

基于改进的排序学习的图片检索算法研究

谭光兴 刘臻晖

(江西财经大学信息管理学院 南昌 330013)

摘要 图片检索是图片共享社会网络中的重要研究内容之一。传统的图片检索方法往往通过对用户输入的关键字和图片的文本描述加以匹配来进行图片检索。由于文本信息存在歧义性,图片的文本描述十分困难,因此检索结果的准确性低。为了提高图片检索的准确性,提出了基于排序学习的图片检索方法。将每幅图片通过多种特征描述符进行描述,当用户的输入为图片时,通过对比查询图片和图片库中图片的相似性进行图片检索。采用支持向量机和关联规则两种学习方法对特征描述符的权重组合进行学习,并提出了相应的学习算法。实验表明,提出的基于学习的图片检索方法与相关图片检索方法相比具有更高的准确性。此外,应用支持向量机和关联规则两种方法对分类函数进行学习时,由于两种算法通过相同的数据实例对图片描述符的权重进行学习,因此得到的结果是相关的。

关键词 图片检索,排序学习,支持向量机,关联规则

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.12.059

Learning to Rank Based Approach for Image Searching

TAN Guang-xing LIU Zhen-hui

(School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330013, China)

Abstract Image searching is one of the most important researches in image sharing based social networks. Traditional image searching methods usually compare the user keywords and the textual description of images in database while searching. Because the textual description is ambiguous, the abstracting of text for images is very hard, and thus the accuracy of image searching is low. In order to improve the accuracy of image searching, this paper proposed a learning to rank based approach. We described each image as a combination of multiple feature descriptors, and compared the similarity of the query and the image in database while users input a query of image. We applied association rules and support vector machine to learn the weight of each feature descriptor, and proposed corresponding learning algorithms. The experiments show that the proposed image searching approach is more accurate than related works while retrieving image for users. In addition, while using association rule and support vector machine to learn the classification functions, the two algorithms use the same instances to learn the weight of each feature descriptor, so they are relevant.

Keywords Image searching, Learning to rank, Support vector machine, Association rule

传统的图片搜索方法主要基于搜索的关键字是否与图片的文本元数据相匹配,其研究面临着严峻的挑战^[1]。应用文本特征描述图片的内容十分困难,此外随着图片数量的增长和图片类型的多样化,该项工作变得越来越困难。大量真实的图片数据应用,尤其是通用图片数据库,很难得到图片内容的文本描述。在这些大型的图片数据库中,可用的上下文文本信息稀少,有些文本信息甚至是不可靠的,对图片的内容进行人工标注文本信息的代价巨大^[2]。此外在查询处理中,用户的需求无法预测,因而搜索结果的可信度不高。

基于图片内容的图片搜索方法^[3]可以有效地解决上述问题,该方法摒弃了图片的文本元数据和关键字搜索^[4,5],取而代之的是根据查询实例图片检索出与之最相似的图片。在基于图片实例的图片搜索中,一个关键的研究任务是待检索的图片进行排序,采用的方法是根据待检索图片与查询图片

的相似性对图片进行排序^[6]。由于图片数据量巨大,用户往往只关心排名靠前的部分图片,因此用户对检索系统的性能感知往往只取决于排名靠前的部分图片。

在基于图片实例(或者称为内容)的图片搜索^[7,8]中,每幅图片用其自身的图片内容来描述,即图片描述符。每个图片描述符包含一个特征向量,图片间的相似性通过描述符之间的相似性进行计算。图片描述符向量将图片的颜色、材质和形状等视觉特征进行编码,描述了该图片在特征向量空间上的分布情况^[8,9]。图片的相似性可以采用简单的相似性度量方法,如欧几里得距离,也可以采用一些复杂算法,如 Earth Mover 距离^[10]。

在对图片进行排序时,针对不同的描述符,需要不同的排序方法,最佳的描述符是数据相关的,在查询之前是未知的。此外,当用描述符向量来表示图片的内容时,不同的图片描述

到稿日期:2015-01-23 返修日期:2015-02-14 本文受江西省博士研究生创新项目科研基金(YC2011-B026)资助。

谭光兴(1954—),男,博士,教授,博士生导师,主要研究方向为数据分析与数据挖掘、高等教育管理;刘臻晖(1977—),男,博士生,主要研究方向为数据分析与数据挖掘, E-mail: liuzh@jxufe.edu.cn(通信作者)。

