基于词频同现与 WordNet 的图像自动标注改善算法研究*

柯 逍[1,2], 李东艳[3], 陈国龙[1,2]

(1. 福州大学 数学与计算机科学学院, 福州 350108; 2. 福建省科学工程计算重点实验室, 福州 350108; 3. 厦门大学 信息科学与技术学院, 福建 厦门 361005)

摘 要: 图像自动标注是模式识别与计算机视觉等领域中的重要问题。针对现有图像自动标注模型普遍受到语义鸿沟问题的影响,提出了基于关键词同现的图像自动标注改善方法,该方法利用数据集中标注词间的关联性来改善图像自动标注的结果。此外,针对上述方法不能反映更广义的人的知识以及易受数据库规模影响等问题,提出了基于语义相似的图像自动标注改善方法,通过引入具有大量词汇、包含了人知识的结构化电子词典 WordNet 来计算词汇间的关系并改善图像自动标注结果。实验结果表明,提出的两个图像自动标注改善方法在各项评价指标上相比以往模型均有所提高。

关键词: 图像自动标注; 标注改善; 词频同现; WordNet

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)07-2796-05

doi:10.3969/j.issn.1001-3695.2012.07.109

Automatic image annotation refinement based on keyword co-occurrence and WordNet

KE Xiao[1,2], LI Dong-yan[3], CHEN Guo-long[1,2]

(1. *College of Mathematics & Computer Science, Fuzhou University, Fuzhou* 350108, *China*; 2. *Fujian Key Laboratory of Scientific & Engineering Computing, Fuzhou* 350108, *China*; 3. *School of Information Science & Technology, Xiamen University, Xiamen Fujian* 361005, *China*)

Abstract: Image automatic annotation is a significant and challenging problem in pattern recognition and computer vision areas. At present, most existing image annotation models are influenced by semantic gap problem. This paper proposed a new image automatic annotation refinement method based on keyword co-occurrence to overcome above problem, which used the correlations between keywords in dataset to improve image annotation result. However, above method did not reflect the generalized knowledge of people and easy influenced by the size of dataset. Aiming at above problem, it proposed a new image automatic annotation refinement method based on semantic similarity to overcome above problem. This method used semantic dictionary WordNet to calculate the correlations between keywords and improve the image annotation results. Experimental results conduct on Corel 5K datasets verify the effectiveness of proposed image annotation method. The proposed automatic image annotation model improves the annotation results on all evaluation methods.

Key words: image automatic annotation; annotation refinement; keyword co-occurrence; WordNet

0 引言

随着数码相机与数码摄像机等数字化设备的普及发展以及互联网相关技术的高速发展,互联网上图像与视频等多媒体信息也相应地飞速增长,这就使得人们难以在海量数据中准确找出所需信息,所以如何对图像和视频信息进行有效的组织与管理就成了当前研究的热点问题。图像与视频检索相关技术是解决以上问题的一个有效途径,而图像自动标注(image automatic annotation)则是实现图像检索的关键步骤。图像自动标注是指根据图像的视觉内容,由计算机系统自动生成图像对应的标注关键词。早期的图像自动标注所关心的大多是图像的整体类别,所以更接近于图像分类。现在的图像自动标注主要是针对图像中不同的视觉内容同时进行标注。图像自动标注的方法目前主要有两大类:一类是利用商业化图像检索引擎(如 Google、Yahoo! 等),即采用自然语言处理相关技术,利用网页中图像的文件名、ALT 标签、URL 以及图像的上下文信息作为图像的标注,这类方法由于没有使用图像视觉特征,加之网络信息的随意性和不确定性,标注效果并不理想;另一类方法主要是通过图像的视觉内容产生相应标注词,可称为基于内容的图像自动标注,这类方法可以很好地构建图像视觉内容与关键词之间的联系。本文所研究的图像自动标注就属于此类。

1 相关工作

图像自动标注[1~3]与目标识别[4~7]有着相似之处,但从本

收稿日期: 2011-11-16; **修回日期:** 2011-12-24 **基金项目:** 国家自然科学基金资助项目(60873179,60803078,10871221);高等学校博士学科点专项科研基金资助项目(20090121110032)

作者简介: 柯逍(1983-),男,福建福州人,讲师,博士,主要研究方向为图像自动标注、模式识别(kex@fzu.edu.cn);李东艳(1985-),女,硕士,主要研究方向为图像自动标注;陈国龙(1965-),男,教授,博导,博士,主要研究方向为智能信息处理、计算机网络信息安全.

