

基于内容检索的图像自动标注方法研究

邓莉琼 郝向宁 夏 鸣 李中宁

(空军大连通信士官学校 大连 116600)

摘 要 图像标注技术是近年来的研究热点。为了更好地解决图像自动标注问题,提出了一个基于检索和重排序的标注方法。在检索阶段,通过使用基于 MSF 的全局特征对待标注图像进行基于内容的检索,从而在图像数据库中得到一系列的相似图像数据集;在重排序阶段,利用随机漫步方法对相似图像数据集的标注信息进行重排序,最终排序后的关键词则为待标注图像的标注信息。该方法不仅跳过了漫长的训练阶段,而且充分利用了网络上那些已经具有标注信息的图像,具有较好的稳定性和可扩展性。实验结果显示了该方法的有效性。

关键词 图像标注,相似度检索,重排序,马尔科夫,基于内容

中图法分类号 TP391 **文献标识码** A

Image Annotation by Similarity Content-based Image Retrieval

DENG Li-qiong HAO Xiang-ning XIA Ming LI Zhong-ning

(Dalian Airforce Communication NCO Academy, Dalian 116600, China)

Abstract Image annotation is an active research topic in recent years. In this paper, we targeted at solving the automatic image annotation problem in a novel search and refinement framework. In the search stage, we performed content-based image retrieval(CBIR) based on the MSF global feature to find similar images from image database. Then in the refinement stage, an algorithm using random walk with restarts is used to re-rank the annotations. The refinement keywords are used to annotate the uncaptioned image. This framework does not impose an training stage, but efficiently utilizes well-annotated images, and is potentially capable of dealing with unlimited vocabulary. The experiment results show the effectiveness and efficiency of the proposed approach.

Keywords Image annotation, Similarity search, Refinement, Markov, Content-based

1 引言

随着网络和数字媒体技术的发展,网络上存在着海量的呈指数级增长的图像数据,例如谷歌图片中就有多达十亿的图像。图像数据的爆炸性增长极大地促进了图像标注技术的发展,高准确度的图像标注方法能够帮助用户快速检索到所需的图像素材。然而目前的图像标注技术通常是建立在分类训练的基础上的,随着图像数据量的增长,这种方法的缺点日益突显出来,一方面是数量有限的训练数据集常常无法识别未知的图像,另一方面则是训练的时间较长严重制约着其实用性,得到的错误标注结果及较慢的标注速度无法使用户满意。

近年来,网络检索技术极大地推广了图像标注技术的应用研究,有不少研究提出了基于网络检索的图像标注技术,其主要是借助网络上的海量图像素材对未知的图像进行标注,这种方法充分利用了网络上已经具有大量语义标注信息的图像数据库,并且这个数据库是在不断地发展和更新当中,相较于传统的基于分类的图像标注方法,这种方法具有更大的可扩展性和稳定性,因此受到了普遍认同。在这一技术当中,

有两个问题一直困扰着研究者们,一个是图像的快速匹配问题,另一个则是对获得的语义标注信息重排序问题。

针对这两个问题,本文提出了一种基于图像内容检索的图像自动标注方法,即首先进行基于内容的相似检索,然后对检索得到的相似图像数据集的标注信息进行重排序,挖掘出合适的标注关键词。

2 相关研究工作

目前图像进行自动标注的方法主要可以分为基于分类的自动标注和基于概率模型的自动标注两类方法。基于分类的自动标注方法是利用支持向量机等机器学习技术对图像素材进行训练分类,从而判定待标注图像类别,这样,待标注图像的标注信息就是该图像所属的类别信息,这一类方法包括基于贝叶斯点机器(Bayes Point Machine, BPM)^[1]、支持向量机(Support Vector Machine, SVM)^[2]以及基于视觉特征分布块的关键词^[3]等标注方法。然而截止到目前,基于分类的自动标注技术大都存在速度过慢的问题,并且其分类的准确度取决于训练图像数据库的大小,如果出现一张在训练图像数据库中没有的图像,那么就无法对这张图像进行准确的分类,

