K, IAPR-TC 12,



(a) $F1$ score when using 100 different λ　(b) $F1$ score when using different τ

(c) $F1$ score when using different δ　(d) $F1$ score when using different μ

(e) $F1$ score when using different k　(f) $F1$ score when using different γ

Fig. 3　The performance of SNLWL with different parameters.

图 3　SNLWL 在不同参数设置下的性能

ESP-GAME,FLICKR 4 个数据集上上分别设置为 4,3,3,2.算法 3 中损失函数正则项系数 μ 用于平衡邻域内样本距离与邻域间距离.检验参数 μ 在{0.1,0.3,0.5,1.0,2.0,10.0}6 种情况下 SNLWL 性能,如图 3(d)所示,最佳范围在 0.5～2.0 附近,实验中设置 $\mu=1$.

检验邻域范围 k 在{10,20,50,100,150,200}5 种情况下 SNLWL 性能,如图 3(e)所示,在实验中设置为 100,过小的取值会导致性能下降.

算法 4 中正则项参数 λ_1 用于限制预测标签和理想标签的一致程度,要求不一致程度要小,所以 λ_1 取值要较大,经过验证集验证其在{50,100,200,300}4 种情况下的性能,λ_1 最优取值为 100.$\gamma=\lambda_2/\lambda_1$ 用于协调理想标签和预测标签的误差的平衡.验证其在{0.01,0.10,0.20,0.50,1.00,2.00}6 种情况下的性能,如图 3(f)所示,γ 最佳范围在 0.2～1.0 之间,实验中设置为 0.3,取值范围过大或者过小均会导致目标函数在训练集标签上过拟合.

4.3　性能度量

图 4 给出了本文方法生成的语义平衡邻域(BN)与以均等贡献组合视觉特征的 JEC[1] 方法生成邻域的实例对比.图 4(a)为来自 IAPR-TC 12[1] 的目标图像,图 4(b)到图 4(e)为其 BN 中近邻,图 4(h)到图 4(i)为其 JEC 近邻,匹配标签用粗体显示.可以看到低频标签(带下划线标签){“field”,“range”,“bush”,“landscape”,“lake”}和高频标签{“tree”,“mountain”}均在 BN 中.而 JEC 方法得到的邻域中大多是高频标签.虽然“creek”,“cloud”,“vegetation”

可以很好地描述目标图像的语义,但是其并没有出现在 JEC 邻域中,BN 中标签分布不平衡和遗漏低频标签的情况已经得到改善.

图 5 给出了目标图像的语义一致邻域(CN)和 JEC 方法形成的邻域的对比. 可以看到 CN 内样本与目标图像具备更好的视觉相似性和语义相似性.

(a) Target image with labels {bush field lake landscape mountain range tree}

(b) Neighbors in BN with labels {**mountain** meadow sky **tree**}

(c) Neighbors in BN with labels {**mountain** meadow **tree** **lake**}

(d) Neighbors in BN with labels {**field** grass Vegetation **bush**}

(e) Neighbors in BN with labels {**Landscape** cloud Creek **range**}

(f) Neighbors by JEC with labels {**mountain** meadow sky **tree**}

(g) Neighbors by JEC with labels {**mountain** meadow **tree** **lake**}

(h) Neighbors by JEC with labels {hill **tree** house sky}

(i) Neighbors by JEC with labels {**mountain** horse **tree** grass}

Fig. 4 Comparison of semantic balanced neighborhood with JEC neighborhood.

