基于语义区域提取的图像重排

陈 曾[1] 侯 进[1] 张登胜[2] 张华忠[1]

摘 要 针对目前图像搜索引擎难以正确把握用户真正意图的问题, 从爬虫 Web 图像搜索引擎检索结果入手, 提出三种聚类算法来提取海量 Web 图像中的语义区域. 这三种聚类算法包括确定初始化中心的 K-means 聚类、确定参数的最大期望聚类以及基于半监督的 K-means 聚类算法. 然后选取显著值较大的显著区域作为语义区域. 实验分析比较了三种聚类算法的有效性, 最终实现的图像重排系统能比网络搜索引擎更好地反馈给用户精确而且有序的查询结果.

关键词 语义区域提取, 半监督聚类, K-means 聚类, 最大期望聚类, 图像重排

DOI 10.3724/SP.J.1004.2011.01356

Image Re-ranking Based on Extraction of Semantic Regions

CHEN Zeng[1] HOU Jin[1] ZHANG Deng-Sheng[2] ZHANG Hua-Zhong[1]

Abstract It is difficult for current image search engines to accurately grasp the real intention of users. Based on the search results, we propose three clustering algorithms to extract semantic regions of Web images. These methods include K-means clustering with determined k centers, expectation maximization clustering with the determined parameters, and semi-supervised K-means clustering. We then select the salient regions with the high salient scores as the semantic regions. We demonstrate the experimental results by comparing the three clustering algorithms. The proposed image re-ranking system can more accurately show the ordered search results than web image engines.

Key words Extraction of semantic regions, semi-supervised clustering, K-means clustering, expectation maximization clustering, image re-ranking

图像重排表示对初次检索反馈显示的结果进行筛选和重排序的过程, 这在基于文本的检索方法中已非常流行了, 而在图像检索领域, 该技术也逐渐成为研究热点. 在不考虑用户参与的情况下, 主要研究结合图像的视觉特征对基于文本检索结果进行重排序. 比如 Cui 等[1] 提出一种实时 Web 检索重排方法, 该方法可以适应相似性, 利用多种特征得到训练模型, 由于采用不同方法处理多种图像, 所以系统很复杂. 而文献 [2] 使用了多实例学习来学习视觉模型. 对于这些方法, 其重排结果都受到图像检索所提供图像总数的限制导致最终的检索结果并不理想. Berg 等[3] 通过从网页搜索而非一个图像检索中下载图像的途径, 解决了下载量限制问题. 该搜索可以产生数万张图像. 这个方法成功实现了比之前方法更大的有效图像库, 但是却需要以大量的人工干预 (Manual intervention) 为代价.

本文从 Web 图像搜索引擎的查询结果入手, 探讨 Web 图像中的语义区域的提取方法, 利用三种聚类方法将语义区域归类并提取出有意义的兴趣区域. 对于聚类算法的研究, 本文分别采用确定初始化中心的 K-means 聚类、确定参数的最大期望 (Expectation maximization, EM) 聚类算法以及基于半监督的 K-means (Semi-supervised K-means, SSK-means) 聚类算法. 得到语义区域之后, 利用机器学习算法进行图像标注, 将标注结果利用倒排文件索引方式以实现最终的图像重排.

1 语义区域的提取算法

为了从图像中提取出精确的语义区域, 本文提出了利用聚类的方法来提取出许多显著区域群 (Salient region group), 具体实现框架如图 1 所示. 首先给定一个查询词作为一个语义概念, 在 Web 搜索引擎中输入该查询词并爬虫得到一定数量的查询结果图像集合. 本文对每个查询词在 Google 图像搜索中爬虫了最靠前的 150 张图像做实验. 本文选择前 100 张最相关的图像集合作为正例候选图像集, 而 100 张以后的结果, 根据 Google 图像搜索的整体

