DOI:10.11992/tis.201606004
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20160808.0831.024.html

基于视觉注意机制和条件随机场的图像标注

孙庆美,金聪
(华中师范大学 计算机学院,湖北 武汉 430079)

摘　要:传统的图像标注方法对图像各个区域同等标注,忽视了人们对图像的理解方式。为此提出了基于视觉注意机制和条件随机场的图像标注方法。首先,由于人们在对图像认识的过程中,对显著区域会有较多的关注,因此通过视觉注意机制来取得图像的显著区域,用支持向量机对显著区域赋予语义标签;再利用 k-NN 聚类算法对非显著区域进行标注;最后,又由于显著区域的标注词与非显著区域的标注词在逻辑上存在一定的关联性,因此条件随机场模型可以根据标注词的关联性校正并确定图像的最终标注向量。在 Corel5k、IAPR TC-12 和 ESP Game 图像库上进行实验并且和其他方法进行比较,从平均查准率、平均查全率和 F_1 的实验结果验证了本文方法的有效性。

关键词:自动图像标注;视觉注意;词相关性;条件随机场
中图分类号: TP391　　**文献标志码**:A　　**文章编号**:1673-4785(2016)04-0442-07

中文引用格式:孙庆美,金聪. 基于视觉注意机制和条件随机场的图像标注[J]. 智能系统学报, 2016, 11(4): 442-448.
英文引用格式:SUN Qingmei, JIN Cong. Image annotation method based on visual attention mechanism and conditional random field[J]. CAAI Transactions on Intelligent Systems, 2016, 11(4): 442-448.

Image annotation method based on visual attention mechanism and conditional random field

SUN Qingmei, JIN Cong
(School of Computer, Central China Normal University, Wuhan 430079, China)

Abstract: Traditional image annotation methods interpret all image regions equally, neglecting any understanding of the image. Therefore, an image annotation method based on the visual attention mechanism and conditional random field, called VAMCRF, is proposed. Firstly, people pay more attention to image salient regions during the process of image recognition; this can be achieved through the visual attention mechanism and the support vector machine is then used to assign semantic labels. It then labels the non-salient regions using a k-NN clustering algorithm. Finally, as the annotations of salient and non-salient regions are logically related, the ultimate label vector of the image can be corrected and determined by a conditional random field (CRF) model and inter-word correlation. From the values of average precision, average recall, and F1, the experimental results on Corel5k, IAPR TC-12, and ESP Game confirm that the proposed method is efficient compared with traditional annotation methods.

Keywords: automatic image annotation; visual attention mechanism; inter-word correlation; conditional random fields

随着互联网的不断发展以及移动终端的迅速发展,图像数据不断扩大。图像数据大规模的增长对图像理解技术提出了更高的要求。如何从巨大的图像库中快速有效地找到想要的图像,已经成为了一个亟待解决且具有很大挑战性的任务。而图像标注技术是数字图像语义文本信息的关键技术,在数字图像处理的各个方面有着广泛的应用[1]。

图像标注技术就是为给定的图像分配相对应的语义关键词以反映其内容[2]。早些年的图像标注

收稿日期:2016-06-02.　网络出版日期:2016-08-08.
基金项目:国家社会科学基金项目(13BTQ050).
通信作者:金聪. E-mail: jinc26@aliyun.com.