图 4 语义平衡邻域与 JEC 邻域对比

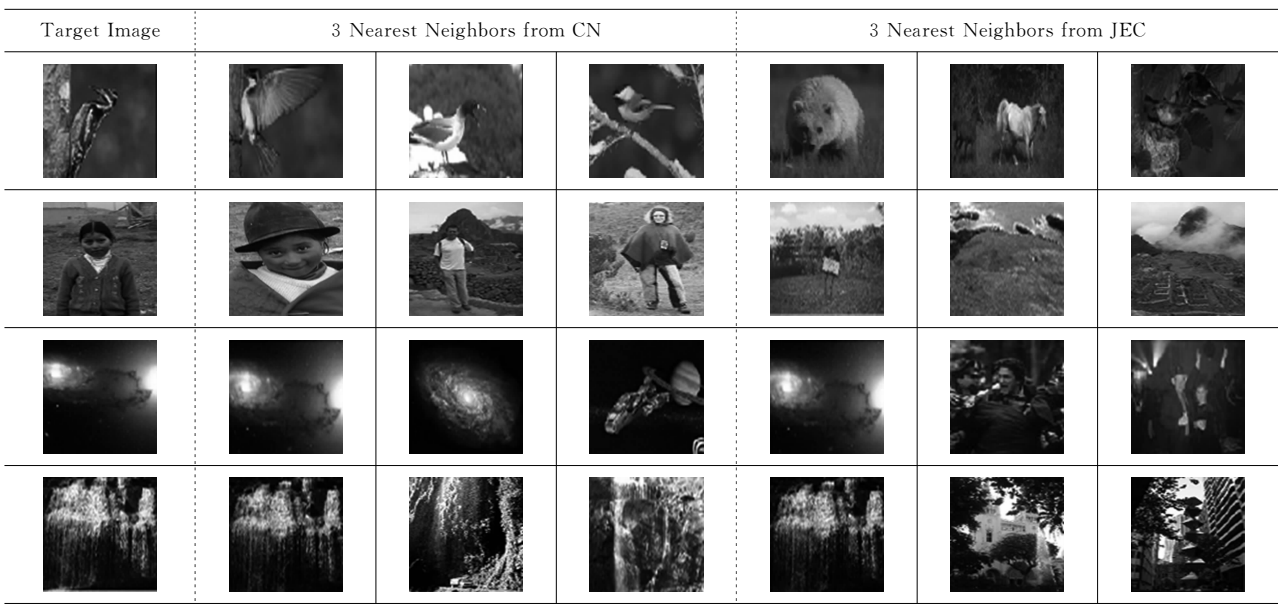


Fig. 5 Comparison of semantic consistent neighborhood with JEC neighborhood.

图 5 语义一致邻域与 JEC 邻域比对

后续实验包含 3 部分:实验 1 是测试在遗漏标签的情况下 SNLWL 性能;实验 2 是在无关标签和遗漏标签情况下进行;实验 3 给出在标准评测集上各种方法的标注性能对比. 对比方法包括文献[1]中采用贪婪方法从近邻进行标签传播的 JEC 方法、基于判别测度学习的方法 Tagprop(ml)和使用 sigmoid 函数的 Tagprop($ml+s$)[6-7]、基于组稀疏性的标注方法(GS)[10]和基于语义类别学习的图像标注方法(SML)[16].

4.3.1 标签不完整情况下的性能比较

在 COREL5K 数据集中模拟用户为图像提供弱标签的情况. 需要说明的是,4 个标准评测数据集原始标签并不完善,但是为了能够对这种遗漏标签的情况进行放大和有一个客观的对比基准,本文根据一定的比例随机选出样本的一部分真实标签作为用户提供的弱标签. 定义标签遗漏率(missing rate)为用户标出的标签与样本真实的完整标签集合大小的比值的均值. 实验中测试了标签遗漏率为 0.1~

0.9 的情况. 图 6 给出了 SNLWL 在 COREL 数据集上不同遗漏率下的 $F1$ 值,从中可以得到 SNLWL 主要步骤对标注性能的影响.

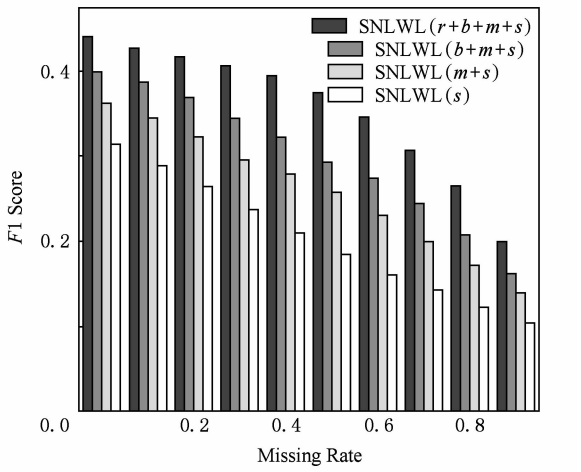


Fig. 6　The performance of SNLWL for different missing rate.

