ISSN 1000-9825, CODEN RUXUEW
*Journal of Software*, Vol.20, No.9, September 2009, pp.2450−2461
doi: 10.3724/SP.J.1001.2009.03380

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

# 基于可判别超平面树的生成模型图像标注方法*

王 梅, 周向东+, 许红涛, 施伯乐

(复旦大学 计算机与信息技术系,上海 200433)

# Effective Image Auto-Annotation via Discriminative Hyperplane Tree Based Generative Model

WANG Mei, ZHOU Xiang-Dong+, XU Hong-Tao, SHI Bai-Le

(Department of Computing and Information Technology, Fudan University, Shanghai 200433, China)

+ Corresponding author: E-mail: xdzhou@fudan.edu.cn

**Wang M, Zhou XD, Xu HT, Shi BL. Effective image auto-annotation via discriminative hyperplane tree based generative model. *Journal of Software*, 2009,20(9):2450−2461.** http://www.jos.org.cn/1000-9825/3380.htm

**Abstract**: Many machine learning methods such as generative model and discriminative model have been applied to image semantic automatic image annotation. However, due to the "semantic gap", the imbalanced training data, and the multi-label characteristic of image annotation, the annotation performance still calls for improvement. In this paper, an image annotation method is proposed which augments the classical generative model with the proposed discriminative hyperplane tree. Based on the high visual generative probability training images (neighborhood) of the unlabeled image, the local hyperplane classification tree is adaptively established. The semantic relevant training image set is obtained through top-down hierarchical classification procedure by exploiting the discriminative information at each level. The joint probability between the unlabeled image and the semantic words is estimated based on the obtained semantic relevant local training set under the proposed framework. This method combines the advantages of both generative model and the discriminative models. From the aspect of generative model: by exploiting the discriminative information of the semantic cluster in the discriminative hyperplane tree, a local generative set is progressively refined, and therefore, improves accuracy. From the aspect of discriminative model: the multiple label assignment can be naturally implement by estimating the joint probability, which reduces the limitation of discriminative model induced by the imbalanced and overlapping training set. The experiments on the ECCV2002 benchmark show that the method outperforms state-of-the-art generative model-based annotation method MBRM and discriminative model based ASVM-MIL with $F1$ measure improving by 14% and 13% respectively.

**Key words**: automatic image annotation; generative model; discriminant classification; discriminative hyperplane tree; hierarchical classification

**摘 要**: 图像语义的自动标注是一个具有挑战性的研究课题,目前常见的机器学习方法,如统计生成模型

* Supported by the National Natural Science Foundation of China under Grant Nos.60403018, 60773077, 90818023 (国家自然科学基金), the National Basic Research Program of China under Grant No.2005CB321905 (国家重点基础研究发展计划(973))

Received 2007-09-29; Accepted 2008-04-15

(generative model)与判别模型(discriminative model)都被用于该问题的研究中.然而由于语义鸿沟的存在、图像训练数据的不平衡性以及图像标注的多标签特性等问题,使得上述方法的性能都有待进一步提高.提出一种基于可判别超平面树的生成模型图像标注方法.该方法根据待标注目标图像的高生成概率邻域,建立局部超平面分类树,进而利用同层类间可判别信息,按自顶向下的层次分类得到待标注图像的语义相关图像集合.由此得到的相关类信息与新的生成模型框架相结合对待标注图像与语义关键词的联合概率进行估计,实现对目标图像的标注.其特点在于生成模型与判别模型方法得到了有效结合,可判别超平面树对隐含语义聚类的判别分析是对待标注图像的生成"邻域"的逐步求精过程,有效地提高了生成模型标注准确度;而对于判别分析难以解决的多标签分类、训练数据不平衡等问题,此方法通过联合概率估计自然地实现目标图像的多标签分配.在常用的包含 5 000 幅图像的 ECCV2002 数据集进行了实验,结果表明,与目前已知的具有较好标注效果的基于生成模型的 MBRM 模型(采用图像分割方法)以及基于辨别分析的 ASVM-MIL 相比,此方法的 *F*1 因子分别提高了 14%和 13%.

**关键词**: 自动图像标注;生成模型;判别模型;可判别超平面树;层次分类

**中图法分类号**: TP311 **文献标识码**: A

图像语义的自动标注是实现图像语义检索的关键,标注就是使用语义关键字或标签来表示一幅图像的语义内容,进而将图像检索转化为文本检索.早期手工标注需要专业人员对每幅图像标出关键字,费时且具有主观性.图像数量的爆炸性增长促使人们利用各种机器学习技术设计出多种图像自动标注模型[1−11].然而,由于存在语义鸿沟(semantic gap),自动获取图像的语义信息仍然非常困难,图像语义的自动标注性能仍亟待提高.

基于统计生成模型(generative model)[1−6]的图像标注方法,如相关模型[2−4]等得到了较为广泛的研究.统计生成模型方法从训练数据集合中估计待标注图像与语义关键词的联合概率进行标注.此类方法对数据规模和语义关键词数量的适应性强,其估计概率的方式为标注结果的选择提供了自然的排序方法.然而,由于语义鸿沟的存在,其标注过程容易受到具有高视觉相似性而语义不同的图像的影响[7],对数据的利用效率也有待提高.将每个语义关键词看作一个类标签,图像标注问题也可使用有指导分类学习的方法解决[7−10],其中判别模型(discriminative model)方法被广泛用于解决图像标注问题[7,8].判别分类方法(如 SVM 等)将给定的训练样本按类标签组织为类,学习最优决策边界,实现数据分类任务,从原理上,生成模型与判别模型具有一定的互补性.但是,可判别分类方法的复杂度往往随着数据规模和语义关键词的数量升高;在全局训练集上使用 SVM 分类器,时间复杂度高且容易受到正负例不平衡的影响[10].使用语义层次结构的学习方法在图像自动标注、对象识别等领域取得了较好成果[9,11−14].如:Srikanth 等人[11]在图像标注中,使用 WordNet 建立文本层次结构,并在该结构下使用混合模型和 Shinkage 方法进行标注.Marszalek 和 Schmid[13]在视觉对象分类任务中,借助于 WordNet 建立语义层次结构,在该结构上进行二元 SVM 分类器的训练,在训练好的层次分类结构中进行视觉对象分类.近年来,把生成模型与可判别分类方法进行结合的统计学习方法已得到了关注[15,16],如 Lasserre 等人为生成模型和判别模型的混合提供理论上的途径以及 Grabner 等人提出的 enginboosting 方法等.然而由于图像语义的内在复杂性,图像标注的多标签学习特性,即每幅图像同时包含多个语义相关的类标签,这样会产生不同语义类之间的重叠,影响分类面的判别能力.目前成功地把判别分析与生成模型相结合的图像标注方法还不多见.