技术需要专业人员根据每幅图像的语义给出关键词,但那样的方法会消耗大量时间并且带有一定的主观性。因此近几年来,有不少的研究者将注意力转移到图像的自动标注技术上来。就当下的自动标注方法而言大致可以分为两类:1)基于生成式的图像自动标注方法[3-4];2)基于判别式的图像自动标注方法[5-6]。前者主要是先对后验概率建模,然后依据统计的角度表示数据的分布情况,以此来反映同类数据本身的相似度。文献[3]就属于该模型,它将标注问题转化成一个将视觉语言翻译为文本的过程,再收集图像与概念之间的关系以此来计算图像各个区域的翻译概率。文献[4]提出的跨媒体相关模型,将分割得到的团块进行聚类,得到可视化词汇,然后建立图像和语义关键词之间的概率相关模型,估计图像区域集合与关键词集合总体的联合分布。与此类似的方法还包括基于连续图像特征的相关模型,该类方法也存在一定的问题,如当遇到图像过分割和欠分割的时候标注性能大大降低,虽然可以通过改进算法来提高标注结果,但这样增加了计算的复杂性,不具备在真实环境应用的条件。另外,可以构建图像特征与标注词之间的关系模型,然而该模型一般情况下复杂度较高,而且无法确定主题的个数。而后者则是通过寻找不同类别之间的最优分类超平面,从而反映异构数据之间的不同。也就是说,该模型为每个类训练一个分类器,以此来判断测试图像是否属于这个类。文献[2]提出了 MRES-VM 算法,即一个基于映射化简的可扩展的分布式集成支持向量机算法的图像标注。为了克服单一支持向量机的局限性,利用重采样对训练集进行训练,建立了一种支持向量机集成方法。在文献[5-6]中提到的方法也属于判别模型。这两者既有优点又有缺点。相比之下,判别式模型可以实现更好的性能。已有的图像标注方法没有得到较好的标注准确率,主要是由于它们使用的图像内容描述方法和人们对图像的理解方式相距甚远。实际上,当人们看一幅图像的时候,不会把注意力平均分配到图像的各个区域,而是会有选择地把注意力集中到显著区域。由此本文提出了一种基于视觉注意机制和条件随机场的图像自动标注方法。

1　显著区域的标注过程

本文使用的图像标注算法,主要是将传统依据底层特征的标注方法和人们认识图像的方式结合在一起,然后又利用标签之间的共生关系对标注词进行校正,得到最终标注词。

在使用本文所提算法之前要先对图像进行预处理,然后使用基于视觉注意机制和条件随机场算法对图像进行标注。算法主要流程如下:

输入　训练图像和测试图像的混合图像集;

输出　所有图像对应的标签集。

1)使用支持向量机对显著区域进行识别并标注;

2)对于非显著区域,结合训练图像库的图像与标签关系进行标注;

3)使用条件随机场模型对每幅图像的标签进行优化。

1.1　显著区域的提取

当人们看一幅图像时,注意力更多地放在显著区域而不是非显著区域。图像的显著区域指的是在一幅图像中最能引起人们视觉兴趣的部分,图像的显著区域和图像要表达的含义往往一致。充分利用这一点能提高图像标注的准确率。基于此,本文选择先对显著区域进行标注,然后标注非显著区域。这种方法可以消除非显著区域对显著区域的影响,由此获得更好的标注效果。

在图像处理方面,很多获取图像显著区域的模型已被提出。例如,文献[7]提出了一种显著区域的获取方法,它主要结合像素特征和贝叶斯算法。文献[8]提出了一种视觉显著性检测算法,它将生成性和区分性两种模型结合在一个统一的框架中。这些区域通常具有较大的共同特征,面积相对较大且亮度更高。因此本文提出一个新的方法来提取显著区域,也就是视觉注意机制。定义如下:

在利用 N-cut 算法对图像分割后,根据视觉注意机制求得图像的每个区域的权重。视觉注意机制模型为

$$W = \omega \cdot \text{Area} + (1 - \omega) \cdot \text{Brightness} \quad (1)$$

式中:W 表示图像中每个区域的显著度;ω 表示权重。为获得图像的显著区域,本文通过大量实验来得到 ω。计算并比较各个区域的显著度 W 的大小,W 值最大的区域就是该图像的显著区域。模型(1)中各参数的意义如下:

a)面积参数 Area。在该模型中,Area 是参数之一,一般情况下,面积越大的区域越能引起人们的注意,但是不能过大,过大面积的区域会使得显著度降低。具体计算式为

