一种基于本体与描述文本的网络图像语义标注方法

陈叶旺 钟必能 王 靖 李海波

(华侨大学计算机科学学院 厦门 361021)

摘 要 网络图像语义自动标注是实现对互联网中海量图像管理和检索的有效途径,而自动有效地挖掘图像语义是实现自动语义标注的关键。网络图像的语义蕴含于图像自身,但更多的在于对图像语义起不同作用的各种描述文本,而且随着图像和描述知识的变化,描述文本所描述的图像语义也随之变化。提出了一种基于领域本体和不同描述文本语义权重的自适应学习的语义自动标注方法,该方法从图像的文本特征出发考查它们对图像语义的影响,先通过本体进行有效的语义快速发现与语义扩展,再利用一种加权回归模型对图像语义在其不同类型描述文本上的分布进行自适应的建模,进而实现对网络图像的语义标注。在真实的Web数据环境中进行的实验中,该方法的有效性得到了验证。

关键词 图像标注,本体,语义扩展,回归模型

中图法分类号 TP301 **文献标识码** A

Semantic Annotation Method for Web Image Based on Domain Ontology and Image Description Texts

CHEN Ye-wang ZHONG Bi-neng WANG Jing LI Hai-bo

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract Semantic auto annotation on Web image is an important method for huge amounts of images management and retrieval on Web, and the semantic mining from the images automatically and effectively is the key. The semantic lies not only in the image itself, but also and more importantly in its description texts. Further, the image semantic varies with the change of images or description knowledge. To address this issue, in this paper, based on domain ontology and different image descriptions, we propose an adaptive semantic annotation method for Web images. This method checks the impacts on image semantic from the description texts feature. It detects the semantic and extends keyword by domain ontology, and then based on a regression model to adaptively model the Web images' semantic distribution on its different description texts. A group of experiments are carried out on a real-world Web image data set and the experimental results show that our proposed method outperforms others is with excellent adaptivity.

Keywords Image annotation, Ontology, Semantic extension, Regression model

1 引言

互联网上存在海量的数据,为了有效地组织、查询与浏览如此大规模的图像资源,图像检索技术应运而生,并且受到了广泛关注。现有的图像检索方式主要分为两种:基于内容的图像检索(CBIR)和基于文本的图像检索(TBIR)。CBIR先按图像的视觉特征(如颜色、纹理和形状等)建立索引,然后根据图像与用户提交的查询样图的视觉相似性度量来实现检索。但因为存在语义鸿沟,使得CBIR不能达到人们的要求。TBIR则是要求用户提交自然查询语言,这也要求对被检索的图像先建立好相应的文本索引,才能把用户的文本检索转化成为对索引的匹配查询。对于普通的用户来说,TBIR方式比CBIR更便捷,也是以Google,Yahoo为代表的主流搜索引擎采用的图像检索的主要方式。对图像建立相关索引,即实现图像的语义标注,是实验图像检索的关键,具有相当的挑战性。早期标注工作主要由专业人员手工完成,费时费力又具有相当的主观性,只能在数据量很小的范围内起作用。面对海量的数据图像,人们开始应用机器学习、统计摸型等多种方法来对图像做自动标注[8,25]。

本文认为图像的描述文本对图像语义表达有重要的作用。这些描述文本按其位置的不同,可以分为不同的类型。如图1所示,这些类型主要是图像文件名、替代文本、周边文本、标题文本、所属页面的标题。这些不同类型的文本对于网络图像的语义内容的重要性都不尽相同,而且随着图像和语义关键词变化,其语义内容也随之变化。现有的多数方法都把Web图像的所有关联文本作为一个整体,忽略了不同类型描述文本对图像语义起到的作用不同,少数方法对不同类型的关联文本赋予固定的权重,忽略了随着图像和描述知识的变化,描述文本所描述的图像语义也随之变化这一事实,这会对图像标注结果造成不良影响。

本文受国家自然科学基金(10901062),福建农业科技重大项目(2010N5008)资助。

陈叶旺 博士,讲师,主要研究方向为语义检索、数据挖掘;钟必能 博士,讲师,主要研究方向为计算机视觉、模式识别、机器学习;王 靖 副教授,主要研究方向为流形学习、人工智能;李海波 博士,副教授,主要研究方向企业信息化、软构件、工作流。

Web图像与其描述文本往往与图像语义表达有重要的联系，因此有些标注工作利用文本挖掘技术对图像的描述文本进行处理。另外，为了能使图像标注工作不致遗漏，往往需要对标注词进行语义扩展。现有的大多数已知标注工作通常是在标注词相互独立这个前提下完成的，这有许多不足。后来，人们利用标注词相互的语义共现性来对这种不足假设进行改进。现有的关键词的语义扩展的方法多是通过统计来估计语义共现性，或通过辞典得到一定的词汇间层次关系。但是这些方法得到的语义扩展信息只是从有限的字面上做扩展，难以挖掘一些深层次的语义。如'厦门在上海的南部，广州在厦门的南部，北京在上海的北部'，从这段文本描述中很难有传统的方法来得到'北京在厦门的北部'、'北京在广州的北部'等知识。