质上讲图像自动标注与目标识别隶属于不同的研究领域。图像自动标注并不关心每个目标在图像中出现的具体位置,比如在图像自动标注中,系统会将“花”“树叶”等词作为某一幅图像的标注,但并不会标出“花”与“树叶”在图像中的具体位置。目标识别一般关心特殊的前景物体,如车辆、行人与人脸等,通常都是针对不同目标分别构建不同的分类器。而在图像自动标注中,背景物体也同等重要。图像自动标注需要一次处理几百个甚至更多的目标,同时学习所有的标注词以及获得每幅图像的若干个标注词。目前,图像自动标注与目标识别都十分重要,且是具有挑战性的研究领域。

近年来,研究者们利用各种机器学习方法与统计模型,建立图像视觉内容与标注关键词之间的联系。同一关键词对应的视觉特征一般具有相似性,如“熊”,其颜色和纹理在视觉特征上基本保持一致。这样,图像可以被分割成一系列具有不同语义的区域,分割后的每个区域对应一个或两个语义对象,通过对这些图像块与语义关键词进行建模,就可以达到对图像进行标注的目的。Duygulu 等人[8] 提出了翻译模型(translation model),将图像自动标注看做是两种语言之间的翻译问题:一种语言由描述图像的视觉内容构成;另一种语言由文本词汇构成。通过 normalized cut[9] 对图像进行分割,并对所有区域进行聚类,图像的标注问题就可以看做是从视觉词汇关键词到语义关键词的翻译过程。翻译模型存在的一个问题是每个区域都对应于某个标注词,而实际上很多区域并不能用某个关键词进行标注。Blei 等人[10] 提出了相关 LDA(relevant latent Dirichlet allocation)模型,将 LDA 分布扩展到词汇与图像,模型假设 Dirichlet 分布可以产生一些隐变量,而这些隐变量可以用于生成图像区域与关键词。Jeon 等人[11] 提出了跨媒体相关模型(cross media relevance model, CMRM),利用视觉关键字与语义关键字之间的联合概率进行标注。CMRM 是一个经典的图像自动标注模型,由于本文提出的模型基于 CMRM,所以下面对 CMRM 进行简要介绍。

假设给定一幅未标注图像 π',$\pi' \in \Omega$。CMRM 使用 blob 来表示每幅图像,即 $\pi' = \{b'_1, b'_2, \cdots, b'_m\}$,希望能找到一系列关键词 $w'_1, w'_2, \cdots, w'_n$,该关键词集可以准确地反映出图像的视觉内容。通过训练集的 Γ 来估计关键词 w 与 $b'_1, b'_2, \cdots, b'_m$ 之间的联合概率。

利用图像训练集中 Γ 来计算关键词 w 与 blob 之间的联合分布,如式(1)所示。

$$P(w|b'_1,\cdots,b'_m) = \sum_{T\in\Gamma} P(T)P(w,b'_1,\cdots,b'_m|T) \tag{1}$$

假设给定了训练图像 T 之后,观察到 w 的事件概率与观察到 $\{b'_1, b'_2, \cdots, b'_m\}$ 的事件概率相对独立,即

$$P(w|b'_1,b'_2,\cdots,b'_m) = \sum_{T\in\Gamma} P(T)P(w|T)\prod_{i=1}^{m} P(b_i|T) \tag{2}$$

