

一种结合多示例学习的图像检索方法^①

王春燕 袁津生 (北京林业大学 信息学院 北京 100083)

摘要: 提出一种基于多示例学习(Multiple-instance learning)的图像检索方法,将多示例学习应用于图像检索中,以有效的处理图像的歧义性。该方法首先将图像作为多示例包,其次采用自适应 k-means 图像分割算法将图像自动分成多个示例,然后根据用户选择的实例图像生成正包和反包,再采用 EM-DD(expectation maximization diverse density)算法进行多示例学习,实现图像检索和相关反馈,最终使用户得到比较满意的结果。

关键词: 多示例学习; 期望最大多样性密度; 基于内容的图像检索; 相关反馈

An Image Retrieval Method with Multi-Instance Learning

WANG Chun-Yan, YUAN Jin-Sheng

(School of Information, Beijing Forestry University, Beijing 100083, China)

Abstract: In this paper, a multi-instance learning-based CBIR (content-based image retrieval) approach is presented, and multi-instance learning is applied in CBIR in order to deal with the inherent ambiguity of images. First of all, the whole image is regarded as a multi-instance bag. Secondly, the image is partitioned into a number of regions using Adaptive k-means image segmentation method. Then query images posed by the user are transformed into corresponding positive and negative bags and a EM-DD algorithm is employed for image retrieval and relevance feedback. Finally, the users can get satisfactory results.

Keywords: multiple-instance learning; expectation maximization diverse density; content-based image retrieval; relevance feedback

1 引言

随着多媒体信息需求的不断增长,基于内容的图像检索(content-based image retrieval, CBIR)得到越来越多的关注和应用,已开始成为一个十分活跃的研究领域。CBIR 通过利用图像的颜色、纹理、形状和对象空间等基本视觉特征建立索引进行相似度匹配。现有的图像检索,例如 IBM 公司的 QBIC^[1]、哥伦比亚大学的 VisualSEEK、由 UCSB(加州大学 San Barbara 分校)开发的 Netra 系统,都是以图像为中心,基于整幅图像或者图像中感兴趣的一部分来检索,在目前相关方面的研究中,图像内容的多样性和歧义性是需要迫切解决的一个问题,而多示例学习结合了以人为中心的思想,可以提高检索效率。

2 多示例学习技术

2.1 多示例学习的概念和算法

最早 Dietterich 等人^[2]在对药物活性预测(drug activity prediction)问题的研究中,提出了多示例学习(multi-instance learning)的概念。多示例学习可以被描述为:假设训练数据集中每个数据是一个包(Bag,即样本),每个包都是一个示例(instance)集合,每个包都有一个训练标记,而包中的示例没有标记;如果一个包被标记为正包(用户感兴趣的图像),则该包中至少存在一个示例是正例;反之一个包被标记为负包(用户不感兴趣的图像),则其中所有的示例均为负例。通过对多个包组成的训练数据集进行学习,从而尽可能正确地对训练集之外的包的概念标记进行预

① 收稿时间:2009-09-16;收到修改稿时间:2009-11-04

测。多示例学习框架可用图1来表示,

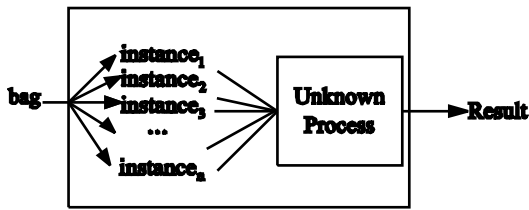


图1 多示例学习框架示意图

自提出多示例学习的概念以来,通过研究已形成了一些有效的算法。O. Maron 和 T. Lozano-Pérez^[3]提出了多样性密度(Diverse Density, 简称 DD)算法。他们认为,多示例学习的目标概念是:示例空间中某个未知区域,满足每个正包中至少有一个示例在该区域内或距离该区域足够近,同时所有来自负包的示例均远离该区域。基于这一假设,他们把属性空间中某个点的多样性密度定义为:有多少个不同的正包拥有距离该点足够近的示例,同时来自负包的示例远离该点的程度的测度。在他们的定义中,在属性空间中某点附近出现的正包数越多,反包示例出现得越远,则该点的多样性密度越大。令 B_i^+ 代表第 i 个正包, B_{ijk}^+ 代表第 i 个正包的第 j 个示例的第 k 个属性的值;反之,令 B_{ijk}^- 代表第 i 个反包的第 j 个示例的第 k 个属性的值;令 t 代表属性空间中多样性密度最大的点。则可以通过最大化