本文提出一种基于可判别超平面树的生成模型图像标注方法.与之前方法使用 WordNet 构建语义层次结构不同,本文的类层次结构通过分析待标注图像的高视觉生成概率"邻域"自动建立,并进一步构建可判别超平面分类树.在待标注图像的分类过程中逐层使用判别信息对其生成邻域中不相关图像进行过滤,并在此基础上估计联合概率获得目标图像的语义标签.

为了弥补全局判别分类的不足,本文在生成模型的基础上,对于给定的待标注图像,获取待标注图像的高视觉生成概率"邻域",从而使本文的判别分类在一个小的局部训练集进行.通过分析该邻域中训练图像的语义相关性,自动构建语义类层次结构,并在同层中进行可判别超平面学习,得到可判别超平面树.对待标注图像进行自上而下的分类,逐层利用隐藏语义类之间的判别信息,减少邻域中与待标注图像语义不相关图像的影响,得到语义相关图像集合.最终,在相关类中估计待标注图像和语义关键词的联合概率,得到待标注图像候选标注词的

自然排序.

本文工作主要贡献如下:

1) 提出基于可判别超平面树的生成模型图像标注方法.在本文标注框架下,使用可判别超平面树对待标注图像的生成邻域逐层划分,得到其语义相关图像集合.在该集合上通过联合概率估计自然地实现目标图像的多标签分配,将生成模型与判别模型方法有效结合.

2) 本文在待标注图像的高生成概率"邻域"上,构造局部类层次树及相应的局部训练样本,克服了全局分类方法的不足,使得分类方法能够有效地应用于基于生成模型的标注方法中;设计自上而下层次分类方法,对待标注图像的生成邻域逐步求精,提高基于生成模型的图像标注性能.

3) 使用基准数据集 ECCV2002 对本文提出的标注算法进行检验,与单纯基于生成模型的 MBRM 方法以及单纯基于判别分类的 ASVM-MIL 方法进地行比较,本文方法的 $F1$ 因子分别提高 14%和 13%.

本文第 1 节讨论图像标注的相关工作.第 2 节介绍本文的标注方法.第 3 节是本文标注算法总结.第 4 节讨论实验结果.第 5 节是总结和展望.

## 1 生成模型标注方法

首先介绍一个标准的相关模型标注过程[2−4]:给定训练集合 $L$,集合的大小记为$|L|$.训练集中每幅已标注的图像 $J_i$ 可使用图像区域和标注词来表示,如 $J_i=\{f_{i,1},f_{i,2},...,f_{i,m};w_{i,1},w_{i,2},...,w_{i,n}\}$,$m$ 和 $n$ 分别表示图像区域的个数和词的总个数.$f_{i,j}$ 是区域特征,维数为 $D$;$w_{i,j}$ 是一个二元变量,表示第 $j$ 个词是否出现在第 $i$ 幅图像中.给定一幅待标注图像 $I=\{f_1,f_2,...,f_t\}$,关键词 $w$ 作为其标注的积分排序函数为:

$$\begin{aligned}P(w\mid I)\propto P(w,I)&=\sum_{i=1}^{|L|}P(w,I\mid J_i)P(J_i)\\&=\sum_{i=1}^{|L|}P(w\mid J_i)P(I\mid J_i)P(J_i)\end{aligned}\tag{1}$$

待求最佳标注 $w^*$为

$$w^*=\arg\max_w P(w,I)\tag{2}$$

在等式(1)中,假设 $P(J)$服从均匀分布,$P(I|J)$表示待标注图像 $I$ 由训练图像 $J$ 生成的概率.对于 $P(W|J)$,我们可认为若 $w$ 在 $J$ 中出现,则对应的概率较高,反之则低.从式(1)可以看出,具有较高生成概率 $P(I|J)$的训练图像 $J$ 中包含的语义关键词和待标注图像容易产生较高的联合概率.然而实际中,由于语义鸿沟的存在,上述图像 $J$ 的语义与 $I$ 很可能并不相同,此时 $J$ 的语义关键词将被错误地赋给 $I$,从而导致错误标注.

如图 1 给出一个使用相关模型标注的例子,图中显示待标注图像(第一排最左侧图像)以及它的前 9 个最高生成概率训练图像组成的邻域,记为 9-HGPN(highest generative probability neighborhood).每个训练图像的语义标签显示在对应图像的下方.由于语义鸿沟问题,在这个例子中,HGPN 中的很大一部分图像与待标注图像并不相关.然而根据公式(1),这部分不相关标签很容易被传递给待标注图像.将每个语义标签记为一类,利用类与类的可判别信息,学习最优分类面,可以从一定程度上对此类问题进行处理.然而,在标注问题中,每个图像对应多个语义相关的标签,在 HGPN 中直接应用上述分类思想,类个数相对较多,而每个类对应的训练图像个数太少,同时,不同的类训练数据会出现重叠,训练分类器依旧是一个非常困难的问题.直观的,若两幅图像包含的语义关键词相似,则它们的语义相关性大.我们利用图像的语义相关性,将 $K$-HGPN 中的训练图像划分为不同的组(主题).同一个主题中包含语义相关的图像.这样,与待标注图像语义相关的图像和不相关图像将被分配到不同的主题中.图 1 矩形区域中显示了本例中的两个主题.将每个主题看作一类,每个类的语义内容相对集中,而类的个数大为减少,同时每类中包含的样例数目较多,应用分类器将得到更好的分类效果.此时,利用主题类之间的可判别信息,设计分类器,相关主题获得较大的后验概率.待标注图像从相关主题中生成,得到的标注更加准确.

