

基于贝叶斯理论的图像标注和检索

张元清¹ 包骏杰¹ 况 夯¹ 钟 洪²

(重庆教育学院 重庆 400067)¹ (中南大学信息工程学院 长沙 410075)²

摘要 图像自动语义标注是基于内容图像检索中很重要且很有挑战性的工作。提出用语义约束的聚类方法对分割后的图像区域进行聚类,在图像标注阶段,使用贪心选择连接(GSJ)算法找出聚类区域的独立子集,然后使用贝叶斯理论进行语义标注。对图像进行标注以后,使用标注的关键词进行检索。在一个包含 500 幅图像的图像库进行实验,结果表明,提出的方法具有较好的检索性能。

关键词 图像分割,贝叶斯理论,约束聚类,图像标注,图像检索

Image Annotation and Retrieval Based on Bayesian Theory

ZHANG Yuan-qing¹ BAO Jun-jie¹ KUANG Hang¹ ZHONG Hong²

(Chongqing Education College, Chongqing 400067, China)¹ (Information Engineering College, Central South University, Changsha 410075, China)²

Abstract Automatic semantic annotation of image is a crucial and highly challenging work in content-based image retrieval. A method of constrained clustering is proposed to cluster the segmented region, during the annotation stage, a greedy selection and joining method is used to find out the independent region cluster subsets. Then, a method of Bayesian is used to image annotation. After the image annotation, we can use the annotation keywords to image retrieval; the system has been implemented and tested on an image database of about 500 images. The experiment results show the effectiveness of the proposed approach.

Keywords Image segmentation, Bayesian theory, Constrained clustering, Image annotation, Image retrieval

1 引言

随着多媒体和网络技术的迅速发展,图像数据的来源不断扩大。为了能从大量的不同类型的图像数据中快速、准确地找到用户所需内容的图像,基于内容的图像检索技术正越来越成为人们研究的焦点^[1,2]。然而,图像检索的根本问题是让计算机基于语义来理解数据库中的图片^[3],例如一张包含人物的图片,计算机要把图片中的人物、位置以及其他物体用语言文字来表述出来。如果能做到这一步,那么现有的图像检索问题实际上就可以转化成技术已经相对成熟的文本检索的问题。图片自动语义标注在基于内容图像检索和计算机物体识别中是相当重要的,它的潜在应用领域包括生物学、商业、军事、教育、数字图书馆和网上检索等。

目前图像语义的表示方法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,但采用手工方法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,工作量太大。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。文献[4]中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。文献[5]则通过利用相关反馈技术构造一个语义模板,在视觉特征和语义概念之间建立关联,用于图像的标注。文献[6]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。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,也用相同的关键词标注,因此标注的精

度比较低。

本文提出用语义约束的聚类算法对图像区域进行聚类。在图像标注阶段,我们使用贝叶斯理论进行标注。在这种方法中,对于给定的一幅图像,计算语义概念的后验概率。由于聚类区域之间不是相互独立的,区域之间存在着依赖关系,我们提出贪心选择连接(GSJ)算法计算出聚类区域的独立子集。经过贪心选择连接算法后,使用贝叶斯理论进行语义标注。对图像进行标注之后,就能以关键字的方式进行检索。

2 图像分割

为了获取有意义的目标区域,我们必须对图像进行分割。目前关于基于图像分割获得目标区域进行图像检索的方法和系统的报道还比较少,原因在于目标区域的获取和特征提取都受到图像分割效果的影响。因为图像分割在理论上仍缺乏简单易行又准确可靠的通用方法,至今没有确切的数学定义能和人的感觉完全一致,从而现有的经典分割方法难以提取具有和人的感觉一致的目标区域。为了克服现有经典分割方法难以提取具有视觉一致性目标区域的问题,我们提出了一种基于颜色与空间的图像分割方法^[7],该方法可以较好地实现图像中有意义的目标区域的提取。图 1 是两幅图像采用这种方法进行分割的结果。

