

## ◎图形、图像、模式识别◎

# 结合相关反馈日志与语义网络的图像标注

杨 珺<sup>1,2</sup>, 王继成<sup>2</sup>, 邢丹俊<sup>2</sup>YANG Jun<sup>1,2</sup>, WANG Ji-cheng<sup>2</sup>, XING Dan-jun<sup>2</sup>

1. 上海电力学院 计算机信息与工程学院, 上海 200090

2. 同济大学 电子与信息工程学院, 上海 201804

1. Shanghai University of Electric Power, Shanghai 200090, China

2. School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

E-mail: yangjun9323@sohu.com

YANG Jun, WANG Ji-cheng, XING Dan-jun. Using relevance feedback log and semantic network for image annotation. *Computer Engineering and Applications*, 2009, 45(21): 99-102.

**Abstract:** Image annotation, a technique for connecting image semantics and visual features, can be used to present image semantics well. A method for image annotation using relevance feedback log and semantic network is presented. Firstly, the image semantics are acquired by users' relevance feedback log; then semantic clustering is carried out based on semantic similarity; lastly automatic image annotation is realized through semantic propagation. The experimental results indicate that increasing images will be annotated as increasing relevance feedback log and the annotation accuracy will be stable with the increment of relevance feedback.

**Key words:** image retrieval; image annotation; relevance feedback; semantic network

**摘要:** 连接高层语义和低层视觉特征的图像语义标注技术能够很好地表示图像的语义, 提出并实现了一种结合相关反馈日志与语义网络的图像标注方法。该方法以收集的用户相关反馈日志为基础获得图像的语义信息, 通过计算图像间的语义相似度进行语义聚类并采用语义传播的方式实现图像的语义标注。实验结果表明, 随着相关反馈日志库的不断扩充, 图像库中越来越多的图像会在反馈的过程中得到标注且标注的准确率会随着反馈次数的增加而趋于稳定。

**关键词:** 图像检索; 图像标注; 相关反馈; 语义网络

DOI: 10.3778/j.issn.1002-8331.2009.21.029 文章编号: 1002-8331(2009)21-0099-04 文献标识码: A 中图分类号: TP391.3

基于内容的图像检索(Content-Based Image Retrieval, CBIR)近十年来得到了研究者的广泛关注<sup>[1-2]</sup>。然而, 尽管图像的视觉内容与语义内容之间存在一定的相关性, 但仅仅依靠视觉特征, 仍然无法准确地表达图像的语义信息, 这种由于图像底层的视觉信息与高层语义存在不一致性, 即所谓的“语义鸿沟”<sup>[3]</sup>, 使得 CBIR 的整体检索性能还差强人意。

图像的标注字<sup>[4-6]</sup>能够很好地表达图像的语义内容, 能够缩小图像的高级语义和低级视觉内容之间的差距。然而, 由于受制于当前计算机视觉技术与图像理解的发展水平, 人们还无法通过计算机自动地提取图像的语义信息, 而人工图像标注不仅费时费力, 而且由于不同的人对同一幅图像有不同的理解, 容易导致“语义偏差”。一个较为可行的方法是先人工地对图像库的部分图像进行标注, 然后通过语义传播<sup>[7]</sup>技术完成对其余图像的语义标注。

提出了一种结合相关反馈日志与语义网络的图像语义标注方法。首先, 对图像库中的部分图像进行人工标注, 建立相应

的语义网络, 收集并存储相关反馈<sup>[8-10]</sup>过程中的用户反馈信息并生成相关反馈日志; 然后, 系统以相关反馈日志为基础, 计算图像间的语义相似度, 并进行语义聚类; 最后, 由于同一语义聚类中的图像具有相似的语义, 对同一语义类中的图像进行语义传播, 并采用一定的策略更新并完善语义网络, 从而实现图像的自动语义标注。

## 1 图像标注框架

提出的图像标注框架如图 1 所示。

首先, 系统对图像库中的每一幅图像提取低层视觉特征并存储于低层特征库中, 采用语义网络作为图像的语义表示方法, 人工地对图像库中的部分图像进行标注, 建立相应的语义网络。