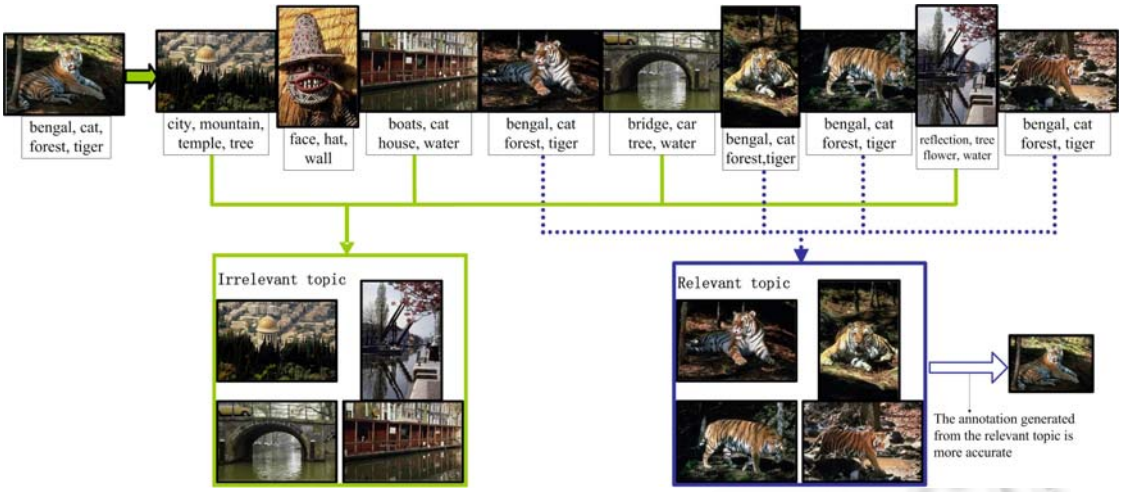



Fig.1 An annotation example by using MBRM

图 1 基于相关模型的标注的例子

## 2 基于可判别超平面树的生成模型图像标注方法

### 2.1 基本框架

如前所述,待标注图像 $I$ 的高生成概率邻域对其标注 $w^*$具有较大影响,因此将式(1)中联合概率 $P(I,w)$重新定义为

$$\hat{P}(I,w)=\sum_{J\in HGPN(I)}P(I,w\mid J)P(J) \tag{3}$$

图像的标注文本描述图像的语义,根据图像的标注,我们将 $K$-HGPN($I$)中语义相似的图像组织成语义聚类(semantic cluster),不同的语义聚类之间语义相差较大.如图 1 中的相关主题和不相关主题构成两个语义聚类.假设在 HGPN($I$)中得到的语义聚类集合为 $\{T_1,T_2,...,T_v\}$,$T_i$ 表示第 $i$ 个语义聚类,$T_i\cap T_j=\varnothing$,$v$ 是语义聚类的个数.概率 $P(I|T_i)$表示待标注图像 $I$ 从 $T_i$ 中生成的概率.$I$ 与 $T_i$ 语义相关性高,则 $I$ 由 $T_i$ 生成的可能性大,对应的 $P(I|T_i)$高.则等式(3)可转化为

$$\begin{aligned}\hat{P}(I,w)&=\sum_{T_i}P_{T_i}(I,w)P(I\mid T_i)\\&=\sum_{T_i}\left(\sum_{J\in T_i}P(I,w\mid J)P(J)\right)P(I\mid T_i)\end{aligned} \tag{4}$$

此时待标注图像 $I$ 的最佳标注为

$$w^*=\arg\max_w\hat{P}(w,I) \tag{5}$$

根据贝叶斯法则有:

$$P(I\mid T_i)=\frac{P(T_i\mid I)P(I)}{P(T_i)} \tag{6}$$

假设 $P(I)$和 $P(T)$符合均匀分布,则估计概率 $P(I|T_i)$的值转化为估计概率 $P(T_i|I)$,即给定待标注图像 $I$,语义聚类 $T_i$ 出现的后验概率.很明显,概率 $P(T_i|I)$的值可以通过判别分类技术获得.因此,公式(4)给出本文生成模型与可判别分类技术相结合的基本模型,通过引入 $P(I|T_i)$并使用可判别分类技术求解其值,在生成模型中增加可判别能力.由公式(3)可知,待标注图像由 HGPN 产生,定义如下一组超平面来分隔整个 HGPN 组成的空间:

$$\{x\mid f_i(x)=w_i^Tx+b_i=0\} \tag{7}$$

其中 $i$=1,…,$v$,$x$ 表示给定训练图像对应在特征空间的向量,$f(x)$为线性决策函数.$P(T_i|I)$可表示为决策值的函数:

$$P(T_i\mid I)=g(f(x_I)) \tag{8}$$

图 2 给出二维空间中对 HGPN($I$)语义聚类的单层线性分类面的示例.实际上,可判别分类面具有多种形式,如线性 logistic 分类面, One-against-all 多类 SVM 分类面,以及层次化分类面等.在不同的情况下,函数 $g$ 各不相同,如在线性 logistic 回归分类中:

$$\left.\begin{aligned} P(T_i \mid I) &= \frac{\exp(f_i(x_I))}{1+\sum_{l=1}^{v}\exp(f_l(x_I))}, \quad i=1,...,v-1 \\ P(T_v \mid I) &= \frac{1}{1+\sum_{l=1}^{v}\exp(f_l(x_I))}, \end{aligned}\right\} \tag{9}$$

而在 One-against-all 多类 SVM 分类面中:

$$\begin{cases} P(T_i \mid I)=1, & \text{if } f_i(x_I)>0 \\ P(T_i \mid I)=0, & \text{otherwise} \end{cases} \tag{10}$$

在图像标注实际应用中,一幅图像包含多个关键词,HGPN 中隐藏的语义聚类关系可能非常复杂,语义聚类包含的语义粒度并不均衡.复杂的语义聚类涵盖的语义覆盖面大,这些类之间的可分性强(strong separation),在分类时应首先对它们进行区分.复杂的语义聚类内部可以进行更细致的分类(weak separation).单一层次的分类面划分显然不能满足要求.根据语义聚类包含的语义粒度大小以及语义强弱,我们很自然地将其组织为层次树结构,在其基础上构造相应的可判别超平面树.求解概率 $P(T_i|I)$时,使用逐步求精的层次化分类方法,逐层在生成模型中引入可判别信息.

图 3 给出本文类层次树的抽象表示形式.树中每个节点代表一个隐藏的语义类.叶子节点为 HGPN($I$)中的单个图像.根节点定义为整个 HGPN($I$).同一节点下的分支描述的语义聚类互不相交.对该类层次树的每一层构造如前所述的单层判别分类面$\langle w,b\rangle$,分类超平面的具体形式可根据实际需要选择,这样就得到本文的可判别超平面树.在计算概率$P(T_i|I)$时,使用自顶向下的分类策略.高层的语义类包含较粗的语义粒度,可分性强,首先进行分类.当前节点的后验概率与父节点的概率相关.父节点的概率值将作为当前节点的先验知识,共同决定最终的后验概率.将当前层由公式(8)产生的后验概率记为 $P_{cur}(T_i|I)$,则最终的后验概率 $P(T_i|I)=P(Parent(T_i)|I)\times P_{cur}(T_i|I)$.从根节点出发逐层向下,直到待分类的子树节点的语义达到一致性,分类过程停止.语义一致,无法继续分类的语义聚类组成集合$\{T_1,T_2,\ldots,T_v\}$.由于利用每一层分类超平面$\langle w,b\rangle$中包含的判别信息,与待标注图像语义相关的语义聚类 $T_i$将获得较大的概率 $P(T_i|I)$,因而 $P(I|T_i)$比较大.这样,根据公式(4)得到的最终标注将取决于与待标注图像语义相关的聚类中的图像,因此所得到的标注也更加准确.