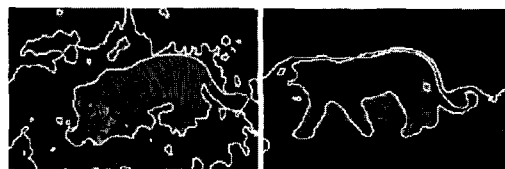


图 1 基于颜色与空间图像分割

3 基于语义约束的图像区域聚类

目前大多数的图像聚类方法都是基于视觉特征,因此不同语义概念但是视觉特征相似的区域可能分到同一类,从而影响聚类效果。解决这个问题的一个办法是在聚类过程中加入一些约束条件。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^[8]引入了 Must-link(相关)和 Cannot-link(不相关)两种约束。

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:

$$R_C(c_i, c_j) = \begin{cases} df(c_i \cap c_j) / df(c_i \cup c_j), & \text{if } df(c_i) > \sigma \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

其中 $df(c_i \cap c_j)$ 表示同时标注 c_i 和 c_j 的图像的数目, $df(c_i \cup c_j)$ 表示标注 c_i 或 c_j 的图像的数目。

我们对于已经标注的图像 x_p 和 x_q , 假设它们的标注分别为 C_p 和 C_q , 则标注 C_p 和 C_q 之间的相互关系定义如下:

$$\text{Rel}(C_p, C_q) = \underset{c_i \in C_p, c_j \in C_q}{\text{argmax}} (R_C(c_i, c_j)) \quad (2)$$

在我们的实验中,如果两个标注之间 $\text{Rel}(C_p, C_q)$ 小于一个预先给定的阈值,则 C_p 和 C_q 及它们的相应区域认为是无关的。如果 C_p 和 C_q 是无关的,则它们之间的所有区域都标记为不相关,表示如下:

$$\forall r_i \in x_p, \forall r_j \in x_q; s(r_i, r_j) = 1$$

其中 s 为区域 r_i 和 r_j 之间的关系函数,对于 cannot-link 为 1, 其它关系为 0。

设有区域集 $\chi = \{r_i\}_{i=1}^n$ 表示整个 n 个图像区域, $\{R_k\}_{k=1}^M$ 表示 M 个聚类区域, $\{\mu_k\}_{k=1}^M$ 表示对应的聚类中心, l_i 为区域 r_i 分配的聚类, $l_i \in \{1, \dots, M\}$ 。对于 k 均值算法来说,就是将每个区域分配到与其最近的聚类中,使得每个区域与其所在聚类的中心距离最小。我们使用约束聚类(CC)方法对 k 均值算法的目标函数进行改进,如下式所示:

$$J_{\text{ackmeans}} = \sum_{i=1}^n \|r_i - \mu_{l_i}\|^2 + \sum_{(r_i, r_j) | l(i) \neq l(j)} s(r_i, r_j) \times p \quad (3)$$

其中 r_j 表示聚类 R_{l_i} 中的区域, $\|g\|$ 表示欧式距离。聚类完成后,得到 M 个聚类 $\Gamma = \{R_1, R_2, \dots, R_M\}$ 。每一个区域都分类到一个聚类中,对于每个聚类区域,为了便于随后贝叶斯理论的处理,我们定义一个倒置的列表记录有区域被分配到这个聚类所有图像:

$$II(R_i) = \{x_j | \exists r \in x_j, l(r) = i\} \quad (4)$$

其中, $R_i \in \Gamma$, x_j 为一幅图像, r 为 x_j 分割得到的区域。

4 基于贝叶斯理论的图像标注

通过图像分割得到分割后的区域,然后利用语义约束的聚类算法进行聚类,接着我们计算聚类区域和语义概念之间的概率。语义标注的核心是通过对训练数据集进行学习,构建图像与语义关键字之间的标注模型,这里关于训练集的学习是一个统计的过程。