$$Pr(x=t | B_1^+, \dots, B_n^+, B_1^-, \dots, B_m^-)$$

确定 t 。假设各包独立,通过(1)确定 t ,

$$\text{argmax}_t \prod_i Pr(x=t | B_i^+) \prod_j Pr(x=t | B_j^-) \quad (1)$$

采用 noisy-or 模型对(1)式中乘积项例化

$$Pr(x=t | B_i^+) = 1 - \prod_j (1 - Pr(x=t | B_{ij}^+)) \quad (2)$$

$$Pr(x=t | B_j^-) = 1 - \prod_i (1 - Pr(x=t | B_{ij}^-)) \quad (3)$$

他们将示例出现在潜在目标处的因果概率定义为该示例与潜在目标之间的距离,

$$Pr(x=t | B_{ij}) = \exp(-\|B_{ij} - x\|^2) \quad (4)$$

在此基础上,使用梯度下降法来寻找最大多样性密度点,因此他们将上述做法不仅可以确定多样性密度最大点,为了求解(1)式,采用梯度下降法求 t ,由于多样性密度空间中存在多个局部极小值点需要将每一个正包示例都作为初始点进行一次搜索。通过梯度下降搜索 t 的同时,还可以挑选相关属性,并用 S_k 表示第 k 个属性的相关度,则(2)式中的距离为:

$$\|B_{ij} - x\|^2 = \sum_k s_k (B_{ijk} - x_k)^2 \quad (5)$$

Q. Zhang 和 S. A. Goldman^[4]将 DD 算法与 EM 算法相结合提出了 EM-DD 算法来求解目标特征 t 的方法。此算法首先给定一个假设的初始目标特征 h ; 接下来在 E 步,从训练集的每个示例包中选出最靠近 h 的示例组成一个集合;然后再 M 步,利用 E 步获得的集合,采用梯度搜索法估算出一个新的使多样性密度最大的目标特征点 h' ,并用来代替 h ;重复 E 步和 M 步直至算法收敛。

EM-DD^[4]算法相比较 DD 算法,在 E 步中对示例的选取,将多示例转变为单一示例,在本文提出的检索方法中,利用特征向量来描述示例,图像作为包含示例的包,由于 DD 算法要进行多次梯度下降搜索,所以计算时间开销很大,不利于提高检索效率,因此本文检索方法中采用 EM-DD 算法。

2.2 多示例学习技术在 CBIR 中的应用

MIT 和 Maron^[5]等人将多样性密度算法应用到自然场景图像的分类和检索^[5]中,他们首先对每幅图像进行高斯滤波并取样,由颜色块(2×2 大小)构成的 8×8 大小的矩阵,每幅图像用 9 个 15 维的属性向量表示,分别是中心彩色块的 R,G,B 均值和四个相邻色块之间的 R,G,B 均值的差值。检索时使用多样性密度算法从正包和反包中学习的模式,其中结合了一种 SBN(Single Blob with Neighbours)方法具有较好的检索效果。Yang 和 Lozano-Pérez^[6]将图像转化为灰度图像,再将图像划分为互相重叠的具有不同大小和位置的 20 个区域,通过反射变换得到 40 幅子图像,对每幅子图像平滑等,将对应的示例组成图像包。

3 基于内容的图像检索

3.1 图像分割方法和包的生成

图像分割是指将图像中具有特殊意义的不同区域

分开,并使这些区域互不相交,并且每个区域满足特定区域的一致性条件。首先将待分割图像的像素值,从 RGB 彩色空间转换成 L^*a^*b 空间的值,计算三个颜色特征 $labL=L/100$, $labA=(a+50)/100$, $labB=(b+50)/100$, 其中 L 、 a 、 b 分别表示图像在 L^*a^*b 空间的原始分量。设图像像素点为 $m \times n$, 像素点 (x, y) 的位置特征为 X 、 Y , 则 $X=x/\max(m, n)$, $Y=y/\max(m, n)$ 。由于纹理是一个空间性质,对纹理的度量应局限于相对均匀的区域。本文抽取了像素点的对比度 con , 极性 p , 各向异性 Ani , 根据 L^*a^*b 彩色空间的 L 分量,通过选用不同的高斯函数的窗口尺寸参数和对各像素极性纹理特性的分析,得到相应较佳的窗口尺寸参数,然后根据这些参数计算各像素点的对比度和各向异性的纹理特征。综上所述,图像中像素点的特征向量表示为 $I(labL, labA, labB, p, con, ani, X, Y)$ 。下一步,利用自适应的 k 均值聚类算法,在像素点特征空间中进行聚类,利用相邻像素相连接的原理,每幅图像聚类自适应生成 K (本文采用 $2 < k < 8$) 个互不重叠的区域,图像分割得到 K 个图像块,最后抽取每个图像块区域的特征向量。在本文中根据测试和分析,综合图像检索速度,为准确地表达分割后区域的语义信息,确定包中每一个示例包含有 9 维特征向量 $I' \{labL, labA, labB, X, Y, con, ani, F_1, F_2\}$ 。

