基于互信息约束聚类的图像语义标注

钟 洪 夏利民
(中南大学信息科学与工程学院,长沙 410075)

摘 要 提出一种基于互信息约束聚类的图像标注算法。采用语义约束对信息瓶颈算法进行改进,并用改进的信息瓶颈算法对分割后的图像区域进行聚类,建立图像语义概念和聚类区域之间的相互关系;对未标注的图像,提出一种计算语义概念的条件概率的方法,同时考虑训练图像的先验知识和区域的低层特征,最后使用条件概率最大的语义关键字对图像区域语义自动标注。对一个包含 500 幅图像的图像库进行实验,结果表明,该方法比其他方法更有效。

关键词 图像检索 互信息 约束聚类 信息瓶颈 图像标注
中图法分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2009)06-1199-07

Semantic Annotations of Image Based on Mutual Information and Constrained Clustering

ZHONG Hong, XIA Li-min
(*Information Engineering College, Central South University, Changsha* 410075)

Abstract An image annotation method based on mutual information and constrained clustering is proposed. We utilized the semantic constraint to improve information bottleneck method, which employed to cluster the segmented region. Then relationships between image semantic concept and clustering regions are established. Toward the un-annotated image, a new method is proposed to calculate the conditional probability of each semantic concept, while considering the prior knowledge of training images and low-level features of the segmented regions. Finally, the image region semantics are automatically annotated by keywords with maximal conditional probability. The proposed method has been implemented and tested on an image database with about 500 images. The experimental results show that the effectiveness of the proposed method outperforms other approaches.

Keywords image retrieval, mutual information, constrained clustering, information bottleneck, image annotation

1 引 言

随着多媒体数据的大规模运用,基于内容的图像检索(CBIR)研究成为当前的一个热点。早期的图像检索系统,多是利用图像的颜色、纹理和形状等低层视觉特征和组合特征进行检索[1-2]。但计算机通过低层特征匹配得到的图像与用户对图像信息的理解存在着不一致性,即存在所谓的“语义差距”。为了克服 CBIR 系统中低层特征无法准确描述高层语义的问题,就需要对图像进行语义解释,这个过程是图像标注最重要的部分。

目前图像语义的表示主要采用手工标注的方法,即用关键字表达图像的部分高层语义,这种方法在某些场合是非常有效的,但采用关键字标注图像语义,由于用户对图像的理解不同,不可避免地存在主观性和不精确性。同时由于图像的广泛应用,存在大量图像需要标注,完全用手工方法标注,工作量

基金项目:国家自然科学基金重大项目(79816101);湖南省自然科学基金项目(05JJ30121)
收稿日期:2007-10-12;**改回日期**:2008-01-16
第一作者简介:钟 洪(1981~),男。2008 年于中南大学获工学硕士学位。研究方向为模式识别、图像检索。E-mail: zhonghong200@163.com

太大,不易实际应用。

如何对图像自动进行语义标注已成为迫切需要解决的问题,目前存在的图像语义标注方法主要有:采用基于子空间聚类算法,用 K-means 算法生成 blob-token,并用统计的方法在 token 和 key-word 之间建立关联,实现图像的标注[3];用分类的方法在 visual terms 和 keyword 之间建立关联,以此构建分类器,将分类器用于后续图像的标注[4]。以上基于视觉特征的分类方法,通常将具有相同视觉特征的区域归为一类,即使区域的语义完全不同,也用相同的关键字标注,因此标注的精确度较低。文献[5]提出一种基于 Boosting 学习的图片自动语义标注方法,其基本思想是:首先构造很多的模型,然后在模型和概念之间建立联系,保持一种多对多的关系。文献[6]提出一种概念索引的方法,采用支持向量机(SVM)的多类分类器的空间映射方法,将图像的低层特征映射为具有一定高层语义的模型特征以实现概念索引。这两种方法在一定程度上提高了标注的精确度,但是受 Boosting 的迭代次数或 SVM 的参数的影响,系统的鲁棒性不强。

本文提出基于互信息约束聚类的图像标注方法,采用语义约束对信息瓶颈算法进行改进,然后使用它对图像区域进行聚类;对图像区域进行聚类以后,确定语义概念和聚类区域之间的相互关系;在语义标注阶段,在给出分割区域的条件下,提出一种计算语义概念的条件概率的方法,它同时考虑训练图像的先验知识和区域的低层特征,使用最大条件概率的语义关键字对分割后的区域进行标注,对图像进行标注之后,就能以关键字的方式进行检索。实验结果表明,该方法比其他方法性能更好,系统具有良好的鲁棒性。