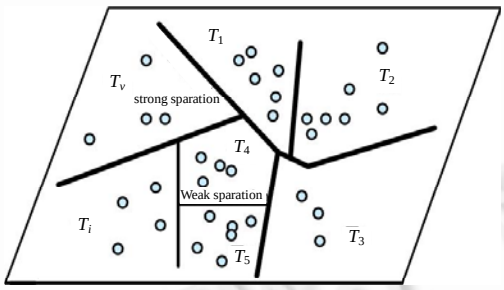



Fig.2 Decision boundaries in two-dimensional subspace of HGPN

图 2 二维空间中 HGPN 的线性决策边界

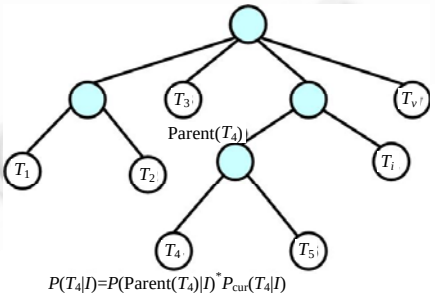



Fig.3 Semantic cluster hierarchy

图 3 语义聚类层次树

### 2.2 可判别超平面树

#### 2.2.1 可判别超平面树的类层次结构

本文的可判别超平面树是在生成的类层次结构基础上构造的,首先通过奇异值分解生成 HGPN($I$)中图像隐藏的类层次结构.奇异值分解(SVD)将 $m\times n$ 矩阵 $A$ 分解为下列 3 个矩阵的乘积:

$$A=U\Sigma V^T \tag{11}$$

其中,对角阵 $\Sigma=diag\{\sigma_1,\sigma_2,...,\sigma_p\}$ ,$p$=min$\{m,n\}$, $\sigma_1\geq\sigma_2\geq...\geq\sigma_p\geq 0$ ,对角元素 $\sigma_i$ 称为矩阵 $A$ 的奇异值.$U\in R^{m\times p}$,$V\in R^{n\times p}$ 为相应的奇异向量构成的矩阵,有 $UU^T=VV^T=I$ ,其中 $\boldsymbol{I}$ 是单位阵.

保留前 $D(D\leq p)$个最大的奇异值,将剩余值置0,可以实现基于SVD的维度消减过程,这可被视为一种隐含语义分析过程[17].将 SVD 的维度消减过程用于本文,使用矩阵 $A$ 描述关键词和图像的共现关系,对矩阵 $A$ 进行标准的 SVD,分别得到矩阵 $U$,$\Sigma$和 $V$.保留前 $D$ 个最大的奇异值,并取该奇异值对应的左右奇异向量,分别得到 $U_D$,$\Sigma_D$ 和 $V_D$.令:$A_1=U_D\Sigma_D.A_1$ 中第 $i$ 行的值为第 $i$ 幅图像在降维后新空间的坐标.

将出现在 HGPN($I$)中图像及其语义关键词组成的共现矩阵使用 SVD 进行降维后,本文通过层次聚类得到类层次树,如图4所示.HGPN($I$)中单个的图像样本 $J_i$ 为叶子节点,这些叶子节点构成单个的聚类.接着,自下而上,将具有最小距离的两个聚类进行合并,形成新的聚类,这些新的聚类就形成了本文的隐藏语义类.这个步骤不断进行,直到所有的对象包含在同一个聚类中.在层次聚类算法中,两个聚类之间的距离有 3 种定义方法,分别是 single,complete 和 average 方法[18],本文采用第 1 种定义方法,即 single 方法.通过层次聚类,图像对象根据它们的上下文相关性被分配到不同的隐藏聚类中.隐藏聚类内部具有较强的语义相关性,而两个隐藏聚类之间的语义不相关.

2.2.2 可判别超平面树生成

由于本文使用层次聚类方法生成类层次树,所有中间节点仅包含左右两个分支.因此,在树结构的中间节点,进行最大边界超平面的学习,得到可判别超平面树,如图 5 所示.鉴于 SVM 良好的泛化能力,本文使用二元 SVM 在每一层进行最大边界分类超平面的学习.

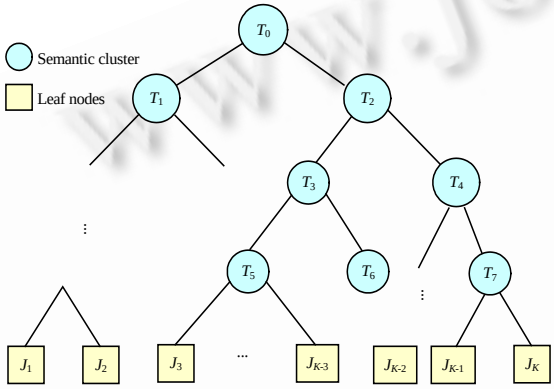



Fig.4 Semantic cluster hierarchy

图 4 语义聚类层次结构

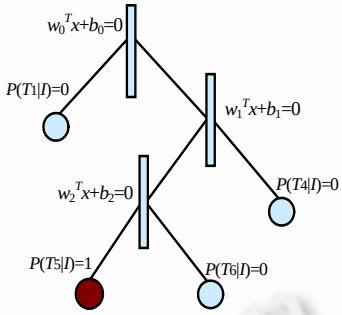



Fig.5 Discriminative hyperplane tree

图 5 可判别超平面树

定义如下记号:$T_i$ 表示类层次树中第 $i$ 个中间节点,$T_{i0}$ 和 $T_{i1}$ 分别表示其左、右孩子节点.图像 $J_k\in T_i$ 当且仅当 $J_k$ 属于 $T_i$,并且有若 $J_k\in T_{i0}$ 或 $J_k\in T_{i1}$,则 $J_k\in T_i$.令 Supp($T_i$)=$\{J_k|J_k\in T_i\}$.