图 6　不同标签遗漏率下 SNLWL 性能

图 6 中,令 r 表示邻域标签填充,b 为语义平衡邻域,m 为测度学习得到的邻域,s 为基于稀疏表示的具有部分相关性的邻域.

可以得出结论,随着标签遗漏率的增加,所有方法的性能均呈现下降趋势. 进一步分析可知,进行标签填充,语义平衡邻域和语义一致邻域构造的 SNLWL($r+b+m+s$)在不同遗漏率下,均取得了最好的标注结果,而且其在遗漏率低于 0.5 的情况下,性能下降趋势缓慢. 未进行标签填充,使用原始数据集标签进行学习的 SNLWL($b+m+s$)相比之下,性能平均下降了 16.7%,虽然填充的过程中会引入一部分噪声,但是从整体性能上可以看到,算法 1 通过正则项协调标签输出与原始数据的近似性与平滑性,通过原始标签的指导降低噪声的策略可以提升系统性能,是有效的. 未构造语义平衡邻域的 SNLWL($m+s$)性能低于 SNLWL($b+m+s$),因为大量低频标签未能有效的参与到测度学习过程中. 直接在样本的 k 邻域内进行稀疏表示并进行标签推论的 SNLWL(s)性能较 SNLWL($m+s$)平均下降了 23.4%,因为其没有保证邻域内样本和目标图像在一个语义空间内,伪邻带来了大量的噪声干扰使得标签推论的性能下降. 图 7 给出 missing rate 取值变化过程中 COREL5K 数据集上的 $F1$ 值曲线.

从图 7 中可以看到,标签的不完整性对标注性能影响很大. SNLWL 性能降低趋势缓慢,其受到遗

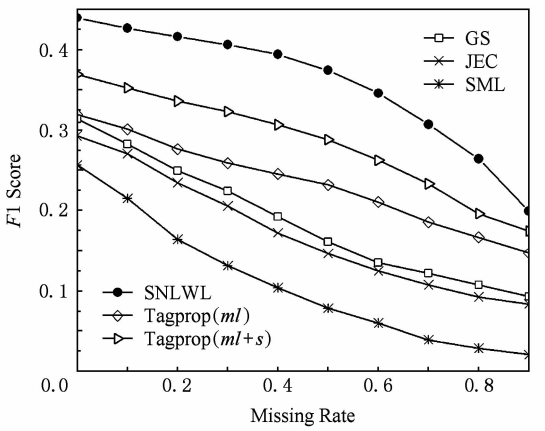


Fig. 7　The performance of various methods for different missing rate.

图 7　不同标签遗漏率下各种方法标注性能

漏率影响小于采用判别式测度学习和 sigmoid 函数处理的 Tagprop($ml+s$)方法. GS 和 SML 方法受影响最大,因为 GS 方法的特征组群的建立依赖于标签的完整性,SML 利用高斯混合模型建立标签语义类的过程也有类似要求. 依据贪婪方法传播标签的 JEC 和 SML,GS 对比,显示出了一定的鲁棒性.

4.3.2　不完整标签和噪声标签下的性能比较

在 COREL5K 训练集随机移除一定比例的标签,再从标准词汇中随机添加相应比例的标签来模拟真实环境下弱标签图像集中的噪声. 比例范围设置在{0.1,0.2,0.3,0.4,0.5},定义其为弱标签率(weak rate). 实验设置与符号表示与 4.3.1 节相同,对比结果如图 8 所示:

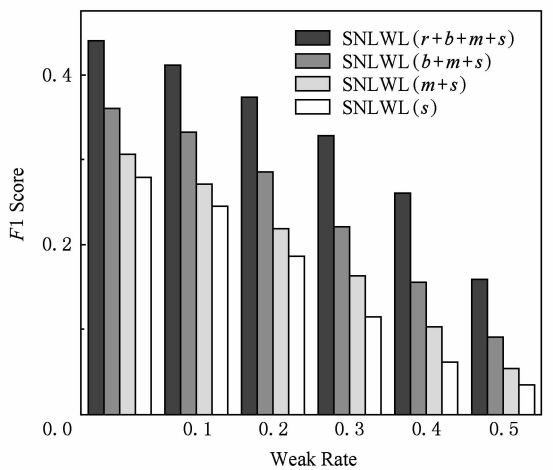


Fig. 8　The performance of SNLWL for different weak rate.