2 信息瓶颈算法

信息论是应用近代数理统计方法研究信息的传输、存储与处理的科学。对于一个事件 X_i,假设其发生的概率为 $p(X_i)$,那么此事件的 X_i 的信息量定义为

$$H(X_i) = -\log p(X_i) \tag{1}$$

在信息论中,熵、互信息和相对熵是 3 个重要的概念。

定义 1(熵) 信息量的单位是比特(bit),用于表示一定时间内的信息的期望值称为熵,记为 $H(X)$。

$$H(X) = -\sum_i p(X_i)\log p(X_i) \tag{2}$$

定义 2(相对熵) 两个概率密度 $p(x)$ 和 $q(x)$ 之间的相对熵,或 Kullback Leibler 距离定义为

$$D_{\mathrm{KL}}(p \| q) = \sum_x p(x)\log\frac{p(x)}{q(x)} \tag{3}$$

定义 3(互信息) 假定存在随机系统有输入变量 $\boldsymbol{X}$ 与输出变量 $\boldsymbol{Y}$,它们的联合概率是 $p(x,y)$,并且它们各自的边缘概率为 $p(x)$ 与 $p(y)$。那么 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的互信息为

$$I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{4}$$

信息瓶颈方法是由 Tishby 等人提出的一种新颖的无监督聚类方法[7]。其基本思想是:对于一个随机变量 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$,其特征变量为 $\boldsymbol{Y}=\{y_1,y_2,\cdots,y_m\}$,通过类 $\boldsymbol{\Gamma}=\{R_1,R_2,\cdots,R_B\}$ 来表示 $\boldsymbol{X}$,使得 $\boldsymbol{\Gamma}$ 尽可能地压缩 $\boldsymbol{X}$ 的信息,同时尽可能保留相关的特征信息 $\boldsymbol{Y}$,即使 $\boldsymbol{X}$ 与 $\boldsymbol{\Gamma}$ 的互信息 $I(\boldsymbol{X};\boldsymbol{\Gamma})$ 最小,$\boldsymbol{Y}$ 与 $\boldsymbol{\Gamma}$ 的互信息 $I(\boldsymbol{Y};\boldsymbol{\Gamma})$ 最大。聚类 $\boldsymbol{\Gamma}$ 可以看作是样本 $\boldsymbol{X}$ 与特征 $\boldsymbol{Y}$ 间的信息瓶颈。根据信息瓶颈方法聚类,实际上就转化为求优化问题,即求从 x 映射到类 R 的概率 $p(R|x)$ 使得:

$$\min_{p(R|x)}\left[I(\boldsymbol{X};\boldsymbol{\Gamma}) - \beta I(\boldsymbol{Y};\boldsymbol{\Gamma})\right] \tag{5}$$

式中,$\beta>0$ 是表示两个目标间平衡关系的参数。$I(\boldsymbol{X};\boldsymbol{\Gamma})$,$I(\boldsymbol{Y};\boldsymbol{\Gamma})$ 分别表示 $\boldsymbol{X}$ 与 $\boldsymbol{\Gamma}$ 的互信息、$\boldsymbol{Y}$ 与 $\boldsymbol{\Gamma}$ 的互信息。

3 基于语义约束信息瓶颈的区域聚类

传统的聚类算法(如 K-means 算法)的性能过分依赖距离函数和聚类中心的选择,如果距离函数选择的不对,则算法的效率很差。信息瓶颈算法不需要定义距离函数,只需要知道一个 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的联合概率分布。与传统的聚类算法相比,信息瓶颈算法考虑了样本与特征的关系。不仅要压缩样本的信息,同时又考虑保留特征信息。但是因为信息瓶颈算法是无监督的聚类算法,它本身存在一定的缺陷,为了得到更好的聚类结果,在聚类过程中加入语义约束对信息瓶颈算法进行改进。

3.1 数据集之间的语义约束

带约束条件的聚类是指特定的领域知识以“约束”的形式表达,并嵌入到聚类过程中,使得聚类算法获得更多的启发式信息,提高了效率和聚类质量。

Wagstaff 引入了 Must-link(相关)和 Cannot-link(不相关)两种约束[8]。我们在区域聚类过程中使用 Cannot-link 约束,对于每一个约束都有一个惩罚代价 P,通过已标注的图像推出 Cannot-link 关系。