设 $T_i$ 为当前中间节点,我们将所有属于 $T_i$ 的图像排列在一起,并对其重新编号如下: $X=\{x_1,x_2,...,x_l\}$ .相应地,给出每个样例对应的标签 $Y=\{y_1,y_2,...,y_l\}$ .若 $J_k\in T_{i0}$,则 $y_k$=1,否则 $y_k$=−1.待学习的超平面为

$$\{x:f(x)=w^T\varnothing(x)+b=0\} \tag{12}$$

这里通过投影函数$\varnothing$,将图像样本投影到高维空间中对线性不可分情况进行处理.对超平面 $w$和 $b$,使用 SVM 解决如下两类分类问题获得:

$$\begin{aligned}\min_{w,b,\xi} \quad & \frac{1}{2}w^T w + C(\sum_t \xi_t) \\ \text{subject to:} \quad & w^T\phi(x_t)+b \le 1-\xi_t, \ \text{if } x_t \in T_{i0}, \\ & w^T\phi(x_t)+b \ge 1-\xi_t, \ \text{if } x_t \in T_{i1}, \\ & \xi_t \ge 0\end{aligned} \tag{13}$$

对于待标注图像 $I$ 对应的样例 $x_I$,根据公式(10),有:

$$\begin{cases} P(T_{i0} \mid I)=1, & \text{if } w^T\phi(x_I)+b>0 \\ P(T_{i1} \mid I)=0, & \text{otherwise} \end{cases} \tag{14}$$

根据公式(6),可得到 $P(I \mid T_{ij}), j \in \{0,1\}$ 的值.若 $P(T_i \mid I)=1$,则 $P(I \mid T_{ij})>0$,否则 $P(I \mid T_{ij})=0$.

随机采样保持数据集平衡. 由于图像标注的多标签特点,HGPN 中语义关系复杂,很难保证生成的隐藏语义树中左右分支完全平衡.语义单一的类节点中包含的图像个数往往较少,此时同层中另一语义类包含的图像个数远远大于该类.在这种情况下,SVM 学习得到的分类超平面会向包含样例较多的类产生偏移,以便在占主导的类中获得较高的分类准确度.然而,这样会损害分类超平面对未知数据分类时的估计能力.因此,在进行超平面学习时,应检查是否有此类不平衡现象产生.如果存在,应采取措施使得学习超平面的左右分支接近平衡.在对第 $k$ 层分类时,若出现样本不平衡问题,将包含图像个数较多的类记为 $T_{ku}$,则同层中另一类为 $T_{k(1-u)}$,本文采用对 $T_{ku}$ 随机采样解决样本不平衡问题.$T_{k(1-u)}$ 中包含的样本数记为 $|\text{Supp}(T_{k(1-u)})|$,从 $T_{ku}$ 中随机选取 $|\text{Supp}(T_{k(1-u)})|$ 个样本,作为新的 $T_{ku}$ 的训练样例.这样既保持了 $T_{ku}$ 的原始分布,又减少 $T_{ku}$ 中包含的图像数量,达到使局部训练数据集均衡的目的.

预定义核. 在公式 5 中,$\phi$是投影函数.通过$\phi$函数,将样本投影到线性可分空间,进行线性分类超平面学习,有 $K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$,其中 $K$ 是对称半正定矩阵,通常称为核矩阵.当图像被分割为区域后,每幅图像由区域序列构成.组织分类样本时,包含给定类的图像的所有区域都被作为正例.但由于图像分割的语义特性,每个区域与类的语义相关性不同,此时得到的正例中往往包含很多不相关样例(区域),影响分类效果.因此,本文将整幅图像作为一个正例,将生成概率看作一类距离度量,使用预定义核矩阵.具体如下:计算给定局部训练集中图像的成对生成概率,得到矩阵 $K'$.对矩阵 $K'$进行处理,使其满足核矩阵的条件,即对称半正定矩阵.

对称化. 由于图像生成概率具有不对称性,因此令 $K=(K'+K'^T)/2$,使其满足对称性.

半正定化. 由于生成概率中三角不等式原则并不能保证成立,因此无法保证 $K$ 矩阵是半正定矩阵.存在多种方法对该问题进行解决,本文使用文献[19]的做法.计算 $K$ 矩阵的特征值,若最小特征值为负,则 $K$ 的对角线元素加上该特征值的绝对值.此时为对角线元素加上一个正常数,相当于增加了自相关性,并不影响不同样本间的相似性.

2.2.3 基于可判别超平面树的分类

由公式(3)可知,本文假设待标注图像 $I$ 由其高生成邻域 HGPN($I$)生成,并且由类层次树的定义知,根节点由属于HGPN($I$)的所有图像组成,因此有 $P(T_0 \mid I)=1$.接下来,从最高层开始,由判别函数 $G(x)$ 判断待标注图像 $I$ 属于 $T_0$ 的左或右分支,得到 $P(T_{0j} \mid I)$的值,$j \in \{0,1\}$.沿 $P(T \mid I)=1$ 的节点 $T$ 包含的子树继续分类,而 $P(T \mid I)=0$ 的分支不被考虑.因此,在同一层中,存在且仅存在一个隐藏类 $T$,使得 $P(T \mid I)=1$.上述分类过程自上向下,直到目标类包含一致语义,分类过程停止.如图 5 中,到第 3 层分类过程停止,得到相关语义聚类 $T_5$,使得 $P(T_5 \mid I)=1$ 且 $T_5$ 中图像语义相对一致.本文使用熵判断目标类包含的语义是否一致.熵越小,说明目标类包含的语义越“纯净”,出现在该类中的词的语义相关性越高.令当前目标类为 $T_{ij}$,即 $P(T_{ij} \mid I)=1$.将所有出现在 $J_k, J_k \in T_{ij}$ 中的语义关键词记为 $\{w_1,w_2,\ldots,w_p\}$,其对应的出现频率记为 $\{q_1,q_2,\ldots,q_p\}$,则停止条件熵使用下式得到:

$$E(T_{ij})=-\sum_{k=1}^{p} q(i)\log q(i) \tag{15}$$

若 $E(T_{ij})$小于给定阈值 $e$,说明此时隐藏类包含的语义较为单一,无须进行下一步的分类.此时得到的隐藏类为待标注图像的目标相关类,将其记为 $T_r$.待标注图像的标注将从其目标相关类中得到.由于 $T_r$ 中图像包含的语

义相对一致且与待标注图像语义相关,因此由 $T_r$ 生成的标注更加准确.

算法 1 对本文可判别超平面树生成方法以及基于可判别超平面树的分类过程进行了总结.

**算法 1**. 可判别超平面树生成及分类.

输入:待标注图像 $I$ 的 $K$-HGPN;

输出:可判别超平面树 $T_0$,$I$ 的语义相关图像集合 $T_r$.