每一幅图像 x 属于一个没有标注的图像集 T , 定义 $\Omega = \{c_1, c_2, \dots, c_N\}$ 为图像标注的概念集合,我们把每一个概念 $c \in \Omega$ 当作一个类,那么 T 中的每一幅图像属于一个或者多个类。每一幅图像 x 对于 $c \in \Omega$ 的类条件概率密度函数为 $f_X(x|c)$ 。定义概率函数 $\{p(c_1), p(c_2), \dots, p(c_N)\}$ 表示概念的先

验知识。在约束聚类后,得到聚类区域的集合 $\Gamma = \{R_1, R_2, \dots, R_M\}$ 。

首先对于每一幅图像 $x \in T$ 被分割成区域 $\{r_1, r_2, \dots, r_m\}$, 然后,对于每一个区域 $r \in \{r_1, r_2, \dots, r_m\}$, 需要找出它的聚类区域。假设每一个区域仅对应一个聚类区域,因此每一幅没有标注的图像 x 可以用聚类区域的集合 $\{R_{K_1}, R_{K_2}, \dots, R_{K_m}\} \subseteq \Gamma$ 进行描述,即 $x = \{R_{K_1}, R_{K_2}, \dots, R_{K_m}\}$ 。根据贝叶斯理论可以推出对于图像 x 的每一个概念 $c \in \Omega$ 的后验概率。

$$p(c|x) \approx p(c|R_{K_1}, R_{K_2}, \dots, R_{K_m}) = \frac{f_\Gamma(R_{K_1}, R_{K_2}, \dots, R_{K_m} | c) p(c)}{f_\Gamma(R_{K_1}, R_{K_2}, \dots, R_{K_m})} \quad (5)$$

其中,分母为聚类区域的边缘概率密度函数,可以把它当作一个常量。因此图像标注的问题表示如下:给出一个聚类区域的集合 $\{R_{K_1}, R_{K_2}, \dots, R_{K_m}\}$, 将图像 x 分类到一个或多个概念类,最终的结论如下:

$$p(c|x) \propto p(c|(R_{K_1}, R_{K_2}, \dots, R_{K_m})) = \frac{f_\Gamma((R_{K_1}, R_{K_2}, \dots, R_{K_m}) | c) p(c)}{f_\Gamma(R_{K_1}, R_{K_2}, \dots, R_{K_m})} \quad (6)$$

4.1 计算聚类区域的独立子集

为了估计类条件密度函数 $f_\Gamma(R_{K_1}, R_{K_2}, \dots, R_{K_m} | c)$, 目前许多方法假设聚类区域之间是相互独立的,类的条件密度函数公式可以等价于

$$f_\Gamma(R_{K_1}, R_{K_2}, \dots, R_{K_m} | c) = \prod_{j=1}^{K_m} f_{\Gamma(j)}(R_j | c) \quad (7)$$

然而,事实上这个假设并不成立。解决这个问题的办法是找出聚类区域 $\{R_{K_1}, R_{K_2}, \dots, R_{K_m}\}$ 之间的依赖关系。在本文中我们提出贪心选择连接(GSJ)算法为 $\{R_{K_1}, R_{K_2}, \dots, R_{K_m}\}$, 计算聚类区域的独立子集。

任何两个聚类区域之间的依赖关系都可以由公式(8)计算得到,它说明两个聚类区域有多大的可能出现在同一幅图像中。本文使用聚类区域的倒置列表计算出区域之间的依赖大小。

$$R_\omega(R_i, R_j) = \frac{|II(R_i) \cap II(R_j)|}{|II(R_i) \cup II(R_j)|} \quad (8)$$

为了限制聚类区域之间依赖级别,减少计算代价,使用参数 t 控制聚类区域的个数,当 $t=1$ 时,就是朴素的贝叶斯模型, t 的值越大,模型越复杂,计算时间越多,这些模型从相似聚类区域中使用更详细的共生信息。贪心选择连接(GSJ)算法为贝叶斯模型找出聚类区域的独立子集。下面给出贪心选择连接算法(GSJ)的详细过程:

- 1 初始化:
 $B = \phi; S = 1$; 随机选择 $R_i \in H$,
 $K_1 \leq i \leq K_m, B_i = \{R_i\}; H = H / \{R_i\}$;
- 2 选择阶段
对于任意的 $R_g \in B_i$,
选择 $R_j = \text{argmax}_{R_k \in H} \sum_{R_g \in B_i} |R_\omega(R_g, R_k)|$,
其中 $|R_\omega(R_g, R_k)| > \epsilon, \epsilon$ 是一个预先定义的阈值;
- 3 连接阶段
If (R_j 存在 and $|B_i| < t$)
 $B_i = B_i \cup \{R_j\}; H = H / \{R_j\}$;
Go to 2 ;
Else if $H \neq \phi$
 $S = S + 1; B = B \cup \{B_i\}$;
Go to 1 ;
Else
Exit
End
End

基于提出的 GSJ 算法,区域聚类集合 H 被分解为相互独立集合 $B = \{B_1, B_2, \dots, B_l\}$, 其中 $\bigcup_{i=1}^l B_i = H$, 且对于任意的 $B_i, B_j \in B, B_i \cap B_j = \phi$. 因此:

$$f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m} | c) = f_B(B_1, B_2, \dots, B_l | c) = \prod_{i=1}^l f_{B(i)}(R_i | c) \quad (9)$$

4.2 图像标注

考虑到在数据库中对于概念 c 存在 n 幅训练图像, 所以我们用以下公式简单估计类的条件概率密度:

$$f_{B(i)}(R_i | c) \approx \frac{vol(B_i, c)}{vol(c)} \quad (10)$$

其中 $|B_i| \leq t, B_i \subseteq H, vol(c) = n, vol(B_i, c)$ 为包含聚类区域 B_i 和概念 c 的图像的个数。

对于给定的待标注的图像 x , 标注每个语义概念的后验概率计算如下:

$$\begin{aligned} p(c|x) &\approx p(c|R_{K_1}, R_{K_2}, \dots, R_{K_m}) \\ &= \frac{f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m} | c) p(c)}{f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m})} \\ &= \frac{f_B(B_1, B_2, \dots, B_l | c)}{f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m})} = \frac{\prod_{i=1}^l f_{B(i)}(R_i | c)}{f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m})} \\ &= \frac{\prod_{i=1}^l \frac{vol(B_i, c)}{vol(c)}}{f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m})} \quad (11) \end{aligned}$$

在公式(11)中, $f_{\Gamma}(R_{K_1}, R_{K_2}, \dots, R_{K_m})$ 为常数, 所以我们可以计算出每一个语义概念的后验概率 $p(c|x)$. 这样在我们标注一幅新的图像时, 选择后验概率 $p(c|x)$ 较大的语义概念进行标注。

5 实验结果分析

为了验证本文所述方法的效果, 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 幅图像作为训练样本, 剩下的 30 幅图像作为测试样本, 这样共有 200 个训练样本。对已标注的训练样本进行分割, 对分割得到的 1124 个区域采用提出的算法进行聚类, 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。在本文的贪心选择连接(GSJ)算法中, 参数 t 的选择是非常重要的。我们令 $t=1, 2, 3$ 分别进行实验。我们发现, 当 t 比较大时, 它捕获的聚类区域的共生信息也多。然而 t 越大, 计算代价也越高。在实验中我们发现 $t=2$ 和 $t=3$ 对实验结果的影响相差不是很大, 所以在本文中我们选择 $t=2$ 。图 2 是本文的方法的标注结果。