其中:$P(T)$ 对于图像集 Γ 中的所有图像都保持不变,$P(w|T)$ 与 $p(b_i|T)$ 的定义分别为

$$P(w|T) = (1-\lambda)\frac{\#(w,T)}{T} + \lambda\frac{\#(w,\Gamma)}{\Gamma} \tag{3}$$

$$P(b_i|T) = (1-\mu)\frac{\#(b_i,T)}{T} + \mu\frac{\#(b_i,\Gamma)}{\Gamma} \tag{4}$$

其中:式(3)和(4)右边第一项是在图像 T 中对应关键词或 blob 出现的概率,反映了局部性特征;第二项则是该关键词或 blob 在整个训练集中出现的概率,反映全局性特征;λ 与 μ 为概率估计中的平滑项。

2　基于关键词同现的图像自动标注

目前大部分的标注方法为了简化标注概率的计算,都假设标注的关键词之间是相互独立的,并没有考虑词与词之间的相关性对标注结果产生的影响,而实际上这是不合理的。例如,关键词集{sky, grass}比关键词集{ocean, grass}更有可能成为一幅图像的标注词。因为在图像中,sky 与 grass 同时出现的概率远大于 ocean 和 grass 同时出现的概率。在大部分的图像自动标注模型中,词与词之间的关联性常常被忽略,因此,针对现有模型存在的问题,本章通过利用标注词间的关联性来提高图像自动标注的结果。

2.1　标注词关联度

本节利用标注词同现(co-occurrence)的全局性方法,即从整个数据集中提取数据。对数据集中所有图像的标注词进行统计,即统计不同关键词同时出现在一幅图像中的次数,对于这些数据,本文以矩阵的形式进行存储。矩阵 co_frequence 的行与列的维数相等,即数据集中关键词的个数。矩阵 co_frequence 中的元素代表行所对应的关键词与列所对应的关键词共同出现的次数。如果行与列所对应的关键词是同一个关键词,那么它们所对应的元素代表这个关键词在数据集中出现的次数。

利用式(5)来计算标注词间的相关度,即

$$C_{w_u,w_v'} = \frac{\mathrm{Num}_{w_u,w_v}}{\mathrm{Num}_{w_u} + \mathrm{Num}_{m_v} - \mathrm{Num}_{w_u,w_v}} \tag{5}$$

其中:Num_{w_u,w_v} 表示关键词 w_u 和关键词 w_v 共同出现的次数,Num_{w_u} 和 Num_{w_v} 分别表示关键词 w_u 和 w_v 在数据集中出现的次数。依据式(5),可以建立一个以数据集中关键词的个数为行和列的矩阵,矩阵中的每一个元素表示的值是行与列所对应的元素的 C_{w_u,w_v} 值。如果两个关键词同时出现的次数越多,它们对应的 C 值就会越大,这也就意味着相对应的两个关键词很可能同时出现在一幅图像中。

式(5)只是通过两个关键词的同现次数来反映关键词之间的关系。为了能够更精确地反映关键词之间的关联度,还需要衡量这两个关键词与数据集中其他关键词的同现度。如果数据集中的其他关键词与这两个关键词的同现度都高,并且数据集中这样的关键词出现次数越多,表示这两个关键词的关联度就越高。关键词间关联度的计算公式如式(6)所示。

$$\mathrm{sim}(w_u,w_v) = \frac{S_{w_u}\cdot S_{w_v}}{|S_{w_u}|\cdot|S_{w_v}|} \tag{6}$$

其中:$S_{w_u} = \{C_{w_u,w_1}, C_{w_u,w_2}, \cdots, C_{w_u,w_n}\}$。在式(6)中,通过关键词与数据集中其他所有关键词的关系来计算词与词之间的关联度,如表 1 所示。

表 1　关键词间的相似度

	W(city)	W(mountain)	W(sky)	W(sun)
W(city)	1.000	0.0325	0.0599	0.1254
W(mountain)	0.0325	1.000	0.2823	0.0616
W(sky)	0.0599	0.2823	1.000	0.1222
W(sun)	0.1254	0.0616	0.1222	1.000