收稿日期 2010-04-25 录用日期 2011-05-31
Manuscript received April 25, 2010; accepted May 31, 2011
教育部留学回国人员科研启动基金, 高等学校博士学科点专项科研基金 (20090184120022), 中央高校基本科研业务费专项资金科技创新项目 (SWJTU09CX036) 资助
Supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, Ministry of Education, the Specialized Research Fund for the Doctoral Program of Higher Education (20090184120022), and the Fundamental Research Funds for the Central Universities (SWJTU09CX036)

1. 西南交通大学信息科学与技术学院 成都 610031, 中国 2. 莫纳什大学吉普斯兰信息技术学院 丘吉尔 维多利亚 3842, 澳大利亚
1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, P. R. China 2. Gippsland School of Info Tech, Monash University, Churchill, Victoria 3842, Australia

评估研究表明, 其相关性和所含的语义信息要明显低于前 100 张的结果. 对于与语义概念相关的显著区域, 本文采用以下步骤从 Web 图像中自动并有效地检测显著区域群.

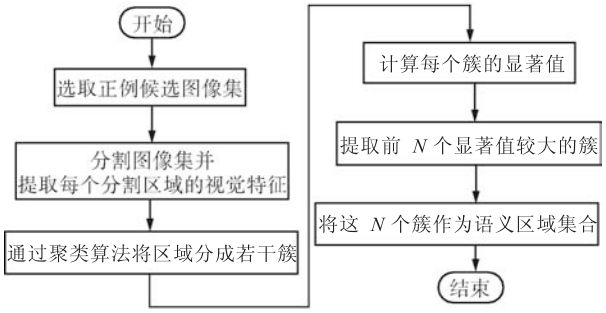


图 1　语义区域的提取算法

Fig. 1　The proposed extraction algorithm of semantic regions

首先, 尽管 Web 页中前 100 张图像具有很高的语义相关性, 但是与语义概念不相干的噪声区域仍然存在. 这是由于 Web 图像的多样性和混杂性, 并且可能会影响聚类结果. 因此, 通过聚类算法将这些正例图像区域特征进行聚类的时候, 需要考虑最终应该分成多少个簇 (Cluster), 也就是区域集合. 在此, 我们定义通过聚类算法可以将正例图像集合中的区域归类成 M 个簇. 那么, 我们需要确定在聚类算法中对于迭代过程的停止规则.

然后, 对于第 m 个区域簇, 它的显著值 (Salient score, SS)[4] 为

$$SS(m) = w_1 \times \frac{N_m}{\sum_{m=1}^{M} N_m} + w_2 \times \frac{R_m}{\sum_{m=1}^{M} R_m} \qquad (1)$$

式中, N_m 为第 m 个区域簇中的区域总数; R_m 为第 m 个区域簇中的区域率的总和, 而区域率指的是该区域在所在图像中的所占面积; $w_{1,2}$ 为权重值, 根据本文实验经验, 分别设定为 0.7 和 0.3.

我们选取显著值较大的前 N 个区域簇作为显著区域群. 由此可见, 与使用单个表示视觉模式相比, 以上这种多个显著区域改进了一个语义概念的表示方式, 这种方式在视觉上更灵活. 对于 N 值的确定, 本文定义前 N 个区域簇的大小应覆盖大约整个正例候选样例大小的 1/3.

1.1　基于 K-means 聚类算法的语义区域提取

本文提出了一种改进的 K-means 聚类算法[5] 以确定 k 个初始化中心. 主要思路是: 找到欧拉距离最近的一对数据点形成一个数据点集, 并从原集中删除; 然后陆续找到与点集最近的点并添加于其中, 直到在点集中的数据点数量达到 $0.75n/k$ (n 为数据点总数); 直到 k 个点集产生, 求出每个点集中的数据点向量的算术均值, 这些均值也就作为我们需要确定的初始化聚类中心. 以此确定 M 值, 也就是预定义簇的个数. 对于本步骤中的 "$0.75n/k$", 其实也就是我们在前面需要讨论的阈值. 这个阈值的确定是根据本实验的数据分布和特点得到的经验值.

