

基于标注词相关度的图像自动标注改善方法

徐功文¹ 廖明海¹ 郑森红² 赵洪奎² 张志军² 赵倩¹ 许春秀¹

(山东协和学院计算机学院 济南 250107)¹ (山东建筑大学计算机学院 济南 250101)²

摘要 随着图像数量的增长,图像检索技术已经成为一个活跃的研究领域。图像标记能够有效地组织和处理大量的图片信息并从中检索出用户需要的信息。由于自动图像标记方法根据图像分割后的区域对图像进行标记,准确率不高,因此提出一种结合标注词相关度的图标标注方法。该方法首先根据区域相似度对图像进行标注,然后利用标注词之间的相关性对标注结果进行改善。实验结果证明,该方法能够有效地对图像进行标注。

关键词 图像标注, k-means, 相似度, 相关度

中图法分类号 TP391 文献标识号 A

Image Automatic Annotation Based on Correlation of Keywords

XU Gong-wen¹ LIAO Ming-hai¹ ZHENG Sen-hong² ZHAO Hong-luan² ZHANG Zhi-jun² ZHAO Qian¹ XU Chun-xiu¹

(School of Computer Shandong Xiehe University, Jinan 250107, China)¹

(School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China)²

Abstract With the growth of the number of images, image retrieval technology has become an active area of research. Image annotation can organize and process large amounts of picture information effectively and retrieve the useful information which the user needs. Image automatic annotation methods tag image based on region after the image segmentation, but the accuracy rate is not high. This paper presented a method to tag images combined the correlation of keywords. Firstly, it tags images according to the similarity of areas. Then it uses the correlation of keywords to improve the results. The experimental results show that this tagging pictures method is effective.

Keywords Automatic annotation, k-means, Similarity, Correlation

1 引言

随着数码相机以及其他数码设备的普及,网络上的图像数量呈几何方式增长,图像检索技术已经成为一个活跃的研究领域。图像检索技术根据检索方法的不同可以分为两类:一种是基于文本的图像检索技术^[1],另外一种是基于内容的图像检索技术^[2]。基于文本的图像检索技术是对图像进行人工标注关键词,根据图像的关键词对图像进行检索,把图像检索转化为相关关键词的查询与匹配。基于文本的图像检索技术的优点是方便、检索速度快,用户只需要输入相关的关键词就可以查询并得到相关的结果。但是该技术需要人工对图像标注关键词,工作量大。基于内容的图像检索技术是根据图像的可视特征在数据集中搜寻相似的图像。虽然基于内容的图像检索技术取得了很多的研究成果,但是该技术根据图像的底层特征标注图像,查询关键词为图像的可视特征,不是自然语言,存在语义鸿沟。很多研究结合语义信息提高基于内容的图像检索技术,语义信息通常为包含描述图像语义属性

的文本关键词。由于人工标注语义信息是一项相当费时费力的工作,因此催生了图像自动标注技术的发展^[3]。

图像标注的方法主要有两种,第一种是基于监督学习方法的人工标注方法^[4,5],利用概率信息对图像进行标注关键词,这些监督方式能够保证高精度率,但是需要大量的时间和人力;第二种是基于图像底层特征的无监督自动标注方法^[6]。图像自动标注是指计算机能够根据图像的内容,自动地为图像标注相关的关键词。通常图像自动标注用于组织图像检索系统中的图像并从中找出与用户查询关键词相关的图像。图像自动标注主要有两种方法,一种方法是基于分类的图像标注方法,另一种方法是基于检索的图像标注方法。基于分类的图像标注方法把图像标注问题看成分类问题^[7],通过人工标注的训练集训练分类器,然后利用分类器标注图像。Jeon等^[8]提出了一种基于概率的学习模型,根据关键词与图像的相关性对图像进行标注。基于检索的图像标注方法将未标记的图像作为查询图像,然后由图像检索系统从注释图像数据库中搜索相关的图像,根据搜索得到的相关图像的关键词对