本文受国家自然科学基金(61201339)资助。

邓莉琼(1986—),女,博士,讲师,主要研究方向为多媒体情报分析技术, E-mail: tigerss1016@163.com; 郝向宁(1983—),男,硕士,助教,主要研究方向为数据挖掘; 夏 鸣 男,硕士,讲师,主要研究方向为数据分析; 李中宁(1983—),男,硕士,讲师,主要研究方向为数据处理技术。

故而实用性不高。基于概率模型的图像标注方法是通过判断图像和标注信息共同出现的联合概率大小来选择具有最大概率的标注信息作为标注结果。具有代表性的工作有多级概率混合模型(Hierarchical Probabilistic Mixture Model, HPMM)^[4]、翻译模型(Translation Model, TM)^[5]以及跨媒体关联模型(Cross-Media Relevance Model, CMRM)^[6]等等。这种图像标注方法大大受限于已有标注信息,若以往的图像标注信息存在大量错误的标注信息,例如网上图片的标注信息常常有很多与内容无关的信息,那么通过基于概率的图像标注方法则非常容易出现错误的标注结果。

针对以上标注方法的缺点,本文所采用的基于检索的自动标注方法是建立在基于内容相似的图像检索基础上的,该方法的主要思想是:首先,以待标注图像作为目标图像,在网络图像数据库中进行相似检索,进而得到与目标图像最相似的一组图像数据集,再对这些相似图像数据集的已有标注信息进行提取并重排序。由于网络图像上容易存在大量错误的标注信息,因此本文采用了基于随机漫步的重排序方法对其进行筛选,从而得到最终的标注结果。这种标注方法的优点是数据量越大,则标注结果越准确。作为网络媒体中的主要传播对象,网络上的图像往往具有丰富的标注信息,这些已有的标注信息给基于检索的图像自动标注方法提供了良好的前提基础。

3 基于内容检索的图像标注

3.1 标注流程

本文提出的基于内容检索的图像自动标注方法的主要流程如图1所示。

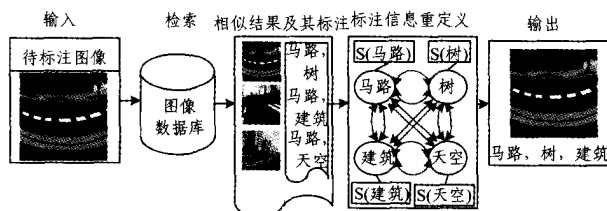


图1 基于内容检索的图像标注流程图

该标注流程主要包含两个关键的步骤。步骤一是从已具有标注信息的参考图像数据库中搜索与待标注图像相似的图像素材;步骤二是从这些相似的参考图像素材的标注信息中挖掘出正确且最具有代表性的关键词作为标注结果。该标注流程有个前提,即已有参考图像的标注信息均已分解为关键词的并联形式,即 $keyword_{i1} \cup \dots \cup keyword_{in}$ 的组成形式。

3.2 基于MSF的颜色全局特征相似匹配

现在常用的图像特征有全局特征和局部特征两种。由于目前的局部特征匹配技术速度较慢,无法适用于大尺度的图像数据库,因此本文提出了一种基于马尔可夫平稳特征(Markov Stationary Features, MSF)的图像颜色全局特征的快速图像匹配方法。作为最基本的图像全局视觉特征,颜色直方图方法是基于内容的图像检索技术中最常用的特征表示方法之一。针对传统的颜色直方图,本文提出了一个可同时描述图像颜色内部组成结构和外部关联结构的全局颜色特征:马尔可夫平稳特征。该特征是利用马尔可夫链的方法从颜色共生矩阵中抽取出的一个稳定特征,是对颜色直方图特征的进一步扩展,它不仅描述了图像的颜色组成结构,还描述