一般来说,具有很强共生关系的概念可能同时用来标注同一幅图像,换句话说,假如两个概念具有共生关系,则可能属于一个概念组。基于共生关系的两个概念 c_i 和 c_j 计算如下:

$$R_C(c_i,c_j)=\begin{cases}\dfrac{dc(c_i\bigcap c_j)}{dc(c_i\bigcup c_j)} & dc(c_i)>\sigma\\ 0 & \text{其他}\end{cases} \tag{6}$$

其中,$dc(c_i\bigcap c_j)$ 表示同时标注 c_i 和 c_j 的图像的数目,$dc(c_i\bigcup c_j)$ 表示标注 c_i 或 c_j 的图像的数目。对于已经标注的图像 $\boldsymbol{I}_p$ 和 $\boldsymbol{I}_q$,假设它们的标注分别为 $\boldsymbol{C}_p$ 和 $\boldsymbol{C}_q$,则标注 $\boldsymbol{C}_p$ 和 $\boldsymbol{C}_q$ 之间的相互关系定义如下:

$$Rel(\boldsymbol{C}_p,\boldsymbol{C}_q)=\arg\max_{c_i\in\boldsymbol{C}_p,c_j\in\boldsymbol{C}_q}(R_C(c_i,c_j)) \tag{7}$$

实验中,如果两个标注之间的 $Rel(\boldsymbol{C}_p,\boldsymbol{C}_q)<\sigma_1$(阈值),则 $\boldsymbol{C}_p$ 和 $\boldsymbol{C}_q$ 及它们的对应区域认为是不相关的。如果 $\boldsymbol{C}_p$ 和 $\boldsymbol{C}_q$ 是不相关的,则它们之间的所有区域都标记为 Cannot-link(不相关)。对于图像中的区域 x_i,x_j,定义相关函数如下:

$$\forall x_i\in\boldsymbol{I}_p,\ \forall x_j\in\boldsymbol{I}_q$$
$$s(x_i,x_j)=\begin{cases}1 & x_i,x_j\text{ 为 Cannot-link}\\ 0 & \text{其他情况}\end{cases} \tag{8}$$

3.2　基于信息瓶颈的聚类算法

利用信息瓶颈算法进行聚类,就是求出 $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$ 的聚类 $\boldsymbol{\Gamma}=\{R_1,R_2,\cdots,R_B\}$,对于给定的约束 $I(\boldsymbol{X};\boldsymbol{\Gamma})$,使得信息损失 $I(\boldsymbol{X};\boldsymbol{Y})-I(\boldsymbol{\Gamma};\boldsymbol{Y})$ 最小,使用凝聚的信息瓶颈的聚类算法,开始时每一个聚类由一个单个的区域组成,为了使聚类过程中的信息损失最小,类的合并的每一步都选择使得信息损失最小的两个类进行合并。假设 R_i 和 R_j 是 $\boldsymbol{\Gamma}$ 中的两个聚类,由于合并 R_i 和 R_j 而造成信息的损失如下:

$$d(R_i,R_j)=I(\boldsymbol{\Gamma}_{\text{before}};\boldsymbol{Y})-I(\boldsymbol{\Gamma}_{\text{after}};\boldsymbol{Y})\geqslant 0$$

式中,$I(\boldsymbol{\Gamma}_{\text{before}};Y)$ 为 R_i 和 R_j 合并前类与特征空间的互信息,$I(\boldsymbol{\Gamma}_{\text{after}};Y)$ 为 R_i 和 R_j 合并后类与特征空间的互信息。由信息理论可知:

$$\begin{aligned}d(R_i,R_j)=&\sum_y p(R_i,y)\log\frac{p(R_i,y)}{p(R_i)p(y)}+\\ &\sum_y p(R_j,y)\log\frac{p(R_j,y)}{p(R_j)p(y)}-\\ &\sum_y p(R_i\bigcup R_j,y)\log\frac{p(R_i\bigcup R_j,y)}{p(R_i\bigcup R_j)p(y)}\\ =&\sum_y p(R_i,y)\log\frac{p(y\mid R_i)}{p(y\mid R_i\bigcup R_j)}+\\ &\sum_y p(R_j,y)\log\frac{p(y\mid R_j)}{p(y\mid R_i\bigcup R_j)}\\ =&\sum_y p(R_i)D_{\text{KL}}(p(y\mid R_i)\parallel p(y\mid R_i\bigcup R_j))+\\ &\sum_y p(R_j)D_{\text{KL}}(p(y\mid R_j)\parallel p(y\mid R_i\bigcup R_j))\end{aligned} \tag{9}$$