当用户通过范例(example)的方式向 CBIR 系统检索图像时, 系统首先根据图像低层视觉特征计算查询图像与图像库中的图像之间的相似度, 并按降序排列向用户显示当前的检索结

**作者简介:** 杨珺(1978-), 女, 博士研究生, 主要研究方向: 图像检索、计算机视觉; 王继成(1958-), 男, 博士, 研究员, 博导, 主要研究领域为图像处理、数据挖掘、软件工程; 邢丹俊(1982-), 男, 硕士研究生, 主要研究方向: 图像恢复。

收稿日期: 2008-04-25 修回日期: 2008-07-21

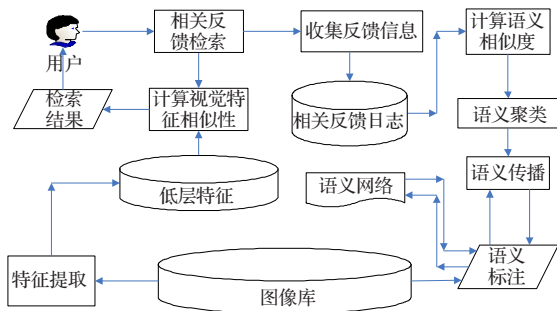


图1 图像标注框架

果。用户根据自己的检索需要做出相关性判断,并将反馈信息提交给系统。系统采用相应的相关反馈算法对用户的反馈信息进行学习,优化查询并返回新的检索结果直至用户满意为止。

如图1所示,用户的相关反馈信息被收集并存储于相关反馈日志库中。这些相关反馈日志记录了用户对图像是否相关性的判断,隐含了用户对图像的理解,揭示了图像间的语义相关性。显然,如果两幅图像在一次查询过程中同时被标记为相关图像,则它们可能具有相同的语义,而且它们被同时标记为相关图像的的次数越多,则它们之间在语义上就越相似。基于以上假设,系统根据相关反馈日志中的反馈信息计算图像的语义相似度,并进行图像语义聚类。由于类中的图像具有相同的语义,系统即可对类中的图像进行语义传播,并采用相应的策略更新并完善语义网络,从而实现自动语义标注。

## 2 图像语义标注

### 2.1 语义网络

在图像检索的过程中首先需要确定图像的语义表示方法。由于不同图像的语义明确程度不一样,每幅图像都可能有一个或几个关键词来对其进行描述,故采用语义网络<sup>[11]</sup>来表示图像的语义,其结构如图2所示。库中每幅图像都以不同的关键词和权重加以描述。关键词用于表达图像的语义,图像语义越简单,则用以描述的关键词就越少(如图像1,3);反之,图像语义越复杂,则用以描述的关键词就越多(如图像2)。每个关键词的权重  $w_{ij}$  反映了该词对图像语义描述的准确程度,权重越大,则该关键词越能清晰地描述该图像。每幅图对应关键词的权重可以根据用户反馈予以调节。利用用户反馈日志完善语义网络的过程实际上就是一个更新关键词并修改相应权重  $w_{ij}$  的过程,并最终使得一个图像库的语义网络能够符合并反映大多数用户对图像的理解。

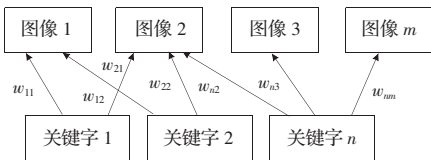


图2 语义网络

### 2.2 语义相似度

如果两幅图像在一次查询中同时被标记为相关图像,则它们可能具有相同的语义,它们被同时标记为相关图像的的次数越多,则它们之间在语义上就越相似<sup>[12]</sup>。系统根据相关反馈日志中的数据计算图像的语义相似度。

定义1 图像  $i$  与图像  $j$  之间的语义相似度  $s_{ij} = \sqrt{(n_{ij}/N)}$ ,

其中  $n_{ij}$  是在检索过程中图像  $i$  和图像  $j$  同时被标记为正例图像的次数,  $N = 1 + \max(n_{ij})$ 。

根据以上语义相似度的定义,可构造相应的语义相似图,如图3所示。图中的每个节点  $i$  代表相应的图像,如果两幅图像在语义上相似,则建立连接这两个节点的边,边的权重为图像间的语义相似度。