本文受山东协和学院校内基金(XHXY201431),山东省教育厅项目(J12LN31, J13LN11, J14LN59),2014 济南市高校院所自主创新计划项目(201303001, 201401214),山东省自然科学基金青年基金项(ZR2012GQ010)资助。

徐功文(1974—),男,博士生,副教授,CCF 高级会员,主要研究方向为机器学习、跨媒体检索等, E-mail: xugongwen@163.com; 廖明海(1946—),男,教授,主要研究方向为计算机网络安全; 郑森红(1994—),女,主要研究方向为计算机信息处理技术; 赵洪奎(1979—),女,博士,副教授,主要研究方向为计算机网络、路由算法; 张志军(1973—),男,副教授,主要研究方向为个性化推荐技术; 赵倩(1984—),女,讲师,主要研究方向为计算机多媒体技术; 许春秀(1980—),女,讲师,主要研究方向为图像处理。

图像进行标注。虽然这种方法不需要预先创建标注模型,但是需要大量的已标注图像作为数据库。

图像自动标注方法通过建立图像分割后的区域与语义信息的关系对图像进行标注。首先提取图像区域的视觉特征,然后根据已标注的图像寻找区域与语义信息的相关性,最后以此为先验知识对未标注的图像进行标注。由于图像自动标注方法只考虑图像的视觉特征,标记准确率不高。因此本文提出一种结合标注词相关度的图像自动标注改善方法。首先根据区域相似度对图像进行标注,然后根据区域相似度得到的标注词的相关度改善标注结果。

2 图像标注

语义信息自动标注图像的方法有两种,第一种通常称为弱标记,关键词标注整个图像,而不是对应于图像中的区域。第二种是区域级别的自动标注。关键词对应于图像中的区域。本文中的图像标注是指第二种标注方法。

2.1 图像分割

基于区域的图像自动标注方法^[9,10]首先需要利用图像分割技术把一幅图像分割为若干个同质区域(Homogenous Region),每一个区域对应于一个对象,由于每个同质区域的语义相对比较单一,因此能够准确地描述图像的语义内容。研究人员已经提出了很多知名的分割算法,其中,N-Cut 算法^[11]和 JSEG 算法^[12]是具有代表性的分割算法。

在基于图论的图像分割方法中,一幅图像被看成无向加权图, $G=\{V,E,W\}$,其中 V 表示节点集,在图像中像素表示节点。 E 表示两两节点之间的连线。 W_{ij} 表示节点之间的权重。权重可以通过计算像素、亮度或者其他信息之间的距离得到。例如,一幅图像分割成两部分 A 和 B : $A \cup B = V, A \cap B = \emptyset$,两个子集之间的相似度由以下公式计算:

$$cut(A,B) = \sum_{i \in A, j \in B} W(i,j) \quad (1)$$

Shi 和 Malikt 提出了 Normalized-Cut 描述两类间的分离度,并且可以得到一个 N_{-cut} 值作为测量标准:

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(B,A)}{assoc(B,V)} \quad (2)$$

其中, $assoc(A,V)$ 表示节点 A 与其他所有节点权重的和。最佳分割方法是最小化以下目标函数:

$$\text{Min } Ncut(A,B) \quad (3)$$

2.2 图像聚类

图像分割之后,对得到的同质区域进行聚类。聚类算法常用于图像标注之前对图像的底层特征进行分类,其中,最常用的几种聚类算法包括 k -means 算法^[13]、期望最大化(EM)^[14]和 Discret Distribution(D2) 聚类算法^[15]。由于 K -means 算法可以通过图像的颜色特征对图像进行分割聚类,因此 K -means 算法已被广泛用于对图像底层特征聚类。