假设对于区域 R , num_R 为区域内像素点的数目, i 为该区域内的像素点的数目, $i \in R$, 则像素点在 L^*a^*b 空间中三个分量的颜色均值、平均对比度、平均各项异性、位置特征(用该示例内象素点在二维空间坐标的平均值)表示为:

$$\overline{labL} = \frac{1}{num_R} \sum_{i=1}^{num_R} L_i / 100, \quad \overline{labA} = \left(\frac{1}{num_R} \sum_{i=1}^{num_R} a_i - 50 \right) / 100 \quad (6)$$

$$\overline{labB} = \left(\frac{1}{num_R} \sum_{i=1}^{num_R} b_i - 50 \right) / 100$$

$$\overline{con} = \frac{1}{num_R} \sum_{i=1}^{num_R} con_i, \quad \overline{ani} = \frac{1}{num_R} \sum_{i=1}^{num_R} ani_i \quad (7)$$

$$\overline{X} = \frac{1}{num_R} \sum_{i=1}^{num_R} X_i, \quad \overline{Y} = \frac{1}{num_R} \sum_{i=1}^{num_R} Y_i \quad (8)$$

在二维复平面中,区域的边界曲线可用一个关于 t ($t \in (0, 2\pi)$) 的周期函数来表示,即 $(z, t) = (x, t) + k(y, t)$, 根据

$$a_k = \frac{1}{p} \int_0^{2p} (x, t) \cos t dt, \quad b_k = \frac{1}{p} \int_0^{2p} (x, t) \sin t dt,$$

$$c_k = \frac{1}{p} \int_0^{2p} (y, t) \cos t dt, \quad d_k = \frac{1}{p} \int_0^{2p} (y, t) \sin t dt \quad (9)$$

则边界曲线可用参数集

$F = (a_0, c_0, a_1, b_1, c_1, d_1, \dots, a_k, b_k, c_k, d_k)$ 来表示。为了计算的方便,采用傅立叶均值 F_1, F_2 来描述形状特征,从而更加准确的表达图像的内容信息。

3.2 图像检索

图像检索的目的就是要寻找包含相同或相似示例的图像。抽取的图像区域的特征向量 I' 生成多示例包,并利用 EM-DD 算法从图像包所组成的训练集中学习,找出最大多样性密度点,依据该点与实例图片中目标特征向量的距离由小到大的顺序返回图像,由于在不同类型图像中有些特征比较明显,而有些特征是冗余的,为此再度量相似性时给每个特征分配一个权值,来表示这个特征的重要度,因此采用加其中, B_{ijk} 是示例 B_{ij} 的第 k 维的值, w_k 是对应的非负权值,使用梯度下降法,可以同时获得目标点 t 和对应的权值 w 。如果用户对检索结果不是很满意,可以在检索结果中再选择感兴趣和不感兴趣的图像加入训练集进行相关反馈,在进行一次或多次检索,直到达到用户满意的为止。

4 测试结果

对本文方法进行实验,选择了包括风景、花、山水、动物、人物、汽车各 100 幅共 600 幅图像作为待检测图像测试库,每一类选择了 4 幅共 24 幅图像,作为查询图像,把某一类中的 4 幅作为用户感兴趣的图片生成正包,从其它 5 类中随机选取 4 幅作为负包组成训练集,使用本文的方法进行学习,排序后返回结果,然后进行相关反馈,将排在最前面 4 幅图像选为正包,排在最后的 4 幅选为负包再进行学习,

本文采用传统的查准率 P (precision) 和查全率 R (recall) 来评价系统的性能,查准率用系统返回的检索结果图像 X_1 中相关图像 Y_1 所占百分比表示

$P = \frac{X_1}{Y_1} \times 100\%$, 而查全率是系统返回的相关图像 X_2 与测试图像库中所有相关图像 Y_2 数目的比例,如 $R = \frac{X_2}{Y_2} \times 100\%$, 下图 2 和图 3, 分别与文献中的 [5, 6] 进行比较