图 8　不同弱标签率下的 SNLWL 性能

可以看到,噪声干扰的增强降低了标注性能,但是 SNLWL($r+b+m+s$)取得了最好的标注结果.

相比之下，SNLWL($b+m+s$)性能平均下降了29.4%，可以得出结论，在噪声干扰下，算法1对标签的填充会大幅增加系统性能，弱化噪声标签的干扰，其处理过程引入的噪声虽然会对部分低频标签带来影响，但是其负面作用影响有限. 另外，弱标签率的增加对语义平衡邻域的构建也会带来噪声干扰，但是由于后续处理过程的噪声处理，加之标签分布的改善，SNLWL($b+m+s$)性能较SNLWL($m+s$)平均提升38.1%.

图9给出了COREL5K数据集上的性能对比. 可以观察到随着弱标签率的增加，各种方法的性能下降速度大于图7，说明噪声标签的引入增加了图像的弱标签性. 弱标签率在0.1～0.2变化过程中，SNLWL受到影响小于Tagprop，基于特征组群的GS方法和SML性能下降趋势最为明显.

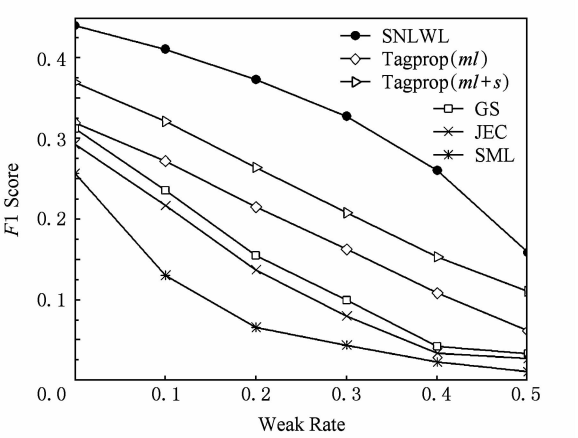


Fig. 9 The performance of various methods for different weak rate.

图9 不同弱标签率下各种方法标注性能

4.3.3 标准评测集下性能

每幅图像取5个标签最为标注结果. 表2记录了4个标准评测集上50次独立随机实验的平均标签准确率(P)、平均标签召回率(R)、$F1$均值和召回标签数($N+$).

Table 2 Performance on Benchmark Dataset of Various Annotation Methods

表2 基准数据集上的标注性能

Method	COREL5K				IAPR-TC 12				ESP-GAME				FLICKR			
	P	R	$F1$	$N+$	P	R	$F1$	$N+$	P	R	$F1$	$N+$	P	R	$F1$	$N+$
SML	0.30	0.33	0.314	146	0.32	0.29	0.304	252	0.36	0.24	0.288	226	0.29	0.22	0.250	375
JEC	0.27	0.32	0.293	139	0.28	0.29	0.285	250	0.22	0.25	0.234	224	0.20	0.21	0.205	355
Tagprop(ml)	0.31	0.37	0.337	146	0.48	0.25	0.329	227	0.49	0.20	0.284	213	0.43	0.17	0.244	363
Tagprop($ml+s$)	0.33	0.42	0.370	160	0.46	0.35	0.398	266	0.39	0.27	0.319	239	0.34	0.25	0.288	403
GS	0.30	0.33	0.314	146	0.32	0.29	0.304	252	0.36	0.24	0.288	226	0.29	0.22	0.250	375
SNLWL	0.43	0.45	0.440	187	0.52	0.37	0.432	276	0.51	0.30	0.378	249	0.48	0.29	0.362	427

可以得出结论，SNLWL在各项指标上均优于对比方法. SNLWL在COREL5K上的准确率和召回率达到了0.43和0.45，分别高于Tagprop($ml+s$)30.0%和6.7%，在ESP-GAME上高于Tagprop($ml+s$)30.8%和11.1%，在FLICKR上高于Tagprop($ml+s$)41.2%和24%. 由于ESP-GAME和FLICKR更接近于真实环境下的弱标签数据集，所以SNLWL在弱标签性强的环境下性能优势更加明显.

