

一种融合图学习与区域显著性分析的 图像检索算法

冯松鹤^{1,2}, 郎丛妍¹, 须 德¹

(1. 北京交通大学计算机与信息技术学院, 北京 100044;

2. 北京邮电大学智能通信软件与多媒体北京市重点实验室, 北京 100876)

摘 要: 为弥合图像低层视觉特征和高层语义之间的语义鸿沟, 改善图像检索的效果, 机器学习算法经常被引入到图像检索问题中. 通常情况下, 机器学习算法是与相关反馈机制相结合, 通过用户的交互操作, 标定出若干正反例图像, 很自然地就可以将图像检索问题转化为模式识别中的分类问题. 目前融合区域显著性分析的区域图像检索算法尚没有与机器学习算法相融合. 本文结合图像区域显著性分析, 并针对用户参与反馈的情况, 分别提出了两种图像检索解决方案. 其一, 在没有用户反馈以及用户只反馈正例图像的情形下, 将图像检索问题转化为直推式学习问题 (Transductive Learning), 改进已有的基于图的半监督学习算法, 提出了融合区域显著性分析的层次化图表示 (Hierarchical Graph Representation) 方式, 用以实现标记传播; 其二, 在用户同时反馈正反例图像的情形下, 利用用户反馈得到的正反例图像构建相似性邻接矩阵, 通过流形排序算法 (Manifold-Ranking) 学习出用户感兴趣的查询目标概念并用相应的特征向量集合表示, 并据此查询图像库返回用户语义相关的图像集合. 实验结果验证了这两种检索策略的有效性.

关键词: 图像检索; 区域显著性; 图学习; 流形排序; 相关反馈

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2011) 10-2288-07

Combining Graph Learning and Region Saliency Analysis for Content-Based Image Retrieval

FENG Song-he^{1,2}, LANG Cong-yan¹, XU De¹

(1. School of Computer & Information Technology, Beijing Jiaotong University, Beijing 100044, China;

2. Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: For the image retrieval task which combines machine learning theory with relevance feedback mechanism, this paper focuses on the graph-based semi-supervised learning algorithm with application to region-based image retrieval. Different schemes which both incorporate the region saliency into the graph-based semi-supervised learning framework are applied to deal with two types of feedback. Firstly, in the case that no sample or only positive samples are available from the user's feedback, the retrieval task can be resolved via a transductive learning manner, a hierarchical graph model which incorporates region saliency information is constructed and the manifold-ranking algorithm is adopted subsequently for positive label propagation. Secondly, in the case that the user provides both positive and negative samples, the region-level adjacency matrix will be constructed via the feedback samples, and the manifold-ranking algorithm is also adopted here to choose instances which truly represent the user's query semantics. The selected instances are then used to retrieve the relevant samples. The experiments have proved the effectiveness of the proposed method.

Key words: content-based image retrieval; region saliency; graph learning; manifold-ranking; relevance feedback

1 引言

随着多媒体技术和计算机网络技术的发展, 人们接触到的图像数据以前所未有的速度增长, 面对海量的图

像资源, 用来有效地分析、组织和管理图像数据的基于内容的图像检索技术 (Content-based Image Retrieval, CBIR) 成为多媒体技术的研究热点^[1]. 研究的难点在于如何使计算机能够从人的认知角度来理解图像语义信

息,最大程度地弥合图像低层特征和高层语义之间的语义鸿沟.基于内容的图像检索系统给出的初始检索结果往往不能很好地满足用户的信息需求,因此在文本检索领域提出的相关反馈(Relevance Feedback, RF)技术被引入到基于内容的图像检索领域^[2].相关反馈技术通过把人的参与引入到信息检索过程中,从而把检索模式从一次进行变成交互式的多次进行,并成为提高检索效能的有效方法.

在基于相关反馈机制的图像检索方面,学者们已经做了大量的研究,如吴洪等人^[3]将现有的基于相关反馈的图像检索算法做了较为全面的梳理和比较.近年来,许多工作把图像检索看成是不同类型的监督学习问题,并针对该学习问题的特点把各种机器学习方法引入到相关反馈算法的研究中.现有的相关反馈机制大多采取由用户标记出部分正例图像和反例图像,由此将图像检索问题转化为一个典型的二分类问题.包括支持向量机(SVM)^[4]、决策树^[5]、主动学习^[6]等机器学习算法都已被应用到图像检索中.由于图像检索应用的特殊性,相关反馈中的学习问题表现出与其他学习问题一些不同的特点,主要存在诸如训练样本少、实时性的要求高、训练样本的不对称性以及存在大量的未标记样本等特性.由于在相关反馈中虽然标记的训练样本很少,但存在大量的未标记样本(整个图像特征库),如何有效利用图像库中大量的未标记样本,在半监督学习框架下解决图像检索问题,是目前图像检索领域一个新的研究热点.半监督学习的主要思想是充分利用数据的总体空间分布特征和原始类别标记信息,使得最终得到的分类结果在数据空间总体上能够充分平滑,同时保证尽可能地拟合训练数据.常用的半监督学习算法包括:协同训练^[8,13]、直推式支持向量机(Transductive SVM)^[9]和基于图的方法^[10,11]等.其中,基于图的半监督学习方法受到越来越多的关注,并且已被应用到了图像检索中^[12].