符之间是相互补充的,因此对不同的描述符进行融合可以更好地描述图片的内容信息,从而可以提高图片的排名质量。虽然对图片的多种描述符进行融合可以更好地描述图片内容,但是描述符的最优组合仍然与具体的图片相关,且在查询处理之前是无法预测的^[11]。

本文采用排序学习方法对图片进行分类,从而进行图片的检索,提出了基于排序学习的图片检索方法。将每幅图片通过多种特征描述符进行描述,当用户的输入为图片时,通过对比查询图片和图片库中图片的相似性进行图片检索。采用支持向量机和关联规则两种学习方法对特征描述符的权重组合进行学习,并提出了相应的学习算法。

1 问题描述

在基于内容的图片检索中,系统将图片按照内容属性编码成向量,然后根据用户预定义的模式检索出相似的图片。在图片的编码过程中,采用图片的内容描述符,每个描述符表达了图片的相关感知质量。

图片的每个描述符由特征提取算法和相似性度量两部分组成。特征提取算法将图片的颜色、形状和材质等特征转化为向量;相似性度量方法通过比较两幅图片的特征向量,计算它们之间的相似性或者距离。在实际的图片数据库中,特征描述符的性能依赖于具体的图片内容,本文采用文献[9]中的18种特征描述符。

不同的特征描述符分别描述图片的不同属性,如颜色、形状和材质等,它们在计算图片间的相似性时相互补充。本文采用学习算法,将18种特征描述符进行融合,从而提高图片排序结果的质量,并将上述特征描述符融合问题表示成分类学习问题。在给定部分查询图片以及相应的检索结果时,通过分类学习算法得到排序函数,并利用该排序函数对给定的未知查询图片进行检索。

在分类学习算法中,输入数据为查询图片以及相应的查询结果图片列表,列表中的图片顺序表明了它们与查询图片之间的相似性。对于每个特征描述符,分类算法都有一个距离矩阵,矩阵的每一行对应一个查询,每一列表示该查询与相应的图片间的相似性。对于预定义的18种特征描述符,分类学习算法根据输入数据对特征描述符的最优组合函数进行学习,并应用学习得到的最优组合函数对未来的图片查询进行处理。

2 排序学习算法

为了对特征描述符的最优组合函数进行学习,本节应用支持向量机和关联规则两种方法,并给出了相应的排序学习算法。

2.1 基于支持向量机的排序学习

支持向量机(Support Vector Machine, SVM)^[12]是一种有监督的分类学习方法,其主要思想是构建一个超平面用于分割 n 维空间中的点,其中维度 n 为被分类数据的属性个数。SVM进行分割的目标是最大化两个类别的边缘(margin)。图1为SVM在二维空间中的分类示意图。

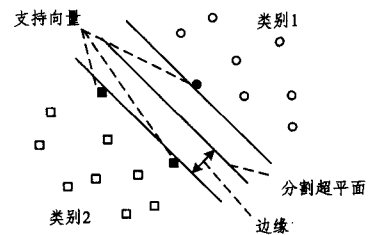


图1 支持向量机示意图

采用支持向量机对图片的特征描述符向量进行排序学习的过程如下:给定输入空间 $X \in \mathbb{R}^n$ (n 为图片的特征个数),图片的排序输出为 $Y = \{r_1, r_2, \dots, r_q\}$,其中下标 $1 \leq i \leq q$ 为图片在排序结果中的位置。在排序结果 Y 中,图片间的顺序关系为 $r_q > r_{q-1} > \dots > r_1$,其中符号 $>$ 表示图片间的偏好关系。