从表2可知，基于判别测度学习方法(Tagprop)因为对每一个语义标签均建立了模型，其标注效果优于GS，JEC和SML. SML因为要求高质量的数据集对语义类建模，所以其在小规模训练集上性能很不理想，特别在弱标签情况严重的ESP-GAME和FLICKR数据集上，性能很低. GS利用了特征组群特性，较JEC方法具有性能优势. SNLWL所取得的高标注性能说明针对弱标签数据集的处理方法是有效的. 将标签按照低频标签和高频标签进行划分，即将标签集划分为50%的高频标签(fg)和50%的低频标签(ifg)两个部分，测试各种方法的平均标签召回率，实验结果如表3所示.

从表3可以得出结论，SNLWL和Tagprop($ml+s$)两种方法在高频标签和低频标签上的召回率差异较小. 因为Tagprop($ml+s$)的sigmoid函数增强了低频标签的权重，所以低频标签上的性能高于Tagprop(ml)，但在高频标签上，Tagprop(ml)的召回率要大于采用了sigmoid函数的Tagprop($ml+s$). SNLWL低频和高频标签上性能更为接近. 同时，低频标签上和高频标签上的召回率都取得了最好的结果. 表4给出了ESPGAME上不同方法的标注结果实例，其中匹配标签为粗体显示.

Table 3　Recall on Label Partition for 3 Datasets

表 3　标签分组后的召回率对比

Datasets	SML		JEC		GS		Tagprop(ml)		Tagprop($ml+s$)		SNLWL	
	fg	ifg	fg	ifg	fg	ifg	fg	ifg	fg	ifg	fg	ifg
COREL5K	0.35	0.23	0.37	0.27	0.38	0.28	0.42	0.32	0.43	0.31	0.46	0.44
IAPR TC-12	0.29	0.17	0.35	0.23	0.36	0.22	0.31	0.19	0.37	0.33	0.38	0.36
ESP-GAME	0.20	0.14	0.28	0.22	0.27	0.21	0.22	0.18	0.29	0.25	0.31	0.29

Table 4　Image Annotation Examples on ESPGAME

表 4　ESPGAME 标注结果实例

Images	Ground truth	JEC	Tagprop	GS	SNLWL
	girl green hair grass smile brown woman picture	**green girl** space sky face **woman** shop metal family machine light truck forest star guy	**green girl** tree people **hair** white dark band leaf **woman picture** blue glasses sky man	**girl green** blue face **woman** people glasses man dress **grass** sky **hair** light chinese eye	sky cloud **green** people **grass woman** tree **blue girl hair picture grass** water **brown** dress
	child family man woman people hair picture red shirt	sky mountain **man** hair **woman** water **red** boy black blue statue sun group brown white	sky mountain **man woman people** white **child** old black cloud dance brown boy **picture** dress	sky **man** white **woman** statue red face **child** mountain table **hair** roof ground statium dance	sky mountain cloud man **woman hair red family shirt child** blue **people** hand brick **family**
	building car city sky street white window	**city window** metal truck **building** room fly line roof water house mountain road helmet **white**	**window city** black **sky** water ocean room **building car** ground middle roof metal man road	**window city** black church white water yellow line tree **sky white** road blue side house	**building city** white **sky** yellow **road window** middle people house cloud **car** ground view **street**

5　结论和展望

因为真实环境下的数据集存在着标签不完整、不准确、分布不均衡的情况,弱标签数据集上的标注方法研究对于真实环境下的图像语义标注具有实际意义,因此,提出了一种基于语义邻域学习的图像自动标注方法,其核心思想是通过邻域信息的指导降低弱标签对学习方法的性能影响.该方法在标签损失误差最小化意义下填补遗漏标签,构建语义平衡邻域.进而通过多标签信息嵌入的邻域测度学习保证邻域内样本的语义一致性,通过稀疏表示获得图像之间的部分相关性,并构建语义一致邻域,保证邻域样本具备全局相似性,部分相关性和语义一致性.最后在邻域标签重构误差最小化意义下进行标签预测,降低噪声标签对性能的影响.实验结果表明本文方法取得了大幅度的性能提升.