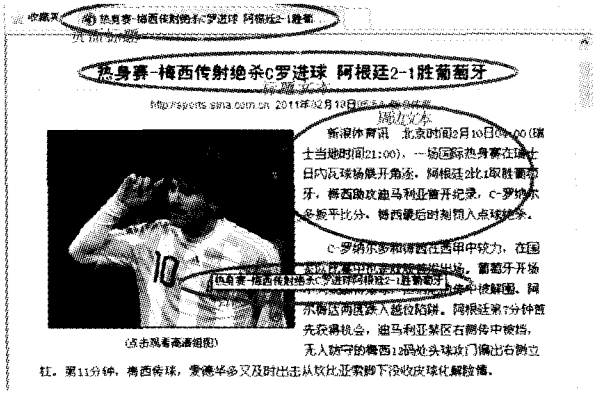

图1 网络图像及其各种描述文本

随着语义网[1]以及本体技术的提出和发展，语义网技术通过本体范地表达领域知识，使得计算机可识别和处理。这也为解决深层语义扩展的困局带来了曙光，因而基于本体的图像检索与语义标注工作随之出现。现有针对Web上的各种文档、图像、视频等提供的语义标注一般是人工方式对小规模文档资源进行标注[2-5]，但是面对海量数据这种方式显得无能为力。因此，自动或半自动工具也相即出现[6,7]，这些工具主要应用传统信息抽取技术的方法、基于本体的信息抽取方法、基于自然语言处理的方法等，取得了不错的效果。通过本体来进行图像语义检索和标注的工作也有一些[20,23,22,26]，但还处在起步阶段，这些基于本体的图像标注与检索技术的研究还不完善，对于图像的标注工作也大多是人工手动进行的，或者简单地对图像进行索引[22]。

因而，针对上述的网络图像语义特点和现有方法存在的不足与问题，本文提出了一种基于领域本体和网络图像不同描述文本语义权重的自适应学习的语义自动标注方法。该方法从图像的视觉特征和文本特征两个方面出发考查它们对图像语义的影响，先通过本体对描述文本进行有效的语义快速发现与语义扩展，再利用一种加权回归模型对图像语义在其不同类型描述文本上的分布进行自适应的建模，进而实现对网络图像的语义标注。这显著提高了网络图像语义自动标注的性能，是实现对互联网中海量图像进行有效管理和检索的重要一步。

本文第2节是国内外的相关工作；第3节介绍介绍Web图像语义标注知识；第4节介绍描述文本位置权重自适应估计方法；第5节是实验与评测；最后总结全文。

2 国内外相关工作

(1)图像标注方面

目前已经有了许多关于图像语义标注的研究工作，并取得了不错的成效。现有的方法对图像的语义标注主要可以分为基于概率模型的方法和基于分类的方法两大类。

基于概率模型的方法是从已标注的训练图像集中估计图像的视觉特征与语义标注词之间的联合概率分布。文献[8]将图像的标注问题看作是从图像视觉特征到关键字的映射，类似的工作有Jeon等人在文献[9]中提出的离散相关模型CMRM，该模型利用视觉特征与语义关键字之间的相关性(联合概率)对图像进行语义标注。CRM[10]和MBRM[11]应用的是基于核的非参数估计方法来估算图像特征生成概率，而对于标注词，则是通过多伯努力分布来进行估计，相比CMRM取得了一定程度的改进。

基于分类的方法将每个语义关键词看作一个类标签，应用有指导的可辨别分类方法来对图像做语义标注。基于模型的方法和基于支持向量机(SVM)是主要的分类方法。基于模型的方法，如Barnard等人[12]和Shi等人[19]分别提出的使用高斯混合模型对每个语义类的分布进行建模和使用层次多项式混合模型对每个语义概念进行建模。基于SVM的标注方法，把经过标注的图像看成类的正例，其余则是负例，以此来做二元分类。Jain[3]及Yang[4]都是基于SVM的标注方法。Carneiro[5]则是对图像做多类标注。

上述这些方法可归结为传统的方法，它们不适合对Web图像做标注。因为Web图像关联丰富的文本信息，且对Web图像语义作标注的词汇可能需要很多，而大多数上述的方法能使用的语义标注词都相当有限。因此，这些传统的标注方法大多在面对海量数据的Web图像时，性能和效果比较差。

Web图像的语义总是与其周边的文本信息有密切关系，因此可以用这些关联文本来作为发现图像语义的重要依据。Google和Yahoo正是如此来进行图像搜索的。Sanderson[24]将所有文本看作一个词集来进行语义标注；而Shen[13]对关联文本做了结构分析，对不同的位置做不同分析。Li[27]利用基于内容的图像检索方法得到一个和待标注图像视觉相似的初始集合，然后利用聚类技术，从这些图像集合中提取最有代表性的关键词作为待标注图像的语义标注词，这种方法受搜索阶段干扰较大且效果不大理想。

国内也有不少的相关工作，许[18]提出了一种自适应的Web图像语义自动标注方法，该方法首先利用Web标签资源自动获取训练数据；然后通过带约束的分段惩罚加权回归模型将关联文本权重分布自适应学习和先验知识约束有机地结合在一起，来实现Web图像语义的自动标注。Li[29]使用条件随机场进行图像标注的半指导学习，其目的是用来解决训练数据不足。Wang等人[14]提出了一个称之为AnnoSearch的图像标注系统，该系统利用搜索和数据挖掘技术，根据用户提供的语义关键词，获得和待标注图像语义相似的图像集，然后再做进一步处理。

前述的方法主要存在以下不足：1)多数方法把Web图像的所有关联文本作为一个整体，这忽略了不同类型描述文本的差别，对标注结果造成不良影响，因为它们对图像语义起到的作用不同；2)少数方法对不同类型的关联文本赋予固定的