2.3 特征提取

特征提取的主要任务是提取能够表示图像可视内容的特征信息。在本文中,主要提取了图像的颜色特征和纹理特征,并通过一定的算法得到图像的颜色特征值和纹特理征值。

由于图像中包含大量的颜色信息,颜色特征的提取方法简单,因此,在图像检索和标注中通常使用颜色特征作为可视内容特征^[16]。颜色信息可以在不同的空间中表示,其中最常见的空间为 RGB 颜色空间。RGB 颜色空间是由红色、绿色

以及蓝色 3 种基色组成,其他颜色都可以由这 3 种基色线性表示。除了 RGB 之外,HSV 也是常见的描述颜色特征的颜色空间。由于 HSV 颜色空间为视觉感知的颜色空间,因此本文使用 HSV 颜色空间。HSV 颜色空间由色调、饱和度和亮度组成。由 RGB 颜色空间转换成 HSV 颜色空间的方法如下:

$$H = \begin{cases} \arccos \frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}, & B \leq G \\ 2\pi - \arccos \frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}, & B > G \end{cases} \quad (4)$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad (5)$$

$$V = \begin{cases} 0, & V \in [0, 0.2] \\ 1, & V \in [0.2, 0.7] \\ 2, & V \in [0.7, 1] \end{cases} \quad (6)$$

其中, $R,G,B \in [0, 1, \dots, 255]$, $H \in [0, 1, \dots, 360]$, $S,V \in [0, 1]$ 。由于转换后的 HSV 颜色值维度过高,因此需要量化转换后的 HSV 颜色值,量化方法参考文献^[18],通过量化,可以得到图像的颜色特征值。

纹理是一种重要的图像特征,纹理特征在图像中普遍存在但难以描述^[18]。纹理特征提取方法的目的是:提取的特征维数较小、提取过程的计算量小、提取的纹理特征有较强的鉴别能力等。纹理特征提取的方法有很多种,其中最具有代表性的方法是 Haralick 基于纹理的分析提出的纹理特征提取方法,即灰度共生矩阵方法,它将灰度值转化为纹理参数。之后基于 MRF 理论提出了 MRF 模型、高斯马尔可夫随机场(GMRF)模型以及隐马尔可夫随机场(HMRF)模型等。小波变换的引入,为纹理特征的提取提供了新的思想。

本文采用灰度共生矩阵方法(GLCM)提取图像的纹理特征。GLCM 方法是建立在估计图像的二阶组合条件概率密度基础上的。GLCM 方法描述两个像元在一定角度和距离上分别具有灰度层 i 和 j 的出现概率。GLCM 是一个对称矩阵,是距离和角度的函数,其阶数由图像中的灰度级决定,由 GLCM 能够导出 14 种纹理参数。本文选取 4 种描述能力强的参数组成图像纹理特征值,分别为:角二阶矩参数、惯性矩参数、熵参数和局部平稳参数。

我们通过 k -means 方法对分割后的区域进行聚类,在每一类中,选择具有代表性的若干区域,对其进行人工标注关键词。我们用所选择区域的颜色特征值和纹理特征值的平均值数值化关键词:

$$K_i = \frac{1}{l} \sum_{j=1}^l X_{ij} \quad (7)$$

其中, K_i 是多维向量,为第 i 个关键词数值化的值, l 为标注词为 K_i 的区域个数, X_{ij} 表示标注词为 i 的第 j 个区域的特征值向量。

我们把所标注的关键词 K_i 作为种子集。对于测试图片,首先计算分割后区域的颜色特征和纹理特征的值向量 T ,然后计算区域特征值向量 T 与关键词向量 K_i 的欧氏距离,如果测试图像分割后的区域特征值向量 T 与关键词向量 K_i 的欧氏距离小于一定的阈值,则就用这个关键词标注测试图像。

3 WUP 改善标注结果

3.1 WUP 方法

WordNet^[19]是著名的数据库之一,可以用来计算两个概念(或者 word sense)之间的语义相似度,它提供了 6 种计算相似度和 3 种计算概念之间的关联度方法,所有的这些方法都是基于 WordNet 词汇数据库的。本文采用 WUP 相似度测量标注词之间的相关性。WUP 方法^[20]是由 Wu 与 Palmer 提出的基于路径结构的相似度度量方法,综合考虑了概念结点、共有父结点、根结点之间的路径关联情况。计算标节点 s_1 和 s_2 相似度的方法如下:

$$sim_{wup}(s_1, s_2) = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3} \quad (8)$$