从表 1 可以看出,对角线元素均为 1,表示数据集中每个

关键词的自相关度。非对角线的矩阵元素表示行与列所对应的关键词的关联度,即式(6)中的 sim 值。

2.2 标注关联模型

关键词关联度的衡量需要考虑词与词之间的关系。这里采取 $k-1$ 层测量法,即每一次标注词选取时,都需要通过计算该词与前 $k-1$ 个标注词的关联度,选出与前 $k-1$ 个词关联度最高的关键词作为第 k 个标注词,具体如式(7)所示。

$$P_{\text{sim}}(w_k|N_{k-1}) = (1-\alpha)\times \text{scale}(w_k,N_{k-1}) + \alpha \times \text{relation}(w_k,N_{k-1}) \tag{7}$$

其中:$\text{scale}(w_k,N_{k-1}) = \dfrac{\#\{\text{co-occurrence}(w_k,N_{k-1})\}}{\#\{\text{co-occurrence}(N_{k-1})\}}$;$\#\{\text{co-occurrence}(w_k,N_{k-1})\}$ 表示关键词 w_k 与前 $k-1$ 个关键词共同出现的次数;$\#\{\text{co-occurrence}(N_{k-1})\}$ 表示前 $k-1$ 个关键词共同出现的次数,在数据集中,如果 $|N_k|$ 的值很大,$\text{scale}(w_k,N_{k-1})$ 的值中会出现很多 0,这是由于很多关键词同时出现的概率比较小。然而,$\text{scale}(w_k,N_{k-1})$ 的值为 0 并不代表第 k 个关键词和前面的 $k-1$ 个关键词的关联度就是 0,所以需要一个平滑公式,即如式(8)所示。

$$\text{relation}(w_k,N_{k-1}) = \begin{cases} \text{relation}(w_k)=1 & |N_k|=1 \\ \text{relation}(w_k,N_{k-1}) = \sum_{w'\in V\setminus N_{k-1}} \text{sim}(w_k,w') & |N_k|>1 \end{cases} \tag{8}$$

其中:$\sum_{w'\in V\setminus N_{k-1}} \text{sim}(w_k,w')$ 表示第 k 个要标注的关键词和前 $k-1$ 个关键词的关联性总和。如果前 $k-1$ 个关键词中与此关键词关联度高的关键词个数越多,则此关键词成为第 k 个标注词的可能性就越大。对式(8)还需要进行标准化,如式(9)所示。

$$\sum_{w_k\in V\setminus N_{k-1}} \text{relation}(w_k,N_{k-1}) = 1 \tag{9}$$

在 CMRM 的基础上,通过引入关键词间的关联度来提升图像自动标注的性能。基于关键词同现的图像自动标注模型的标注公式如下:

$$P(w_k|N_{k-1},I) = \beta^{P_{\text{sim}}(w_k|N_{k-1})} \times P_{\text{CMRM}}(w_k|b'_1,\cdots,b'_m) \tag{10}$$

其中:$P_{\text{sim}}(w_k|N_{k-1})$ 由式(7)进行计算;$P_{\text{CMRM}}(w_k|b'_1,\cdots,b'_m)$ 由式(2)进行计算;β 为参数,主要用于调节关键词间关联度的权重。本文提出的基于关键词同现的图像自动标注算法如下。

算法 1 基于关键词同现的图像自动标注算法

输入:两个词之间的相关度 $\text{sim}(w_k,w_v)$、关键词同现次数 $\#\{\text{co-occurrence}(w_k,N_{k-1})\}$。