权重，这种也有缺陷，因为图像不同类型的描述文本表达的语义会随着图像和语义关键词的改变，其语义内容也会随之改变。

(2) 图像描述文本处理与关键词语义扩展方面

Web图像与其描述文本往往对图像语义表达有重要的关系，但由于Web页面变化相当复杂，因此有些标注工作利用文本挖掘技术对图像的描述文本进行处理。文献[14,27]通过提取关键词，或者文献[17]提取专有名词短语，如时间、地点、人物等，来过滤掉语义不相关的内容。

前述现有的关键词的语义扩展的方法多是通过统计来估计语义共现性，或通过辞典可以得到一定的词汇间层次关系。但是这些方法得到的语义扩展信息仍然比较狭隘，只能从有限的字面上做扩展，对于一些深层次的语义无法被挖掘。如'厦门在上海的南部，广州在厦门的南部，北京在上海的北部'，从这段文本描述中很难用传统的方法得到'北京在厦门的北部''北京在广州的北部'等知识。利用本体的知识描述结构与推理可以解决这方面的困局，因而基于本体的图像检索与语义标注工作随之出现。

在利用本体进图像检索方面，文献[20]中作者提出基于本体的图像检索，其主要是借助本体改进关键字的方式进行图像标注；Hyvonen等[26]开发的Promoottori工具可支持基于本体和元数据的图像浏览与检索，但是要人工方式进行图像标注，应用范围也比较小，主要用于博物馆的图像检索，另外，其语义查找的相似度匹配用的是精确匹配，导致查全率不高；Mezaris[23]提出一种应用本体对象对图像进行标注的高层语义标注与查询方法，该方法需要把图像先分割，然后提取每个区的图像特征，再把这些底层特征与本体对象进行映射，最后才能利用本体对象来进行所谓的语义查询；SooVW[28]等开发出历史本图像检索系统，该系统基于共享领域本体和专业词典，所用的语义内容比较简单，标注工作是手工完成的，其首先对图像作文字注释，再对注释作语义标注。这些方法总的来说仍只是初步应用了本体对知识的描述信息，远未发挥本体的深层知识挖掘与推理功能。

3 Web图像语义标注

3.1 Web图像语义标注概念

简单来说，Web图像语义标注就是把在本体中的知识点和Web图像之间建立关联索引，表示这些图像是这个知识点的一个扩展描述。经过本体知识标注后的图像和知识构成一种关系，这种关系从另一个角度说明了图像与本体知识的相关程度。

标注工作按自动化程度划分，目前可分两种，一是完全的人工手动标注，也可以由工具实现半/全自动化标注。人工标注的可信度固然高，但是这种方法有很大的主观性，且在面对海量资源时，靠人工标注的方法几乎是不可能的。因而自动标注工具也必不可少，虽然其标注准确度不高，但更具有生命力。根据经验，现有的信息抽取(IE)技术能提高语义半自动化标注工具的准度和性能。本文也将只关注于这种自动化的方法。虽然信息抽取(IE)方面有很多的研究和进展，但是总的来说其缺乏公共的形式化知识标准与整合格式。这使得其标注的结果被其它系统所重用的代价相当的大，这也间接地阻碍了自然语言处理技术的进步。利用本体可以为领域知识定义类型、描述关系，建立一个知识库及其知识管理基础设施，并可为传统信息抽取技术提供格式参考，同时也解决了无类型问题，这为上述困难提供了解决方案。

3.2 标注框架

记 L_{train} 为图像训练数集，这个集合中的一张图像 J_i 是由它的标注实体与图像的描述文本特征组成，记作 $J_i=\{E_i, T_i\}$，其中 $E_i=\{e_{i1},\cdots,e_{il}\}$ 含有 l 个本体实体，e_{ij} 表示第 j 个本体实体与第 i 幅图像的相关度，取值在0到1之间；$T_i=\{T_{i1},\cdots,T_{in}\}$ 表示图像的不同位置的描述文本特征，T_{ij} 表示第 j 个描述文本向量，n 是特征个数。

未知图像 I，记 $p(e|T)$ 表示语义标注实体 e 在描述文本 T 中的生成概率，则不同类型的描述文本对图像语义描述起到的作用，可以用概率的形式来表示，即记 $X=\{p(e|T_1,\cdots,T_n)\}$；本文认为不同类型的描述文本对标注实体 e 标注图像的过程中起的作用也不尽相同，这可以通过各类型描述文本权重来表示，记为 $\omega=\{\omega_1,\cdots,\omega_N\}$，$N$ 是该幅图像的文本类型总数。因而一个侯选语义标注实体 e_i 作为图像 I 的语义标注实体的可能性 $p(e_i,I)$ 定义为如下线性模型：

$$p(e_i|I)=\sum_{j=1}^{N}\omega_j(e_i)X_{ij} \tag{1}$$

式中，$\omega(e_i)$ 表示对应的描述文本权重，X_i 为文本类型对图像语义的作用。图像 I 的最佳语义标注实体为：

$$e^*=\arg\max_e p(e|I) \tag{2}$$

假设实体 e 的各个三元组在描述文本 T_j 中是相互独立的，则 e 从文本 T_j 中生成的概率 $p(e|T_j)$ 可以估计如下：

$$p(e|T_j)=\begin{cases}\frac{1}{n}\sum_{i=1}^{n}p(tp_i|T_j), & Type(T_j)=TEXT\\ p(label(e),T_j), & \text{else}\end{cases} \tag{3}$$

