

一种基于概念层次的图像检索方法^{*})

陈世亮 李战怀 袁柳

(西北工业大学计算机学院 西安 710072)

摘要 针对现有图像检索方法中存在的不同用户间因对图像的理解差异而存在的歧义问题,通过引入概念抽象的思想,将图像标注和检索中语义相关的关键词联系起来,统一不同用户对图像语义的理解和描述;结合相关反馈技术,确定检索关键词所代表语义概念信息,并实现概念层次的图像检索。实验结果表明,该方法能有效地提高图像的检索性能。

关键词 概念语义, 图像检索, 语义层次

A Novel Concept-based Image Retrieval Method

CHEN Shi-Liang LI Zhan-Huai YUAN Liu

(School of Computer Science, Northwestern Polytechnical University, Xi'an 710072)

Abstract To bridge the semantic gap between user's description and annotated labels, much interesting work has been attempted. Relationships between image annotation keywords and user used in retrieval are built by introducing the abstraction of concept, and then the description of image annotation is unified. By adopting correlative feedback technologies, the concept of the query keyword is selected, and then the concept-based image retrieval method is proposed. The experiments show that this method can improve the performance of retrieval.

Keywords Concept semantic, Image retrieval, Semantic hierarchy

1 引言

随着多媒体设备和互联网技术的发展,数字图像数量飞速增长。如何从海量图像数据库中快速找到用户感兴趣的图像,实现快速、高效的图像检索,是图像数据管理的核心任务之一。20世纪90年代以来,研究人员已进行了大量有关图像管理相关技术的研究。利用图像底层视觉特征,实现基于内容的图像检索(CBIR)是当前图像管理技术的研究热点之一^[1],并已取得许多成果,如 QBIC^[2]、Virage^[3]、VisualSeek^[4]等。利用图像视觉特征的检索方法主要采用基于例子(Query-By-Example (QBE))的查询策略,由用户指定查询所用的例子图像,然后根据特征距离函数计算检索图像与数据库中图像间的视觉特征距离,将满足条件的图像作为查询结果返回给用户。这种基于图像底层特征内容的检索方式一般要求用户指定例子图像,或给出期望检索图像的底层视觉特征的相关信息。而用户的习惯通常是以自然语言或概念的形式用关键词描述查询的语义信息,特别是当用户对查询的图像只有一个简单的语义轮廓时,很难直接提供查询计算所需要的底层特征内容的相关数据信息,使得这种方法的柔性(inflexible)较差。

相比较而言,采用基于关键词标注的方法更易于表达用户所要查询的概念。关键词不仅能描述图像中视觉对象的名称和属性等语义信息,还可以部分描述视觉上的内容信息。因此,通过关键词对图像进行标注,利用文本检索的相关技术实现图像检索,既便于用户的查询表示,也可以获得较好的查询效果。

基于关键词标注的图像检索技术最早出现在20世纪70年代,早期的方法通常采用手工标注图像的方式实现。这种方法的缺点在于:它完全依赖于人对图像语义的主观理解,以手工标注方式实现。由于用户的理解存在差异,不可避免会出现标注的歧义现象;另外,手工标注图像的工作量太大,难以满足快速增长的图像数量的需要。本文将 WordNet 本体应用到图像检索中,利用 WordNet 自身的概念层次结构,着重解决图像检索中由于用户主观理解和使用习惯不同而产生的对图像语义表述的不一致而产生的歧义问题。

2 相关工作

图像的语义信息通常采用文字描述,传统的方法以半自动或手动输入的方式提取语义特征,并用关键词标注图像的语义信息。在传统的基于关键词的信息检索中,通常采用准确的基于字符串匹配方式,即在查询比较时,检查查询关键词与被检索关键词是否完全一致。而图像检索则通常是一种基于相似性匹配的检索,是一种非精确检索的形式,只要求被检索图像与检索图像在视觉上相似即可。另一方面,这种关键词标注的方法只能实现对图像的简单标注,难以描述蕴藏在图像中丰富的语义信息和图像内容所描述的视觉信息。人们常说,一幅图胜过千言万语,图像的许多信息用文字是难以描述清楚的。而且,不同用户对图像语义的主观理解又不尽相同,受主观性影响较大,容易产生歧义^[5]。因此,图像的高层语义信息的表示需要综合考虑不同用户的理解,在检索时也不必强求检索关键词与标注关键词完全一致,这大大增加了基于语义图像检索技术实现的难度。