算法过程:

1. 建立共现矩阵 $A=\{a_{ij}\}_{K\times L}$,描述出现在 $K$-HGPN 中的图像和关键词的关系,其中图像和关键词的个数分别是 $K$ 和 $L$.$A$ 的每个元素 $a_{ij}$ 表示单词 $w_j$ 在第 $i$ 幅图像中的出现次数.若 $w_j$ 在第 $i$ 幅图像中出现,则取值为 1;否则值为 0.
2. 针对共现矩阵 $A$,使用标准的 SVD,可得:

$$A=U\Sigma V^T,$$

保留前 $D(D\le p)$个最大的奇异值,将剩余值置 0,进行维度削减,分别得到 $U_D$, $\Sigma_D$ 和 $V_D$ .令

$$A_1=U_D\Sigma_D,$$

得到每幅图像在降维后空间的坐标.

3. 在新空间中,进行层次聚类,得到隐藏语义类层次树,根节点记为 $T_0$,令 $l=0$.
4. 从根节点出发,根据第 3.2.2 节所述方法构造训练集 $X$ 和样本标签集 $Y$,若 $J_k\in T_{l0}$,则 $y_k=1$;若 $J_k\in T_{l1}$,$y_k=-1$.

判断是否出现样本集不平衡,即$||\text{Supp}(T_{l0})|-|\text{Supp}(T_{l1})||>C$?

若是,转第 5 步;否则,转第 6 步.$C$ 为预先指定大于 0 的阈值.

5. 按第 3.2.2 节介绍的方法随机采样,平衡数据集,重新生成训练集 $X$ 和样本标签集 $Y$.
6. 根据 $X$ 和 $Y$,生成预定义核矩阵,按照公式(13)进行分类超平面学习.
7. 对待标注图像 $I$ 进行分类,若 $f(x_I)$,则 $I\in T_{l0}$,否则 $I\in T_{l1}$.将分类得到的目标类记为 $T_r$,有 $P(T_r\mid I)=1$.

计算 $E(T_r)$,若 $E(T_r)$小于给定阈值,则停止;否则,$l=l+1$,令 $T_{l0}$,$T_{l1}$ 分别等于 $T_r$ 的左、右分支,转第 4 步,进行下一层分类.

### 2.3 基于局部相关类的标注

经过可判别超平面树分类后,语义一致不可再分的语义聚类中,仅 $T_r=\{J_{T_{r1}},J_{T_{r2}},...,J_{T_{rs}}\}$ 对应的概率 $P(I|T_r)>0$.在对 $\hat{P}(w,I)$ 排序时,对标注单词表中的词 $W$,$P(I|T_r)$值相同.因此,等式(4)转化为估计语义关键词和待标注图像 $I$ 从 $T_r$ 中生成的联合概率:

$$\hat{P}(w,I)\Leftrightarrow\sum_{i=1}^{|T_r|}P(w,I\mid J_{T_{ri}})P(J_{T_{ri}})=\sum_{i=1}^{|T_r|}P(w\mid J_{T_{ri}})P(I\mid J_{T_{ri}})P(J_{T_{ri}}) \tag{16}$$

其中,假设 $P(J)$均匀分布,图像分割后各区域相互独立.$P(I|J)$等于 $I$ 中各区域生成概率的乘积,即 $P(I\mid J)=\prod_j P(f_j\mid I)$.区域 $f_j$ 由图像 $J$ 生成的概率 $P(f_j|J)$使用核密度非参数估计[3],如下:

$$P(f_j\mid J)=\frac{1}{m}\sum_{k=1,g_k\in J}^{m}\frac{\exp\{-(g_i-f_j)^T\Sigma^{-1}(g_i-f_j)\}}{\sqrt{2^D\pi^D\mid\Sigma\mid}}, \tag{17}$$

其中,$g_k$ 表示训练图像 $J$ 的第 $i$ 个区域的特征,$m$ 是 $J$ 中区域个数.

对词的估计 $P(w|J)$我们使用如下二重平滑:

$$P(w\mid J)=\lambda_1P_M(w\mid J)+\lambda_2P_{T_r}(w\mid J)+(1-\lambda_1-\lambda_2)P_B(w\mid J) \tag{18}$$

其中 $P_M(w|J)$是指对 $P(w|J)$的极大似然估计, $P_{T_r}(w\mid J)$ 和 $P_B(w|J)$是分别把目标相关类 $T_r$ 和整个训练集作为背景集得到的概率估计值,作为对极大似然估计的平滑项,$\lambda_1$ 和 $\lambda_2$ 是平滑因子.

## 3 标注算法

**算法 2**. 基于可判别超平面树的生成模型图像标注方法(DHTG).

输入:训练集 $L$,待标注图像 $I$;

输出:待标注图像 $I$ 的语义标注.

算法过程:

1. 计算 $I$ 从 $L$ 中每幅图像生成的概率 $P(I|J),J\in L$.
2. 选择前 $K$ 个最高生成概率训练图像,得到 $I$ 的 $K$-HGPN.
3. 根据算法 1,生成可判别超平面树,并对待标注图像 $I$ 进行自上而下分类,得到最终相关的目标主题 $Tr$.

根据公式(16),计算语义关键词 $w$ 和 $I$ 的联合概率 $\hat{P}(w,I)$ ,选择令 $\hat{P}(w,I)$ 最大的前 $t$ 个语义关键词作为图像 $I$ 的语义标注.

时间复杂度分析:令 $n$ 表示训练集的个数,则算法 2 前两步时间复杂度均为 $O(n)$.第 3 步的时间复杂度由以下几部分组成:建立共现矩阵 $A$+SVD 的时间复杂度为 $O(K^3)$+层次聚类的时间复杂度 $O(K^2)$+超平面学习时间复杂度.令 $t$ 表示固定标注长度,邻域中图像个数为 $K$,则建立共现矩阵 $A$ 的时间复杂度为 $O(tK)$.学习分类超平面的时间复杂度由:每层学习时间×分类层数.由于该复杂度受语义主题树的结构影响,而不同数据集得到的主题树结构变化可能较大.同时,实际分类的层数受测试数据本身影响,因此很难给出一个准确的界.由于每层的训练样例不超过 $K$,因此每层学习时间不超过 $O(K^2)$.而在最坏情况下语义主题树层数为 $K-1$ 层.因此,学习分类超平面总的时间复杂度不超过 $(K-1)O(K^2)$,即 $O(K^3)$.假设 $m$ 为单词表的个数,则算法第 4 步的时间复杂度为 $O(m)$.由于 $K$ 和 $m$ 相对 $n$ 非常小,如在本文实验中 $K/n$=25/5000=0.005.因此,算法 2 的复杂度主要由 $O(n)$主导.