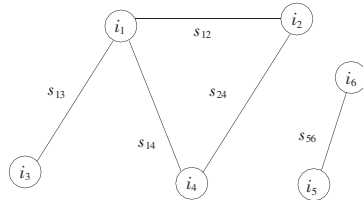


图3 语义相似图

### 2.3 语义聚类

在获得了图像的语义相似度后,可以对图像进行语义聚类。整个聚类过程分为两个步骤<sup>[13]</sup>:首先,对图像进行语义预分类,将图像库中的每一幅图像分到且仅分到一个语义组中。其次,在预分类的基础上对图像进行语义聚类。在聚类过程中,一个语义组中的一幅图像可以被聚集到一个簇中,也可以被聚集到另一个簇中,即一幅图像可以属于多个语义簇,这样的聚类结果符合图像语义的多重性这一客观事实。当然,一幅图像也可能不属于任何一个语义簇。

定义2 对于数据集  $T$ , 图像  $i$  的邻居为:  $N(i, T) = \{j | j \in T, s_{ij} > 0, j \neq i\}$ 。

定义3 图像  $i$  与数据集  $T$  的关系为:  $R(i, T) = \sum_{j \in N(i, T)} s_{ij}$ 。

$R(i, T)$  的值反映了图像  $i$  与数据集的关系,即  $R(i, T)$  越大,则图像  $i$  与数据集  $T$  的关系越紧密,反之,则关系越松散。

定义4 聚类  $C$  是整个图像集  $U$  的一个子集,图像  $i$  隶属于一个聚类  $C$  的隶属度为:

$$A(i, C) = \sqrt{\frac{R(i, C) \cdot R(i, C)}{\sum_{i \in C} R(i, C) \cdot R(i, U)}}$$

以上定义不仅考虑了图像  $i$  与聚类  $C$  之间的关系,还考虑了图像  $i$  与整个图像集之间的关系。

定义5  $C$  是一个聚类,其聚类密度定义为:

$$D(C) = \sqrt{\frac{\sum_{i \in C} R(i, C)}{|C| \times (|C| + 1)}}$$

具体的聚类算法描述<sup>[13]</sup>如下:

```

Set  $U$  to the whole image set
Set  $\varepsilon$  to the affinity threshold
WHILE  $U \neq \emptyset$ 
     $m = \arg \max_{x \in U} R(x, U)$ 
     $C = N(m, U) \cup \{m\}$ 
    REPEAT
         $C' = C$ 
         $C = \{x | A(x, C') \geq \varepsilon, x \in U\}$ 
    WHILE  $C \neq C'$ 
    Save  $C$  as a class
     $U \leftarrow U - C$ 
END
FOR each image  $x$  in class  $C$ 

```

```

 $C_x = \{x\}$ 
REPEAT
   $C_m = C_x$ 
  FOR each image  $y$  in class  $C$ 
    IF  $y \in C_m$  THEN
       $C_y = C_m - \{y\}$ 
    ELSE
       $C_y = C_m \cup \{y\}$ 
    END IF
  IF  $D(C_y) > D(C_x)$  THEN
     $C_x = C_y$ 
  END IF
END FOR
WHILE  $C_x \neq C_m$ 
  Save  $C_x$  as a cluster
END FOR

```

算法一次建立一个语义分类,在分类过程中一幅图像属于且仅属于唯一的语义类,在一次迭代过程中可以一次加入或移出多幅图像。在经过了粗略的语义预分类后,一个语义预分类中可能包含着多种不同的语义类,因此有必要在语义预分类的基础上再进行语义聚类,以便将同一语义的图像聚集在一起。算法的聚类原则是:一幅图像是否应该加入或者移出一个聚类主要取决于该图像的加入或移出是否改善了整个聚类的密度。算法在语义预分类中任意选择一幅图像作为聚类种子并进行语义聚类,不断地更新聚类直至该聚类的密度最大。由于语义预分类中的每一幅图像都有可能成为聚类种子,理论上聚类的个数应该等于预分类中图像的个数。但事实上,许多以不同图像作为聚类种子所得到的聚类结果都是相同的,此外还有一些聚类间有所重叠,因此,有必要引入一些启发式规则去除那些冗余的聚类。对于那些相同的聚类,只保留其中的一个聚类即可;而对于那些有重叠的聚类,只保留具有最大密度的聚类,移去其他的聚类。经过这些后续处理,只有少量的聚类被保留了下来,这些类中的图像具有相同的语义。