本文针对用户参与相关反馈的情况,结合图像区域的显著性分析,分别提出了两种解决方案.在无用户反馈和用户只反馈正例样本图像的情况下,算法将图像检索问题视为一个封闭的直推式学习(Transductive Learning)问题.通过将半监督学习策略与区域显著性信息融合起来,提出了一种层次化的图模型构建算法,并采用流形排序算法(Manifold-Ranking)实现标记的传播,完成图像检索的过程;而在用户同时反馈正反例图像的情况下,算法将用户标记的正反例图像的区域集合构建一个相似性邻接矩阵,并融合区域的显著性利用流形排序进行相似性的传播,迭代计算后得到最接近用户感知的区域集合将被作为检索查询目标区域集,并以此用于检索.这种融合的优点在于:相关反馈策略

采用的是一种自上而下的机制,将用户的主观感知引入到检索系统中;而图像显著性分析则提供了一种自下而上的表示机制.如果将这两个不同的图像语义提取机制有机地融合起来,则既考虑了用户的主观需求,又考虑到了图像本身的特性.

2 基于图学习算法与区域显著性分析相融合的检索

本文将显著性分析与基于图的半监督学习算法有机融合起来,研究基于区域的图像检索中的学习问题,并针对用户参与反馈的情况,分别提出了两种检索策略.一种是针对用户无反馈和只提供正例反馈图像的情况,我们提出了基于显著性分析和层次化图表示的半监督学习算法用于检索.另一种则针对用户同时提供正例反馈图像和反例反馈图像的情况.受文献^[14]的启发,我们提出了融合区域显著性信息的目标区域选取算法,根据用户反馈的正反例样本图像学习得到用户的查询概念,并基于选取的目标区域实现图像检索.需要指出的是,后一种算法既考虑了图像中区域本身的显著性信息(自底向上机制),又融入了用户的主观查询意图(自上而下机制),实现了显著区域和用户感兴趣区域两者概念的有机融合.

2.1 基于流形排序的图像检索算法

流形排序算法作为一种基于图学习算法,它既是一种半监督学习算法,同时也是一种直推式学习算法(Transductive learning),即已标记训练样本和未标记的测试样本都将参与到算法的学习过程中.其理论基础是:(1)特征空间上距离相近的两个样本点,其具有相似标记的概率较高;(2)特征空间上具有相似流形结构的样本点,其具有相似标记的概率较高.第一个假设属于局部相似性问题,而第二个假设属于全局相似性问题.文献^[10,11]采用了图的方法来逼近流形,并通过图理论来学习流形空间中各数据点的分类情况.这种图学习的过程可以简单地理解为基于数据点间的关系进行类标记传播的过程(由于篇幅限制,关于流形排序算法的详情请参阅文献^[10,11]).同理,若我们将每幅图像作为图节点,以图像间的相似关系作为边,通过图学习算法就可以实现将待检索图像的标记传播到未标记图像中,从而完成图像检索的任务^[15,16].在基于图的半监督学习算法中,邻接矩阵 $\mathbf{W}^{N \times N}$ 是整个算法的核心,它反映了特征向量空间中两点之间的相似性度量,也决定了标记传播的概率.构建 $\mathbf{W}^{N \times N}$ 需要考虑两部分的内容:第一,构成邻接矩阵的节点和边的连接方式;第二,边权值的度量定义.邻接矩阵的节点通常由图像的全局特征或者图像中的区域特征表示;对于边的连接方式,则有全连通(full connection)和 K 近邻两种方式.

2.2 融合层次化图模型与区域显著性分析的检索

本算法的设计主要针对以下两种情形:(1)用户不参与相关反馈,即用户只提供一幅待检索图像;(2)用户只反馈正例图像,即系统每次返回检索结果图像集合时,用户只标记出符合查询意图的正例图像集合;针对上述两种反馈情形,提出了将图像区域显著性分析融入到流形排序算法中,通过一种层次化的图表示来构建相似邻接矩阵,以实现标记传播并完成图像的语义检索;算法的优势在于,既考虑融入了大量的未标记样本,同时又在注意力机制的指导下自适应地分配图像区域的权重.算法采用层次化图表示的方式,分别构建区域级邻接图和图像级邻接图,有效地融合了上述两种邻接图构建的优点,且融入了区域显著性的度量,符合用户的查询感知.