输出:图像自动标注后的结果。

a)使用 CMRM 标注图像并得到每个词的标注概率。

b)计算 $\text{scale}(w_k,N_{k-1}) = \dfrac{\#\{\text{co-occurrence}(w_k,N_{k-1})\}}{\#\{\text{co-occurrence}(N_{k-1})\}}$。

c)利用式(8)计算 w_k 与前 $k-1$ 个关键词的关联度。

d)利用式(10)计算 w_k 的标注概率。

e)取标注概率最大的 K 个标注词作为最终的标注结果。

3 基于语义相似的图像自动标注

在计算词与词间的关系时,一种方法是利用上一章的词频同现方法,即计算训练数据集中已标注词汇间的同现概率。该方法基于已标注信息,相对准确,但它不能反映更广义的人的知识以及存在易受数据库规模的影响等问题。本章提出利用具有大量词汇的、包含了人的知识的结构化语义词典 WordNet 来计算词汇间的关系。与上一章的方法相比,WordNet 包括了更加完整的语义信息。

3.1 基于 WordNet 的相关度度量

本节度量关键词间相关度的方法,主要是将 Jiang 等人[12]提出的 JNC 方法和 Banerjee 等人[13]提出的 BNP 方法相结合。JNC 方法从一定程度上解决了 RIK 方法存在的问题[14],即 RIK 度量方法不能区分一个关键词与含有相同信息内容关键词之间的相关度。JNC 方法通过引入被度量关键词的根节点来解决上述问题,即计算两个关键词的关联度时,不仅要考虑这两个关键词的信息内容,同时还需要考虑这两个关键词的根节点中信息内容最大根节点的信息内容,这样就能有效地区分包含相同信息内容的关键词间的相关度。然而在图像自动标注中,JNC 方法也存在一定的缺点,如 JNC 方法对数据集较敏感以及存在部分关键词无法找到根节点等问题。这是由于 WordNet 是一个庞大的语义词典,在 WordNet 中,数据的存储形式是以树状形式存储的,即每对关键词的根节点基本上都能找到,然而图像自动标注使用的数据集规模要远小于 WordNet 的规模。也就是说,在 WordNet 中找到的根节点,在图像标注的数据集中不一定存在,这就是 JNC 度量方法存在的一大缺陷。为了更精确地度量关键词间的相关度,本文同时考虑了 BNP 方法。BNP 方法主要是通过关键词间注释集的重叠部分来计算词与词间的相关度。也就是说,如果两个关键词注释集中重叠部分越多,这两个关键词之间的关联度就越大。

JNC 的具体计算方法如下:

$$\text{sim}_{\text{JNC}}(w_1,w_2) = \max_{c_1\in \text{sen}(w_1),c_2\in \text{sen}(w_2)}[\text{sim}_{\text{JNC}}(c_1,c_2)] \tag{11}$$

其中:$\text{sim}_{\text{JNC}}(c_1,c_2) = \dfrac{1}{IC(c_1)+IC(c_2)-2\times IC(\text{lcs}(c_1,c_2))}$,通过此公式计算出两个概念的相关度,两个关键词的相关度为所有概念相关度中的最大值。信息内容的计算方法如式(12)所示。

$$IC(\text{concept}) = -\log \text{prob}(\text{concept}) \tag{12}$$

其中:prob 的计算公式为 $\text{prob}(c) = \text{freq}(c)/N$,$\text{freq}(c) = \sum_{n\in \text{word}(c)} \text{count}(n)$,$\text{freq}(c)$ 表示语义概念 c 在数据集中出现的总次数。

BNP 方法主要度量关键词注释集中的重叠部分。在 WordNet 中,名词间的关系主要有上位词(hypernym)、下位词(hyponym)、部分词(meronym)以及整体词(holonym)。此外,注释集(gloss)也是衡量关键词间相关度的一个重要指标。衡量关键词间相关度的关系集如下所示。

$$\text{related-pairs} = \{(\text{gloss},\text{gloss}),(\text{hype},\text{hype}),(\text{hypo},\text{hypo}),(\text{hype},\text{gloss}),(\text{gloss},\text{hype})\} \tag{13}$$

BNP 方法中重叠部分的提取如下:当比较两个注释集(gloss, gloss)时,重叠部分为两个注释集中最长的一个或多个连续的关键词序列,这个最长的重叠序列中的第一个关键词和最后一个关键词不能为代词、介词、冠词和连接词。如果两个注释集中包含多个长度及内容相同的重叠序列,那么只考虑第一次出现的重叠序列。将两个注释集中重叠部分的序列或单个关键词提取出来,并做好标记,然后继续在注释集剩下的部分寻找重叠部分。重复上述操作,直到没有重叠部分为止。