3.3　基于语义约束信息瓶颈算法的区域聚类

用集合 $\boldsymbol{X}=\{x_i\}_{i=1}^n$ 表示 n 个图像区域,为了应用信息瓶颈算法(IB)对图像区域进行聚类,必须首先定义图像区域和其特征的联合概率分布。假设对于给定图像区域的先验概率为 $p(x)$,图像特征分布用条件密度函数 $f(y|x)$ 描述,则可得到图像区域-特征联合分布 $p(x,y)$。$f(y|x)$ 是一个高斯混合模型(GMM)[9]:

$$f(y\mid x)=\sum_{j=1}^{k(x)}a_{x,j}N\left(\boldsymbol{\mu}_{x,j}\boldsymbol{\Sigma}_{x,j}\right) \tag{10}$$

式中,$k(x)$ 为 $f(y|x)$ 中高斯混合成分的个数,$\boldsymbol{\mu}_{x,j}$ 表示均值,$\boldsymbol{\Sigma}_{x,j}$ 表示方差。

对于每一个聚类 $\boldsymbol{R}$,其特征分布 $f(y|R)$,表示聚类 R 每个区域 x 的 GMM 的平均值。

$$\begin{aligned}f(y\mid R)=&\frac{1}{|\boldsymbol{R}|}\sum_{x\in\boldsymbol{R}}f(y\mid x)\\ =&\frac{1}{|\boldsymbol{R}|}\sum_{x\in\boldsymbol{R}}\sum_{j=1}^{k(x)}a_{x,j}N\left(\boldsymbol{\mu}_{x,j},\boldsymbol{\Sigma}_{x,j}\right)\end{aligned} \tag{11}$$

式中,$|\boldsymbol{R}|$ 为聚类 $\boldsymbol{R}$ 中元素的个数,因为 $f(y|x)$ 是一个 GMM 分布,对于每个聚类 $\boldsymbol{R}$ 的密度函数 $f(y|\boldsymbol{R})$ 是 $f(y|x)$ 的线性组合,因此它也是一个 GMM。

设 $f(y|\boldsymbol{R}_i)$ 和 $f(y|\boldsymbol{R}_j)$ 分别为图像区域聚类 $\boldsymbol{R}_i$ 和 $\boldsymbol{R}_j$ 的高斯混合模型,则 $\boldsymbol{R}_i$ 和 $\boldsymbol{R}_j$ 合并后的高斯混合模型为

$$\begin{aligned}f(y\mid\boldsymbol{R}_i\bigcup\boldsymbol{R}_j)=&\frac{1}{|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j|}\sum_{x\in\boldsymbol{R}_i\bigcup\boldsymbol{R}_j}f(y\mid x)\\ =&\frac{|\boldsymbol{R}_i|}{|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j|}f(y\mid\boldsymbol{R}_i)+\\ &\frac{|\boldsymbol{R}_j|}{|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j|}f(y\mid\boldsymbol{R}_j)\end{aligned} \tag{12}$$

根据式(9)聚类 R_i 和 R_j 之间的距离表示如下:

$$d(\boldsymbol{R}_i,\boldsymbol{R}_j)=\frac{|\boldsymbol{R}_i|}{n}D_{\mathrm{KL}}(f(y|\boldsymbol{R}_i)\,\|\,f(y|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j)+\frac{|\boldsymbol{R}_j|}{n}D_{\mathrm{KL}}(f(y|\boldsymbol{R}_j)\,\|\,f(y|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j) \tag{13}$$

式中，n 表示图像区域的数目，$d(\boldsymbol{R}_i,\boldsymbol{R}_j)$ 表示合并两个聚类所造成的信息损失。

因为信息瓶颈算法不能保证得到最优的聚类结果，提出约束聚类(CC)方法对它进行优化，在这个方法中对数据集 $\boldsymbol{X}=\{x_i\}_{i=1}^{n}$ 引入约束，对信息瓶颈的目标函数进行改进，改进后的目标函数为聚类区域之间的距离及约束惩罚代价的总和。

$$d_{\mathrm{new}}(\boldsymbol{R}_i,\boldsymbol{R}_j)=\frac{|\boldsymbol{R}_i|}{n}D_{\mathrm{KL}}(f(y|\boldsymbol{R}_i)\,\|\,f(y|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j))+\frac{|\boldsymbol{R}_j|}{n}D_{\mathrm{KL}}(f(y|\boldsymbol{R}_j)\,\|\,f(y|\boldsymbol{R}_i\bigcup\boldsymbol{R}_j))+\sum_{|(x_i,x_j)|l(i)=l(j)|}s(x_i,x_j)\times p \tag{14}$$