每个实例 $\vec{x}_i \in X$ 表示一个查询-图片对 (α, β) ,其中 α 为数据库中的一个图片, β 表示查询图片。每个实例 \vec{x}_i 用一个特征向量表示,其中每个特征为一个图片特征描述符。特征描述符的含义如上文所述。每个 \vec{x}_i 关于 β 都有相应的排序位置,因此不同的实例关于相应的查询 β 存在着偏好关系 $>$ 。在不同的实例之间,如果 \vec{x}_i 的偏好大于 \vec{x}_j ,那么有 $\vec{x}_i > \vec{x}_j$ 。

对于每个实例 $\vec{x}_i \in X$,可以应用排序函数 $f \in F$ 计算出相应的值,因此可以应用排序函数集 F 计算出不同实例间的偏好关系:

$$\vec{x}_i > \vec{x}_j \Leftrightarrow f(\vec{x}_i) > f(\vec{x}_j) \quad (1)$$

上述排序问题可以转化为分类学习问题:对于每组实例对 (\vec{x}_i, \vec{x}_j) ,将二者之间的偏好关系 $\vec{x}_i > \vec{x}_j$ 表示为向量 $\vec{x}_i - \vec{x}_j$;当排序结果正确时,令分类结果 $z = +1$ (即 $y_i > y_j$);当排序结果错误时,令分类结果 $z = -1$ ($y_j > y_i$)。具体描述如式(2)所示。

$$\left(\vec{x}_i - \vec{x}_j, z = \begin{cases} +1, & y_i > y_j \\ -1, & y_j > y_i \end{cases} \right) \quad (2)$$

接下来,对每个图片对 (\vec{x}_i, \vec{x}_j) 进行分类。在 (\vec{x}_i, \vec{x}_j) 中,当 \vec{x}_i 的排序结果在 \vec{x}_j 之前时,排序结果正确,并且分类算法的输出结果为 $z = +1$;当 \vec{x}_i 的排序结果在 \vec{x}_j 之后时,排序结果错误,此时分类算法的输出结果为 $z = -1$ 。对于所有的原始数据实例 $\vec{x}_i \in X$,根据式(2)生成一个训练数据集 S' ,并基于 S' 学习得到相应的分类函数。

2.2 基于关联规则的排序学习

关联规则^[13]用于描述数据项间的共存关系。在关联规则 $x \rightarrow y$ 中, x 为前件, y 为后件。关联规则不同于逻辑蕴含关系。在逻辑蕴含关系中,如果 x 发生,那么 y 一定发生;在关联规则 $x \rightarrow y$ 中,如果 x 发生,那么 y 的发生具有一定的倾向性。

在应用排序学习对图像的内容进行排序时,根据训练数据集中数据描述符之间的相似性来描述数据之间的相关性。首先,对数据描述符的相似性进行离散化处理,然后构建形如 $X \rightarrow r_i$ 的关联规则,其中关联规则的前件是相似性值的集合(相似性值来源于不同的描述符),后件为相关性水平。本文采用如下两个测度来估计关联规则的质量。

1)支持度: $X \rightarrow r_i$ 的支持度用 $\sigma(X \rightarrow r_i)$ 表示,代表训练数据集中包含特征集合 X 和相关性 r_i 的数据所占的比例。

2)置信度: $X \rightarrow r_i$ 的置信度用 $\theta(X \rightarrow r_i)$ 表示,代表给定特征集合 X 的前提下相关性 r_i 的条件概率。置信度越高, X 和 r_i 之间的关联越强。

在规则提取时,为了避免计算规模爆炸式增长,需要设定最小的支持度阈值。为了估计图片的相关性,需要对采用不同规则得到的预测值进行组合。在给定特征集合 X 以及相关性水平 r_i 下,每个关联规则 $X \rightarrow r_i$ 解释为对 r_i 的一次投票。不同的关联规则具有不同的投票权重,投票的权重取决于该规则的置信度 $\theta(X \rightarrow r_i)$ 。对相关性 r_i 的不同规则进行加权求和并求其平均值,其计算如式(3)所示。

$$s(r_i) = \frac{\sum_{X \rightarrow r_i \in R} \theta(X \rightarrow r_i)}{|R|} \quad (3)$$