(1) 显著性定义

这里给出区域平均显著性的定义.给定图像库中的一幅图像 I ,使用 JSEG 分割算法将其分成若干个同质区域 $I = \{r_i\}, i = 1, \dots, |I|$,其中 r_i 表示图像 I 的第 i 个区域, $|I|$ 表示图像 I 的区域总数.对于每一个分割区域 r_i ,其所对应的区域平均显著性:

$$SR(r_i) = \frac{\sum_{x \in r_i} SP(x)}{Area(r_i)} \quad (1)$$

其中 $SP(x)$ 表示区域 r_i 中每个像素点 x 的显著值, $Area(r_i)$ 是区域 r_i 的面积.像素点 x 的显著值 $SP(x)$ 的定义如下:

$$SP(x) = \sum_{l=1}^L \sum_{y \in \theta_x} (\gamma_{cl} S_{cl}^l(x, y) + \gamma_0 S_0^l(x, y)) \quad (2)$$

其中, $l = [1 \dots L]$ 表示第 l 层高斯金字塔,这里取 $L = 3$ 表示我们建立的是三层高斯金字塔. $y \in \theta_x$ 表示像素点 x 的 θ_x 中的任意一个像素点,这里我们取 θ_x 为以 x 为中心的 3×3 邻域作为感受野的大小. γ_{cl} 和 γ_0 分别表示颜色亮度特征对比值和方向特征对比值的权重,这里取 $\gamma_{cl} = \gamma_0 = 1$. $S_{cl}(x, y)$ 和 $S_0(x, y)$ 分别表示像素点 x 与 y 之间的颜色-亮度以及朝向对比差值.最后,为了去除显著图中的噪声点,我们用一个高斯滤波器对得到的显著图进行滤波, $SP(x) = SP(x) \otimes G(x)$,其中, $G(x)$ 为标准方差 $\sigma_G = 1$ 的高斯滤波器(显著图生成算法详见参考文献[17]).计算得到图像库中每幅图像各区域的显著值并且归一化到 $[0 - 1]$ 后,用来度量区域在标记传播时的重要度.

(2) 区域级邻接矩阵构建

受文献[18]的启发,算法首先给出区域级邻接矩阵中标记传播的准则:(a)正例图像中的显著区域赋予较高的标记传播概率;(b)未标记样本图像中与正例图像显著区域相似的区域有较高的标记传播概率;(c)限制图像中不显著区域的标记传播概率.

算法首先构建抽象的区域级邻接矩阵.这里所谓的抽象,是指算法无需真实地构建区域级邻接矩阵,而只是给出其定义.邻接矩阵中的每个节点表示为一个区域,节点的属性值由区域所对应的低层视觉特征向量(如区域的颜色矩特征等)和区域的显著值共同表示.接下来我们给出区域级邻接矩阵中边权值的定义.依据上述准则,算法将区域的视觉特征和显著性信息融入到区域的相似性度量中,并以此作为边的权值.具体的,对于邻接矩阵中任意两个节点 r_i, r_j ,其对应的边权值 $w_R(r_i, r_j)$ 定义为:

$$w_R(r_i, r_j) = \frac{1}{2} (SR'(r_i) + SR'(r_j)) \exp(-dist^2(r_i, r_j)/\sigma^2) \quad (3)$$

其中, $SR'(r_i)$ 和 $SR'(r_j)$ 分别表示区域 r_i 和 r_j 归一化后的显著值. $dist(r_i, r_j)$ 表示区域 r_i 和 r_j 对应低层特征的欧式距离.从式(3)中可以看出,不同于传统的邻接矩阵中单纯考虑节点间视觉特征的相似性,本算法中邻接矩阵的边权值定义同时考虑了区域间的视觉特征相似性和区域的显著性.式(3)可理解为,当且仅当两个区域视觉特征相似,且这两个区域的显著值都较高的情况下,其边权值较高;若仅是视觉特征相似,而两个区域中任意一个区域的显著值较低,也会导致其边权值较低.这种边权值的定义方式较好地体现了面向感知的图像检索,当且仅当图像库中的待检索图像的某区域与查询图像中某区域相似,且两者都具有较高的显著值时,才以较高的概率进行标记的传播;反之,若待检索图像的某区域只与查询图像中显著值较低的区域视觉特征相似,标记的传播概率较低.当且仅当两个区域互为 k 近邻时,在这两个区域节点间建立一条边,并采用式(3)来衡量两个区域之间的相似性.