两个关键词相关度的计算公式如式(14)所示。

$$\mathrm{sim}_{\mathrm{BNP}}(w_1,w_2)=\sum_{\alpha,\beta\in\text{related-pairs}}\mathrm{score}(\alpha(w_1)+\beta(w_2)) \quad (14)$$

BNP 方法通过关系集找到重叠部分以后,开始计算重叠部分的分数 score。具体计算方法如下:如果重叠部分是单个关键词,那么重叠部分的 score = 1;如果重叠部分是多个关键词序列,那么 score 为序列中关键词数的平方。

3.2　基于 WordNet 的图像自动标注改善

将 JNC 与 BNP 方法相结合,如式(15)所示。

$$\mathrm{sim}(w_A,w_B)=\lambda\times\mathrm{sim}_{\mathrm{BNP}}(w_A,w_B)+(1-\lambda)\times\mathrm{sim}_{\mathrm{JNC}}(w_A,w_B) \quad (15)$$

通过上述公式,可以得到数据集中所有关键词间的关联度,对每一个关键词与其他所有关键词间的关联度进行度量,选取关联度值最大的关键词作为最终的标注词,剔除关联度小的关键词。

在 CMRM 的基础上,通过引入基于 WordNet 的相关度度量来提升图像自动标注的结果。本文提出的基于语义相似的图像自动标注改善算法如下:

算法 2　基于语义相似的图像自动标注改善算法

输入:两个词之间的相关度 $\mathrm{sim}(w_A,w_B)$。

输出:图像标注改善后的结果。

a)使用 CMRM 标注图像并得到每个关键词的标注概率。

b)取前 N 个标注概率最大的关键词作为候选标注词。

c)利用式(11)(14)分别计算候选标注词间的 JNC 度量与 BNP 度量。

d)利用式(15)计算某个候选标注词与其他所有候选标注词的相关度之和。

e)对标注结果进行重新排序,选取相关度最高的 K 个标注词作为最终的标注词。

4　实验

4.1　实验设置

为了验证本文提出的标注改善模型的有效性,使用了图像自动标注领域普遍使用的 Corel 5K 图像库。这个数据集共包含5 000幅图像,每幅图像有 1 ~ 5 个标注词,标注词总数为 374 个。将数据集分为 4 000 幅训练图像、500 幅验证图像以及 500 幅测试图像三个部分。其中,验证图像库主要用于确定模型参数,待参数确定后,将验证图像库中的所有图像加入训练库中形成新的训练图像库。这样就与其他模型普遍采用的 4 500幅训练图像、500 幅测试图像相一致,每幅图像固定返回 5 个标注词。

同其他图像标注模型一样,本文使用查准率、查全率与 F 度量来验证标注结果。假设 w 为某个关键词,$\#(s)$ 为返回的标注结果中包含 w 的图像数,$\#(c)$ 为正确标注的图像数,$\#(t)$ 为测试图像库中包含标注词 w 的图像数,则

$$Pr(w)=\frac{\#(c)}{\#(s)} \quad (16)$$

$$Re(w)=\frac{\#(c)}{\#(t)} \quad (17)$$

$$F(w)=\frac{2\times Pr(w)\times Re(w)}{Pr(w)+Re(w)} \quad (18)$$

对所有出现在测试库中的关键词进行计算,最后将得到每个词的查准率、查全率以及 F 度量的平均值作为最终的评价指标。此外,笔者还统计了至少被正确标注一次的关键词数量,记做 NZR。它反映了模型对标注词的覆盖程度,这也是一个很重要的标注性能评价指标。