其中, R 为投票过程应用的规则集合。在式(3)中,与相关性 r_i 相关联的值 $s(r_i)$ 为预测 r_i 的所有关联规则的加权平均置信度。

最后,对所有的与相关性 $r_i (i \in \{0, 1\})$ 相关联的值 $s(r_i)$ 进行规范化处理,并进行线性组合得到最终的相关性计算式如式(4)所示。

$$relevance = \sum_{i \in \{0, 1\}} (r_i \times \frac{s(r_i)}{\sum_{j \in \{0, 1\}} r_j}) \quad (4)$$

3 实验结果与分析

3.1 图片数据集

实验采用 Corel^[14] 和 Caltech^[15] 两个公开的图片数据集。Corel 数据集包含两万多张图片,本实验筛选出 Lib2 的部分子数据集,包含 3906 张图片和 123 条查询。Corel 数据集的图片分为 85 个类别,每个类别包含 7 至 98 张不等数量的图片。Caltech 数据集包含 8677 张彩色图片和 122 个查询,这些图片被分为 101 个类别,每个类别包含的图片数量在 40 到 800 之间。在这两个数据集中,所有类别之间是相互不包含的,即一张图片仅仅属于一个类别。当检索到的图片与查询图片属于一个类别时,认为两张图片是相关的。图 2 为两个数据集中对每张查询图片进行查询得到的相关图片的比例。当其它条件相同的情况下,其包含的潜在检索结果越少,那么图片检索的挑战越大。从图 2 可以看出,对于大多数查询来说,其检索结果中图片百分比不足 2%,因而图片检索面临着巨大的挑战。

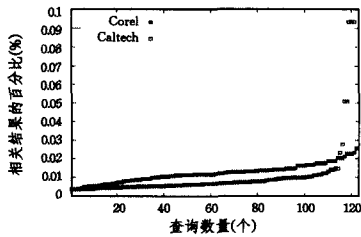


图 2 数据集中相关图片检索结果的分布

3.2 评价指标

为了评价排序结果的性能,实验采用准确率和平均准确率两种度量指标。准确率(Precision)为返回的 Top k 个结果中正确结果的比例。在准确率 \times 召回率的分布图上,平均准确率(Mean Average Precision, MAP)可以解决单点值的局限性,其公式为 $MAP(Q) = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} Precision(R_{jk})$, 其中 Q 是查询集合, m_j 表示第 j 次查询对应的所有相关图片的数目, R_{jk} 表示返回结果中直到检索到相关图片 k 所在位置前的所有图片集合,即 $Precision(R_{jk}) = \text{相关图片 } k \text{ 在返回的相关文档中的位置} / \text{该文档在返回的所有文档中的位置}$ 。

3.3 实验结果

将本文提出的基于关联规则的图片检索方法记为 ARIR,将提出的基于支持向量机的图片检索方法记为 SVMIR,并且将这两种算法与文献[9]中的遗传算法(GP)和文献[5]中的基于视觉特征的算法(VF)进行了对比。

首先,实验对比了 4 种算法在 Corel 和 Caltech 两个数据集上的 MAP,实验结果如表 1 所列。在该组实验中,我们进行了 5 次实验,观察了每种算法在每次实验下的 MAP,并且计算了 5 次实验的平均 MAP。从表 1 可以看出,VF 算法的 MAP 值在 5 次实验中都是最小的,其次是 GP 算法,ARIR 算法的 MAP 值和 SVMIR 算法的 MAP 值相近,并且明显大于 VF 和 GP 两种算法。

表 1 算法的 MAP 性能对比

数据集		Corel			
实验次数	ARIR	SVMIR	GP	VF	
1	0.405	0.404	0.379	0.279	
2	0.352	0.353	0.315	0.298	
3	0.366	0.366	0.356	0.341	
4	0.362	0.364	0.354	0.290	
5	0.352	0.351	0.246	0.193	
MAP 均值	0.367	0.368	0.330	0.280	
数据集		Caltech			
实验次数	ARIR	SVMIR	GP	VF	
1	0.151	0.159	0.051	0.058	
2	0.206	0.210	0.038	0.018	
3	0.190	0.189	0.098	0.093	
4	0.146	0.149	0.069	0.061	
5	0.194	0.198	0.092	0.063	
MAP 均值	0.177	0.181	0.070	0.059	