近期研究工作表明,利用标签相关性和特征相关性可以进一步提高标注性能,但是其模型建立是一个更加困难的问题.下一步工作主要考虑如何利用邻域标签相关信息和特征相关信息来提高弱标签图像集上的标注性能.

参　考　文　献

[1] Makadia A, Pavlovic V, Kumar S. A new baseline for image annotation [C] //Proc of the 10th European Conf on Computer Vision. Berlin: Springer, 2008: 316-329

[2] Lu Jing, Ma Shaoping. Region based image annotation using heuristic support vector machine in multiple instance learning [J]. Journal of Computer Research and Development, 2009, 46(5): 864-871 (in Chinese)
(路晶, 马少平. 使用基于多例学习的启发式 SVM 算法的图像自动标注[J]. 计算机研究与发展, 2009, 46(5): 864-871)

[3] Ke Xiao, Li Shaozi, Cao Donglin. Automatic image annotation based on relevant visual keywords [J]. Journal of Computer Research and Development, 2012, 49(4): 846-855 (in Chinese)

(柯逍, 李绍滋, 曹冬林. 基于相关视觉关键词的图像自动标注方法研究[J]. 计算机研究与发展, 2012, 49(4): 846-855)

[4] Nguyen N, Caruana R. Classification with partial labels [C] //Proc of the 14th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2008: 551-559

[5] He Xuming, Zemel R S. Learning hybrid models for image annotation with partially labeled data [C] //Proc of the 22nd Conf on Neural Information Processing Systems. New York: Curran Associates, 2008: 625-632

[6] Guillaumin M, Mensink T, Verbeek J. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation [C] //Proc of the 12th Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2009: 309-316

[7] Verbeek J, Guillaumin M. Image annotation with TagProp on the MIRFLICKR set [C] //Proc of the 2nd ACM SIGMM Int Conf on Multimedia Information Retrieval. New York: Association for Computing Machinery, 2010: 537-546

[8] Fan Jianping, Shen Yi, Zhou Ning. Harvesting large-scale weakly-tagged image databases from the web [C] //Proc of the 23rd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2010: 802-809

[9] Bucak S S, Jin Rong. Multi-label learning with incomplete class assignments [C] //Proc of the 24th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2011: 2801-2808

[10] Zhang Shaoting, Huang Junzhou, Huang Yuchi. Automatic image annotation using group sparsity [C] //Proc of the 23rd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2010: 3312-3319

[11] Wang Hua, Huang Heng. Image annotation using bi-relational graph of images and semantic labels [C] //Proc of the 24th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2011: 793-800

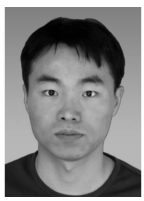
[12] Zhang Hao, Berg A. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition [C] // Proc of the 19th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Computer Society, 2006: 2126-2136

[13] Shaler-Shwartz S, Singer Y. Pegasos: Primal estimated sub-gradient solver for svm [J]. Mathematical Programming, 2011, 127(1): 3-30

[14] Sadd Y, Schultz M H. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems [J]. SIAM Journal on Scientific and Statistical Computing, 1986, 7(3): 856-869

[15] Huiskes M J, Lew M S. The MIR Flickr retrieval evaluation [C] //Proc of the 1st ACM Int Conf on Multimedia Information Retrieval. New York: ACM, 2008: 39-43

[16] Carneiro G, Chan A B. Supervised learning of semantic classes for image annotation and retrieval [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410



Tian Feng, born in 1980. PhD in the State Key Laboratory of Virtual Reality Technology and Systems at Beihang University. Associate professor in Northeast Petroleum University. His main research interests include multimedia semantic analysis, pattern recognition and computer vision.

Shen Xukun, born in 1965. Professor and PhD supervisor in the State Key Laboratory of Virtual Reality Technology and Systems at Beihang University. His main research interests include computer vision and multimedia content management.