式中，$p(tp_i|T_j)$ 表示三元组 tp_i 在文本 T_j 中生成概率；$label(e)$ 表示实体 e 的本体标签；$Type(T_j)$ 表示 T_j 的描述文本类型，对于TEXT(周边文本)类型的描述本文，我们按实体 e 的三元组形式检查它在文本中的生成概率，其余类型的描述文本(如标题、替代文本)则检查实体的标签词在文本中的生成概率。之所以这样处理是因为图像的周边文本是对图像详细的描述，而实体 e 的三元组集合也正是实体 e 的具体知识关系展开，其它类型的描述文本如标题是对图像的精练简要描述，实体 e 的标签值也具有同样的功率。具体实现如下：

假设图像的非周边描述文本中关键词的分布符合多项式分布，则可以利用极大似然估计方法来估计概率 $p(label(e,T_j)$：

$$\hat{p}(label(e)|T_i)=p_M(label(e)|T_i)=\frac{tf(label(e))}{|T_i|} \tag{4}$$

式中，$p_m(label(e)|T_i)$ 是对 $p(label(e)|T_i)$ 的极大似然估计，$tf(label(e))$ 是实体标签词在文本 T_i 中出现的词频，$|T_i|$ 是文本中出现的关键词的总个数。

对于图像周边描述文本中，实体 e 的三元组在文本中的生成概率计算如下：

$$\hat{p}(tp|T)=p_M(tp|T)=tf(label(tp.subject)\wedge label(tp.object)\wedge label(tp.predicate))/T \tag{5}$$

式中，$tp.subject$ 表示三元组的主语，$tp.object$ 表示三元组的宾语，$tp.predicate$ 表示三元组的谓语，$tf(label(tp.subject)\wedge label(tp.object)\wedge label(tp.predicate))$ 表示三元组的主语谓

语和宾语同时出现在文本 T 中的个数。

3.3 基于本体的语义扩展

按式(4)和式(5)来计算一个实体与图像描述文本的相关性可能会出现遗漏标注的情况，如实体的标签值并未在文本 T_i 中出现，则计算的结果为 0，这并不能说明二者完全没有关系。如关键词“树”和“鸟”都是某一张图像的语义标注词，且一起出现的频率很高，当某文本 T_i 只含“鸟”时，T_i 也应在一定程度上对“树”有语义蕴含。为尽量避免这种情况，本文可以通过本体推理功能进行扩展，如下：

本体实体分为类、属性、实例 3 种，其扩展方式有所不同，扩展规则如表 1 所列。

(1)类(概念)扩展：对一个概念实体 e，可以从知识库中选取与之相关的概念作为 e 的扩展。如上下位概念、等同概念(包括 owl:sameAs 和 owl:equivalentTo)等。

(2)属性扩展：与类的扩展类似，对一个属性实体 e，可以从知识库中选取与之相关的属性作为 e 的扩展。如上下位属性、等同属性(包括 owl:sameAs 和 owl:equivalentTo)等。

(3)实例扩展：除了可以通过等同关系、参照关系实现扩展外，还可以通过一些特殊属性关系推导出与其有等同关系的实体。在我们的农业领域本体中常用到这些特殊属性，主要有 FunctionalObjectProperty、SymmetricObjectProperty 和 Transtivity Property：

· FunctionalObjectProperty(功能性属性)：对于一个 owl:FunctionalObjectProperty $\boldsymbol{P}$ 而言，如果 $\boldsymbol{P}(X,Y)$ 与 $\boldsymbol{P}(X,Z)$ 都成立，$Y=Z$。那么根据这种逻辑关系，可以使用 owl:FunctionalObjecproperty 关系推导出与实体 e 等同的其它实体作为 e 的扩展集；

· SymmetricObjectProperty(对称性属性)：与 FunctionalObjectProperty 类似，可以使用 owl:SymmetricObjectProperty 关系推导出与实体 e 等同的其它实体作为 e 的扩展集；

Transtivity Property(传播性属性)：对于一个 owl:TranstivityProperty ObjectProperty $\boldsymbol{P}$ 而言，如果 $\boldsymbol{P}(X,Y)$ 与 $\boldsymbol{P}(Y,Z)$ 都成立，则 $P(Y,Z)$ 成立。根据这种逻辑关系，可以使用 owl:TranstivityProperty 关系推导出与实体 e 相关的其它实体作为 e 的扩展集。

表 1 扩展规则表

扩展规则	适用类别			优先级
	类	属性	实例	
Owl:sameAs	√	√	√	++++
Owl:equivalentTo	√	√	√	++++
Functional Property Same			√	++++
Symmetric Property Same			√	++++
Transtivity Property Extender			√	+++
下位实体	√	√		+++
上位实体	√	√		++
兄弟结点	√	√	√	+