## 2.4 语义传播及权值更新

在经过了语义聚类之后,由于同一聚类中的图像具有相同的语义,因此可以在类中进行语义传播并进行标注。具体的语义传播及标注过程如下:

(1)如果类中的所有图像均无被任何关键字标注过,则由用户对其共同赋予一个关键字进行标注。由于该关键字唯一地表示了图像的语义,故将该关键字的初始权重设置为1,即  $w_j = 1$ 。

(2)如果类中的所有图像均被不同的关键字标注过,则采取以下的策略更新各关键字的相应权重:

①计算各关键字的出现频率,得到前  $n$  个具有最大频率的关键字  $f_i, i=1, 2, \dots, n$ 。

②建立该  $n$  个关键字与类中图像的关联。如果图像已经被该关键字标注过,更新关键字权重  $w_j = w_j + f_i$ , 否则初始化关键字

权重为  $w_j = f_i$ 。归一化关键字权重  $w_j = w_j / \sum w_j$ 。

③归一化后,如果某关键字的权重小于系统设置的最小阈值  $\zeta$ , 则删除该关键字与图像的连接。

(3)如果类中的部分图像被标注过,另外一些图像没有被标注过,则采用语义传播的方法对其进行标注。

①类中共有  $N$  个关键字,其中  $M$  幅图像被标注过,  $m_i$  为拥有关键字  $k_i$  的图像个数。由于一个聚类中的图像应该拥有相

似的语义,因此对于类中没有被标注过的图像,采用语义传播的方式对其进行标注。即建立该  $N$  个关键字与未标注图像的连接  $\{k_i, m_j / \sum_{i=1}^N m_i\}$ , 其中  $m_j / \sum_{i=1}^N m_i$  为相应关键字的权重。

②对于那些已经被标注过的图像,更新其各关键字的相应权重。其更新策略同(2)中的①和②。

③如果某关键字的权重小于系统设置的最小阈值  $\zeta$ , 则删除该关键字与图像的连接。

系统在初始状态下只有少数图像事先被人工标注,因此相应的语义网络还不完善。然而,在收集了用户的相关反馈信息后,系统根据用户的反馈信息,获得了图像的语义相关性,通过计算图像的语义相似度并进行语义聚类。由于同一类中的图像拥有相似的语义,可以将类中被标注过的图像的语义传播给那些没有被标注过的图像,迅速建立起图像和关键词的关联。此外,随着相关反馈日志库中用户信息的不断更新和积累,系统可以获得越来越多的语义信息,更新并丰富语义网络,经过一段时间后,系统的语义网络将日趋完善,所有的图像都将得到标注。提出的语义传播及相应的权重更新策略充分考虑了相关反馈日志库的动态变化,反映了大部分用户对图像的理解,不仅可以扩充描述图像的关键词,充分反映图像语义的多样性,而且整个权值更新过程将使得那些能够准确描述对应图像的关键词得到更大的权重,而权值小于系统阈值的关键词将被删除。

## 3 实验

从 Internet 上得到了文献[14]所用 1 000 幅 24-bit 真彩色 JPEG 图像。原图像大小为  $384 \times 256$  或  $256 \times 384$ , 它们被分为 10 大类,分别为印第安人、海滨、建筑、公共汽车、恐龙、大象、花卉、马、高山以及食物。选择海滨、大象、马共 300 幅图像作为实验的语义标注对象。首先,对每一类图像随机地选择 20 幅,共 60 幅图像进行人工语义标注。由于同一幅图像含有几个不同的对象,因此可以用不同的关键词对其进行标注,如海滨类的图像含有大海、蓝天、沙滩等对象,大象(马)类的图像含有大象(马)和森林。在人工标注状态下,对一幅图像只用一个关键词进行标注并将其权值设置为 1, 该关键词反映的是当前用户对这幅图像所关心的语义信息。