$$\text{Area} = S_i / S \quad (2)$$

式中:S_i 表示每幅图像中第 i 个区域的像素个数;S 表示整幅图像的像素个数。

b)亮度参数 Brightness。亮度参数是获得显著区域最重要的参数。HSV 颜色模型比较直观,在图

像处理方面是一种比较常见的模型。定义一个区域的亮度为该区域和图像其他区域 HSV 值的方差,用式(3)计算。也就是说,先计算图像中所有区域 HSV 的平均值,然后计算每个区域 HSV 的值,最后取得各个区域的亮度值。具体公式为

$$\text{Brightness} = (m_i - \bar{m})^2 \tag{3}$$

式中:$\bar{m}$ 是图像的各个区域亮度的平均值,m_i 是第 i 个区域的亮度值。

1.2 显著区域的标注

每一幅图像中都包含不等个数的区域,这些区域或简单或复杂、或大或小,而它们都有不一样的语义。传统的标注方法中,对图像的各个区域同等对待,而事实上人们往往把更多的注意力集中在显著区域。所以可以利用式(1)求出每幅图像的显著区域进行单独标注,对非显著区域的区域在后续的步骤中进行标注。

在对显著区域进行标注时,用一组训练图像训练 N 个支持向量机分类器 $C = \{c_1, c_2, \cdots, c_n\}$。具体来说,对一组训练图像利用视觉注意机制提取显著区域,再对每个显著区域提取它们的底层特征构成特征向量,并作为输入训练支持向量机。

近年来支持向量机已经被广泛地应用于图像标注中,像文献[2]和[9]。在最简单的情况下,支持向量机是线性可分的支持向量机,这时必须满足数据是线性可分的。但是在实际应用中,线性可分的情况很少,绝大多数问题都是线性不可分的。在遇到线性不可分的问题时,可以通过非线性变换将它映射到高维空间中,从而转化为线性可分问题。SVM 的学习策略就是最大间隔法,可以表示为一个求解凸二次规划的问题。设线性可分样本集为 $(x_i, y_i), i = 1,2,\cdots,n, , i = 1,2,\cdots, n, x_i \in R^n$, $y_i \in \{+1, -1\}$ 是类别标号。通过间隔最大化或等价地求解相应的凸二次规划问题学习得到的分离超平面为 $\boldsymbol{w}^{\mathrm{T}} \cdot x + b = 0$,线性判别函数为 $g(x) = \boldsymbol{w}^{\mathrm{T}} \cdot x + b$。然后将判别函数进行归一化,使两类中的所有样本都必须满足条件 $|g(x)| \geqslant 1$,即让距离分类面最近的样本的 $|g(x)|$ 值等于 1,这样分类间隔就等于 $2/\|\boldsymbol{w}^{\mathrm{T}}\|$,因此使间隔最大等价于使 $\|\boldsymbol{w}^{\mathrm{T}}\|$ 最小;分类线若要对所有样本都能正确分类,那么它必须满足以下条件:

$$y_i(\boldsymbol{w}^{\mathrm{T}} \cdot x_i + b) \geqslant 1, i = 1,2,\cdots,n \tag{4}$$

因此,要求得最优分类面,可以将问题转化成约束化问题,也就是在式(4)的约束下,使目标函数 $\max \dfrac{1}{\|\boldsymbol{w}^{\mathrm{T}}\|}$ 的值达到最小。为此定义拉格朗日函数:

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\|\boldsymbol{w}^{\mathrm{T}}\|^2 - \sum_{i=1}^{n} \alpha_i (y_i(\boldsymbol{w}^{\mathrm{T}} x_i + b) - 1) \tag{5}$$

通过对 $\boldsymbol{w}$ 和 b 求解,计算出拉格朗日函数的极小值。再利用 KKT 条件对分类决策函数求出最优解,最终结果为

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n} \alpha_i^* y_i (x_i \cdot x) + b^*\right) \tag{6}$$