其中,将 s_3 记为 s_1 和 s_2 最深上层父节点, N_1 表示 s_1 到 s_3 路径上节点的数量, N_2 表示 s_2 到 s_3 路径上节点的数量, N_3 表示 s_3 到根节点路径上节点的数量。

3.2 提高标注准确率

将关键词种子集中的关键词编号,利用 WUP 相似度计算关键词之间的相关度。通过计算关键词之间的相关度得到一个对称相关度矩阵 T :

$$T = \begin{bmatrix} \tau_{11} & \tau_{12} & \dots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \dots & \tau_{2n} \\ \vdots & \vdots & \dots & \vdots \\ \tau_{n1} & \tau_{n2} & \dots & \tau_{nn} \end{bmatrix} \quad (9)$$

其中, n 表示数据集中标注词的个数, τ_{ij} 表示第 i 个标注词与第 j 个标注词之间的相关度。对于某一个关键词 i ,按照相关度从大到小得到一个向量 $X = [\tau_{i1}, \tau_{i2}, \dots]$, l, m 为标注词的序号。对于一幅图像,若其通过基于区域的标注方法得到 y 个标注关键词,计算标注词数据库中与该图像标注关键词相关度得到 y 个相关度向量。设定一个阈值,如果 y 个相关度向量中大于阈值的标注词编号中有重复,则用编号重复的标注词对图像进行二次标注,提高标注的准确率。

4 实验结果与分析

本节将介绍数据集以及在数据集上所做的实验,通过分析实验结果说明提出的方法的有效性。

4.1 数据集

本节中实验以及得到的结果所使用的数据集为 Corel 图像库。该图像库中包含 10 类图像集,分别为建筑、花、人、马、大象、海洋、恐龙、车、食物以及山。每类图像集中有 100 幅图像,我们选取每个类中的 70 幅图像作为训练集。每一幅图像通过图像分割算法被分为 1-5 部分,然后利用 k-means 算法对分割后的区域聚类。选择每一类中具有代表性的区域进行人工标注。剩下的 300 幅图像作为测试集。每一幅图像的标注词个数为 1-4 个。

4.2 度量标准

我们用查全率、查准率和 $F1$ 值来衡量方法的有效性。对于指定标注词 K ,测试集中包含标注词 K 的图像个数为 K_i 。使用本文提出的图像自动标注方法得到的标注结果中包含标注词 K 的图像个数为 K_r , K_r 表示标注正确的个数。

查全率表示标注方法对于单个标注词的完整性;

$$Recall = \frac{K_r}{K_i} \quad (10)$$

查准率表示标注方法标注的精确度;

$$Precision = \frac{K_r}{K_s} \quad (11)$$

$F1$ 值实际上是查全率和查准率的调和平均,它将查全率和查准率综合为一个指标;