图 4 和图 5 分别给出了  $\zeta$  取不同阈值对标注准确率的影响以及反馈次数对图像标注数量的影响。在图 4 中,  $P(0.15)$ 、 $P(0.25)$  分别表示  $\zeta$  取 0.15 和  $\zeta$  取 0.25 时标注的准确率,  $R$  表示图像标注的覆盖率,即被标注图像的数量占总的图像数量的比率。在实验中,判定一幅图像是否被正确标注的标准是该图像没有被不正确的关键词标注,如一幅马的图像,如果随着语义的传播它被标有三个语义“马”、“森林”、“大象”,则被认为是标注错误。

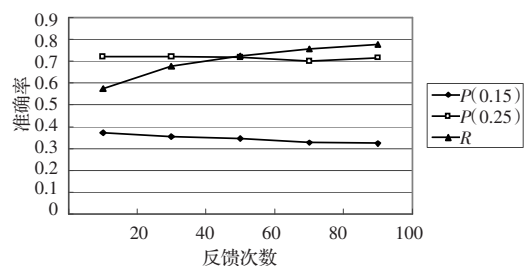


图 4 不同阈值对标注准确率的影响



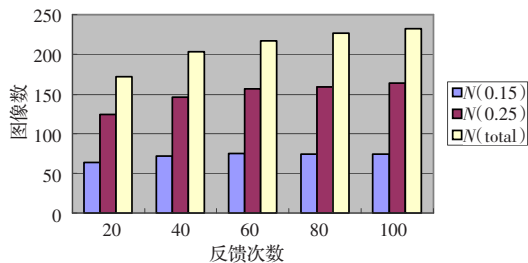


图5 反馈次数对图像标注数量的影响

从图4可以看出,随着反馈次数的增加,由于越来越多的图像被加入到反馈日志中去,覆盖率 $R$ 随着反馈次数的增加而不断增大;另外, $\zeta$ 的取值对标注的正确率有比较大的影响, $\zeta$ 取0.25时的检索准确率比 $\zeta$ 取0.15时有明显改善,这是因为适当地增大阈值可以去除不同图像间拥有的共同语义而造成的错误语义传递。如:马和大象这两类图像都拥有“森林”的语义,当用户进行“森林”的语义查询时,会同时选择含有森林背景的马和大象的图像,因此会造成在语义传播过程中,“马”的语义传给了大象,“大象”的语义传给了马,适当地增大阈值可以减少这类错误的标注。

在图5中, $N(0.15)$ 、 $N(0.25)$ 分别表示 $\zeta$ 取0.15、0.25时被正确标注的图像数量, $N(\text{total})$ 表示被标注的图像数量。从图中可以看出,无论是 $N(0.15)$ 、 $N(0.25)$ 还是 $N(\text{total})$ 都伴随着反馈次数的增加而逐步增加,但是相比于 $N(\text{total})$ , $N(0.15)$ 、 $N(0.25)$ 的增长缓慢,这就解释了虽然图1中 $P(0.15)$ 、 $P(0.25)$ 随着反馈次数的增加而略微有所下滑,但却随着 $N(\text{total})$ 的增速趋于平缓, $P(0.15)$ 、 $P(0.25)$ 也趋于稳定。

#### 4 结语

提出并实现了一个结合相关反馈日志与语义网络的图像标注方法。该方法以收集的用户相关反馈日志为基础获得图像的语义信息,通过计算图像间的语义相似度进行语义聚类,然后采用语义传播的方式实现图像的语义标注。实验结果表明,随着相关反馈日志库的不断扩充,库中越来越多的图像会在反馈的过程中得到标注且标注的准确率会随着反馈次数的增加而趋于稳定。此外,关键字更新策略及相应的权重更新策略丰富和完善了语义网络,适当地调整阈值 $\zeta$ 的大小能明显提高标注的准确率。

#### 参考文献:

[1] Smeulders A W M, Worring M, Santini S, et al. Content-based im-

age retrieval at the end of the early years[J]. IEEE Transactions on Pattern and Machine Intelligence, 2000, 22(11): 1349-1380.