(3) 基于层次图表示的流形排序算法

给定区域级邻接矩阵的定义,我们以此构建图像级邻接矩阵.这里定义的图像级邻接矩阵,即是用于流形排序算法的加权邻接矩阵 $W^{N \times N}$.图中每个节点表示图像库中的任一幅图像,两个节点间的边权值定义为其对应的两幅图像之间的语义相似度,其定义如下:

$$w_I(I_u, I_v) = \sum_{r_i \in I_u, r_j \in I_v} w_R(r_i, r_j) \quad (4)$$

这里,边权值定义为给定两幅图像中所有区域的边权值之和.需要指出的是,我们采用了较为简单的全匹配策略,而没有利用比较经典的区域匹配策略如 IRM^[19],这主要是基于节省计算复杂度的考虑.给定图像级邻接矩阵后,我们可以将其应用到流形排序算法中.具体的,若检索过程没有用户参与反馈,则将用户提交的查询图像作为唯一的正例样本图像,即初始概念标记为 1.图像库中其他的待检索图像作为未标记样本,初始

概念标记为 0. 根据构建好的图像级邻接图将查询图像的标记传播到图像库中的待检索图像, 算法迭代收敛到稳定态后, 根据图像的结果标记值, 按照由大到小的顺序排列, 取出标记值最高的前若干幅图像返回给用户, 即完成了图像检索整个过程. 同理, 若用户参与反馈并只提交正例样本图像, 则只需要将用户反馈的正例样本图像的初始概念标记设为 1, 其他步骤同上即可.

2.3 结合相关反馈机制与区域显著性分析的检索

上述算法解决的是用户不参与反馈以及只反馈正例图像的情形. 但其存在一个前提假设: 即用户对图像中区域的感兴趣程度与基于注意力机制得到的区域显著性分析基本一致. 但是用户感兴趣区域是一个自上而下的概念. 由于人类视觉感知的主观性, 不同的用户针对同一幅图像, 得到的感兴趣区域不尽相同; 即使是同一用户针对同一幅图像, 在不同的查询条件下其感兴趣的区域也并不总是一致. 而显著区域的概念更多的是基于自下而上的机制, 其取决于图像内容本身, 而与用户无关. 因此, 本算法旨在同时考虑到用户的感兴趣区域与显著区域, 即在显著性分析的指导下学习用户的感兴趣区域.

本节算法考虑的用户参与反馈的情况是: 用户同时反馈正反例样本图像, 即系统每次返回检索结果图像集合时, 用户根据是否与查询概念相似, 选择部分图像分别标记为正例样本图像和反例样本图像. 标记为正例样本的图像集合中的每个正例样本一定存在用户感兴趣的区域; 而标记为反例样本的图像集合中, 每个反样本均不存在用户感兴趣区域. 据此本节主要分析如何根据给定的正反例图像来学习用户的潜在查询语义概念.

(1) 基于流形排序的目标区域选取

具体地, 给定用户反馈标记的正反例图像集合 $L = L^+ \cup L^-$, 其中 $L^+ = \{I_1^+, I_2^+, \dots, I_{|P|}^+\}$, $L^- = \{I_1^-, I_2^-, \dots, I_{|N|}^-\}$; $|P|$ 和 $|N|$ 分别表示正例图像和反例图像的数目. $R = \{r_1, r_2, \dots, r_{|R|}\}$ 表示集合 L 中的所有图像的区域集合, $|R|$ 为集合 L 中各图像所包含的区域数目总和. 将 R 中每个元素 $r_i \in R$ 作为邻接矩阵 \mathbf{W} 中的一个节点, 组成区域级相似性邻接矩阵 $\mathbf{W}^{|R| \times |R|}$, 矩阵中任意一个元素 w_{ij} 反映了区域 r_i 和 r_j 之间的视觉特征相似度. 我们将区域的显著性信息体现在初始标记向量上. 具体地, 由于用户标记的反例图像中不存在用户感兴趣区域, 则若任一区域 $r_i \in L^-$, 则将 r_i 的初始类别标记 y_i 可设为 -1 ; 而对于用户标记的正例图像, 由于其标记信息是标记于整幅图像上, 而正例图像中通常含有部分反例噪声区域, 因而系统并不能获取正

例图像中每个区域的初始类别标记. 基于此, 算法假定区域 $r_i \in L^+$ 的初始标记值根据其区域显著性大小, 赋予其 $[0-1]$ 之间的实数标记, 标记值越高, 表示区域 r_i 反映用户查询概念的概率就越大. 设初始类别标记向量 $\mathbf{Y} = [y_1, y_2, \dots, y_{|R|}]^T$, 则

$$y_i = \begin{cases} SR^+(r_i), & r_i \in L^+ \\ -1, & r_i \in L^- \end{cases}, i = 1, \dots, |R| \quad (5)$$