了图像颜色所处的位置及颜色之间的关联关系。下面将详细介绍该方法的具体实现步骤。

假设将颜色直方图中每种颜色当作马尔可夫链中的一个状态,则共生程度可以理解颜色之间的转换概率。这样一来,颜色直方图的比较就可以转换为相应地马尔可夫链的比较。若在二维图像空间中,像素所处的空间位置越近,则可以认为该像素之间的相关性就越强,相应地,在其马尔可夫链模型中也具有同样的特性,对于图像的状态矩阵,如果状态所处的空间位置越近,则该状态之间的转换概率就越高。因此,使用马尔可夫链模型来描述图像的颜色状态空间是具有合理性的。基于MSF的图像全局特征的主要计算步骤可以归纳如下:

步骤1 将图像视觉输入量化为 K 层的颜色直方图;

步骤2 给定一个距离 d , 计算其空间共生矩阵 $C = (c_{ij})_{K \times K}$;

步骤3 计算共生矩阵 $C = (c_{ij})_{K \times K}$ 的马尔可夫转移矩阵 $P = (p_{ij})_{K \times K}$;

步骤4 根据式(1)和式(2)计算其平稳分布 π ;

步骤5 将自转移矩阵进行标准化,作为初始分布 $\pi(0)$,

两者融合作为完整的MSF颜色特征 $\vec{h}_{MSF} = [\pi(0), \pi]^T$ 。

$$A_n = \frac{1}{n+1} (I + P + P^2 + \dots + P^n) \quad (1)$$

$$\pi \approx \frac{1}{K} \sum_{i=1}^K \vec{a}_i, \text{ 其中 } A_n = [\vec{a}_1, \dots, \vec{a}_K]^T \quad (2)$$

图2展示了一个MSF颜色特征的计算流程,图中第一列为包含3种颜色A,B,C的两幅图像,图中第二列为其生成的颜色空间共生矩阵,图中第三列为基于两幅图像颜色共生矩阵计算得到的马尔可夫转移矩阵,第四列是计算得到的MSF颜色特征。

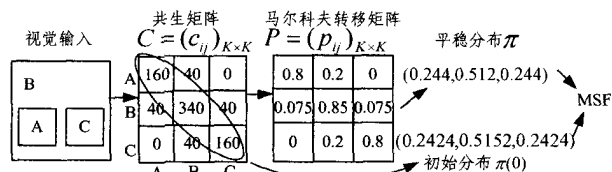


图2 图像MSF特征的生成过程

从上述计算步骤中可以看出,MSF特征的维数仅为 $2 \times K$,既避免了过大的维数,又有着鲁棒的效果。MSF特征既保留了图像的颜色组成信息,又包含了图像的空间位置关系,并且由于MSF来自于平稳分布的计算,因此该特征对噪声是鲁棒的,在该方法中, K 值的取值会影响到整个特征计算的速度,本文通过实验发现 K 最适宜的取值范围为 $25 \sim 50$ 。上述步骤中的相关参数在本文中的取值分别为 $K=36, d=1, n=50$ 。衡量两幅图像在MSF全局特征上的相似度,本文采用地球移动距离(The Earth Mover's Distance, EMD)^[7]来计算两者之间的相似度。EMD距离定义为从一个特征变换为另一个特征的最小消耗。消耗的概念是基于用户定义的基准距离,即两个特征之间的距离。这里取基准距离为欧式距离。EMD距离不仅计算了特征分量之间的相似度,还考虑了特征向量中各个维度之间相对应的相似度,即同时考虑了全局与局部的相似度。

3.3 基于随机漫步的标注信息重排序

针对检索得到的相似图像的标注信息,本文提出了基于

随机漫步的图像标注重排序的方法。该方法是利用随机漫步算法对候选的标注信息进行重排序,在此过程中不仅定义了共生相似度,还对原始的标注信息集合进行了基于可信度的排序。下面重点介绍该算法的实现步骤。