[2] Vasconcelos N, Kunt M. Content-based retrieval from image databases: Current solutions and future directions [C]// Proceedings of International Conference on Image Processing, Thessaloniki, Greece, 2001: 6-9.

[3] Veltkamp R C, Tanase M. Content-based image retrieval systems: A survey[R]. Utrecht University, 2000.

[4] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models [C]// 26th Annual Int ACM SIGIR Conference, Toronto, Canada, 2003: 119-126.

[5] 茹立云, 马少平, 路晶. 基于 Boosting 学习的图片自动语义标注[J]. 中国图象图形学报, 2006, 11(4): 486-491.

[6] 于林森, 张田文. 基于视觉与标注相关信息的图像聚类算法[J]. 电子学报, 2006, 34(7): 1265-1269.

[7] Hare J S, Lewis P H. Saliency-based models of image content and their application to auto-annotation by semantic propagation [C]// Proceedings of Multimedia and the Semantic Web. European Semantic Web Conference, Heraklion, Crete, 2005.

[8] Rui Y, Huang T S, Mehrotra S. Relevance feedback techniques in interactive content-based image retrieval [C]// Proceedings of IS&T and SPIE Storage and Retrieval of Image and Video Databases, 1998: 25-34.

[9] Zhang H J, Su Z. Relevance feedback in CBIR [C]// Proceedings of 6th IFIP Working Conference on Visual Database Systems, Brisbane, 2002: 21-35.

[10] 王崇骏. 基于高层语义的图像检索算法[J]. 软件学报, 2004, 15(10): 1461-1469.

[11] Lu Y, Hu C H, Zhu X Q, et al. A unified framework for semantics and feature based relevance feedback in image retrieval systems [C]// Proceedings of the 8th ACM International Conference on Multimedia. New York: ACM Press, 2000: 31-58.

[12] M Li. A statistical correlation model for image retrieval [C]// Proceedings of 3rd International Workshop on Multimedia Information Retrieval, Ottawa, Canada, 2001: 42-45.

[13] Yin Xiaoxin, Li Mingjing, Zhang Lei, et al. Semantic image clustering using relevance feedback [C]// Proceedings of International Symposium on Circuits and System, 2003: 904-907.

[14] Wang J, Wiederhold G. Simplicity: Semantics sensitive integrated matching for picture libraries [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(8): 1-17.

(上接 90 页)

[2] Landauer T K, Foltz P W, Laham D. Introduction to latent semantic analysis [J]. Discourse Processes, 1998, 27(25): 259-284.

[3] Deerwester S, Dumais S T A. Indexing by latent semantic analysis [J]. Journal of the Society for Information Science, 1990, 41(6): 391-407.

[4] Mao Qinjiao, Feng Boqin, Pan Shanliang. Latent semantic analysis for query interfaces of deep web sites [J]. Journal of Southeast University: English Edition, 2008, 24(3): 312-314.

[5] 余正涛, 樊孝忠, 郭剑毅, 等. 基于潜在语义分析的汉语问答系统答案提取 [J]. 计算机学报, 2006, 29(10): 64-69.

[6] Luo Si, Rong Jin. Adjusting mixture weights of gaussian mixture model via regularized probabilistic latent semantic analysis [J]. Lecture Notes in Computer Science, 2005, 3518: 622-631.

[7] Bellegarda J R. Fast update of latent semantic spaces using a linear transform framework [C]// Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE ICASSP'02. New Jersey USA: IEEE Press, 2002, 1: 769-772.

[8] Akita Y, Kawahara T. Language model adaptation based on PLSA of topics and speakers for automatic transcription of panel discussions [J]. IEICE Transactions on Information and Systems, 2005, E88-D(3): 439-444.

[9] Artes-Rodriguez A, Figueiras-Vidal A R. Wiener extrapolation of sequences and the expectation-maximization algorithm [J]. IEEE Signal Processing Letters, 1996, 3(9): 260-262.

[10] KDD Project [EB/OL]. [2005-10-28]. <http://www.ecn.purdue.edu/KDDCUP/>.