将上述相似性邻接矩阵 $\mathbf{W}^{|R| \times |R|}$ 和初始类别标记向量 \mathbf{Y} 代入到流形排序算法中, 经迭代收敛后得到标记样本图像中各个区域的最终标记 y_i^* ($i = 1 \dots |R|$). 需要指出的是, 传统的流形排序是将类别标签从已标记样本传播到未标记样本. 而本算法的一个假设是, 用户所提交的正例样本图像在特征空间中能够形成比较紧的聚类, 而反例样本图像则往往分散地分布在其余的空间中. 这种假设也符合目前图像检索的相关反馈机制的特性. 经过标记传播算法进行标记传播以后, 表征用户查询语义概念的正例图像中的感兴趣区域往往会得到较高的感兴趣值. 接下来, 选取标记值较高的区域集合来表征用户的查询语义概念. 这里通过阈值 y_{thres} 来选取满足条件的目标区域集合, 参考文献[14]其定义如下:

$$y_{thres} = \max \{ y^* \mid \forall I_i^+ \in L^+, \exists r_i \in I_i^+, \text{ s.t. } y_i^* \geq y^* \} \quad (6)$$

式(6)表示从满足当且仅当每个正例图像中至少有一个区域的标记值高于给定阈值的候选阈值列表中, 选择最大值作为最终的阈值 y_{thres} . 其出发点在于, 用户反馈标记为正例的图像中至少存在一个区域符合用户的主观查询概念.

(2) 基于目标区域的检索算法

定义迭代后选取的符合用户查询语义的感兴趣区域集合为 $R^* = \{r_1^*, r_2^*, \dots, r_{|R^*|}^*\}$, 系统利用反馈学习得到的 R^* , 从图像库中选择与 R^* 相近的图像返回给用户. 设图像库中的待检索的图像集合表示为 $U = \{I_1, I_2, \dots, I_{|U|}\}$, 其中每个图像 $I_i = \{p_1^i, p_2^i, \dots, p_{|I_i|}^i\}$. 由于每幅图像由若干同质区域组成, 这里采用最小 Hausdorff 距离来定义待检索图像 $I_i \in U$ 与目标区域集合中任一目标区域 $r_j^* \in R^*$ 的距离:

$$d(r_j^*, I_i) = \min_{k=1, \dots, |I_i|} \text{dist}(r_j^*, p_k^i) \quad (7)$$

其中 $\text{dist}()$ 为表征两个区域的视觉特征间的欧式距离. 则待检索图像 $I_i \in U$ 符合用户查询概念的概率为:

$$l_i = \frac{1}{|R^*|} \sum_{j=1}^{|R^*|} e^{-d(r_j^*, I_i)^2} \quad (8)$$

图像检索的结果则根据图像 I_i 的概率值 l_i 按从大到小

的顺序排列返回给用户.若用户需要进一步的反馈,则按照如上算法循环执行,直到用户满意为止.

3 实验结果分析

3.1 实验设置

我们选用图像检索领域较为常用的两个基准图像库来验证算法的有效性.其一,采用了 COREL 图像集,其中包括 5000 幅彩色图像,它们分别属于 50 类,每类 100 幅图像;其二,采用了 SIVAL 图像集^[20],共 1500 幅彩色图像,它们分别属于 25 类,每类的数量为 60 幅.我们给出本文提出的两种检索算法在三种不同的反馈情形下的检索结果.所有实验在 MATLAB2008b, 3.0G/2G RAM 实验平台下进行.

3.2 融合层次化图模型与区域显著性分析的算法性能分析

该小节待验证的学习算法即为融合区域显著性信息的层次化图表示模型.本算法在 COREL 图像库上进行验证.实验中图像数据库已知(即为 COREL 数据集),用户所提交的查询图像也是从该数据库中随机抽取,即算法的实验是在封闭的数据集内进行.算法的优点在于,无需用户反馈反例样本图像即可实现标记信息的传播.用户提交的查询图像通常被看做是标记正样本,而待检索的图像库中的图像集合看作是未标记样本.算法的目标就是将标记信息通过构建好的相似性邻接矩阵从标记样本传播到未标记样本中.如上文所述,该算法适用于无用户反馈以及用户只反馈正例样本图像的检索情形.由于针对以上两种检索情形,算法只需调整初始状态向量,因此实验中重点关心在无用户反馈的情形下,算法与其他现有的基于流形排序理论的检索算法性能的比较.算法任取 COREL 图像库中的 1 幅图像作为查询图像,剩余的 4999 幅图像作为待查询图像集合.检索的查准率由这 5000 次查询结果的平均值决定.这里将本文所提算法(记为“Hierarchical-MR”)与以下几种典型的图像检索算法进行比较:(1)基于全局特征的流形排序算法^[12],记为“Global-MR”;(2)基于块特征的流形排序算法^[15],记为“Block-MR”;(3)基于融合区域匹配的图像检索系统 SIMPLcity^[19],记为“IRM”.