1.2　基于最大期望聚类算法的语义区域提取

EM 主要是为了求得最终的簇, 以得到每个簇参数的最大似然估计[6]. 其中, 未知参数主要由均值向量 $\boldsymbol{\mu}_y$ 和协方差矩阵 Σ_y 构成, $y = 1, \cdots, m$, m 是最后聚类生成的簇的个数. 本文利用 K-means 算法来初始化确定模型参数.

对于均值向量 $\boldsymbol{\mu}_y$ 的初始化, 首先利用上一节提出的改进型的 K-means 算法求得最新的簇中心 (聚类中心), 然后我们把得到的这些最新的簇中心也就作为 $\boldsymbol{\mu}_y$ 的初始值. 对于协方差矩阵 Σ_y 的初始化, 可以利用输入的样本数据 (或观察数据) $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 和前面求得的均值向量 $\boldsymbol{\mu}_y$ 通过式 (2) 来求得, 其中 d 表示维数.

$$\Sigma_{ij} = \sum_{k=1}^{N} (\boldsymbol{x}_k^i - \boldsymbol{\mu}_k^i)(\boldsymbol{x}_k^j - \boldsymbol{\mu}_k^j), \ \ i, j = 1, \cdots, d \ \ (2)$$

1.3　基于半监督的 K-means 聚类算法的语义区域提取

对于前两种聚类算法, 从理论上讲, 本文提出的 EM 算法要比 K-means 算法得到的聚类结果更准确, 但 EM 算法的时间复杂度要远远高于 K-means 算法, 需要降维处理. 为了解决这个问题, 本文借鉴了半监督聚类的思想[7]. 该算法利用已标记数据构成一个 seed 集来初始化 K-means 聚类中心. 本文从 Google 上爬虫了 10 个类的相关图像, 对于每个类, 我们人工选取了 20 个相对应的语义区域. 因此我们拿这 200 个已标记的样本区域作为一个 seed 集, 并且使用这个 seed 集进行 K-means 聚类中心的初始化选取. 得到初始化中心之后, 便可以利用传统的 K-means 聚类算法得到最终的簇.

2　实验分析

为了验证这三种聚类算法的可行性和有效性, 本文将其应用于 Web 图像重排系统中. 我们从图像资源最丰富的 Google 搜索引擎中爬虫了 10 个关键词的图像集合, 每个关键词对应有 150 张图像, 共 1 500 张图像. 对于每个关键词, 均是依序从 Google 上爬虫下来的最靠前的 150 张图像. 以该网页排序结果作为基础, 我们利用本文算法在此排序结果基

础上加以提炼, 使得重排的查询结果令用户更满意. 从 Google 上爬虫的 10 个关键词包括: baby, bear, bird, car, cat, deer, fish, boat, elephant 和 beach 等. 图 2 给出了本文实现的 Web 图像重排系统, 以 “boat” 查询词为例, 由图可以明显看出, Google 的原始搜索结果含有一些卡通图、素描图和抽象图等, 而这些都与用户的查询本意是不相符的, 因此利用本文所提出的算法可以有效地过滤掉这些图像.

通过用户点击相关图像, 然后提交后便可以观察和分析相关算法的重排性能对比. 图 3 给出了 “elephant” 查询词的重排性能比较曲线. 为了简化操作和方便实验分析, 本系统曲线以前述的重排结果为样本绘制而成的. 曲线所表示的含义是排在前 k 张图像集合中该查询词的查准率, 本文对 k 的定义为每 10 张图像计算一次查准率. 由图 3 可知, 通过本文提出的聚类算法, 有效地提炼和排序了 Web 图像检索结果, 其重排性能明显优于 Google 的检索结果. 基于 EM 算法的重排结果优于 K-means 算法, 是因为 EM 的初始化参数是由 K-means 算法确定的, 即集成了 K-means 算法的优点, 若随机初始化 EM 参数, 则会明显增加运行时间. 而 SSK-means 聚类算法优于其他两种算法, 这是因为半监督算法引入了已标记的样本数据, 然后通过 K-means 进行聚类, 又弥补了 EM 算法运行时间较长的问题, 因此 SSK-means 算法可得到最好的重排结果.