算法 1 列出扩展伪代码。以 e 为原点，先找出与实体 e 等同的实体，再根据实体 e 的类别，以不同的策略实现扩展，并赋以不同的相关度值。假设取定一个阈值 H，则记 e 的扩展集合为：

$$EXTEND(e)=\{e' \mid \mathrm{sim}(e,e')>H\}$$

考虑到扩展后的实体集合，对式(3)做修改，则有：

$$p(e \mid T)=\max_{e'}(\mathrm{sim}(e',e) * p(e' \mid T)) \tag{6}$$

$\mathrm{sim}(e',e)$ 为实体 e' 与 e 的相似度，其中 $e' \in EXTEND(e)$。

算法 1 语义扩展算法

输入数据：本体知识库 O，本体实体 e

输出数据：扩展实体集合

0. 初始化实体扩展集合 EXTList={}
1. 找出所有与 e 存在 owl:s ameAs 关系的实体，相似度设为 1，加入 list 中
2. 找出所有与 e 存在 owl:equivalentTo 关系的实体，相似度设为 1，加入 list 中
3. 判断 e 的类别
4. 如果 e 是实例
   ①找出所有可以用 Functional Property 判断与 e 存在等同关系的实体，相似度设为 1，加入 list 中
   ②找出所有可以用 Symmetric Property 判断与 e 存在等同关系的实体，相似度设为 1，加入 list 中
   ③找出所有可以用 Transtivity Property 判断与 e 存在相关关系的实体，相关度设为 0.7^n，n 为传播距离，加入 list 中
5. 如果 e 是类或属性
   ①找出所有 e 的下位实体，相似度设为 0.8^n，n 为距离，加入 list 中
   ②找出所有 e 的上位实体，相似度设为 0.6^n，n 为距离，加入 list 中
   ③找出所有 e 的兄弟实体，相似度设为 0.5^n，n 为距离，加入 list 中
6. 返回 EXTList

3.4 描述文本位置权重的自适应估计

如前所述，图像的不同类型描述文本的语义随标注实体与图像的变化而变化，也就意味着描述文本的权重 $\omega(e)$ 应该是随之变化的。为估算 $\omega(e)$，本文作了如下假设：(1)一张网络图像的文本特征之间相互独立；(2)Web 图像若相似，则其描述文本也具有相似性。因而，图像的语义在描述文本上具有一定的统计规律性。所以，为找到这个规律，即估算描述文本的位置权重 $\omega(e)$，本文的方法是先找出与某张图像 I 相似的图像集合，即它的相似邻域 $neighbor(I)$，再在这个邻域中进行训练，对 $\omega(e)$ 进行加权回归估计。

3.4.1 描述文本权重的加权回归估计

给定一个语义标注实体 e，有可能会出现一种情况，即在图像 I 的相似邻域中，e 没有出现在某些训练图像的标注实体集合中。为了提高对应于图像 I 的不同位置描述文本和实体 e 之间的权重估计的准确性，记在相似邻域中受 e 标注过的训练图像为 $neighbor(I,e)$，并以它作为真实的训练数据集进行加权回归分析，估计 $\omega(e)$。

对于给定的一张在 $neighbor(I,e)$ 出现的图像 J_i，本文假设语义标注实体 e 在图像的描述文本上具有一定的规律性，即标注实体 e 与一组描述文本权重参数 $\omega'(e)$ 相对应。记标注实体 e 在图像的描述文本 T 中出现的概率为 $X=\{p(e \mid T_1,\cdots,T_n)\}$；记 $y_i=p(e \mid J_i)$ 表示语义标注实体 e 作为图像语义标注的概率，若 e 是训练图像 J_i 的语义标注实体，则 $y_i=1$；另记图像 J_i 和 I 二者之间的相似度为 $\mu_i=p(I \mid J_i)$，那么 e 在 J_i 中对应的 $\omega'(e)$ 与 e 在 I 中的 $\omega(e)$ 相同的可能性为 μ_i，因而，回归系数加权罚残差平方和：

$$Q=\sum_{i=1}^{|neighbor(I,e)|}\mu_i(y_i-\omega_0(e)-\sum_{j=1}^{n}(X_j\omega_j(e))^2 \tag{7}$$

式中，$\omega_0(w)$ 以 $\bar{y}=\sum_1^K y_i/K$ 估计($K=|neighbor(I,e)|$)，本文通过最小化 Q 来完成对权重分布参数的近似求解，分别对剩余的 $\omega_1(e)\cdots\omega_n(e)$ 求偏导，令它们等于 0，可得到一组正规方程组，并可将其化为矩阵形式：

$$\mu X^T X\omega = X^T y \tag{8}$$

则回归模型的解为：

$$\bar{\omega} = (\mu X^T X)^{-1} X^T y \tag{9}$$

3.4.2 自动标注算法

本节针对前述的所有过程给出最后的自动标注算法，具体如算法 2 所示。

算法 2 自动标注算(Anno_PWAL)

输入数据：训练图像集合 L，待标图像 I，领域知识本体 O

输出数据：待标注图像 I 的语义标注 anno

算法过程：

1. FOR i=1,…,|L|
2. 选择 neighbor(I)中包含语义标注实体 e_i 的图像组成 neighbor(I,e_i)；
3. 按式(3)计算描述文本生成标注实体 e_i 的概率；
4. 按式(9)估计 e_i 对应的描述文本权重分布 $\omega(e_i)$；
5. 按式(1)计算 e_i 可用于标注图像 I 的概率；
6. ENDFOR
7. 选择前 k 个实体作为图像 I 的语义标注实体

4 实验

为验证与比较本文提出的标注算法的性能，我们把本文方法与两个基准(baseline)算法作比较：(1) ModelFMD[21]，该方法基于文本特征进行分析后，再作加权来实现自动标注，在我们的实验中，二者的权重分别取 0.3 和 0.7；(2) ModelNTC[30]，该方法不区分图像描述文本的不同位置上对图像的语义影响的差别。

4.1 实验环境

我们在 Java 环境下，使用 Eclipse 作为开发平台。实验运行平台配置为 Intel Centrino Duo T2400 1.83GHz PC+2 GB 内存+WindowsXP SP2。