式中，$x_i\in\boldsymbol{R}_i$，$x_j\in\boldsymbol{R}_j$。在聚类过程中，每次选择 $d_{\mathrm{new}}(\boldsymbol{R}_i,\boldsymbol{R}_j)$ 最小的两个聚类进行合并，直到 $d_{\mathrm{new}}(\boldsymbol{R}_i,\boldsymbol{R}_j)>\sigma_2$(阈值)，最后得到 B 个聚类。

4 图像语义标注

由于图像的语义大都通过区域对象来体现，采用基于区域的图像标注是一种合理的方法，即通过区域的语义标注图像。本文提出一种新的图像标注方法，它同时考虑训练图像的先验知识和区域的低层特征，首先通过已标注的图像求出关键字和聚类区域的关联度，将待标注图像的区域特征与聚类区域特征进行比较，建立待标注图像区域与关键字间的关联，实现待标注图像语义的自动标注。

4.1 关键字与聚类区域关联度的计算

通过图像分割和区域聚类后，得到 B 个聚类 $\boldsymbol{R}=(\boldsymbol{R}_1,\boldsymbol{R}_2,\cdots,\boldsymbol{R}_B)$，定义 $\boldsymbol{C}$ 为图像标注的关键字的集合，$\boldsymbol{C}=\{c_1,c_2,\cdots,c_W\}$，设图像库中共有 N 幅训练图像，通过图像分割和区域聚类后，需要统计出每一个语义关键字和聚类之间的概率。为了确定关键字和聚类之间的联系，首先建立一个概率表，假设总共有 W 个关键字，B 个图像聚类区域($\boldsymbol{R}_1,\boldsymbol{R}_2,\cdots,\boldsymbol{R}_B$)，和 N 幅图像，然后用一个矩阵 $\boldsymbol{M}_{N\times(W+B)}$ 来表示数据集，在矩阵 $\boldsymbol{M}$ 中，行 N 表示图像的个数，前 W 列表示 W 个语义关键字，最后的 B 列表示 B 个聚类。将矩阵 $\boldsymbol{M}_{N\times(W+B)}$ 进行划分如下：

$$\boldsymbol{M}_{N\times(W+B)}=[\boldsymbol{M}_{N\times W}|\boldsymbol{M}_{N\times B}]=[\boldsymbol{M}_{W1}|\boldsymbol{M}_{B1}] \tag{15}$$

$\boldsymbol{M}_{W1}[i,j]$ 为第 j 个语义关键字出现在第 i 幅图像中的权值，$\boldsymbol{M}_{B1}[i,j]$ 为第 j 个聚类出现在第 i 幅图像中的权值。

先后确定聚类区域的权值和关键字的权值。例如 w_{il} 为图像 l 中聚类 R_i 的权值，设 N 为总的图像的数目，n_i 为在聚类 R_i 中出现的图像的数目。定义归一化频率为

$$tf_{il}=\frac{freq_{il}}{\max\limits_{h}freq_{hl}} \tag{16}$$

式中，$freq_{il}$ 是聚类 R_i 在图像 l 中出现的次数，$\max\limits_{h}freq_{hl}$ 是聚类 $\{R_h\}$，$i=1,\cdots,B$ 在图像 l 中出现的最大的次数。假设当一个聚类出现在大多数的图像中，则它在区分相关图像和不相关图像时是不重要的。所以需要为 R_i 定义一个倒排文档频率：idf_i：