此外, 表 1 给出了对于所有 10 个查询词, 利用单个查询词的平均准确率 (Mean average precision, MAP) 性能指标[8] 来比较分析了 Google 搜索、K-means、EM 和 SSK-means 四种方法的重排性能. MAP 反映的是检索结果的排序情况, 是基于每个关键词前 k 个检索结果, 并定义为在该排序下, 所有关键词的平均查准率. 一个优良的检索系统所提供的排在靠前的检索结果接近用户的期望. 从该表也可看出, 本文提出的聚类算法对于海量的 Web 图像库可以有效地提取语义区域, 在倒排文件索引的协助下, 整个重排性能也明显得到提高.

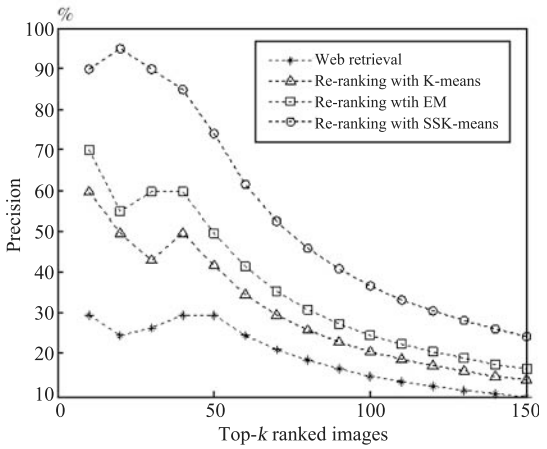


图 3 基于不同聚类算法的 Web 图像重排性能比较

Fig. 3 Performance comparison among different methods for web image re-ranking

3 结论

本文在 Web 图像搜索结果的基础上, 提出了基于语义区域提取的图像重排方法. 对于语义区域的提取, 为了解决人工选取的庞大工作量问题, 本文提

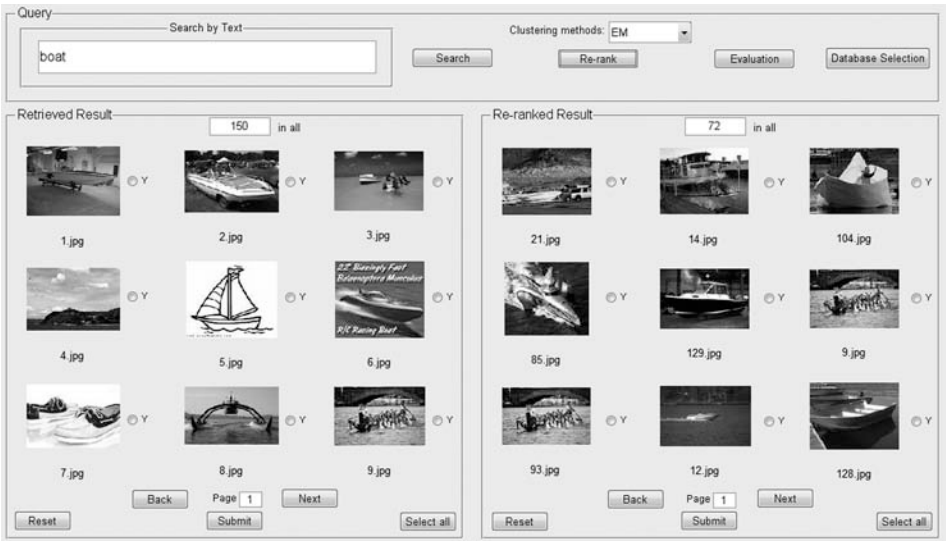


图 2 基于聚类算法的图像重排系统

Fig. 2 Image re-ranking system based on clustering methods

表 1 在 Google 图库中对于 10 个查询词不同方法的 MAP 比较

Table 1 Comparison of MAP among different methods for 10 terms in the Google dataset