^{*})国家自然科学基金项目资助(60373108)。陈世亮 博士生,主要研究方向:多媒体信息管理、多媒体信息检索;李战怀 博士,教授,博士生导师,主要研究方向:数据库理论与技术;袁柳 博士生,主要研究方向:语义 Web 及信息检索。

从上面的分析,可以发现基于关键词的检索仅是一种字符串形式的匹配,是仅停留在语法层次的关键词匹配模式,未将标注图像所用的关键词间的语义关联关系纳入考虑范围,在检索中未能解决好图像自身特点所带来的问题。在实际应用中需要统一因用户对图像的理解及使用习惯上的不同而产生的差异,对于不同用户采用不同的关键词标注的图像语义,只要所表达的语义一致或相接近,所采用关键词本身并不一致,也不应对最后的查询产生影响。

目前,基于概念的信息检索技术已广泛应用于文本信息检索领域,并取得了较好的效果^[6,7]。将基于概念的信息检索技术应用到图像检索中的相关研究也已经出现,文[8]中采用基于 k 最邻近距离的聚类方法得到与概念相关的图像训练集,并通过基于支持向量机的学习方式,用概念表示图像的语义信息,产生图像的概念语义表示,实现基于概念的图像检索。基于统计分析的方法也是图像检索经常采用的方法之一,这种方法通常是在手工标注关键词的基础上,通过学习获取关键词与图像之间的关联关系,并将此关联关系传递给待标注的图像。由于用于训练学习的手工标注的图像数量有限,已标记的图像数据库为学习过程提供的相关信息不够充分,数量众多的关键词没能与图像正确的关联。为了解决上述问题,文[9]在基于内容的图像检索技术和统计分析的基础上,提出了一种改进的图像标注概率模型,根据关键词在不同语义层次的语义距离,确定标注新图像所采用的关键词。其中的语义层次包括图像层、关键词层和概念层。为了避免遗漏相关的关键词,在标注新图像时,该模型不是直接用关键词标注图像,而是用与关键词关联的概念去标注图像。

3 基于概念的图像检索方法

通过观察和分析用户的检索意图,可以发现用户检索时所提供的关键词表面上看是一个文本形式的符号,但在用户心理上却代表了与该关键词相关的概念信息,即把与此关键词相关的概念作为检索条件,期望系统返回与该关键词在语义上关联的所有图像。对用与查询关键词具有相同概念的其它关键词标注的图像,也是用户希望得到的结果。因此引入抽象的概念思想,将基于关键词标注的图像检索扩展到基于概念的检索形式,在概念层次实现图像的语义检索,可以较好地解决因用户理解不同存在的歧义问题。为了解决这个问题,需要一部能够表达图像概念的语义词典,这是基于概念层次图像检索处理工作的一项不可或缺的基础性资源。Princeton 大学的 WordNet 就是一个非常好的语义词典的范例,本文采用 WordNet 的同义词集合将关键词抽象为概念,并将标注图像的关键词扩展到概念层次,实现图像的检索。

3.1 关键词的概念层次

WordNet 的独特之处在于它是依据词义而不是词形来组织词汇信息的,它用同义词集合(Synset)代表概念(Concept)。将英语词汇组织为同义词集合后,每个同义词集合都有唯一的 ID 索引号,用于标明一个词汇概念,可看作是意义明确唯一的概念。WordNet 构造的核心是如何表示词汇概念节点,以及在概念节点之间建立起各种语义关系。同时力图在概念间建立不同的指针,实现同义词集合间的关联,便于表示单词之间的各种语义关系,如同义关系(synonymy)、反义关系(antonymy)、上下位关系(hyponymy)、部分/整体关系(meronymy)、形态关系(morphological relation)等,这样就构成了一个比较完整的词汇语义网络系统。