4.2 实验数据

4.2.1 实验数据的获得

为了考察我们方法的有效性和正确性，实验中建立两个不同规模的本体。第一个是农作物病虫害领域本体 CropDisease，第二个足球本体 Soccer。统计信息如表 2 所列。

表 2 本体统计数据

领域本体	本体概念总数	实例总数
新浪国际足球新闻(Soccer)	155	2533
农作物病虫害知识(CropDisease)	367	5725

对应于两种不同的本体，相应采用的检索文档集有两个，一个是新浪足球新闻网国际足球新闻；第二个是来自中国农科院作物品种资源研究所依据《中国粮食作物、经济作物、药用植物病虫原色图鉴》、《中国农业百科全书》所制作的农作物病虫害知识[1)]。我们使用一个“轻量级”的页面解析程序包 HTMLParser[15]将 Web 页面集合 P 中的页面转换成 DOM Tree，然后通过对 DOM Tree 的遍历获得初始的图像集 L 及其对应的描述文本。最后，我们对图像集 L 进行过滤以去除集合中的噪音图像，如长宽都非常小的 logo 图像、长宽比失调的图像等，最终得到包含 3200 幅图像的图像数据集 L。图像的描述文本特征共包含来自于 HTML 页面结构中 5 个不同位置的文本，分别是图像文件名文本、图像的替代文本、标题文本、周边文本、所属页面的标题文本，记为 $T_1 \cdots T_5$。转化后的文档集的统计信息如表 3 所列。

表 3 文档集的统计数据

领域文档集	文档数	图像数	T_1	T_2	T_3	T_4	T_5
新浪国际足球新闻(Soccer)	703	873	873	873	703	1378	703
农作物病虫害知识(CropDisease)	1119	3578	3578	3578	1119	5538	1119

4.2.2 训练数据集的自动获取

获取有效的训练数据是本文方法的一个前提，传统的方法是人工挑选训练数据，面对海量数据时，这种方法就显得不太有效。因此，本文提出一个自动获取训练数据的方法，其基本思想类似于 TF/IDF 的思想，通过考察实体 e 的扩展集合中的各实体标签或三元组成员标签在图像描述文本中出现的情况，来估计 e 作为图像语义标注实体的可能性。考查扩展实体 e 的标签值 $w=label(e)$ 与图像的描述文本之间的关系，我们利用的启发式规则如下：

规则 1 一个词汇 w 在越多图像的描述文本中出现，则它与单个其中的图像的相关度越低。

从词汇在知识空间的分布情况来分析，一个词语与越多的知识关联，它对概念的区分性就越不明显，它与单个知识的相关程度也就越低。

规则 2 一个词汇 w 在一个图像的描述文本中的词频越高，它与该图像之间的相关程度越高。

在与某个本体知识相关的文档空间中，对词汇进行词频统计可使统计粒度细，区分性强，更准确地刻画这个词对概念的所属程度。

规则 3 一个词汇 w 在一幅图像的越多的不同类型的描述文本中出现，它与该图像之间的相关程度越高。

记 T_i 为图像 I 的第 i 个类型的描述文本，记 $WS(I)$ 为图像 I 的描述文本的所有关键词集合，则给定一个词汇 $w \in WS(I)$，w 与图像 I 的语义相关度定义如下：

$$rel(w,I)=\begin{cases} EXP[-(cf(w)-1)]\times\dfrac{df(w)}{n}\times \\ \quad \sum_{i=1}^{n}\alpha_i \dfrac{tf(w,T_i)}{|T_i|}, & cf(w)>0 \\ 0, & \text{else} \end{cases} \tag{10}$$

式中，$cf(w)$ 表示描述文本中包含 w 的图像的数量，$df(w)$ 是包含关键词 w 的不同类型的描述文本的个数，$tf(w,T_i)$ 是 w 在描述文本 T_i 中的词频；$|T_i|$ 是描述文本 T_i 中包含的词汇总个数；α_i($\sum\alpha_i=1$)是描述文本 T_i 的权重。

给定置信度阈值 η，图像 I 的语义标签集合定义如下：

$$anW@(I)=\{w \mid w\in WS(I), rel(w,I)>\eta\} \tag{11}$$

那么对于图像 I 的实体标签集合定义如下：

$$anE@(I)=\{e \mid LabelSet(EXPEND(e)) \cap anW@(I) \neq \text{NULL}\} \tag{12}$$

式中，$LabelSet(EXPEND(e))$ 表示 e 的扩展集合中的所有实体的标签集合。

记 L 为初始图像数据集，L_{train} 为训练集，L_{test} 为测试集：

$$L_{train}=\{I \mid anE@(I)\neq\text{NULL}, I\in L\} \tag{13}$$

1) http://icgr.caas.net.cn/disease

$$L_{test}=L\backslash L_{train} \tag{14}$$

4.3 实验设置

我们将图像集 L 平分为训练、验证两个子集，目的是确定本文模型中的相似度变化阈值 ξ。通过反复验证，取 $\xi=0.8$。再把验证数据重新放回训练集中。评价方法性能通过查全率(Recall)、查准率(Precision)以及 F1 因子来度量。其中，F1 为：

$$F1=\frac{2*Recall*Precision}{Recall+Precision} \tag{15}$$

4.4 实验结果

本文所提基于本体的自适应学习的网络图像语义自动标注算法(Anno_PWAL)的优点在于依靠领域本体对关键词进行有效的语义扩展，并为不同图像及这些扩展语义关键词自适应地学习其对应的特定的描述文本位置权重分布。为了检验本文的方法，我们设计了以下几组实验：