具体来说,对于“Global-MR”算法,我们提取 64 维 HSV 颜色直方图特征和 6 维小波变换纹理特征组成每幅图像的低层视觉特征,并采用 L1 距离度量两幅图像之间的相似性,并以此作为两幅图像间边的权值.对于“IRM”算法,由于融合区域匹配策略是 SIMPLcity 系统的核心,因此这里只实现了其融合区域匹配策略,而用 JSEG 图像分割算法替代其原有的自适应 k -均值算法进行图像分割;分割后的每个区域我们提取其颜色矩信

息作为分割区域的低层特征.对于“Block-MR”算法,我们采用与文献^[15]中相同的特征表示.实验中定义 $k = 50$, $\alpha = 0.99$,这里 k 值表示算法在构建区域级邻接矩阵时采用了基于 k -NN 的方式.由于本算法属于直推式学习算法,因此邻接矩阵的构建只需要执行一次即可,每次检索时只需改变初始标记向量的值.图 1 给出了各个算法平均检索查准率的比较结果.图 1 横坐标为检索返回的图像数量,纵坐标为各算法的平均查准率.从图中可以看出,基于区域显著性分析的层次化图模型检索算法,其检索平均查准率要明显高于上述其他三种算法.

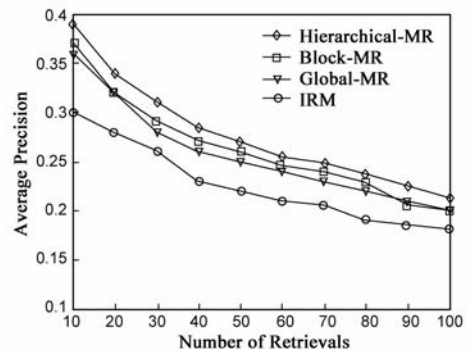


图1 Hierarchical-MR与现有算法的检索查准率的比较

3.3 结合相关反馈机制与区域显著性分析的算法性能分析

由于本算法主要处理用户同时反馈正反例样本图像的检索情形,因此为验证算法的有效性,我们分别在 COREL 图像库和 SIVAL 图像库中进行实验.在基于 COREL 图像库的实验中,我们主要采取以下两种算法与之进行比较:即文献^[7]提出的采用 EMD 核函数的 SVM 学习算法,记为 EMD-SVM.该算法解决了已有的 SVM 中的核函数不适用于图像区域级表示的问题,通过对高斯核函数进行推广实现了将 SVM 算法应用于区域级图像表示的检索反馈机制中.文献^[14]提出的半监督多示例检索算法,记为 SSMI;而在基于 SIVAL 图像库的实验中,本算法与基于多示例学习的 ACCIO 算法^[21]进行比较.

与上述算法相比,本算法的优势在于:首先该问题不再局限于求解直推式学习问题.虽然本算法也应用了流形排序这一基于图的半监督学习算法,但算法的应用点则已经不再是将标记从正样本传播到未标记样本,而是将正例图像中的区域标记和反例图像中的区域标记通过相似性邻接矩阵进行标记的转移.由于该算法改进了文献^[14]的算法,因此其本质思想仍然是一个求解多示例学习中的目标示例的问题.只是该目标示例的求解不是采用多样性密度算法,而是通过基于图学习的半监督学习算法来实现.但不同的是,本算法

可以将直推式学习推广到判别式学习问题中.因为本算法选取的是若干个符合用户查询语义的目标区域集合.利用该目标区域集合再次对图像库进行查询,即可获得查询图像集合.通过多次迭代以后,即能改进图像检索的性能.而[14]则是一个直推式学习问题,即它采用了二次流形排序算法,这样就限制了问题求解的扩展,而且每次迭代均要重新计算相似性邻接矩阵,系统响应时间过长,影响了检索性能.这不同于上一个子算法只需要执行一次相似性邻接矩阵的构建.

(1) 基于 COREL 图像库的算法性能比较

实验首先采用最简单的检索算法进行初始检索.由于图像库图像数量较多,通过用户真实地进行标注反馈耗时耗力,因此这里通过系统对用户的反馈过程进行仿真.具体说,为了模拟算法相关反馈的过程,在每次查询后通过判别检索图像与待查询图像是否属于同一语义类别而将其划分为正例图像和反例图像,系统自动地从初始检索结果中选取 n 幅正例图像和 n 幅反例图像交予系统进行学习.在系统得到反馈的正反例图像后,EMD-SVM 算法可以学习得到分类器,并用于下轮图像检索.图 2 给出了首轮反馈提交 20 幅正反例图像后经过学习得到的检索结果,其中,纵坐标为平均查准率,用以衡量算法的检索性能.