4.2　实验对比

4.2.1　参数确定

第 2 章提出的基于关键词同现的图像自动标注模型中需要确定参数 α 的取值。通过实验确定参数 α 的最佳值,如图 1 所示。其中横坐标表示参数 α 的取值,纵坐标表示精确率与召回率的值。通过图 1 可以看出,当 α 取 0.1 时标注效果最好。

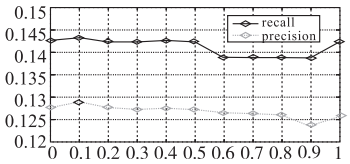

图 1　参数 α 对标注结果的影响

4.2.2　实验结果

本节将验证本文提出的两种图像自动标注改善方法的有效性,实验结果如表 2 所示。CMRM + C 表示在 CMRM 的基础上加入了本文提出的基于关键词同现的图像自动标注改善方法;CMRM + W 表示在 CMRM 的基础上加入了本文提出的基于语义相似的图像自动标注改善方法;CMRM + C + W 表示同时加入基于关键词同现与基于语义相似的图像自动标注改善方法。

表 2　标注改善模型的标注结果

方法	precision	recall	F	NZR
CMRM	0.094	0.108	0.101	77
CMRM + C	0.113	0.128	0.120	83
CMRM + W	0.097	**0.141**	0.115	90
CMRM + C + W	**0.117**	0.140	**0.127**	**91**

从表 2 可以看出,基于关键词同现的图像自动标注改善方法在精确率、召回率、F 度量及 NZR 上比基本的 CMRM 分别提高了 20.2%、18.5%、18.8% 和 7.8%;基于语义相似的图像自动标注改善方法在精确率、召回率、F 度量及 NZR 上比基本的 CMRM 分别提高了 3%、30.6%、13.9% 和 16.9%;基于关键词同现与基于语义相似的图像自动标注改善方法在精确率、召回率、F 度量及 NZR 上比基本的 CMRM 分别提高了 24.5%、29.6%、25.7% 和 18.2%。

4.2.3　结果分析

表 3 是几幅示例图的自动标注结果。将本文提出的图像标注改善模型与 CMRM 进行比较,加粗的词表示该词属于原始标注,即正确标注的关键词。这里,并没有选用标注结果最好的那些图像,而是选取了一些可以较好地反映本文模型特点的图像。本文提出的图像标注改善模型可以通过关键词间的关联度,筛选出相关度较大的关键词作为最终的标注词,这样就可以剔除一些噪声词。表 3 中的第二幅和第四幅图像在使用 CMRM 进行标注时,所有的标注词均标注错误,而本文的模型正确标注了 2 ~ 3 个关键词。尤其最后一幅图像,使用本文的模型可以正确标注所有 4 个关键词。由表 3 中几幅图的标注结果可以看出,通过 WordNet 考虑关键词间的关联度可以较

好地改善图像自动标注的标注结果。因此,本文提出的图像标注改善模型是有效的。

表3 标注结果对比

图像	原始标注结果	CMRM	本文模型
	jet sky plane smoke	sky water grouper moss aerial	sky plane jet aerial water
	grass bear meadow grizzly	tree jet plane field horses	bear grizzly horses mare foals
	sky sun clouds land	sky mountain sunset silhouette Scotland	sun clouds boats people sunset
	mountain water bear grizzly	clouds field grouper moss aerial	water bear grizzly meadow aerial
	rocks sand valley canyon	rocks sand grouper moss aerial	rocks sand valley canyon aerial