在 WordNet 的名词集中,共有 75804 个名词概念按层次方式组织单词之间的语义关系;共包括了 25 个基本类别,分别代表 25 个不同的语义类,处于较高层次的概念代表更抽象的意义,而处于较低层次的概念则代表较具体的意义,可以表示概念在不同抽象层次的语义信息。如:

shetland pony @ → pony @ → horse @ → equid @ → odd-toed ungulate @ → placental mammal @ → mammal @ → vertebrate @ → chordate @ → animal @ → organism @ → entity

其中:@ → 读作“is a”或“is a kind of”,表示单词间的层次语义信息。

从这个例子中可以看出,单词之间具有层次关系,即上下位关系。但单词的语义层次一般都是比较浅的,部分层次比较深的情况通常是由于专业词汇造成的,而不是日常语言中的用词。如上例有关“horse”不同抽象层次的语义描述中,虽然包含了 12 个层次,但其中有 6 个层次是专业词汇,而非日常用词。因此,在用 WordNet 实现基于概念的图像检索时可以主要考虑单词的同义关系。

3.2 概念的定义

通过把具有同义(或近义)关系的单词抽象为概念的形式,可以将标注图像所用的关键词扩展到概念层次。本文用关键词的同义词集合表示与该关键词相关的语义概念,将分离的单词联系起来,实现基于概念层次的图像检索。具体的定义如下:

定义 1(概念) 如果单词 $K_1, K_2, K_3, \dots, K_n$ 彼此间具有相同或相近的语义,即它们同属一个同义词集合,用概念 C 表示该同义词集,记为 $K_i \in C(i=1, 2, \dots, n)$ 。

引入概念表示后,形式上看仍然用关键词标注图像,但在实际检索中关键词已经不仅仅是一个文本符号,而是代表了与关键词相关的语义概念信息。通过这种概念抽象的形式可以将语义相关的不同关键词联系起来,解决图像标注和检索中由于不同用户的理解不同而存在的歧义问题。

经过这样的处理后,原本具体的单词被具有更抽象语义的概念所取代,而抽象的概念则被形式化了,概念之间还可以建立多种语义关系的联系和推理,便于用户的使用。

3.3 图像的语义层次

将标注图像的关键词扩展到概念语义层次后,图像的语义表示分为两层:关键词表示层和概念表示层,如图 1 所示。在基于概念的图像检索中,利用概念的这种层次结构将分离的词条联系起来。例如:对检索词 K_1 ,一方面,词 K_1 本身标注的图像满足检索条件,另一方面与 K_1 同属概念 C_1 的其它关键词如 K_i, K_j 所标注的图像也满足用户的检索条件。由此可见用概念(WordNet 中的同义词集合)来代替检索关键词 K ,从而将检索扩展到概念语义层次。

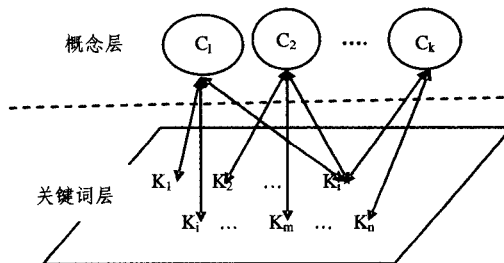


图 1 图像语义的抽象层次

3.4 基于概念的图像表示

引入概念抽象表示后,图像的标注划分为两个不同的抽象语义层次:关键词层和概念层。图像可以用以下三元组表示。

定义 2 图像(I)用三元组: $\langle id, K, C \rangle$ 表示。

其中 id 是图像的标识符; $K=(K_1, K_2, \dots, K_n)$ 是标注图像的所有关键词集合; $C=(C_1, C_2, \dots, C_m)$ 表示与 K 中关键词关联的概念集合,能表示图像的概念信息。

3.5 一词多义的处理

由于存在一词多义的情况,在不同的语义环境中,单词所代表的语义也将不同,即代表的概念也不同。在利用概念实现基于概念语义层的图像检索时,存在一个检索词与多个同义词集合关联的情况,此时往往无法判断用户的检索词与哪个同义词集合关联,比如用户输入“bank”,很难判断出用户是想查询关于“银行”方面的图像还是查询关于“河岸”方面的图像,这势必带来查准率的下降。因此采用基于概念的图像检索方式,将当前的查询从关键词扩展到同义词集时,首先必须明确查询单词在当前检索条件下所代表的确切语义,即在当前语义条件下,检索词所属的同义词集合,此时需要考虑用户的检索意图。