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (12)$$

4.3 实验结果及分析

利用 WUP 改善图像标注时,需要确定关键词数据库中 与图像标注词相关度的阈值,阈值的选择与查准率、查全率以及 $F1$ 值的关系如图 1 所示。

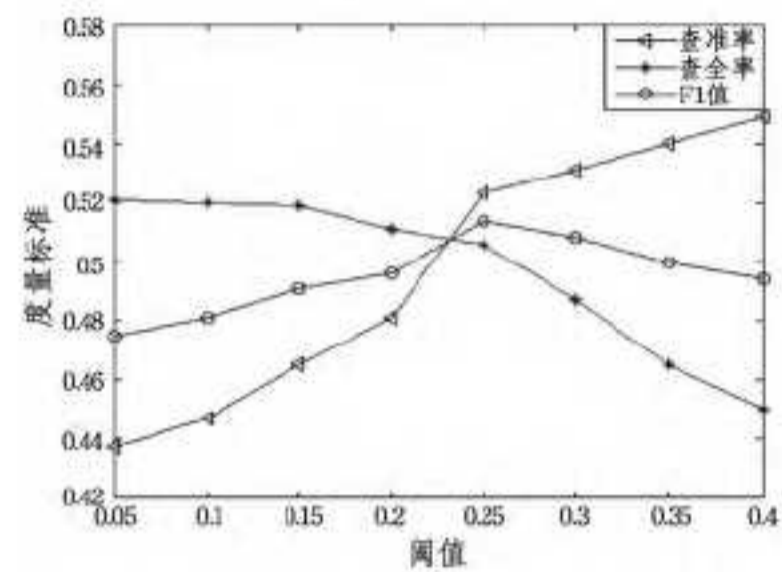


图 1 阈值对度量标准的关系

通过图 1 可以看出阈值越大,查准率越大,查全率越小。当阈值选择 0.25 时,实验效果最好,查准率为 0.523,查全率为 0.506, $F1$ 值为 0.514。

对图像进行标注时,我们使用了两种方法作比较,第一种方法只使用区域相似度对图像进行标注,第二种方法同时使用区域相似度与标注词相关度对图像进行标注。计算标注词相关度时,阈值选择 0.25,实验结果如表 1 所列。

表 1 实验对比

	查准率	查全率	F1 值
只使用区域相似度	0.461	0.484	0.472
结合标注词相关度	0.523	0.506	0.514

从表 1 中可以看出,只使用区域相似度标注图像时,查准率为 0.461,查全率为 0.484, $F1$ 值为 0.472;当结合了标注词相关度标注图像时,查准率为 0.523,提高了 13.45%;查全率为 0.506,提高了 4.55%; $F1$ 值为 0.514,提高了 8.90%。实验结果显示,结合区域相似度与标注词相关度能够有效地对图像进行标注。

图 2 为提出方法的部分实验结果。



图 2 部分图像标注结果

- bling innovation in campus networks [J]. ACM SIGCOMM Computer Communication Review, 2008, 38(2): 69-74
- [3] Meyer D, Lewis D. The Locator/ID Separation Protocol (LISP) [J]. IETF RFC, 2013, 6830
- [4] Lougheed K, Rekhter Y. A border gateway protocol 3 (BGP-3) [R]. RFC 1267, Cisco Systems, TJ Watson Research Center, IBM Corp, 1991
- [5] IEEE Standard. LLDP [OL]. <http://standards.ieee.org/get-ieee802/download/802.1AB-2009.pdf>
- [6] LISPmob 2014 [OL]. <http://lispmob.org/>
- [7] Kataoka K, Agarwal N, Kamath A V. Scaling a broadcast domain of Ethernet: Extensible transparent filter using SDN [C] // 2014 23rd International Conference on Computer Communication and Networks (ICCCN). IEEE, 2014: 1-8
- [8] Feamster N, Rexford J, Zegura E. The road to sdn: an intellectual history of programmable networks [J]. ACM SIGCOMM Computer Communication Review, 2014, 44(2): 87-98
- [9] Medved J, Varga R, Tkacik A, et al. OpenDaylight: Towards a Model-Driven SDN Controller architecture [C] // 2014 IEEE 15th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE, 2014: 1-6
- [10] Pongrácz G, Molnár L, Kis Z L, et al. Cheap silicon: a myth or reality? picking the right data plane hardware for software defined networking [C] // Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking. ACM, 2013: 103-108
- [11] OpenDaylight 2014 [OL]. <http://www.opendaylight.org/>
- [12] 陈鸣, 赵广松, 邢长友, 等. 基于 Openflow 的 SDN 技术研究 [J]. 软件学报, 2013, 24(5): 1078-1097
- [13] Nadeau D T, Gray K, 等. 软件定义网络 [M]. 北京: 人民邮电出版社, 2014
- [14] 江国龙, 付斌章, 陈明宇, 等. SDN 控制器的调研和量化分析 [J]. Journal of Frontiers of Computer Science and Technology, 2014, 8(6): 653-664