1)本文所提网络图像语义自动标注方法从两方面对已有的标注方法进行改进，一方面是利用本体对关键词进行有效的语义扩展，另一方面，自适应地学习不同的图像和语义关键词对应的描述文本位置权重分布。因此，第一组实验主要用于检验本文提出的描述文本位置权重自适应学习(Adap)的有效性。

2)第二组是对语义内容在图像不同位置的描述文本上的分布情况作分析，其结果可作为分析图像语义的依据。

4.4.1 实验一：描述文本位置权重自适应学习的有效性验证

为了验证基于本体的描述文本位置权重自适应学习(Ont-Based)和基于关键字的描述文本位置权重自适应学习(Keyword-Adap)的方法对于提高网络图像语义自动标注性能的有效性，本组实验比较了它与两个基准方法的标注性能。图 2 给出了标注性能的比较结果。

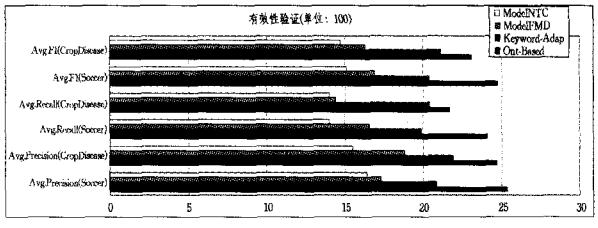

图 2 描述领域本体的文本位置权重自适应学习的有效性验证

由图 2 可以看出，Ont-Based、Keyword-Adap 两种方法和标注方法 ModelFMD、ModelNTC 相比，Ont-Based、Keyword-Adap 在 Recall、Precision 和 F1 上均有不同程度的提高。这说明：(1)领域本体对关键字的语义扩展起到了明显的成效，能有效地改进关键字的语义不够所造成的不足；(2)与将所有描述文本作为一个整体，或者仅仅根据先验知识对不同位置的描述文本赋予固定权重的标注方法相比，自适应地学习每一幅待标注图像和每一个候选语义标注词对应的描述文本位置权重分布的做法更加合理，可以更准确地估计描述文本对预测网络图像语义内容的贡献，从而有效地提高了网络图像语义自动标注的性能。

4.4.3 实验二：图像描述文本类型权重平均分布分析

图 3 对 PPWR 与 RR 两个模型作了比较，对两者独立估计出来的图像描述文本的不同位置权重平均分布情况，如前文所述，5 个位置分别是文件名文本、图像的替代文本、周边标题文本、周边文本、页面标题文本。由图可得：(a)由 RR 算法得到的权重 b 除替代文本(T_2)比 PPWR 低外均高于 PPWR；(b)图像的 5 个不同位置文本对语义的重要性排序依次如下：标题文本(T_3)、页面标题文本(T_5)、周边文本(T_4)、替代文本(T_2)、文件名(T_1)。

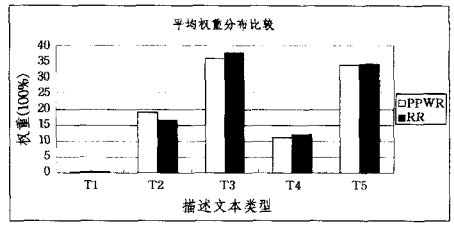

图 3 PPWR 与 RR 算法的平均权重分布比较

其原因可归结为以下几点：(1)精心取好名字的文章标题与页面标题是对语义的提纲挈领，能准确地表达页面语义，因而这两种类型的描述文本所占的比重较大；(2)替代文本一般指的是用于代替图像的语义描述文字，往往也能比较准确地描述语义内容，但不是每个页面上的图像都有相应的替代文本，因而其权重相对来说要比标题文本还低；(3)周边文本是对图像的展开描述，一般比较泛，且有许多不直接与之相关的内容，所以相关度次之；(4)在我们抓取的网页内容中，网络图像的文件名绝大多数是各种数字与符号合写、字母缩写等，基本不能表达语义，因而权重最低。

结束语 针对现有的图像标注方法的不足，本文提出了一种基于领域本体和描述文本位置权重自适应学习的网络图像语义自动标注方法，其通过本体对标注实体进行有效的语义快速发现与语义扩展，对不同类型的网络图像的描述文本进行分类考查，根据描述文本的不同自适应地学习与确定语义权重，再利用加权回归模型完成对网络图像语义分布的自适应建模。从各项实验中可以得到：(1)本体的语义扩展与推理对图像语义标注有较大帮助；(2)不同类型的图像描述文本对图像的语义影响不同，按重要性排序依次如下：标题文本(T_3)、页面标题文本(T_5)、周边文本(T_4)、替代文本(T_2)、文件名(T_1)；(3)不同类型的描述文本对图像的语义影响程度不是固定的，随图像以及标注实体的不同而呈现不同的分布，因而需要自适应地进行区别对待；(4)本文提出的标注方法显著提高了网络图像语义自动标注的性能。

参 考 文 献

[1] Mayfield J, Finin T. Information retrieval on the Semantic Web: Integrating inference and retrieval[C]// Workshop on the Semantic Web at the 26th Intl. ACM SIGIR Conference on Research and Development in Information Retrieval. 2003

[2] Tsarkov D, HorrocksI. FaCT++ Description Logic Reasoner: System Description[C]// Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJCAR). 2006

[3] Vires A, Roelleke T. Relevance Information: A Loss of Entropy but a Gain for IDF? [C]// SIGIR'05. Salvador, Brazil, 2005