$$idf_i=\log\frac{N}{n_i} \tag{17}$$

平衡上面的两个因素，得到聚类区域的权值为

$$w_{il}=tf_{il}\times idf_i=tf_{il}\times\log\frac{N}{n_i} \tag{18}$$

应用相似的方法确定图像中每一个关键字的权值。通过上面的方法得到 $\boldsymbol{M}_{N\times(W+B)}$ 的概率值后，得到两个矩阵 $\boldsymbol{M}_{W1}$ 和 $\boldsymbol{M}_{B1}$，将 $\boldsymbol{M}_{W1}$ 的转置矩阵和 $\boldsymbol{M}_{B1}$ 相乘，即 $\boldsymbol{M}_{W1}^{\mathrm{T}}\times\boldsymbol{M}_{B1}$，得到一个 $W\times B$ 维的概率矩阵，然后对矩阵中的每一列都归一化，这样就得到一个概率表 T_{coor}，它表示关键字和聚类区域的关联度，$T_{\mathrm{coor}}[i,j]$ 是给定聚类区域 R_j 的条件下，关键字 c_i 的条件概率，即 $p(c_i|R_j)$。

4.2 图像语义标注

在图像标注阶段，首先对待标注的图像 $\boldsymbol{I}'$ 进行分割，得到区域的集合 $\boldsymbol{x}'=\{x_1',x_2',\cdots,x_k'\}$，图像标注的任务就是：对所有的 $x_j'\in\boldsymbol{x}'$，计算条件概率 $p(c_i|x_j')$，$i=1,\cdots,W$，以 $p(c_i|x_j')$ 最大的 c_i 作为图像区域的语义标注，给定区域 x_j' 和候选关键字 $c_i\in\boldsymbol{C}$，条件概率 $p(c_i|x_j')$ 计算如下：

$$p(c_i|x_j')=\sum_{k=1}^{B}(sim(x_j',R_k)\times p(c_i|R_k)) \tag{19}$$

其中，先验概率 $p(c_i|R_k)$ 是已标注的聚类区域和关键字间的关联程度，$sim(x_j',R_k)$ 表示待标注区域 x_j' 与已标注聚类 R_k 间的相似性。

$$sim(x_j',R_k)=1-norm(\|F_{x_j'}-F_{R_k}\|) \tag{20}$$

式中，$\|\cdot\|$ 表示欧氏距离，$F_{x_j'}$、F_{R_k} 分别为区域

x'_j,R_k 的低层特征向量,$norm(\| F_{x'_j} - F_{R_k} \|)$ 表示 $\| F_{x'_j} - F_{R_k} \|$ 归一化的结果。一般说来,$p(c_i|x'_j)$ 值越大,则关键字 c_i 与区域 x'_j 关联的程度越密切,c_i 越能表达区域 x'_j 的语义,所以选择 $\underset{c_i}{\operatorname{argmax}}\, p(c_i|x'_j)$ 且 $p(c_i|x'_j) > \lambda$(阈值)的关键字 c_i 标注该图像区域的语义。

5　实验结果分析

为了验证本文方法的效果,在 Intel P4-2.8G CPU 的微机,Windows2000 平台上采用上述方法对一个具有 500 幅来自 Corel 图像库的图像进行实验,该图像库分为 10 类,每类有 50 幅图像。选择每类中的 20 个图像作为训练样本,剩下的 30 个图像作为测试样本。这样共有 200 个训练样本。对已标注的训练样本使用文献[1]的分割方法进行分割,对分割得到的 1 124 个区域采用改进的信息瓶颈算法进行聚类,对未标注的样本采用提出的语义标注算法进行标注,整个数据集共有 24 个不同的关键字。

图 1 是采用文献[3]、文献[6]、文献[10]和本文方法对 4 幅不同图像的标注结果,文献[3]采用基于子空间聚类算法对图像进行标注。文献[6]采用支持向量机(SVM)的多类分类器的空间映射方法,将图像的低层特征映射为具有一定高层语义的模型特征以实现概念索引。文献[10]采用 K-means 聚类和贝叶斯模型进行语义自动标注,从图 1 中可以看出前几种方法都将第 1 幅图像中的 sky(天空)标注成了 water(水),而本文方法得到了正确的标注结果。

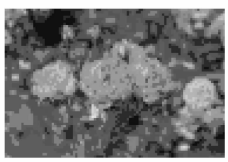
方法				
文献[10]	elephant leaf water grass	horse leaf grass	flowers grass tress rose	grass water building
文献[3]	elephant water field grass	horse trees grass	flowers leaf grass rose	grass sky building
文献[6]	elephant leaf water grass	horse leaf grass	flowers grass leaf rose	grass water mountains
本文方法	elephant sky field grass	horse trees grass	flowers leaf grass rose	grass sky mountains