式中:α^* 为最优解,b^* 为分类的阈值。

分类时先提取测试图像的显著区域,然后提取图像显著区域的特征值,构成特征向量输入到训练好的支持向量机分类器中,得到每个显著区域的标注词。

2 非显著区域的标注过程

对图像的非显著区域进行标注时,本文将带有标签的图像区域引入对其进行标注。本文将未被标注的非显著区域和带有标注词的图像区域混合在一起,使用 k 近邻法(k-nearest neighbor, k-NN)聚类算法进行聚类,最终求得非显著区域的标注词。k-NN 算法的思路:假设给定一个训练数据集,里面的实例都有确定的类别,对测试实例,根据其 k 个最近邻的训练实例的类别,通过多数表决方式进行预测。具体的流程如下:

输入 待标注的非显著区域和带标签的图像区域;

输出 非显著区域的标注词。

1)在带有标签的图像区域中找出与每个待标注的非显著区域相似的 K 个样本,计算公式为

$$\text{Sim}(d_i, d_j) = \left(\sum_{k=1}^{M} W_{ik} \times W_{jk}\right) \Big/ \sqrt{\left(\sum_{k=1}^{M} W_{ik}^2\right)\left(\sum_{k=1}^{M} W_{jk}^2\right)} \tag{7}$$

2)在每个非显著区域的 k 个近邻中,分别计算出每个类的权重,计算公式为

$$p(\boldsymbol{x}, C_j) = \sum_{\boldsymbol{d}_i \in kNN} \text{Sim}(\boldsymbol{x}, \boldsymbol{d}_i) y(\boldsymbol{d}_i, C_j) \tag{8}$$

式中:$\boldsymbol{x}$ 为待标注区域的特征向量,$\text{Sim}(\boldsymbol{x}, \boldsymbol{d}_i)$ 为相似性度量计算公式,与上一步骤的计算公式相同,而 $y(\boldsymbol{d}_i, C_j)$ 为类别属性函数,即如果 $\boldsymbol{d}_i$ 属于类 C_j,那么函数值为 1,否则为 0。

3)比较类的权重,将待标注区域划分到权重最大的那个类别中。这样非显著区域就得到了相应的标注词,同时也得到了获得该标注词的概率。

3 标注词校正

设每一幅待标注图像分割为 n 个子区域 D_i

($i = 1, 2, \cdots, n$)。在得到一幅图像的显著区域标签和非显著区域标签集合后,将这些标签整合成图像的标签向量:

$$\boldsymbol{r} = [p(a_{\text{focus}}) \quad p(a_1) \quad \cdots \quad p(a_{n-1})]$$

式中:$p(a_n)$表示该图像的第 n 个区域获得标注词 a_n的概率。本文使用条件随机场对图像已获取的标注向量进行校正,最终获得图像的标注词。自从条件随机场被提出以来,已有很多研究者把它引入图像标注问题的研究中[10],为了提高图像标注性能,本文根据标注词之间的关系构建合适的条件随机场模型。条件随机场可以用在很多不同的预测问题上。图像标注问题属于线性链条件随机场。本文条件随机场模型是一个无向图模型,图中的每一个点代表一个标注词,而两个点之间的边则代表两个标注词之间的关系。

条件随机场算法对标注词的校正除了涉及到标注词之间的共生关系之外,还将标注词的概率向量作为标注词的先验知识,然后建立标注词关系图并重新计算图像的标注词概率向量。该算法构建所有标注词的关系无向图,在该无向图中除了包含有边势函数(即式(9))之外还包含有点势函数(即式(10)),其中标注词概率向量确定图中点的势函数,而边的势函数则由学习训练集中标注词的关系所得到。例如标注词"马"出现了 k_1 次,标注词"草地"出现了 k_2 次,两者同时出现在同一幅图像的次数为 k_3 次。那么两个标注词的联合概率为

$$\boldsymbol{\Phi}(a_i, a_j) = -\log(P(a_i, a_j)) \tag{9}$$

$$\boldsymbol{\Theta}(a_i) = \frac{1}{1 + \mathrm{e}^{p(a_i)}} \tag{10}$$

$$P(a_1, a_2) = \frac{k_3}{\min(k_1, k_2)} \tag{11}$$

式中 $p(a_i)$是前面得到的图像被标注为 a_i的概率。

获得无向图中所有点势和边势之后,求取最优的图结构就能得到最终的图像标注集$\{a_{\text{focus}}, a_1, \cdots, a_{n-1}\}$。当图势函数值达到最小时,就得到了最优图结构,即式(12)中 M 的值最小的图结构:

$$M = \sum_{a_i \in S} \boldsymbol{\Psi}(a_i) + \lambda \sum_{(a_i, a_j) \in E} \boldsymbol{\Phi}(a_i, a_j) \tag{12}$$