从已往的研究来看,考虑用户检索意图的最好方法就是采用相关反馈的形式。在图像检索中,用户的反馈通常采用查询点移动的方法,通过调整查询点的位置,使之更接近期望的结果,或通过修改相关特征的权值,从而提高查询的效率。与前面两种形式的反馈不同,本文采用反馈技术的目的是:为了在检索关键词扩展到概念层次时,将查询限制在单个同义词集内,即由用户根据检索意图从系统返回多个同义词集合中选定与关键词的语义相关的同义词集合。

例如用户输入的查询词是“horizon”时,系统将返回四个同义词集,分别代表四个不同的概念,如下所述。

第 1 个同义词集的语义:the line at which the sky and Earth appear to meet。该同义词集包含的所有单词为:horizon、apparent horizon、visible horizon、sensible horizon、sky-line。

第 2 个同义词集的语义:the range of interest or activity that can be anticipated;“It is beyond the horizon of present knowledge”。该同义词集包含的所有单词为:horizon、view、purview。

第 3 个同义词集的语义:a specific layer or stratum of soil or subsoil in a vertical cross section of land。该同义词集包含的所有单词为:horizon。

第 4 个同义词集的语义:the great circle on the celestial sphere whose plane passes through the sensible horizon and the center of the Earth。该同义词集包含的所有单词为:horizon、celestial horizon。

如果用户的实际检索意图在于检索与语义“the line at which the sky and Earth appear to meet”有关的概念时,则选择第一个同义词集合,并将其中的所有单词作为检索词,在数据库中查找满足条件的图像。

在实际检索中,概念通过同义词集合来体现。上例中,单词“horizon”作为查询关键词时,由于与其同属一个同义词集合的单词 horizon、apparent horizon、visible horizon、sensible horizon、skyline 共同组成同一个概念,因此除去用“horizon”本身标注的图像外,用“apparent horizon”和“skyline”等属于

同一个同义词集合中的单词标注的图像均可认为是满足检索条件的图像返回给用户。这样将单个词扩展到概念层次,即将标注图像的关键词扩展到了概念层次。

3.6 基于概念的图像检索算法

在图像检索中,对于用户脑子里已有的概念,可以找到一个对应的同义词集合表达该概念的有关信息,用户用于检索的关键词只需与此概念关联,无需直接在标注图像的关键词中出现,与此概念相关的关键词所标注的图像均可作为检索结果返回给用户。因此形式上用关键词标注的图像,实际上代表了图像的概念语义,同时还包含了图像在不同抽象层次的语义,从而解决了用户因对概念的理解不同,而选择不同的单词表达同一个概念的问题。因此只要所表达的概念一致,就可以将不同关键词统一到概念层次,为采用基于概念的图像检索提供支持。

基于概念的图像检索算法描述如下:

```

输入:查询关键词  $K_q$ ; 已标注的图像库  $I$ 
输出:与查询关键词概念相关的图像集合  $I_{kq}$ 
procedure  $img\_find(K_q, I)$ 
{
 $I_{kq} = \emptyset$ ;
Input  $K_q$ ;
 $S_{synset} = getSenses() \text{ of } K_q$ ;
//得到  $K_q$  的所有同义词集合
Select an proper set  $S_i$  from  $S_{synset}$ ;
//选择与检索语义相关的同义词集
For each  $K_i$  in  $S_i$  do
{
 $I_i = image \text{ annotated with } K_i$ ;
//查找用  $K_i$  标注的所有图像
 $I_{kq} = I_{kq} \cup I_i$ ;
}
return  $I_{kq}$ ;
}

```

4 实验结果

为了验证基于概念语义图像检索方法的有效性,本文的实验数据选自“Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary”^[10]的标注结果,该文使用的图像选自 Coral 图像库,共包括 4500 幅图像,根据图像所代表的语义信息不同,每幅图像采用 1 至 5 个不同的关键词标注,实际用于标注图像的关键词一共有 374 个。在本实验中我们比较同一个的关键词符号分别采用基于关键词检索和基于概念检索时的查询结果。