随机漫步算法(Random Walk with Restarts, RWR)最早出现于文献[8],它是一种不规则的变动形式,在这种变动过程中,每一步的变化都纯粹是随机性的,就如同一个人酒后乱步,其向前走一步的概率和向后退一步的概率相同,这种随机变动经过 N 步以后将达到一个稳定状态。随机漫步过程可用图 3 的节点转移图来表示。在图 3 中,随机漫步中的邻接矩阵反映的是节点之间是否有到达关系,即视频片段之间是否存在相似性;转移矩阵对应的是节点 i 转移到节点 j 的转移概率,转移概率对应视频片段之间的相似度。

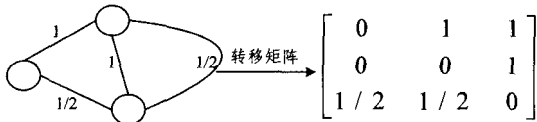


图 3 随机漫步的邻接矩阵与转移矩阵示意图

利用随机漫步原理对标注信息进行重排序的实现步骤:首先,建立标注信息的节点图,计算每个节点的权重值。假设在基于相似度检索的步骤中得到 m 个与待标注图像相似的匹配结果,对这 m 个图像的标注信息进行合并得到一组标注集合 W ,该标注集合是由 n 个不同的关键词 w_i 所组成的并集。以每个关键词作为一个节点,依据式(3)计算每个关键词节点的权重值 $weight(w_i)$:

$$weight(w_i) = \frac{num'(w_i)}{|W|} \quad (3)$$

式中, $num'(w_i)$ 为关键词 w_i 在集合 W 中出现的次数, $|W|$ 为集合 W 中关键词的个数。

第二步需要建立节点之间的连线,并计算每条连线的权重, $weight(w_i, w_j)$ 即表示由节点 w_i 和 w_j 形成的连线的权重值,计算方法见式(4):

$$weight(w_i, w_j) = \begin{cases} \frac{num(w_i, w_j)}{\min(num(w_i), num(w_j))}, & num(w_i, w_j) > 0 \\ 0, & num(w_i, w_j) = 0 \end{cases} \quad (4)$$

式中, $num(w_i, w_j)$ 表示匹配的相似图像中同时具有关键词标签 w_i 和 w_j 的图像个数, $num(w_i)$ 为标注信息里具有关键词 w_i 的图像个数。由此得到由 n 个节点组成的图 $G = \{V, E\}$, V 为所有节点的集合, E 为所有边的集合。根据随机漫步原理,假设有一个醉汉在该图中行走,他行走的起始点为 w_i 的概率为 p_i ,每当醉汉要继续往下走时,他都有两个选择,一个是随意选择一条邻接的边行走走到另一个节点,另一个选择是跳到另一个节点 w_j ,假设最后的稳定状态为 u ,则稳定状态矩阵 u 满足如下条件:

$$u = (1-c)Au + cr \quad (5)$$

式中, A 是标准化后的图 G 的邻接矩阵, c 是跳到另外一个节点的概率,本文将其设为 0.3, r 是初始的标注节点权重,其值为对向量 $\{weight(w_1), \dots, weight(w_m)\}$ 进行标准化计算后的结果。式(5)可以解释为在稳定状态 u 上继续往下走一步之后形成的状态矩阵必定与 u 相等,于是根据式(5)可以计算

得到矩阵 u 为:

$$u = c(I - (1-c)A)^{-1}r \quad (6)$$

式中, I 为 $n \times n$ 的单位矩阵。对于稳定状态矩阵 u ,其第 i 个元素 $u(i)$ 即为关键词 w_i 在最终标注信息中的概率,在此基础上,将矩阵 u 中的概率值从大到小进行排序,则位置靠前的 h 个关键词形成的组合即为最终的标注信息。以上就是基于内容检索的图像自动标注方法的实现过程。