## 4 实 验

由于本文算法结合统计生成模型和分类方法的优点,因此为了验证本文标注方法的有效性,分别进行以下两组实验:与基于统计生成模型的标注方法比较,与基于分类的标注方法比较.

### 4.1 实验建立

我们实验中使用的 Corel 数据集取自 ECCV 2002 基准数据集[1].该数据集包括 5 000 幅图像,来自 50 个 Corel Stock Photo CDs.每个 CD 目录下包含同一主题的 100 幅图像.每幅图像和 1~5 个标注词关联,共有 374 个词.我们将数据集分为 3 部分:训练集 4 000 幅,验证集 500 幅,测试集 500 幅图像,其中验证集包括每个目录下的 10 幅图像,主要用来确定模型参数.参数确定后,验证集的数据加到训练集中形成新的训练集重新训练模型.与之前的方法一样,我们主要用检索单个词的查全率、查准率以及 $F1$ 因子来度量标注的性能好坏.给定查询词 $w$,若存在测试集中手工标注结果中包含 $w$ 的图像个数为 $|W_G|$,使用自动标注模型的标注结果中包含该词的图像个数为 $|W_M|$,其中 $|W_C|$ 个是正确的,则 $Recall=\frac{|W_C|}{|W_G|}$,$Precision=\frac{|W_C|}{|W_M|}$,$F1=\frac{2\times Recall\times Precision}{Recall+Precision}$.$Recall$ 度量出对单个词查询的完整性,$Precision$ 度量查询的精度、平均的查准和查全率则反映标注整体的性能.我们使用相同的训练集对这些模型进行训练,并在相同的测试集上进行测试.固定标注长度 $t$ 设为 5.

### 4.2 参数调整

本文所提算法中包含以下可调节的参数:$K$,待标注图像的邻域大小,即最高生成概率邻域中训练图像的个数;$D$,生成类层次结构树时,使用 SVD 降维得到的隐藏语义空间的维数;$e$,自顶向下层次分类的终止条件;$C$,判断是否出现分类样本不平衡的条件阈值.本文通过验证集估计这些参数的最优值.其中,对于 $K$ 的取值,$K$ 值过大,容易在 HGPN 中引入过多对待标注图像的标注作用较小的图像及语义关键词,从而影响 SVD 和 SVM 的效果.$K$ 取值过小,对标注有用的训练图像有可能被排序在 HGPN 之外,从而降低 SVD 得到的主题与待标注图像的相关性.图 6 左图给出调整 $K$ 时,算法在测试集上的平均查全和查准率.容易看出,$K$ 值变化对标注效果影响较大.$K$ 取

25 时获得最优实验结果.对于 $D$,由于 SVD 降维后的空间应刻画 HGPN 中的最主要的信息,而在 HGPN 中,包含的图像个数较少,并且这些图像与待标注图像都具有较高的生成概率,因此本文在实验中选择 $D$=2.同时实验结果表明,$D$ 取不同值时,对标注性能的影响并不明显.考虑 $e$,若 $e$ 选择过大,则分类终止时的目标主题中可能包含的语义较为分散,此时容易降低语义聚类中图像语义的内聚度.而较小的 $e$,容易导致目标主题过小,产生过拟合的效果.图 6 右图给出调整 $e$ 时的实验结果.可以看出,$e$ 的改变对标注性能的影响较小,其取值在 2.0 附近获得较好的标注效果.在下面的实验中,实验结果基于如下参数设置: $K=25, e=2.0, D=2, C=6$ .本文使用 LibSVM[20] 实现单层 SVM 分类器.

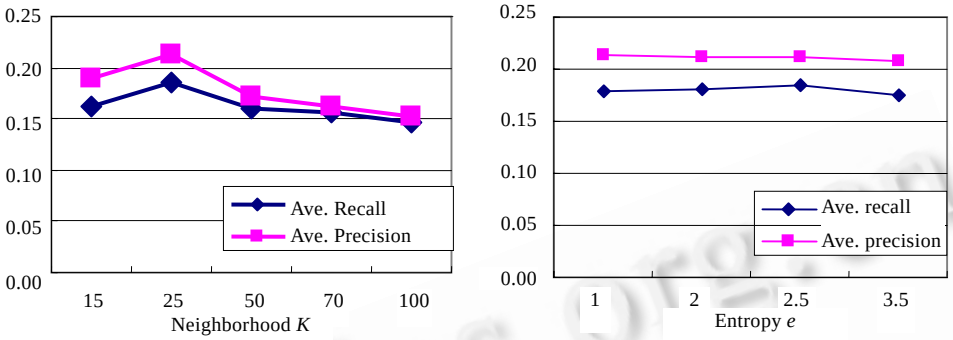



Fig.6 Average recall and precision for different settings of parameters $K$ and $e$

图 6 参数 $K$ 和 $e$ 取不同值时的标注性能比较

### 4.3 性能度量

#### 4.3.1 与基于生成统计生成模型的算法性能比较

在基于生成模型的图像标注方法中,MBRM 取得了较好的实验结果[3].将本文标注方法与 MBRM 比较,见表 1.由于本文研究工作是建立在图像分割的基础上,故下面的实验比较如无特殊声明,MBRM 基于图像分割.非参数核密度估计均选择高斯核.可以看到和 MBRM 相比,在测试集中出现的所有 263 个关键词上统计,本文提出标注方法在平均 *Recall* 和 *Precision* 上均有所提高,分别由 16.1%和 19.0%提高到 18.5%和 21.2%,而 $F1$ 提高了 14%.这时,由于 MBRM 的无指导标注过程无法区分错误图像(具有较高生成概率但与待标注图像语义不相关的图像)与相关图像,容易受错误图像影响从而导致错误标注.而本文使用分类方法对基于生成模型标注方法进行改进,逐层使用可判别超平面树中超平面的可辨别信息对错误图像与相关图像进行区分,减小了错误图像对标注的影响,从而有效地提高了标注性能.表 5 同样给出数据集上最常出现的 49 个关键词的实验结果,实验结果表明,本文算法在出现频率较高的词上性能较 MBRM 提高更多,*Recall* 和 *Precision* 分别由 42.2%和 37.1%提高到 45.5%和 39.4%.