实验环境:P4(R) 2.8GHzCPU、1GB 内存;Windows XP 操作系统,程序用 JAVA 语言实现。

实验结果如图 2 所示,从中可以看出,采用基于概念的图像检索,能找到更多的满足条件的图像,这主要得益于采用基于概念的图像检索时,将查询关键词扩展到了与该检索关键词语义相关的同义词集中所有的单词。如查询词“ground”作为概念查询时,如选择的相关概念语义为“the solid part of the earth's surface”,则相应的同义词集合包括的单词有:land、dry land、earth、ground、solid ground、terra firma,除用“ground”直接标注的 60 幅图像外,其它相关同义词,如用“land”标注的 15 幅图像与“ground”在语义上也是相关的,同样也满足检索条件,因此也作为检索结果返回给用户。

当查询关键词未出现在已标注图像所用的关键词中时,效果更明显。如查询关键词“woodland”,在基于关键词的检索时,由于没有用“woodland”直接标注的图像,因此返回的结果为零。但由于在其同义词集中包含有词“forest”,而“forest”已经标注了部分语义相关的图像,因此在基于概念的检

(下转第 153 页)

在结合各自属性选择方法后的入侵检测分类结果上,快速属性选择算法与朴素贝叶斯分类器有更出色的表现。由于C4.5算法需要在内存中保存全部数据以方便计算,且构建的决策树较庞大,而本文提出的方法可以连续读入数据再进行处理,分类模型仅在内存中保留一种表的结构,因而本文方法的内存消耗在两个数据集上分别仅有C4.5算法内存消耗的1/12和1/42。由于朴素贝叶斯的条件独立假设简化了计算,使得其在分类时间上有一定优势。C4.5的高内存消耗是由于其采用了牺牲空间以换取时间的方法,但是我们注意到这样的做法并没有使得其分类时间优于朴素贝叶斯分类器,两者的分类精度都是令人满意的。

总结与展望 通过以上实验及分析,本文提出的快速属性选择方法可以高效地从大规模入侵检测数据中获得适合朴素贝叶斯分类器的属性。随着训练数据量的增加,训练模型精度进一步提升且时间及空间开销仅为线性增加,因而具有较好的实用性。当测试数据同训练数据的概率分布出现较大变动时,朴素贝叶斯分类器的准确率会受到严重影响,因而将朴素贝叶斯分类器应用于真实网络环境时需要不断地去探测网络数据分布是否发生了重大的改变,进而更新分类模型以

适应这种改变。由于快速属性选择可以高效地选择代表当前数据特征的属性,使得贝叶斯分类模型的更新代价进一步降低。在今后的工作中,将找寻如何度量当前网络数据是否发生变化的方法,进一步增强贝叶斯分类模型的适应性。

参考文献

- 1 Han J W, Kamber M. Data Mining: Concepts and Techniques. San Francisco, CA: Morgan Kaufmann, 2000
- 2 Ayoade J. Feature deduction and ensemble design of intrusion detection systems. Article Computers & Security, 2005, 24(6)
- 3 史忠植. 知识发现[M]. 北京: 清华大学出版社, 2002
- 4 Mitchell T M. 机器学习[M]. 曾华军, 张银奎, 等译. 北京: 机械工业出版社, 2003
- 5 王国胤, 于洪, 杨大春. 基于条件信息熵的决策表约简[J]. 计算机学报, 2002, 25(7): 759~766
- 6 Liu Huan, Setiono R. Feature selection and classification—a probabilistic wrapper approach[C]. In: Proceedings of the Ninth International Conference on Industrial and Engineering Applications of AI and ES
- 7 翟素兰, 郑诚. 用于入侵检测的基于粗糙集的贝叶斯分类器. 计算机技术与发展, 2006, 01-0226-02
- 8 Information and Computer Science University of California. Irving KDD cup 1999 Data [EB/OL]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
- 9 胡学钢, 郭亚光. 一种基于粗糙集的朴素贝叶斯分类算法. 合肥工业大学学报(自然科学版), 2006(2)

(上接第 141 页)