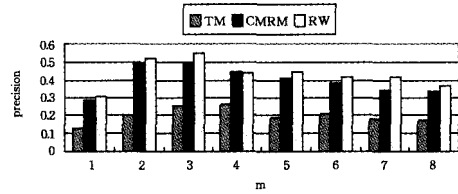
4 实验

4.1 实验设计

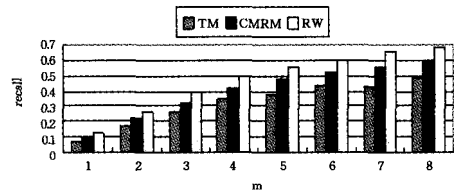
为了验证本文算法的准确性,我们分别对非网络图像和网络图像进行了验证。非网络图像的实验数据库来自于文献[9]的 Corel 图像数据库,该数据库包含了 5000 幅图像;对于网络图像,本文首先建立了包含 2 万张从网络上下载的图像作为参考图像数据库,然后随机选取了 4 幅图像进行标注实验。

4.2 非网络图像的实验结果

为了证明本文方法的有效性,实验选取了 3 种方法进行比较,分别是本文提出的基于随机漫步的标注方法(RW)、基于文献[5]的翻译模型标注方法(TM)和基于文献[6]的跨媒体关联模型方法(CMRM),评价指标为查全率和查准率,图 4 显示了实验对比结果。



(a) 查准率对比结果



(b) 查全率对比结果

图 4 非网络图像的实验对比结果

图 4 显示了利用 3 种方法进行自动标注的查全率和查准率对比结果,其中横轴 m 表示的是标注关键词的个数,纵轴表示的是前 m 个标注关键词的准确度。从该实验结果可以看出,本文提出的自动标注方法具有良好的效果,尤其是在前 5 个标注关键词中具有较高的准确率。

4.3 网络图像的实验结果

由于网络图像数据量过大,要完整地评价本文方法对网络图像的标注效果难度较大,因此本文随机从网络图像中选取了 4 幅图像作为示例来演示本文算法的标注效果。网络图像的标注信息一般包含标题、类别以及评价 3 个组成部分,在本文算法中分别将这 3 个标注部分的权重设为 3、2、1。表 1 显示的是这 4 幅图像的标注结果示例,从结果来看本文的标注算法对网络图像具有较好的准确度。

表1 网络图像标注结果示例



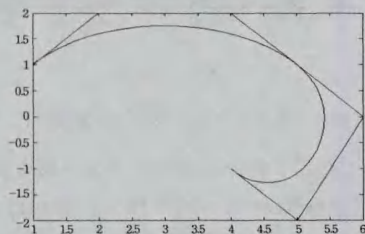
结束语 本文提出了一种基于内容检索的图像自动标注方法,即通过挖掘相似图像的标注信息来获得待标注图像的标注结果。本文算法包含了两个关键技术,分别是在基于内容检索时提出的基于 MSF 的图像全局特征,另一个则是在进行标注信息重排序时提出的基于随机漫步算法。本文提出的标注算法具有良好的稳定性和可扩展性,不需要做漫长的训练及其他先验知识,并且参考图像的数据库越大,本文算法的标注结果准确率越高。实验结果充分表明了本文算法的有效性。在下一步的工作中,我们将进一步研究如何将局部特征引入图像检索阶段以提高检索准确率,并同时保证检索速度的快速性。

(上接第 102 页)

将上面两式代入式(9)化简得到两拟 Bézier 曲线拼接时曲率连续的条件为:

$$\frac{1}{\lambda_1(\lambda_1 - 1)} = \frac{mt_2(n-1)(t_1-1)}{nt_1(m-1)} \cdot \frac{|W_0W_1 \times W_1W_2|}{|W_0W_1|} \cdot \frac{|V_{m-1}V_m|}{|V_{m-2}V_{m-1} \times V_{m-1}V_m|}$$