接下来,对比了 4 种算法在 Corel 和 Caltech 两个数据集上的 Precision,实验结果如表 2 所列。在 Precision 的对比中,分别令返回结果 Top k 的 k 为 1 到 10 的整数。从表 2 可以看出,在 k 取不同值的情况下,VF 算法的 Precision@ k 是最小的,其次是 GP 算法,ARIR 算法和 SVMIR 算法的 Precision@ k 结果相近,并且明显于 GP 和 VF 两种算法。

表 2 算法的 Precision 性能对比

数据集		Corel			
Precision@k	ARIR	SVMIR	GP	VF	
1	0.772	0.751	0.710	0.633	
2	0.699	0.698	0.668	0.606	
3	0.651	0.651	0.622	0.569	
4	0.625	0.626	0.597	0.545	
5	0.602	0.604	0.581	0.537	
6	0.593	0.593	0.564	0.515	
7	0.582	0.584	0.556	0.495	
8	0.571	0.573	0.545	0.481	
9	0.559	0.560	0.531	0.476	
10	0.550	0.550	0.521	0.464	
数据集		Caltech			
Precision@k	ARIR	SVMIR	GP	VF	
1	0.273	0.276	0.206	0.137	
2	0.247	0.246	0.184	0.196	
3	0.230	0.233	0.182	0.178	
4	0.221	0.225	0.175	0.169	
5	0.197	0.194	0.166	0.155	
6	0.193	0.195	0.155	0.149	
7	0.182	0.180	0.151	0.140	
8	0.175	0.173	0.147	0.134	
9	0.166	0.166	0.146	0.132	
10	0.161	0.163	0.141	0.127	

[7] 刘贵松, 王晓彬. 采用自适应 GHA 神经网络的分类器设计[J]. 电子科技大学学报, 2007, 36(6):1241-1244
Liu Gui-song, Wang Xiao-bin. Classifier Design Using Adaptive GHA Neural Networks[J]. Journal of University of Electronic Science and Technology of China, 2007, 36(6):1241-1244

[8] 杨存祥, 朱琛, 解豪杰. 基于 RPROP 神经网络算法的异步电动机故障诊断[J]. 电力自动化设备, 2012, 32(1):80-83
Yang Cun-xiang, Zhu Chen, Xie Hao-jie. Fault Diagnosis Based on RPROP Neural Network for Asynchronous Motor[J]. Electric Power Automation Equipment, 2012, 32(1):80-83

[9] Liu X, Chen T, Kvijava Kumar B V. Face authentication for

multiple subjects using eigenflow [J]. Pattern Recognition, 2003, 36(2):313-328

[10] Sander T D. Optimal unsupervised learning in a single-layer linear feed forward neural network [J]. Neural Networks, 1989, 2:459-473

[11] Alex L, Sandor C, Der Nasser Z, et al. Dual-band passive infrared imagery for automatic clutter rejection [M]. ICIP, 2000

[12] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 1997, 19(7):711-720

(上接第 277 页)

在表 1 和表 2 中, 当应用关联规则和支持向量机两种方法对分类函数进行学习时, 其分类的准确性(包括 Precision@k 和 MAP)结果都是相近的, 因此通过实验分析了两种方法的相关性。最后对基于关联规则的图片检索方法和基于支持向量机的图片检索方法两种方法的相关性进行分析, 实验结果如图 3 所示。图中的每个点的横坐标为采用 SVMIR 算法时的 MAP 值, 纵坐标为采用 ARIR 算法时的 MAP 值。从图 3 可以看出, 所有的点都分布在斜率为 $CC=0.978$ 的直线周围。因此可以认定, 应用关联规则和支持向量机两种方法对分类函数进行学习时, 由于这两种算法都通过相同的数据实例对图片的 18 种描述符的权重进行学习, 因此得到的结果是相关的。