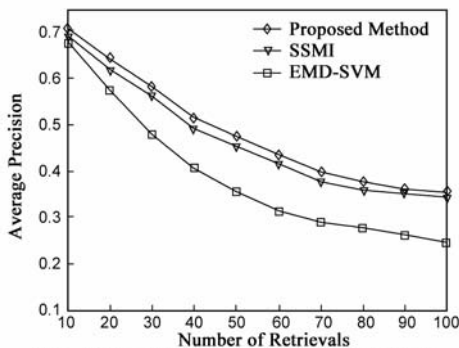


图2 基于20幅正反例图像的检索性能比较

从图 2 中可以看出,本算法的检索性能要明显优于基于 EMD-SVM 的检索算法;而且与 SSMI 算法相比,本算法通过在初始状态向量中融入正例图像各区域的显著性后,检索性能有了一定提高.此外,由于 SSMI 算法采用二次流形排序用于标记传播,因此其属于直推式学习问题.而本算法由于通过流形排序后得到若干

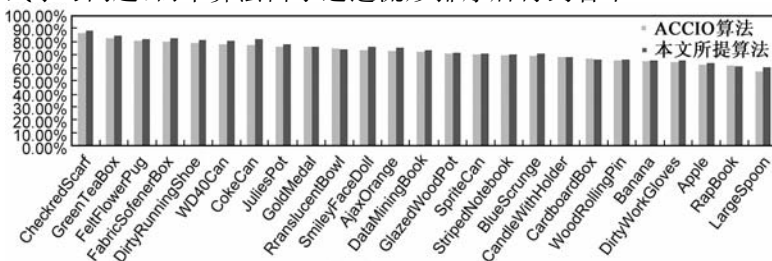


图3 基于SIVAL图像库的检索性能对比

表征用户查询语义的区域集合,并以此区域集合对图像进行检索,因而属于判别式学习问题,可以将其扩展到未知图像库中.另外,由于文献[7]和[14]均没有考虑到不同区域的显著性差异,因此本算法在检索性能上更加贴近用户的感知,因而也取得了更好的效果.

(2) 基于 SIVAL 图像库的算法性能比较

这里我们将本节算法与文献[20]在 SIVAL 图像库上进行比较.该图像库侧重于验证局部化图像检索(Localized CBIR)的思想,即如何通过多示例学习算法学习到用户的查询概念,并利用该查询概念实现检索.为了客观地比较两个算法之间性能的差异,SIVAL 库中每幅图像采用改进的层次化分割算法进行分割,每个分割区域采用一个 30 维的特征向量来表示.文献[21]提出的 ACCIO 算法,采用多样性密度算法在给定的正反样本图像上进行学习以得到目标示例集合,并以此作为用户的查询语义概念进行检索.而本节算法采取的是通过将用户标记的正反例样本构建区域级邻接矩阵,并采用融合区域显著性信息的基于图的半监督学习算法学习出用户查询语义.从本质上分析,两者均将图像检索问题作为歧义性问题来处理,即均假定正例图像中存在用户不感兴趣的伪区域.对于每个语义类别进行 5 轮实验,每轮实验通过系统仿真反馈随机标记出 8 个正例样本和 8 个反例样本.由于每个语义类别包含有 60 幅图像,这里采用每个语义类别的 5 轮检索后的平均查准率(即系统检索返回的前 60 幅图像中正例图像的数目)来反映算法的检索性能.图 3 给出了两种算法的检索结果性能比较,从中可以看出,由于同时融入了区域的显著性信息和用户反馈的高层语义,本算法的检索性能比单纯使用反馈机制的 ACCIO 算法[21]的检索性能有所提高.

4 结论

本文主要研究基于区域的图像检索中的学习问题,重点分析如何在基于图的半监督学习框架下,通过融合区域的显著性信息和相关反馈策略来改进图像检索的性能.针对用户参与反馈的情况,分别提出了两种检索方案,并相应地将区域显著性信息分别融入邻接矩阵的构建和初始标记向量中,实现了自下而上的图像显著性分析机制和自上而下的用户反馈机制的有机融合.通过在 COREL 图像库和 SIVAL 图像库上的实验仿真,我们得出的结果验证了上述两种检索策略的可行性和有效性,且检索算法也更加贴近用户的感知.

参考文献

- [1] Datta R, Joshi D, Li J, Wang J Z. Image retrieval: ideas, influences, and trends of the new age[J].