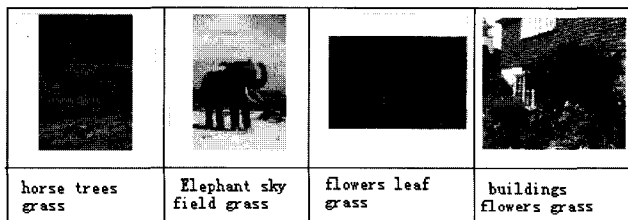


图 2 基于贝叶斯理论的图像标注(部分结果)

对检索算法的性能评价, 重要指标是查准率(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 测度值用公式表示如下:

$$F1 = \frac{2 * recall * precision}{recall + precision} \quad (12)$$

在图 3 中, 通过标注的关键字(horse)进行检索。数据库中相关图像是 50, 实验中检索输出图像数目为 12, 检索到的相关图像为 10, 所以查准率为 $10/12=83\%$, 查全率 $10/50=20\%$, F1 测度值 0.323。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, 分别以马(horse)、人(people)、花(flower)、大象(elephant)、建筑物(building)、山(mountain)进行检索, 并与 CMRM^[9] 的性能进行比较, 得到的平均 F1 测度值如表 1 所示。结果表明, 我们的方法性能更好。

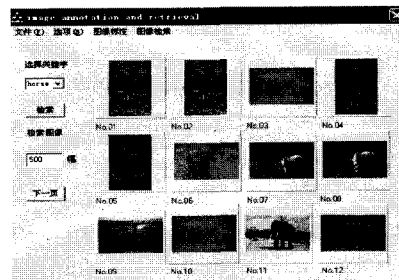


图 3 基于标注关键字的检索结果

表 1 提出的方法和 CMRM 方法的 F1 测度值对比

关键字	平均 F1 测度值		性能提高
	CMRM	Proposed	
horse	0.252	0.421	
people	0.242	0.401	
flower	0.254	0.387	
elephant	0.263	0.398	
building	0.226	0.375	
mountain	0.232	0.369	

结束语 本文提出用语义约束聚类算法对区域进行聚类。在图像标注阶段, 使用贝叶斯理论进行标注, 在这种方法中我们计算语义概念的后验概率。由于区域之间存在着依赖关系, 因此使用贪心选择连接(GSJ)算法找出聚类区域的独立子集。经过 GSJ 后, 使用贝叶斯理论进行语义标注, 然后以标注的关键字进行检索, 并与目前认为比较好的标注方法 CMRM^[9] 比较。实验结果表明, 提出的方法的检索性能更好。

参考文献

- [1] Swain M J, Ballard D H. Color indexing. International Journal of Computer Vision, 1991, 7(1), 11-32
- [2] Huang J. Image indexing using color correlograms // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1997; 762-768
- [3] Ma Xiaohang. Semantics Modeling based Image Retrieval System Using Neural Networks[C] // IEEE Conf. Image Processing. 2005(1), 1165-1168
- [4] Wang lei, Liu Li, Latifu. Automatic image annotation and retrieval using subspace clustering algorithm[C] // Proceedings of the 2nd ACM International Workshop on Multimedia Databases. 2004; 263-274
- [5] Yeung M M, Boon-Lock Y, Bouman C A. Apply semantic template to support content-based image retrieval[C] // Proceedings of SPIE Storage and Retrieval for Media Databases. 2000; 442-449
- [6] Li Wei, Sun Mao-song. Automatic image annotation based on WordNet and hierarchical ensembles[C] // LNCS 3878; CICLing 2006. Berlin Heidelberg; Springer-Verlag, 2006; 417-428
- [7] 钟洪. 基于本体的图像检索[J]. 计算机工程与应用, 2007, 43(17); 37-40
- [8] Wagstaff K, Cardie C, Rogers S, et al. Constrained K-Means clustering with background knowledge // Proc. of 18th Intl. Conf. on Machine learning(ICML-2001). 2001
- [9] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models[C] // ACM SIGIR'03. July-Aug. 2003