当 $t=2, \lambda=3$ 时两条拟 Bézier 曲线的 G^2 光滑拼接如图 4 所示。

图4 当 $t=2, \lambda=3$ 时两条拟 Bézier 曲线的 G^2 光滑拼接图

由上述定理可知,当拟 Bézier 曲线拼接时由于有调节因子 λ 以及参数 t 的调节可能很好地达到一阶、二阶连续和曲率的连续,使得拼接时有更大的优越性。

结束语 本文主要将 Bernstein 基函数中的变量 u 用函数 $f(u) = \frac{(t-1)u^3 + u}{t}$ 代替,将 Bernstein 基函数进行了推广,使其生成了新的拟 Bernstein 基函数和拟 Bézier 曲线;并且讨论了拟 Bernstein 基函数的构造和性质以及拟 Bézier 曲线的一些特性,也讨论了两条拟 Bézier 曲线的拼接。

参考文献

- [1] Chang E, Goh K, Sychay G, et al. CBSA: content-base soft annotation for multimodal image retrieval using bayes point machines [C]// Proc. CVPR. 2004
- [2] Cusano C, Ciocca G, Schettini R. Image annotation using SVM [C]// Proc. Internet imaging. 2004:330-338
- [3] Carneiro G, Vasconcelos N. A database centric view of semantic image annotation and retrieval [C]// Proc. ACM SIGIR. 2005: 559-566
- [4] Mori Y, Takahashi H, Oka R. Image-to-word transformation based on dividing and vector quantizing images with words [C]// MISRM. 1999
- [5] Barnard P D, de Freitas J F G, Forsyth D A. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary [C]// Proc. ECCV. 2002:97-112
- [6] Blei D M, Jordan M I. Modeling annotated data [C]// Proc. ACM SIGIR. 2003:127-134
- [7] Yossi R, CAarlo T, Leonidas J G. The Earth Mover's Distance as a Metric for Image Retrieval [C]// Proceedings of the 2000 IEEE International Conference on Computer Vision. Bombay, India, 2000:59-66
- [8] Tong H H, Faloutsos C, Pan J Y. Fast Random Walk with Restart and Its Applications [C]// Proc. ACM SIGKDD. 2006
- [9] Barnard P D, de Freitas J F G, Forsyth D A. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary [C]// Proc. ECCV. 2002:97-112

基函数的构造法可以用在曲线中,还可以应用在曲面中,有关这一方面的内容还需要进一步的研究。

参考文献

- [1] 李重,黄敏,韩丹夫. Bernstein 多项式推广及其所生成的拟 Bézier 曲线[J]. 计算机工程与应运, 2003(2):15-17
- [2] 朱心雄. 自由曲线曲面造型技术[M]. 北京: 科学出版社, 2000
- [3] 苏步青, 刘鼎元. 计算几何[M]. 上海: 上海科学技术出版社, 1981
- [4] 刘根洪, 何仁杰. 微分几何与计算几何[M]. 成都: 四川大学出版社, 2000
- [5] 韩旭里, 刘圣军. 二次 Bézier 曲线的扩展[J]. 中南工业大学学报: 自然科学版, 2003(2)
- [6] 吴晓勤, 韩旭里. 三次 Bézier 曲线的扩展[J]. 工程图学学报, 2005, 26(6):98-102
- [7] 吴晓勤. 带形状参数的 Bézier 曲线[J]. 中国图象图形学报, 2006, 11(2):269-274
- [8] 张贵仓, 赵菲, 意海英. 带三个形状参数的四次 Bézier 曲线[J]. 西北师范大学: 自然科学版, 2014, 50(2):26-29, 34
- [9] 吴荣军, 彭国华, 罗卫民, 等. 四次带参 Bézier 曲线的形状分析[J]. 计算机辅助设计与图形学报, 2009, 23:146-156
- [10] 师利红, 张贵仓. 三次 TC-Bézier 曲线的新拓展[J]. 计算机工程与应用, 2011, 47:201-204
- [11] Goldman R. The rational Bernstein bases and the multirational blossoms[J]. Computer Aided Geometric Design, 1999, 16:701-738