[4] 熊文新，宋柔. 信息检索查询语句的表述分析[C]// 中国应用语言学会(筹)，第四届全国语言文字应用学术研讨会论文. 成都：

四川大学,2005

[5] Bozsak E, Ehrig M, Handschuh S, et al. Kaon-towards a large scale semantic Web[C]// Proceeding of the Third International Conference on E-Commerce and Web Technologies 2002. Springer,2002:304-313

[6] Sure Y, Erdmann M, Angele J, et al. On-toEdit: Collaborative ontology development for the semantic Web [C]// Proceedings of the first International Semantic Web Conference 2002. Sardinia, Italia, Springer, LNCS 2342, June 2002

[7] Farquhar A, Fikes R, Pratt W, et al. Collaborative ontology construction for information integration[R]. Stanford University, 1995

[8] Duygulu P, Barnard K, de Freitas J F G, et al. Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary [C] // Proceeding of European Conference on Computer Vision. Berlin: Spring-Verlag, 2002: 97-112

[9] Jeon J, Lavrenko V, Manmatha R. Automatic Image Annotation and Retrieval using Cross-Media Relevance Models [C] // Proceeding of International. ACM SIGIR. Toronto, ACM Press, 2003: 119-126

[10] Lavrenko V, Manmatha R, Jeon J. A Model for Learning the Semantics of Pictures[C]// Proceeding. of Neural Information Processing Systems (NIPS). Vancouver and Whistler: MIT Press, 2004: 553-560

[11] Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli Relevance Models for Image and Video Annotation[C]// Proceeding of the IEEE Conference. Computer Vision and Pattern Recognition. Washington DC, 2004: 1002-1009

[12] Barnard K, Forsyth D A. Learning the Semantics of Words and Pictures[C]// Proceeding International Conference on Computer Vision, Vancouver. Canada: IEEE Computer Society, 2001: 408-415

[13] Shen H T, Qoi B C, Tan K L. Giving meaning to WEB images [C]// Proceedings of ACM International Conference on Multimedia. LA, USA, 2000: 39-47

[14] Wang X J, Zhang L, Jing F, et al. AnnoSearch: Image auto-annotation by search[C]// Proceeding of the Conference on Image and Video Retrieval. 2006: 1483-1490

[15] http://htmlparser. sourceforge. net

[16] Jin R, Chai J Y, Si L. Effective Automatic Image Annotation via A Coherent Language Model and Active Learning[C]// Proceeding of International Conference on ACM Multimedia. New York: ACM Press, 2004: 892-899

[17] Ding J, Gravano L, Shivakumar N. Computing Geographical Scopes of Web Resource [C] // Proceeding of the. 26th Intel. Conference on Very Large Data Bases (VLDB 2000). Cairo: Morgan Kaufmann, 2000: 545-556

[18] Hua Z, Wang C, Xie X, et al. Automatic Annotation of Location Information for Web Images[C]// Proceeding International Conference on Mulitimedia and Expo (ICME). Amsterdam: IEEE Computer Society, 2005: 771-774

[19] Li T, Zhang C L, Zhu S H. Empirical Studies on Multi-label Classification[C]// Proceeding of 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI06). Washington: IEEE Computer Society, 2006: 86-92

[20] Hu Bo, Dasmahapatra S, Lewis P, et al. Ontology-based medical image annotation with description logics[C]// Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence. 2003: 77-82

[21] Tseng V S, Su J H, Wang B W, et al. WEB Image Annotation by Fusing Visual Features and Textual Information[C]// Proceedings of the 2007 ACM symposium on Applied computing, Symposium on Applied Computing. New York: ACM Press, 2007: 1056-1060

[22] Jiang Shu-qiang, Du Jun, Huang Qing-ming, et al. Visual ontology construction for Digitized Art image retrieval[J]. Journal of Computer Science and Technology, 2005, 20(6): 855-860

[23] Mezaris V, Kompastsiaris I, Strintzis M G. An Ontology Approach to Object Based Image Retrieval[C]// Proc. 2003 Int. Conf. on Image Processing. 2003: 511-514

[24] Sanderson H M, Dunlop M D. Image Retrieval by Hypertext Links[C]// Proceeding of the 20th Annual International ACM SIGIR. Philadelphia: ACM Press, 1997: 296-303

[25] Li J, Wang J Z. Real-Time computerized annotation of picture [C]// Proceeding of the ACM Int'l Conf. on Multimedia. Santa Barbara: ACM Press, 2006: 911-920

[26] Hyvonen E, Styrman A, Saarela S. Ontology-based Image Retrieval[C]// Proceedings of XML Finland 2002 Conference. Helsinki, 2002: 15-27

[27] Li X R, Chen L, Zhang L, et al. Image Annotation by Large-scale Content-based Image Retrieval [C] // Proceeding of the 14th ACM International Conference on Multimedia. Santa Barbara: ACM Press, 2006: 607-610

[28] Soo V W, Lee C Y, Li C C, et al. Automated semantic annotation and retrieval based OD sharable ontology and case-based learning Techniques[C]// Proc. 2003 Joint Conference on Digital Libraries. Washingtong DC. IEEE Computer Society, 2003: 61-72

[29] Li W, Sun M S. Semi-Supervised learning for image annotation based on eonditional random fields[C]// Proceeding of the conf on Image and video Retrieval. LNCS. 2006: 463-472

[30] 许红涛,周向东,向宇,等. 一种自适应的 Web 图像语义自动标注方法[J]. 软件学报,2010,21(9):2183-2195