**Table 1** Annotation performance comparison between our method (DHTG) and MBRM

**表 1** 本文方法(DHTG)与 MBRM 标注性能比较

| | Results on all 263 keywords | | | Results on 49 mostly used keywords | | |
|---|---|---|---|---|---|---|
| | Avg.Recall (%) | Avg.Precision (%) | F1 | Avg.Recall (%) | Avg.Precision (%) | F1 |
| MBRM | 16.1 | 19.0 | 0.174 | 42.2 | 37.1 | 0.395 |
| DHTG | 18.5 | 21.2 | 0.198 | 45.5 | 39.4 | 0.422 |

#### 4.3.2 与基于分类的标注方法性能比较

为了进一步验证本文标注方法的有效性,将本文所提方法与基于分类的标注方法进行比较.目前在图像标注中广泛使用的分类器为 SVM.然而根据文献[7]所述,将图像分割为区域后,直接运用 SVM 的标注性能往往低于基于多实例学习标注性能.ASVM-MIL[7]将标注问题作为多实例学习问题,并提出非对称 SVM 对其进行处理,获得了较好的标注效果.将本文方法与 ASVM-MIL 相比较,表 2 列出实验结果.由于 ASVM-MIL 实验中使用的

数据集与本文相同,因此关于 ASVM-MIL 的实验结果直接取自文献[7].同时,为了具有可比性,本文统计了基于最频繁出现的 70 个词的平均标注性能,见表 4.容易看出,本文所提方法的标注性能远远高于 ASVM-MIL,*Recall* 和 *Precision* 分别由 39.7%和 31.2%提高到 40.9%和 38.2%,F1 提高了 13%.

**Table 2** Annotation performance comparison between our method and ASVM-MIL

**表 2** 本文方法与 ASVM-MIL 标注性能比较

| | Results on 70 mostly used keywords | | |
|---|---|---|---|
| | Avg.Recall (%) | Avg.Precision (%) | F1 |
| ASVM-MIL | 39.7 | 31.2 | 0.349 |
| DHTG | 40.9 | 38.2 | 0.395 |

## 5 总结和展望

生成模型与辨别分析方法各有优缺点,为了将两者的优势进行结合,本文提出一种新的图像标注方法.该方法首先在待标注图像的高视觉生成概率领域中,构造可判别超平面树,并设计自顶向下层次分类算法得到待标注图像的目标相关类.在得到的相关类上生成待标注图像的最终标注.由于重新构造用于辨别分析的类,并在分类过程中逐层引入可辨别信息,过滤与待标注图像语义不相关图像的影响,本文所提方法有效地克服了单纯的生成模型与可辨别分类在标注问题中的不足,实现了两者的优势互补.实验结果表明了本文所提方法使得标注性能有显著提高.鉴于在本文的工作中存在多个参数最优值需通过验证集设定,在下一步的工作中,将考虑对参数最优值的自动估计方法.另外,除图像标注属多标签分类问题,还有其他多标签分类问题,下一步工作是将本文研究成果应用于其他多标签分类问题上,并考察其效果.

**References**:

[1] Duygulu P, Barnard K, de Freitas JFG, Forsyth DA. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: Heyden A, ed. Proc. of the European Conf. on Computer Vision. Berlin: Springer-Verlag, 2002. 97−112.

[2] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models. In: Proc. of the Int'l. ACM SIGIR. Toronto: ACM Press, 2003. 119−126.

[3] Feng SL, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2004. 1002−1009.

[4] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures. In: Sebastian T, Lawrence KS, Bernhard S, eds. Proc. of the Neural Information Systems (NIPS).Vancouver, Whistler: MIT Press, 2004. 553−560.

[5] Monay F, Gatica-Perez D. On image auto-annotation with latent space models. In: Lawrence AR, Harrick MV, Thomas P, Prashant JS, John RS, eds. Proc. of the ACM Int'l Conf. on Multimedia. Berkeley: ACM Press, 2003. 275−278.

[6] Monay F, Gatica-Perez D. PLSA-Based image auto annotation: Constraining the latent space. In: Henning S, Nevenka D, eds. Proc. of the Int'l Conf. on ACM Multimedia. New York: ACM Press, 2004. 348−351.

[7] Yang CB, Dong M, Region-Based image annotation using asymmetrical support vector machine-based multiple-instance learning. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2006. 2057−2063.

[8] Gao YL, Fan JP, Xue XY, Jain R. Automatic image annotation by incorporating feature hierarchy and boosting to scale up SVM classifiers. In: Klara N, Matthew T, Yong R, Wolfgang K, Ketan MP, eds. Proc. of the ACM Int'l Conf. on Multimedia. Santa Barbara: ACM Press, 2006. 901−910.

[9] Shi R, Chua TS, Lee CH, Gao S. Bayesian learning of hierarchical multinomial mixture models of concepts for automatic image annotation. In: Hari S, ed. Proc. of the Conf. Image and Video Retrieval. Tempe: Lecture Notes in Computer Science, 2006. 102−112.

[10] 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 Intellience, 2007,29(3):394−410.

[11] Srikanth M, Varner J, Bowden M, Moldovan D. Exploiting ontologies for automatic image annotation. In: Ricardo ABY, Nivio Z, Gary M, Alistair M, John T, eds. Proc. of the SIGIR. Salvador: ACM Press, 2005. 552−558.

[12] Fan JP, Ga YL, Luo HZ. Hierarchical classification for automatic image annotation. In: Proc. of the SIGIR. Amsterdam: ACM Press, 2007. 111−118.

[13] Marszalek M, Schmid C. Semantic hierarchies for visual object recognition. In: Proc. the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Minneapolis: IEEE Computer Society Press, 2007.

[14] Sun A, Lim EP. Hierarchical text classification and evaluation. In: Proc. of the IEEE Int'l Conf. on Data Mining. IEEE Computer Society Press, 2001. 521−528.

[15] Lasserre JA, Bishop CM, Minka TP. Principled hybrids of generative and discriminative models. In: Proc. the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. New York: IEEE Computer Society Press, 2006. 87−94.

[16] Grabner H, Roth PM, Bischof H. Eigenboosting: Combining discriminative and generative information. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Minneapolis: IEEE Computer Society Press, 2007.

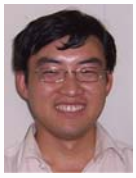
[17] Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. Journal of the American Society for Information Science, 1990,41:391−407.

[18] Hastie T, Tibshirani R, Friedman J. The element of statistical learning; data mining, inference, and prediction. New York: Springer-Verlag, 2001. 472−480.

[19] Zhang H, Berg AC, Maire M, Malik J. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2006. 2126−2136.

[20] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. Software available. 2008. http://www.csie.ntu.edu.tw/~cjlin/libsvm

**王梅**(1980－),女,陕西汉中人,博士生,主要研究领域为多媒体数据库,信息检索.



**许红涛**(1980－),男,硕士生,主要研究领域为多媒体检索.

**周向东**(1969－),男,博士,副教授,主要研究领域为数据库,信息检索.

**施伯乐**(1935－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库理论与应用.