

# 基于 t 混合模型的自动图像标注

江 欢<sup>1</sup>, 汤 进<sup>1</sup>, 郭玉堂<sup>1,2</sup>, 罗 斌<sup>1</sup>

(1. 安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039;

2. 合肥师范学院 计算机科学与技术系, 安徽 合肥 230061)

**摘 要:** 图像自动标注的实质是通过对图像视觉特征的分析来提取高层语义关键词用于表示图像的含义, 从而使得现有图像检索问题转化为技术已经相当成熟的文本检索问题, 在一定程度上解决了基于内容图像检索中存在的语义鸿沟问题。采用 t 混合模型在已标注好的训练图像集上计算图像区域类与关键字的联合概率分布, 在此基础上, 对未曾观察过的测试图像集, 利用生成的模型根据贝叶斯最小错误概率准则实现自动图像标注。实验结果表明, 该方法能有效改善标注结果。

**关键词:** t 混合模型; 图像标注; 贝叶斯

中图分类号: TP277

文献标识码: A

文章编号: 1673-629X(2010)10-0086-04

## Automatic Image Annotation Based on t Mixture Model

JIANG Huan<sup>1</sup>, TANG Jin<sup>1</sup>, GUO Yu-tang<sup>1,2</sup>, LUO Bin<sup>1</sup>

(1. Ministry of Education Key Lab. of Intelligent Computing and

Signal Processing, Anhui University, Hefei 230039, China;

2. Department of Computer Science and Technology, Hefei Normal College, Hefei 230061, China)

**Abstract:** The essence of automatic image annotation is to obtain the semantic key words of images from visual features and to support the semantic level search, then image retrieval can be transformed into text retrieval, which is fairly mature. To a certain extent, it can solve the semantic gap existing in content-based image retrieval. This paper was based on t mixture model, and computed a joint probability distribution for image regions classes and keywords. On this basis, it annotated unseen set of test images by used model according to Bayes minimum error probability criterion. The experiments results show that this method can significantly improve the labeling results.

**Key words:** t mixture model; image annotation; Bayes

## 0 引 言

随着现代图像数据的急剧增长, 由于人工标注方式的局限性, 基于自动图像标注技术近年来备受关注而成为重点研究领域之一。当图像被标注后, 图像检索实际上就可以转化成技术已相当成熟的文本信息检索<sup>[1]</sup>, 从而比较容易快速高效地实现。

近年来, 围绕如何对图像自动标注, 已经吸引了计算机视觉、图像处理以及统计学习等领域几十年的研究成果, 各种图像标注算法应用而生。如文献[2]中对

每一个词汇训练对应的二元分类器, 当遇到待标注图像时, 通过分类器来寻找词汇对图像进行标注。

目前, 通过建立图像和语义概念的统计概率模型已成为自动图像标注的重点研究方向<sup>[3]</sup>。基于统计概率模型的图像标注的基本思想是, 将图像分割成众多子块, 通过模型训练建立各子块与标注词间的统计关联关系, 对待标注图像, 利用建立的统计关联模型实现对图像标注。文献[4]提出了跨媒体相关模型(CM-RM), 该模型假设一张图像可以由众多子块(blob)聚类组合, 它使用已标注的图像训练集, 采用相关模型(Relevance Model)分析了子块和关键字间的联合概率分布。其核心思想在于建立图像内容和标注词之间的概率相关性, 找到一组语义使之与图像内容之间的条件概率最大, 这组语义就是最后的标注结果。文献[5]提出了连续相关模型(CRM), 该模型针对 CMRM 模型中以 blob 来表示图像会损失掉一些有用的图像信

收稿日期: 2010-02-10; 修回日期: 2010-05-21

基金项目: 国家自然科学基金(60772122); 安徽省自然科学基金重点项目(KJ2008A033, KJ2009A1501)

作者简介: 江 欢(1985-), 男, 安徽安庆人, 硕士研究生, 研究方向为模式识别与图像处理; 罗 斌, 教授, 博导, 研究方向为模式识别与图像处理。

息这一问题,直接利用图像中各分割区域的连续特征向量的组合来表示该幅图像,然后通过区域间的关系联合来具体估计概率。

文中构建基于感兴趣区域的自动图像标注,首先将已标注图像集分割为图像区域,提取各区域的颜色、纹理及形状作为综合特征,采用 t 混合模型进行语义聚类,建立图像特征与语义词汇之间的概率关系。当有待标注图像时,同样进行区域分割后提取各区域综合特征,利用生成的模型计算区域特征与各词汇间的概率分布,根据贝叶斯最小错误率准则,取概率最高的若干词汇作为图像标注结果。