- ACM Computing Surveys, 2008, 40(2): 1 – 60.
- [2] Rui Y, Huang T S, Ortega M, Mehrotra S. Relevance feedback: A power tool for interactive content-based image retrieval[J]. IEEE Trans on Circuits and Systems for Video Technology, 1998, 8(5): 644 – 655.
- [3] 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[J]. 计算机学报, 2005, 28(12): 1969 – 1979.
Wu H, Lu H, Ma S. A survey of relevance feedback techniques in content-based image retrieval [J]. Journal of Computer, 2005, 28(12): 1969 – 1979. (in Chinese)
- [4] 邬俊, 鲁明羽, 刘闯. 基于混合学习框架的 SVM 反馈算法研究[J]. 电子学报, 2010, 38(9): 2101 – 2106.
Wu J, Lu M, Liu C. SVM-feedback scheme within hybrid learning framework for image retrieval [J]. Acta Electronica Sinica, 2010, 38(9): 2101 – 2106. (in Chinese)
- [5] 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.
- [6] Zhang C, Chen T. An active learning framework for content-based information retrieval [J]. IEEE Trans on Multimedia, 2002, 4(2): 260 – 268.
- [7] Jing, F, Li, M, et al. An efficient and effective region-based image retrieval framework [J]. IEEE Trans on Image Processing, 2004, 13(5): 699 – 709.
- [8] Zhou Z H, Chen K J, et al. Enhancing relevance feedback in image retrieval using unlabeled data[J]. ACM Trans on Information Systems, 2006, 24(2): 219 – 244.
- [9] 陈毅松, 汪国平, 董士海. 基于支持向量机的渐进直推式分类学习算法[J]. 软件学报, 2003, 14(3): 451 – 460.
Chen Y, Wang G, Dong S. A progressive transductive inference algorithm based on support vector machine[J]. Journal of Software, 2003, 14(3): 451 – 460. (in Chinese)
- [10] Zhou D, Bousquet O, Lal T N, Weston J, Scholkopf B. Learning with local and global consistency[A]. Proc of Advances in Neural Information Processing Systems (NIPS'03)[C]. Vancouver: Curran Associates Inc, 2003. 321 – 328.
- [11] Zhou D, Weston J, Gretton A, Bousquet O, et al. Ranking on data manifolds[A]. Proc of Advances in Neural Information Processing Systems (NIPS'03)[C]. Vancouver: Curran Associates Inc, 2003. 169 – 176.
- [12] He J, Li M, Zhang H J, Tong H H, Zhang C. Generalized manifold-ranking based image retrieval [J]. IEEE Trans on Image Processing, 2006, 15(10): 3170 – 3177.
- [13] Zhou Z H, Chen K J, Dai H B. Enhancing relevance feedback in image retrieval using unlabeled data[J]. ACM Trans on Information Systems, 2006, 24(2): 219 – 244.
- [14] Tang J, Hua X S, Qi G J, Wu X. Typicality ranking via semi-supervised multiple-instance learning [A]. Proc of ACM Int Conf. on Multimedia[C]. Augsburg: Association for Computing Machinery, 2007. 297 – 300.
- [15] Wan X. Content based image retrieval using manifold-ranking of blocks[A]. Proc of IEEE Int Conf on Multimedia and Expo (ICME'07)[C]. Beijing: IEEE Computer Society, 2007. 2182 – 2185.
- [16] Li F, Dai Q, Xu W, Er G. Correlated probabilistic label propagation for region-based image retrieval[A]. Proc of IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP'07) [C]. Hawaii: IEEE Computer Society, 2007. 789 – 792.
- [17] S H Feng, D Xu, X Yang. Attention-driven salient edge(s) and region(s) extraction with application to CBIR[J]. Signal Processing, 2010, 90: 1 – 15.
- [18] Rahmani R, Goldman S A. MISSL: multiple-instance semi-supervised learning[A]. Proc of Int Conf on Machine Learning (ICML'06) [C]. Pittsburgh: Association for Computing Machinery, 2006. 705 – 712.
- [19] Wang J Z, Li J, Wiederhold G. SIMPLiCity: semantics-sensitive integrated matching for picture libraries[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(9): 947 – 963.
- [20] Rahmani R, Goldman S A, Zhang H, Cholleti S R, Fritts J E. Localized content based image retrieval [J]. IEEE Trans on Pattern Recognition and Machine Intelligence, 2008, 30(11): 1902 – 1912.
- [21] Rahmani R, Goldman S A, Zhang H, Krettek J, Fritts J. Localized content based image retrieval[A]. Proc of ACM Int Conf on Multimedia Information Retrieval (ACM MIR'05) [C]. Singapore: Association for Computing Machinery, 2005. 227 – 236.

作者简介



冯松鹤 男, 1981 年生于江苏盐城, 北京交通大学计算机与信息技术学院讲师, 主要研究方向为图像语义检索及自动标注.

E-mail: shfeng@bjtu.edu.cn



郎丛妍 女, 1978 年生于内蒙古呼和浩特, 北京交通大学计算机与信息技术学院副教授, 主要研究方向为视觉信息认知计算.

E-mail: cylang@bjtu.edu.cn