图 1　部分图像的标注结果比较

Fig. 1　Annotation results of part image

对检索算法的性能评价的重要指标是查准 precision,查全率 recall 和 $F1$ 测度值。查准率定义为检索出的图像中相关图像的数目占的比例,查全率定义为检索出的相关图像的数目占数据库中所有相关的图像数目的比例。查全率反映系统检索相关图像的能力,而查准率则反映系统拒绝无关图像的能力。$F1$ 测度值是综合考虑了查全率和查准率的性能评价指标,$F1$ 测度值表示如下:

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \tag{21}$$

以关键字进行检索,数据库中相关图像是 50,本文按照检索出的图像数目分别为 12,30,45,60,75,90 时 6 种情形进行实验,分别以 horse(马),people(人),flower(花),elephant(大象),building(建筑物),mountain(山),bus(车)进行检索,并与文献[3]~文献[6]和 CMRM[10] 的性能进行比较,得到的平均 $F1$ 测度值如表 1 所示。文献[3]采用基于子空间聚类算法,用 K-means 算法生成 blob-token,在聚类的过程中,没有考虑区域的语义,只将具有相同视觉特征的区域归为一类,而且 K-means

算法的性能过分依赖距离函数和聚类中心,因此文献[3]的聚类效果不是很理想,这样影响了标注的准确度,也就会影响检索的性能,从表 1 中可以看出,文献[3]以大象(elephant)为关键字进行检索的 *F*1 测度值为 0.286,以其他关键字进行检索的 *F*1 测度值更低。本文采用改进的信息瓶颈算法进行聚类,使聚类算法获得更多的启发式信息,提高了效率和聚类质量,同样以大象为关键字检索,本文方法的 *F*1 测度值为 0.421,和文献[3]对比,*F*1 值提高了 47.2%。

文献[4]用分类的方法在 visual terms 和 keyword 之间建立关联,以此构建分类器,将分类器用于后续图像的标注。该方法通常将具有相同视觉特征的区域归为一类,即使区域的语义完全不同,也用相同的关键字标注,因此标注的精确度较低,从表 1 中可以看出,文献[4]的方法以 bus(车)检索时 *F*1 测度值最大,值为 0.361,和文献[4]相比,本文方法在图像标注时考虑了训练图像的先验知识和区域的低层特征,克服了文献[4]中方法的缺陷,同样以 bus(车)检索,*F*1 测度值为 0.432,和文献[4]相比,提高了 19.7%。文献[10]首先采用 Normalized-cut 方法将图像分割成区域,然后使用 K-means 聚类算法对分割的区域进行聚类,最后使用贝叶斯模型进行图像标注。它存在和文献[3]、[4]同样的问题,因此标注的精确度低,从表 1 中可见,*F*1 测度值最大为 0.267,本文方法和它比较,*F*1 测度值提高了 61.8%。文献[5]采用 Boosting 算法对图像进行标注,它首先构造很多的模型,然后在模型和概念之间建立联系,保持一种多对多的关系,它受 Boosting 迭代次数的影响,系统的鲁棒性不强,标注的精确度也较低,例如以 flower(花)检索的 *F*1 测度值为 0.373,而本文方法和它比较,*F*1 测度值提高了 10.5%。文献[6]提出一种概念索引的方法,采用支持向量机(SVM)的多类分类器为空间映射方法,将图像的低层特征映射为具有一定高层语义的模型特征以实现概念索引,这种方法在一定程度上缩小了低层特征和高层语义之间的语义鸿沟,提高了标注的精确度,但受 SVM 的参数的影响,对于不同类别的图像,标注的效果相差较大,例如,对于 building(建筑物),*F*1 测度值只有 0.371,而对于 bus(车),*F*1 测度值为 0.423。在本文中,通过语义约束对信息瓶颈算法进行改进,在图像标注时同时考虑训练图像的先验知识和区域的低层特征,缩小了语义鸿沟,系统具有良好的鲁棒性,因此较好地克服了以上 5 种方法的缺陷,实验结果表明,本文的方法性能更好。