## 1 多维 t 混合模型及其 EM 算法

设  $X = \{X_1, \dots, X_n\}$  为随机观测数据集,其中  $X_i = [X_{i1}, \dots, X_{ip}]^T \in R^p$ , 且  $X_i$  间相互独立。如果每个  $X_i$  的概率密度函数表示如下:

$$f(x_i, \Theta) = \sum_{j=1}^g \pi_j \Psi_j(x_i; \theta_j) \quad (1)$$

则说  $X$  服从有限混合分布,其对应的模型就为有限混合模型<sup>[6]</sup>。

其中,  $g$  为混合分量个数;  $\pi_j$  称为混合参数,即第  $j$  个成分的先验概率,并且满足如下等式:  $\pi_j \geq 0$ ,  $\sum_{j=1}^g \pi_j = 1$ ;  $\Psi_j(x_i; \theta_j)$  是给定  $X_i$  来自第  $j$  个分量的类条件概率密度;  $\theta_j$  是第  $j$  个成分的分布参数;  $\Theta = (\pi_1, \dots, \pi_g, \theta_1, \dots, \theta_g)$  是所有参数的集合。

### 1.1 t 分布的混合模型

式(1)中,假如分量密度  $\Psi_j(x_i; \theta_j)$  满足 t 分布,则所对应的模型称为 t 混合模型。因为 t 分布比高斯分布具有较重的尾部,它提供了一个比高斯分布更鲁棒的方法,自然地人们倾向于用 t 分布实现数据聚类。

$p$ -维 t 分布的概率密度的形式为:

$$\Psi(x; \theta) = t_p(x; \mu, \Sigma, \nu) = \frac{\Gamma(s) | \Sigma |^{-1/2}}{(\pi \nu)^{p/2} \Gamma(\frac{\nu}{2})} \left( 1 + \frac{\delta(x; \mu, \Sigma)}{\nu} \right)^{-(\nu+p)/2} \quad (2)$$

其中:  $\delta(x; \mu, \Sigma) = (x - \mu)^T \Sigma^{-1} (x - \mu)$ ;  $\theta = (\mu, \Sigma, \nu)^T$ ,  $\mu$  为均值向量,  $\Sigma$  为方差矩阵,  $\nu$  为自由度;  $\Gamma(s)$  为 Gamma 函数, 定义为  $\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} dx, (s > 0)$ 。现在考虑  $g$  个分量的  $p$  维 t 混合模型,每个分量的类条件概率密度函数都为  $p$  维 t 分布密度函数,则式(1)就变为:

$$f(x_i, \Theta) = \sum_{j=1}^g \pi_j \Psi_j(x_i; \theta_j) = \sum_{j=1}^g \pi_j t_p(x_i; \mu_j, \Sigma_j, \nu_j)$$

$$\sum_j, \nu_j) \quad (3)$$

此时,  $p$  维 t 混合模型的参数  $\theta_j$  实际上由混合参数  $\pi_j$ 、均值向量  $\mu_j$ 、方差矩阵  $\Sigma_j$ 、自由度  $\nu_j$  所决定。

通过对样本建立极大似然估计,采用 EM 算法可获得极大似然估计的混合参数。引入  $x_i$  分量的标记向量  $Z_i = (z_{i1}, z_{i2}, \dots, z_{ig})^T, (z_{ij} \in [0, 1])$ ,  $z_{ij}$  表示  $x_i$  来自于第  $j$  个混合分量的概率,此时,完全数据向量集为  $X^c = (X^T, Z_1^T, Z_2^T, \dots, Z_n^T)^T$ , 其中  $X = [X_1^T, \dots, X_n^T]^T$ 。相应的完全对数似然函数可以写为:

$$L(\Theta | X^c) = \sum_{i=1}^n \sum_{j=1}^g z_{ij} (\ln \Psi_j(X_i; \theta_j) + \ln \pi_j) \quad (4)$$

### 1.2 利用 EM 算法进行参数估计

EM 算法是一种针对概率模型的迭代优化技术,通过多步迭代,使得似然函数收敛于某个最优解。对于 t 混合模型,进行 EM 算法参数估计的过程为<sup>[7]</sup>:

1) 首先初始化参数  $\Theta$ 。

2) E 步:  $Q(\Theta, \Theta^{(t)}) = E[L(\Theta | X^c) | X, \Theta^{(t)}]$  (5)

3) M 步: 最大化式(5), 得到新的参数值  $\Theta$ 。即:

$$\Theta^{(t+1)} = \arg \max_{\Theta} Q(\Theta, \Theta^{(t)}) \quad (6)$$

4) 利用式(5)、式(6) 循环迭代直至参数收敛,得到参数  $\Theta$  的估计值。

## 2 t 混合模型的自动图像标注

### 2.1 综合特征提取

特征提取是基于区域的自动图像标注的前提。由于颜色、纹理和形状特征<sup>[8]</sup>是图像中最重要的内容信息,因此文中选取了颜色、纹理、形状作为图像语义提取的底层特征,在此基础上构建综合特征矢量作为区域特征。

1) 颜色特征。

RGB 颜色空间是一种常见的颜色表示法<sup>[9]</sup>,但是 RGB 颜色空间与人眼的感知差异很大。因此文中采用较能符合人眼感知特性的 HSV 颜色空间。下面简要介绍如何由 RGB 转化为 HSV:

给定 RGB 颜色空间的值  $(r, g, b)$ ,  $r, g, b \in [0, 1, \dots, 255]$ , 则转换到 HSV 空间的  $h, s, v$  值计算如下:

定义  $r^*, g^*, b^*$  为:

$$r^* = \frac{\max(r, g, b) - r}{\max(r, g, b) - \min(r, g, b)},$$

$$g^* = \frac{\max(r, g, b) - g}{\max(r, g, b) - \min(r, g, b)},$$

$$b^* = \frac{\max(r, g, b) - b}{\max(r, g, b) - \min(r, g, b)}$$

$$\text{则 } v = \frac{\max(r, g, b)}{255}, s = 1 - \frac{\min(r, g, b)}{\max(r, g, b)}$$

$$h' =$$

$$\begin{cases} 5 + b^*, r = \max(r, g, b) \text{ 且 } g = \min(r, g, b) \\ 1 - g^*, r = \max(r, g, b) \text{ 且 } g \neq \min(r, g, b) \\ 1 + r^*, g = \max(r, g, b) \text{ 且 } b = \min(r, g, b) \\ 3 - b^*, g = \max(r, g, b) \text{ 且 } b \neq \min(r, g, b) \\ 3 + g^*, b = \max(r, g, b) \text{ 且 } r = \min(r, g, b) \\ 5 - g^*, \text{其它} \end{cases} \quad (7)$$

$$h = 60 * h'$$

在这里,  $h \in [0, 360], s \in [0, 1], v \in [0, 1]$ 。

根据色彩的不同范围,对 HSV 进行量化<sup>[10]</sup>,量化后的色调(H)、饱和度(S)、亮度(V)表示为:

$$H = \begin{cases} 0, h \in [0, 25] \\ 1, h \in [26, 41] \\ 2, h \in [42, 75] \\ 3, h \in [76, 156] \\ 4, h \in [157, 201] \\ 5, h \in [202, 272] \\ 6, h \in [273, 285] \\ 7, h \in [286, 330] \end{cases}$$

$$S = \begin{cases} 0, s \in [0, 0.2] \\ 1, s \in [0.2, 0.7] \\ 2, s \in [0.7, 1] \end{cases}, V = \begin{cases} 0, v \in [0, 0.2] \\ 1, v \in [0.2, 0.7] \\ 2, v \in [0.7, 1] \end{cases} \quad (8)$$

## 2) 纹理特征。

采用基于统计分析的灰度共生矩阵来描述图像的纹理特征<sup>[11]</sup>。从灰度共生矩阵中,可以提取 14 种特征来表征图像的纹理,但考虑到计算量问题,在这里只提取其中的 4 种特征,即反差 CON、能量 ASM、熵 ENT、相关 COR 来分析。

构造  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  4 个方向上的灰度共生矩阵,分别计算每个共生矩阵上的上述 4 个纹理特征向量,组成 16 维的特征向量。

## 3) 形状特征。

形状特征是刻画图像中物体和区域特点的重要手段,是描述高层视觉特征的重要手段。由于矩表征图像形状时具有平移、缩放与旋转不变等特点,从而广泛使用于描述形状特性。在此,采用七个不变矩来描述图像的形状特征<sup>[12]</sup>。