(上接第 240 页)

图 2 中, 第三幅图像标注词“草”是根据计算标注词数据集中与“大象”和“树木”相关度之后添加的标注。

结束语 本文提出了一种利用标注词相关性改善标注结果的方法。首先, 我们通过 N-Cut 方法把图像分割成若干个同质区域, 然后用 k-means 方法对分割后的区域进行聚类。根据已标注图像计算出区域与标注词的相关性, 根据相关性对未知图像进行标注, 最后利用 WUP 方法计算标注词之间的相似度改善标注结果, 实验结果显示, 该方法能够有效地对图像进行自动标注。

参 考 文 献

- [1] Tmura H, Yokoya N. Image systems: A survey [J]. Pattern Recognition, 1984, 17(1): 29-43
- [2] Kato T. Database architecture for content-based image retrieval [C] // Proc. Of SPIE Int. Conf. on Image Storage and Retrieval System, San Jose, CA, USA, May, 1992: 112-123
- [3] Jeon J, Manamatha R. Automatic image annotation of news images with large vocabularies and low quality training data [C] // ACM Multimedia, New York, USA, Oct, 2004
- [4] Carneiro G, Chan A B, Moreno P J, et al. Supervised Learning of Semantic Classes for Image Annotation and Retrieval [C] // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29: 394-410
- [5] Hollink L, Schreiber G, Wielemaker J, et al. Semantic Annotation of Image Collections [C] // Workshop on Knowledge Markup and Semantic Annotation (KCAP 2003), 2003
- [6] Jeon J, Lavrenko V, Manmatha R. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models [C] // Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003
- [7] Escalante H J, Hernández C A, Gonzalez J A, et al. The segmented and annotated IAPR TC-12 benchmark [C] // Computer Vision and Image Understanding, 2010, 114: 419-428
- [8] Jeon J, Lavrenko V, Manmatha R. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models [C] // 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003: 119-126
- [9] Ko B C, Byun H. Frisp: a region-based image retrieval tool using automatic image segmentation and stepwise Boolean and matching [C] // IEEE Trans. on multimedia, 2005, 7: 105-113
- [10] Liu Y, Zhang D, Lu G. Region-based image retrieval with high-level semantics using decision tree learning [J]. Pattern Recognition, 2008, 41(8): 2554-2570
- [11] Yining D, Manjunath B S, Shin H. Color image segmentation [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999: 446-451
- [12] Jianbo S, Jitendra M. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905
- [13] Çavuş, Aksoy S. Semantic Scene Classification for Image Annotation and Retrieval [J]. LNCS, 2008, 5342: 402-410
- [14] Wang L, Liu L, Khan L. Automatic Image Annotation and Retrieval using Subspace Clustering Algorithm [C] // 2nd ACM International Workshop on Multimedia Databases, 2004: 100-108
- [15] Li J, Wang J Z. Real-Time Computerized Annotation of Pictures [C] // ACM Multimedia Conference, 2006: 911-920
- [16] Z Xin, W Xian-bin, M Qing-xia. Research of Image Retrieval Method Based on Color Feature [J]. Computer Science, 2012, 39(11): 243-260
- [17] 王娟, 孔兵, 贾巧丽. 基于颜色特征的图像检索技术 [J]. 计算机系统应用, 2011, 20(7): 160-164
- [18] Mikolajczyk K, Cordelia S. A performance evaluation of local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630
- [19] Deselaers T, Ferrari V. Visual and Semantic Similarity in ImageNet [C] // CVPR 2011, 2011
- [20] Wu Z, Palmer M. Verb Semantics and Lexical Selection [C] // Proceedings of the 32nd annual meeting on Association for Computational Linguistics (ACL 1994), 1994: 133-138