4.3 模型时间复杂度分析

利用CMRM计算词汇生成概率所需的时间为 $T(n)=l\cdot n$,其中 l 为标注词总数,n 为训练图像的数量,由于 $l<<n$,所以计算词汇生成概率的复杂度为 $O(n)$。在基于关键词同现的图像自动标注方法中,计算所有关键词间关联度的时间为 $T(l^2)$,迭代累积计算前 k 个标注词关联度的时间为 $T(k\cdot l^2)$,所以基于关键词同现的图像自动标注方法所需的总时间为 $T(n+l^2+k\cdot l^2)$,由于 $k<<l^2$,因此该方法总的时间复杂度为 $O(n+l^2)$。在基于语义相似的图像自动标注方法中,利用JNC方法计算两个概念相关度的时间为 $T(m\cdot l)$,而计算所有概念相关度的时间为 $T(m^2\cdot l^2)$,其中 m 为每个概念在WordNet中的层数。利用BNP方法计算任意两个关键词重叠注释集所需的时间为 $T(c\cdot l)$,而计算所有关键词重叠注释集的时间为 $T(c^2\cdot l^2)$,其中 c 为注释集的单词个数,所以基于语义相似的图像自动标注方法所需的总时间为 $T(n+m^2\cdot l^2+c^2\cdot l^2)$,由于 $m,c<<l$,所以该方法总的时间复杂度为 $O(n+l^2)$。结合基于关键词同现的图像自动标注方法与基于语义相似的图像自动标注方法的时间复杂度也同样为 $O(n+l^2)$,所以结合本文提出的方法相比传统的CMRM在时间复杂度上增加了 $O(l^2)$。

5 结束语

现有的基于内容的图像自动标注模型普遍受到语义鸿沟问题的影响,即每幅图像的标注结果很难保证在语义上保持一致。针对上述问题,本文提出了基于关键词同现的图像自动标注改善方法,利用标注词间的关联性来提高图像标注的结果。此外,本文还提出了基于语义相似的图像自动标注改善方法,通过引入具有大量词汇、包含了人知识的结构化电子词典WordNet来计算词汇间的关系。实验结果表明,本文提出的两个图像自动标注改善模型均有效。下一步的研究工作主要考虑如何在确保图像自动标注效果的前提下,将基于内容的图像自动标注用于大规模的网络图像标注。

参考文献:

[1] ZHAO Yu-feng, ZHAO Yao, ZHU Zhen-feng. TSVM-HMM: transductive SVM based hidden Markov model for automatic image annotation [J]. Expert Systems with Applications, 2009, 36(6): 9813-9818.

[2] QI Xiao-jun, HAN Yu-tao. Incorporating multiple SVMs for automatic image annotation [J]. Pattern Recognition, 2007, 40(2): 728-741.

[3] 路晶,马少平. 使用基于多例学习的启发式SVM算法的图像自动标注[J]. 计算机研究与发展,2009,46(5): 864-871.

[4] ANDRILUKA M, ROTH S, SCHIELE B. People-tracking-by-detection and people-detection-by-tracking [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2008: 1-8.

[5] MUTCH J, LOWE D G. Object class recognition and localization using sparse features with limited receptive fields[J]. International Journal of Computer Vision, 2008, 80(1): 45-57.

[6] KALANIT G, RORY S. Object recognition: insights from advances in fMRI methods [J]. Current Directions in Psychological Science, 2009, 17(2): 73-79.

[7] DOLLÁR P, WOJEK C, SCHIELE B, *et al*. Pedestrian detection: a benchmark[C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2009: 304-311.

[8] DUYGULU P, BARNARD K, FREITAS J, *et al*. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary [C] //Proc of the 7th European Conference on Computer Vision. Berlin: Springer, 2002: 97-112.

[9] SHI J, MALIK J. Normalized cuts and image segmentation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.

[10] BLEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3(3/1): 993-1022.

[11] JEON J, LAVRENKO V, MANMATHA R. Automatic image annotation and retrieval using cross-media relevance models[C]//Proc of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2003: 119-126.

[12] JIANG J J, CONRATH D W. Semantic similarity based on corpus statistics and lexical taxonomy[C]// Proc of International Conference on Research Computational Linguistics. 1997: 19-33.

[13] BANERJEE S, PEDERSEN T. Extended gloss overlaps as a measure of semantic relatedness[C]//Proc of the 18th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2003: 805-810.

[14] JIN Yo-han, KHAN L, WANG Lei, *et al*. Image annotations by combining multiple evidence & WordNet[C]//Proc of the 13th Annual ACM International Conference on Multimedia. New York: ACM Press, 2005: 706-715.