## 2.2 标注模型训练

结合上面的介绍,构建基于 t 混合模型的标注模型训练算法如下:

1) 设集合  $I = \{I_1, I_2, \dots, I_n\}$  表示已标注的图像,集合  $K = \{K_1, K_2, \dots, K_g\}$  表示标注所使用的关键字。利用 Normalized Cut 算法<sup>[13]</sup>将  $I_i$  分解成若干区域,

分割后的区域用集合  $R = \{R_1, R_2, \dots, R_p\}$  表示。利用人工标注的方式,将区域  $R_i$  与关键字  $K_j$  建立映射。

2) 抽取区域  $R_i$  的底层特征(包括区域的颜色、纹理、形状等),用特征向量  $X_i$  表示,  $X_i$  为 26 维的特征向量。根据式(1),构建如下的 t 混合模型:

$$f(x_i, \Theta) = \sum_{j=1}^g \pi_j \Psi_j(x_i; \theta_j) = \sum_{j=1}^g \pi_j t_p(x_i; \mu_j, \sum_j, v_j) \quad (9)$$

其中混合分量数为关键字个数  $g$ 。通过极大似然估计来建立对应关键字的类条件概率密度函数。

引入  $X_i$  分量的标记向量  $Z_i = (z_{i1}, z_{i2}, \dots, z_{ig})^T$ , ( $z_{ij} \in [0, 1]$ ), 其中  $z_{ij}$  代表随机样本  $X_i$  与关键字  $K_j$  的关联程度,初始值全为 0。则完全数据向量集为  $X^c = (X^T, Z_1^T, Z_2^T, \dots, Z_g^T)^T$ , 其中  $X = [X_1^T, \dots, X_p^T]^T$ , 相应的完全对数似然函数可以写为:

$$L(\Theta | X^c) = \sum_{i=1}^p \sum_{j=1}^g z_{ij} (\ln \Psi_j(X_i; \theta_j) + \ln \pi_j) \quad (10)$$

利用 EM 算法,进行参数估计,得到参数  $\Theta$  以及每个区域属于各关键字的后验概率  $z_{ij}$ , 根据后验概率找到对应关键字的类概率密度函数。

由式(5)可得:

$$z_{ij}^{(t)} = \frac{\pi_j^{(t)} \Psi_j(x_i; (\mu_j^{(t)}, \sum_j^{(t)}, v_j^{(t)})^T)}{\sum_{l=1}^g \pi_l^{(t)} \Psi_l(x_i; (\mu_l^{(t)}, \sum_l^{(t)}, v_l^{(t)})^T)} \quad (11)$$

以及  $W_i$  的条件期望为:

$$w_{ij}^{(t)} = E[W_i | x_i, z_{ij} = 1] = \frac{v_j^{(t)} + p}{v_j^{(t)} + \delta(x_i; \mu_j^{(t)}, \sum_j^{(t)})} \quad (12)$$

由式(6)可得:

$$\text{参数 } \pi_j^{(t+1)} = \frac{1}{p} \sum_{i=1}^p z_{ij}^{(t)} \quad (13)$$

$$\text{均值 } \mu_j^{(t+1)} = \frac{\sum_{i=1}^p z_{ij}^{(t)} w_{ij}^{(t)} x_i}{\sum_{i=1}^p z_{ij}^{(t)} w_{ij}^{(t)}} \quad (14)$$

方差  $\sum_j^{(t+1)} =$

$$\frac{\sum_{i=1}^p z_{ij}^{(t)} w_{ij}^{(t)} (x_i - \mu_j^{(t+1)}) (x_i - \mu_j^{(t+1)})^T}{\sum_{i=1}^p z_{ij}^{(t)}} \quad (15)$$

自由度  $v_j$  第  $(t+1)$  次迭代后的估计  $v_j^{(t+1)}$  是非线性方程

$$-\Psi\left(\frac{v_j}{2}\right) + \ln\left(\frac{v_j}{2}\right) + 1 + \frac{1}{n_j^{(t)}} \sum_{i=1}^p z_{ij}^{(t)} (\ln w_{ij}^{(t)} - w_{ij}^{(t)}) + \Psi\left(\frac{v_j^{(t+1)} + p}{2}\right) - \ln\left(\frac{v_j^{(t+1)} + p}{2}\right) = 0 \quad (16)$$

的解,其中  $n_j^{(t)} = \sum_{i=1}^p z_{ij}^{(t)}$ ,  $\Psi(t) = \frac{d \ln \Gamma(t)}{dt}$ 。