索时,用“forest”标注的图像也作为满足条件的图像返回,此时返回的图像数目为 71。

如果已标注的图像库中存在用查询的关键词标注的图像,而与该查询关键词同属一个同义词集合的其它单词没有

用于图像标注,则用该关键词检索时,基于概念的图像检索和基于关键词的图像检索,得到的图像数量是相同的。如检索词“wall”用于基于关键词和基于概念的检索时,由于不存在“wall”同义词标注的图像,因此满足检索条件返回的图像数量均为 98。

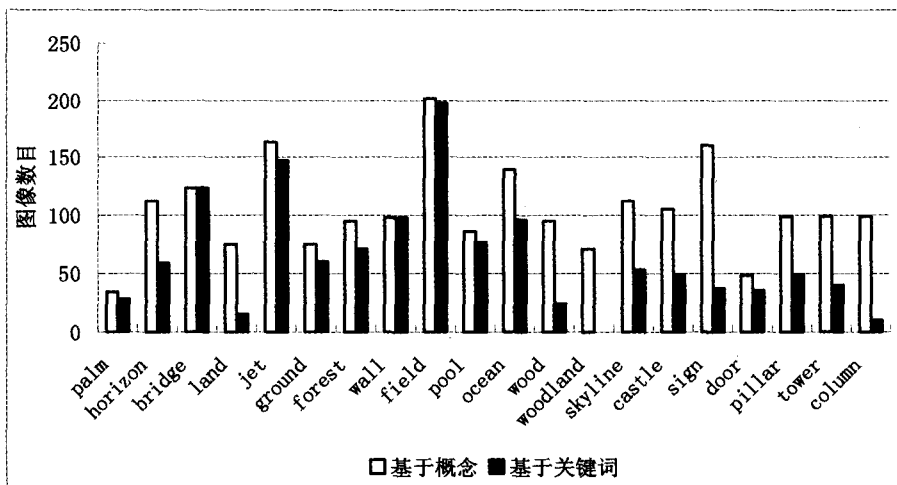


图 2 实验结果

结论与展望 本文通过采用 WordNet 同义词集合,将图像检索所用的关键词扩展到概念层次的方法,实现了基于概念语义的图像检索,能部分解决因用户的理解不同而产生的对图像语义理解的歧义问题,从实验结果可以看出,能较好地提高检索的性能。下一步的工作考虑在概念关系的定义中引入单词间的“上下位关系”、“部分与整体的关系”等复杂的语义关系,进一步提高图像检索的性能。

参考文献

- 1 Smeulders A, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years. IEEE Trans Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349~1380
- 2 Flickner M, et al. Query by image and video content. The QBIC system. IEEE Computer, 1995, 28(9): 23~32
- 3 Bach J R, Fuller C, Gupta A, et al. Virage image search engine: an open framework for image management. SPIE Storage and Retrieval of Image and Video DataBases, 1996, 4: 76~874
- 4 Smith J, Chang S F. VisualSEEK: A fully automated content-based image query system. In: Proceedings of the 4th ACM Mul-

- 5 timedia Conference, Boston MA, USA, 1996. 87~98
- 6 Rui Y, Thomas S H, Chang S F. Image retrieval: Past, present and future. Journal of Visual Communication and Image Representation, 1999, 10(1): 39~62
- 7 Woods W. Conceptual Indexing: a better way to organize knowledge. [Technical Report, SMLI TR-97-61]. Sun Microsystems Laboratories, Mountain View, USA, 1997
- 8 Clark P, Thompson J, Holmback H, et al. Exploiting a Thesaurus-based Semantic Net for Knowledge-based Search. In: Proceedings of the Twelfth Conference on Innovative Applications of AI (AAAI/IAAI'00), 2000. 988~995
- 9 Web Image Learning for Searching Semantic Concepts in Image Databases. WWW2004, New York, USA, ACM, May 2004 1-58113-912-8/04/0005
- 10 Cheng P J, Chien L F. Effective Image Annotation for Search Using Multi-level Semantics. Journal of Digital Libraries: Special Issue on Asian Digital Libraries, 2004. 258~271
- 11 Duygulu P, Barnard K, de Freitas N, et al. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: the Proceeding of Seventh European Conference on Computer Vision (ECCV2002), Copenhagen, Denmark, May - June 2002