表 1 6 种图像标注方法的平均 *F*1 测度值比较结果

Tab. 1 *F*1 measure values results of image annotation six methods

关键字	平均 *F*1 测度值					
	文献[3]	文献[4]	文献[5]	文献[6]	文献[10]	本文方法
horse	0.273	0.347	0.368	0.415	0.252	0.442
people	0.264	0.325	0.357	0.402	0.242	0.407
flower	0.279	0.354	0.373	0.407	0.254	0.412
elephant	0.286	0.357	0.379	0.412	0.263	0.421
building	0.252	0.317	0.349	0.371	0.226	0.394
mountain	0.258	0.316	0.354	0.373	0.232	0.397
bus	0.285	0.361	0.382	0.423	0.267	0.432

为了评价标注关键字的个数对系统性能的影响,对标注关键字的个数分别为 3,4,5,6 这 4 种情况进行实验,同样以标注的关键字进行检索,以查全率、查准率和 *F*1 测度值评价检索的性能。实验结果如表 2 所示。

表 2 标注关键字的个数对系统检索性能的影响

Tab. 2 Influence of annotation keywords number on performance of system retrieval

关键字个数	查全率	查准率	*F*1 测度值
3	0.39	0.44	0.413
4	0.46	0.41	0.434
5	0.48	0.36	0.411
6	0.52	0.31	0.388

从表 2 可以看出,随着标注关键字个数的增加,查准率越来越小,查全率越来越大,*F*1 测度值先增大后减少,当标注关键字的个数为 4 时,*F*1 测度值最大,系统的检索性能最好。

6 结 论

对图像的关键字自动标注技术进行了研究,提出了一种改进的信息瓶颈聚类算法,并用改进的信息瓶颈算法对分割后的图像区域进行聚类,然后计算出语义概念和聚类区域之间的关联度,在语义标注阶段,对未标注的图像进行分割,然后计算给出分割区域的条件下标注语义概念的概率,使用条件概率最大的语义关键字对分割后的区域进行标注,和

目前认为标注效果较好的几种方法进行比较,实验结果表明,本文方法具有更好的检索性能,系统具有良好的鲁棒性。下一步,将在系统中增加主动学习机制,并考虑与相关反馈技术相结合,以达到更高的检索性能。

参考文献(References)

1 Ouyang Jun-lin, Xia Li-min. Image retrieval based on color and shape features[J]. Journal of Chinese Computer Systems, 2007, **28**(7): 1262-1266. [欧阳军林,夏利民. 基于二值信息的颜色和形状特征的图像检索[J]. 小型微型计算机系统,2007,**28**(7):1262-1266.]

2 Zhong Hong, Xia Li-min. Ontology-based image retrieval [J]. Computer Engineering and Applications, 2007, **43**(17): 37-40. [钟洪,夏利民. 基于本体的图像检索[J]. 计算机工程与应用,2007,**43**(17):37-41.]

3 Wang lei, Liu Li, Latifu. Automatic image annotation and retrieval using subspace clustering algorithm[A]. In: Proceedings of the 2nd ACM International Workshop on Multimedia Databases [C], Washington, DC, USA, 2004: 100-108.

4 Li Wei, Sun Mao-song. Automatic image annotation based on WordNet and hierarchical ensembles [A]. In: Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Text Processing[C], Mexico City, Mexico, 2006: 417-428.

5 Ru Li-yun, Ma Shao-ping. Boosting-based automatic linguistic indexing of pictures [J]. Journal of Image and Graphics, 2006, **11**(4): 486-491. [茹立云,马少平. 基于 Boosting 学习的图片自动语义标注[J]. 中国图象图形学报,2006,**11**(4):486-491.]

6 Lu Jing, Ma Shao-ping. Automatic image annotation based on concept indexing[J]. Journal of Computer Research and Development, 2007, **44**(3): 452-459. [路晶,马少平. 基于概念索引的图像自动标注[J]. 计算机研究与发展,2007,**44**(3):452-459.]

7 Slonim N, Tishby N. Agglomerative information bottleneck [A]. In: Solla S A, Leen T K. Advances in Neural Information Processing Systems[C], Cambridge, USA: MIT Press, 1999: 617-623.

8 Kiri Wagstaff, Claire Cardie, Seth Rogers. Constrained K-means clustering with background knowledge [A]. In: Proceedings of the 18th International Conference on Machine Learning [C], Williams College, USA, 2001: 577-584.

9 Yang M, Ahuja N. Gaussian mixture model for human skin color and its application in image and video database[A]. In: Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases[C], San Jose, USA, 1999: 458-466.

10 Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models[A]. In: Proceedings of the 26th Annual Internation ACM SIGIR Conference on Research and Development in Information Retrieval[C], Toronto, Canada, 2003: 119-126.