循环执行式(11)~式(16)直到参数收敛为止,最终得到参数  $\Theta$  及标记向量  $Z_{ij}$  的值,从而得到对应关键

字的类概率密度函数。

2.3 判别规则

对待标注图像利用 Normalized Cut 算法<sup>[13]</sup>进行图像分割,提取各区域底层特征  $X_i$ ,其中  $X_i = [x_{i1}, \cdots, x_{i26}]$ 。通过以下两步来实现自适应图像标注。

- 1) 计算  $X_i$  与所有关键字间的概率分布,即  $Y_{ij} = \pi_{j|p}(X_i; \mu_j, \sum_j v_j), j = 1, \cdots, g$ 。
- 2) 根据贝叶斯最小错误率准则对图像进行标注。即如果  $Y_{ik} = \max_{j=1, \cdots, g} (Y_{ij})$ ,则判断  $X_i$  属于  $Y_{ij}$  即第  $k$  类,用对应关键字标注。

3 实验结果与分析

实验数据集采用通用的 Corel 数据库。图像数据集由 5000 幅图像组成,其中 4500 幅图像作为训练集,500 幅图像作为测试集。每幅图像分割成 1~10 个区域,共形成 42000 多个图像区域。每幅图像标注 4~5 个关键字。图像数据集中一共有 374 个标注关键字。




实验中使用平均查准率、平均查全率作为图像标注的性能指标。表 1 给出了文中算法与 TM、CMRM 算法在标注性能上的比较。表 2 是一些标注结果比较。

实验表明,该方法能有效改善标注结果。

表 1 自动图像标注性能比较

方法	平均查准率	平均查全率
TM	0.08	0.06
CMRM	0.12	0.11
文中方法	0.15	0.13

表 2 图像标注结果比较

图像			
标注真值	grass, bears water	grass, tiger water	buildings, locomotive, smoke
TM	sheep, sky, water pig, plane	grass, cat, sun, water, flower	train,sky,snow, sheep, bridge
CMRM	water, pig, snow, flower, cat	grass,cat, people, flower,panda	car,building, snow mountain,smoke
文中算法	sun, water, grass , cat, sheep	people,water, flower, cat, grass	train,sky,rocks smoke, snow

4 结束语

文中采用 t 分布聚类方法,建立图像视觉特征与

关键字间的混合模型,利用 EM 算法求解该模型。在此基础上,对待标注图像,进行图像分割提取区域特征,利用该模型建立各区域与关键字间的概率分布,根据贝叶斯最小错误概率准则实现图像标注。实验结果表明,该方法能有效改善标注结果。

参考文献:

[1] 向友君,谢胜利. 图像检索技术综述[J]. 重庆邮电学院学报:自然科学版,2006,16(3):348-354.

[2] Cusano C,Ciocca G,Schettini R. Image annotation using SVM [J]. Proceeding of SPIE- IS & T Electronic Imaging,2004, 5304:330-338.

[3] 朱文球,刘 强. 一种新的自动图像语义标注与检索算法 [J]. 计算机应用研究,2007,24(7):318-320.

[4] Jeon J,Lavrenko V,Mnmatha R. Automatic image annotation and retrieval using cross- media relevance models[C]//Proc. of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Toronto: [s. n. ],2003:119-126.

[5] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures [C]//Proc. of the 17th International Conference on Neural Information Processing Systems. Vancouver:[s. n. ],2003:553-560.

[6] McLachlan G J, Peel D. Finite Mixture Models[M]. New York:Wiley,2000.

[7] McLachlan G J, Krishna T. The EM Algorithm Extensions [M]. New York: Wiley, 1977.

[8] 黄祥林,沈兰荪. 基于内容的图像检索技术研究[J]. 电子学报,2002,30(7):1065-1071.

[9] 彭群生. 计算机真实感图形的算法基础[M]. 北京: 科学出版社,1999:428-448.

[10] 田玉敏,林 高. 全基于颜色特征的彩色图像检索方法[J]. 西安电子科技大学学报,2002(2):43-46.

[11] 曹莉华,柳 伟,李国辉. 基于多种主色调的图像检索算法研究与实现[J]. 计算机研究与发展,1999,36(1):96-100.

[12] Hu M K. Visual pattern recognition by moment invariants[J]. IEEE Trans. on Information Theory, 1962, IT-8: 179-187.

[13] Shi J B, Malik J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,22(8):888-905.