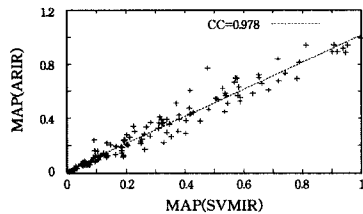


图 3 ARIR 和 SVMIR 两种方法的相关性分析

结束语 本文基于排序学习的相关理论研究了图片共享社会网络中的图片检索方法。每张图片表示为多个特征描述符的组合。当用户输入的查询为图片时, 对比查询图片和图片库中图片的相似性来完成图片的检索。采用支持向量机和关联规则两种学习方法对特征描述符的权重组合进行学习, 并提出了相应的学习算法。实验表明, 本文提出的基于学习的图片检索方法与相关图片检索方法相比具有更高的准确性。此外, 应用支持向量机和关联规则两种方法对分类函数进行学习时, 由于两种算法通过相同的数据实例对图片描述符的权重进行学习, 因此得到的结果是相关的。

参 考 文 献

[1] Rui Y, Huang T S, Chang S F. Image retrieval: Current techniques, promising directions, and open issues[J]. Journal of Visual Communication and Image Representation, 2009, 10(1):39-62

[2] 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 Informaion Retrieval. ACM, 2013:119-126

[3] Akgül C B, Rubin D L, Napel S, et al. Content-based image re-

trieval in radiology: current status and future directions[J]. Journal of Digital Imaging, 2011, 24(2):208-222

[4] 李晓燕, 陈刚, 寿黎但, 等. 一种面向协作标签系统的图片检索聚类方法[J]. 中国图象图形学报, 2010, 15(11):1635-1643
Li Xiao-yan, Chen gang, Shou Li-dan, et al. An image clustering algorithm in collaborative tagging system[J]. Journal of Image and Graphics, 2010, 15(11):1635-1643

[5] Jing F, Li M, Zhang H J, et al. A unified framework for image retrieval using keyword and visual features[J]. IEEE Transactions on Image Processing, 2005, 14(7):979-989

[6] He J, Li M, Zhang H J, et al. Manifold-ranking based image retrieval[C]//Proceedings of the 12th annual ACM international conference on Multimedia. ACM, 2014:9-16

[7] 唐朝霞, 章慧, 徐冬梅. 一种改进的粒子群算法和相关反馈的图像检索[J]. 计算机科学, 2011, 38(10):278-280
Tang Zhao-xia, Zhang hui, Xu dong-mei. Image Retrieval Based on Improved PSO Algorithm and Relevance Feedback[J]. Computer Science, 2011, 38(10):278-280

[8] 沈新宁, 王小龙, 杜建洪. 基于颜色自相关图和互信息的图像检索算法[J]. 计算机工程, 2014, 40(2):259-262
Shen Xin-ning, Wang Xiao-long, Du Jian-hong. Image Retrieval Algorithm Based on Color Autocorrelogram and Mutual Information[J]. Computer Engineering, 2014, 40(2):259-262

[9] Liu G H, Li Z Y, Zhang L, et al. Image retrieval based on micro-structure descriptor[J]. Pattern Recognition, 2011, 44(9):2123-2133

[10] Rubner Y, Tomasi C, Guibas L J. The earth mover's distance as a metric for image retrieval[J]. International Journal of Computer Vision, 2010, 40(2):99-121

[11] da Silva Torres R, Falcão A X. Content-Based Image Retrieval: Theory and Applications[J]. RITA, 2006, 13(2):161-185

[12] Chen Y, Zhou X S, Huang T S. One-class SVM for learning in image retrieval[C]//Proceedings 2001 International Conference on Image Processing, 2001. IEEE, 2011:34-37

[13] Ribeiro M X, Bugatti P H, Traina Jr C, et al. Supporting content-based image retrieval and computer-aided diagnosis systems with association rule-based techniques[J]. Data & Knowledge Engineering, 2009, 68(12):1370-1382

[14] Tang J, Lewis P H. A study of quality issues for image auto-annotation with the corel dataset[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(3):384-389

[15] Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories[J]. Computer Vision and Image Understanding, 2007, 106(1):59-70