方法	MAP
基于 Google 元数据搜索	0.45
基于 K-means 的重排	0.62
基于 EM 的重排	0.78
基于 SSK-means 的重排	0.80

出了三种改进的聚类算法来提取显著值较大的显著区域作为语义区域. 本文以 Google 图像为实验对象, 实验表明基于本文所提出的聚类算法的重排结果明显优于 Google 搜索, 其中基于半监督的 K-means 聚类算法拥有最好的重排性能. 然而, 对于本文提出的语义区域提取模型, 目前主要考虑区域面积和区域个数这两个因素, 在未来还需要添加更多的因素来求取区域显著值, 如区域位置、区域相关性等. 另外, 面对海量的 Web 图像, 对于一些抽象的语义查询词的搜索, 本系统还难以提取其中的语义区域, 这也是未来需要解决的问题.

References

1 Cui J, Wen F, Tang X. Real time Google and live image search re-ranking. In: Proceedings of the 16th ACM International Conference on Multimedia. Vancouver, Canada: ACM, 2008. 729−732

2 Vijayanarasimhan S, Grauman K. Keywords to visual categories: multiple-instance learning for weakly supervised object categorization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1−8

3 Berg T L, Forsyth D A. Animals on the web. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 1463−1470

4 Sun Y, Shimada S, Taniguchi Y, Kojima A. A novel region-based approach to visual concept modeling using web images. In: Proceedings of the 16th ACM International Conference on Multimedia. Vancouver, Canada: ACM, 2008. 635−638

5 Napoleon D, Lakshmi P G. An enhanced K-means algorithm to improve the efficiency using normal distribution data points. *International Journal on Computer Science and Engineering*, 2010, **2**(7): 2409−2413

6 Govaert G, Nadif M. An EM algorithm for the block mixture model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(4): 643−647

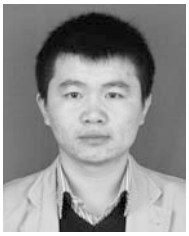
7 Gao Yan-Yu, Yin Yi-Xin, Uozumi T. A hierarchical image annotation method based on SVM and semi-supervised EM. *Acta Automatica Sinica*, 2010, **36**(7): 960−967
(高彦宇, 尹怡欣, Uozumi T. 一种基于支持向量机和半监督期望最大化算法的分级图像标识方法. 自动化学报, 2010, **36**(7): 960−967)

8 Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2004. 1002−1009

陈 曾 西南交通大学信息科学与技术学院硕士研究生. 主要研究方向为图像检索与计算机视觉.
E-mail: chenzeng_000@163.com
(**CHEN Zeng** Master student at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers image retrieval and computer vision.)

侯 进 西南交通大学信息科学与技术学院副教授. 主要研究方向为多媒体检索与人机交互. 本文通信作者.
E-mail: jhou@swjtu.edu.cn
(**HOU Jin** Associate professor at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers multimedia information retrieval and human-computer interaction. Corresponding author of this paper.)

张登胜 澳大利亚莫纳什大学信息技术学院副教授. 主要研究方向为音乐识别与语义图像检索. E-mail: dengsheng.zhang@infotech.monash.edu.au
(**ZHANG Deng-Sheng** Associate professor at the Faculty of Information Technology, Monash University, Australia. His research interest cover music recognition and semantic image retrieval.)



张华忠 西南交通大学信息科学与技术学院硕士研究生. 主要研究方向为图像标注与机器学习.
E-mail: zhz_233@yeah.net
(**ZHANG Hua-Zhong** Master student at the School of Information Science and Technology, Southwest Jiaotong University. His research interest covers image annotation and machine learning.)