参考文献

- 1 Niblack W, Barber R, Equitz W, et al. The qbic project: Querying images by content using color texture, and shape Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases. San Jose, CA, 1993. 173 – 187.
- 2 Dietterich TG, Lathrop RH, Lozano-Pérez T. Solving the multiple-instance problem with axis-parallel rectangles. Artificial Intelligence, 1997, 89(1-2): 31 – 71.
- 3 Maron O, Lozano-Pérez T. A framework for multiple-instance learning. In: Jordan MI, Kearns MJ, Solla SA eds. Advances in Neural Information Processing Systems 10, Cambridge, MA: MIT Press, 1998. 570 – 576.
- 4 Zhang QS. A Goldman, EM-DD: an improved multiple-instance learning technique. Neural Information Processing Systems (NIPS'01), 2001: Cambridge: MIT Press: 1073 – 1080.
- 5 Maron O, Raturi A L. Multiple-instance learning for natural scene classification. Proc. of the 15th International Conference on Machine Learning. Madison, WI, 1998. 341 – 349.
- 6 Yang C, Lozano-Pérez T. Image database retrieval with multiple-instance learning Techniques. Proc. of the 16th International Conference on Data Engineering. San Diego, CA, 2000. 233 – 243.
- 7 Dai HB, Zhang LM, Zhou ZH. A Multi-Instance Learning Based Approach to Image Retrieval. Pattern Recognition and Artificial Intelligence, 2006, 19: 179 – 185.
- 8 Michael S. Content-based Multimedia Information Retrieval: State of the Art and Challenges. ACM Transactions on Multimedia Computing, Communications and Applications. LEW Leiden University, The Netherlands and NICU SEBE, Feb. 2006. 389 – 402.
- 9 林开颜, 吴军辉, 徐立鸿. 彩色图像分割方法综述. 中国图象图形学报, 2005, 10(1): 1 – 10.
- 10 李苏梅, 韩国强. 基于 K-均值聚类算法的图像区域分割方法. 计算机工程与应用, 2008(16): 163 – 167.
- 11 吴洪, 卢汉清, 马颂德. 基于内容图像检索中相关反馈技术的回顾. 计算机学报, 2005, 28(12): 1969 – 1979.
- 12 Liu W, Xu WD, Li LH. Two new bag generators with multi-instance learning for image retrieval 2008 IEEE.
- 13 Han FQ, Wang DC, Liao XF. An Improved Multiple-Instance Learning Algorithm ISSN 2007, Part I, LNCS 4491, 1104 – 1109.
- 14 Rahmani R. Localized Content-Based Image Retrieval IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11).

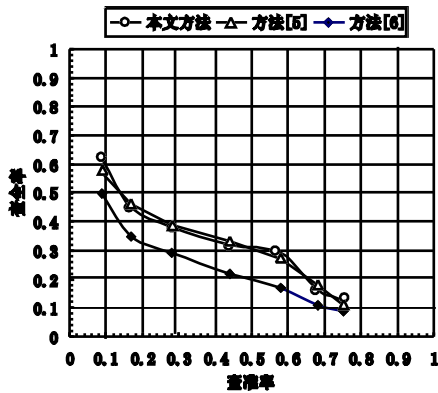


图2 初始检索结果

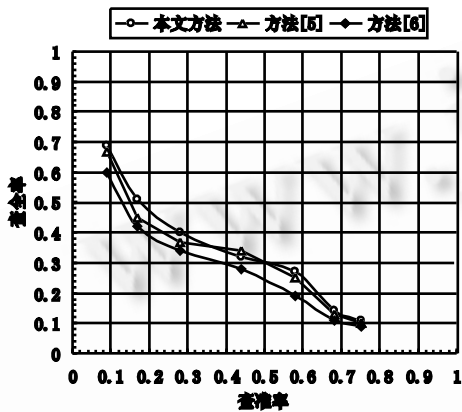


图3 相关反馈检索结果

表1 初始检索和相关反馈的平均时间

	方法[5]	方法[6]	方法[7]	本文方法
初始检索的平均时间(s)	3.12	7.63	1.92	2.32
相关反馈的平均时间(s)	9.43	8.51	2.67	2.89

从表1可以看出，本文方法的初始检索和反馈时间明显优于方法[5]的9个15维特征向量和方法[6]的20个多为特征向量，值得注意的是本文将每幅图像看作一个包，自适应分割为2至8个(示例)9维的特征向量，虽然比方法[7]速度慢，但查准率和查全率有所提高，具有较高的检索效率。

5 结论

本文提出的一种基于多示例学习的图像检索和相关反馈方法，该方法与已有的方法实验进行比较，取得了较好的效果，可以提高检索的效率，具有一定的应用价值。笔者正进一步尝试其它的方法实现包中示例的生成，并尝试改进多示例学习算法应用于图像检索中。