式中:λ 表示点势函数和边势函数的权重关系,本文通过交叉验证的方法确定 $\lambda = 0.3$。

4　实验结果

4.1　图像库

为了验证本文算法的图像标注性能,使用 3 个图像库。第 1 个图像库是 Corel5K,该库被许多图像处理研究人员使用。它在许多文献中都有提及。Corel5k 数据集有 5 000 幅图像,其中包括 4 500 个训练样本和 500 测试样本。每一幅图像平均有 3.5 个关键词。在训练数据集中有 371 个标签,在测试数据集中有 263 个标签。另一个数据集是 IAPR TC-12。删除一部分图像后,留有 100 类的 10 000 幅图像。在实验过程中,使用 80% 幅图像用于训练,20% 幅图像用于测试。所使用的第 3 个数据集是 ESP Game。总共包含 21 844 幅图像。其中,19 659 张图像用作训练集,2 185 张图像用作测试集。

4.2　实验设置

为了验证图像标注性能,采取 3 种评估方法:召回率、查准率和 F-measure 值。假设一个给定的标签的图像数量是$|W_1|$,$|W_2|$为有正确标注词 w 的图像数量,$|W_3|$是由图像标注方法得到标签的图像的数量。召回率和查准率可计算如下:

$$\text{recall} = \frac{|W_2|}{|W_1|}, \text{precision} = \frac{|W_2|}{|W_3|}$$

平均查准率(AP)和查全率(AR)可以反映整体标注性能。F-measure 可以定义为

$$F\text{-measure} = \frac{2 \cdot \text{AP} \cdot \text{AP}}{\text{AR} + \text{AP}}$$

在本文实验中选择了 3 种视底层觉特征进行测试,它们分别为颜色直方图、纹理特征和 SIFT。这 3 种底层特征从不同的角度描述图像的底层信息,同时使用会使标注性能更好。然后在 3 个数据库 Corel5k、IAPR TC-12 和 ESP Game 上,将 VAMCRF 算法和其他著名算法进行比较。这些算法已表现出了良好的性能,并且取得了很好的标注结果。因此与它们的比较将能证明 VAMCRF 算法的性能。表 1 列出了这些算法和相应的标引。

表 1　实验中用到的算法

Table 1　Algorithms used in the comparison experiments

算法	描述	文献
2PKNN+ML	2-pass K-nearest neighbor+ML	[11]
CCD (SVRMKL+KPCA)	Canonical contextual distance	[12]
MBRM	Multiple-Bernoulli relevance	[13]
JEC	Joint equal contribution	[14]
TagProp+ML	Nearest neighbor models+ML	[15]
TagProp+ σ ML	Nearest neighbor models+ σ ML	[15]

4.3　实验结果和比较

4.3.1　参数影响

在视觉注意机制中有一个参数 ω。该参数对显著区域的提取有着重要的影响,需要通过实验来确定它的值。

首先从图像库中选取 100 幅有代表性图像,根

据经验人眼对亮度的敏感度比面积大一些,所以对 ω 取这样不同的一组值{0.30, 0.32, 0.34, 0.36, 0.38,0.40, 0.42, 0.44, 0.46, 0.48, 0.50}。通过实验发现当 ω = 0.42 时,提取图像显著区域效果最好。图 1 说明了当 ω = 0.42 时一些类的显著区域提取的实例。从表中可以看到,VAM 算法能够预测并很好地提取图像的显著区域。

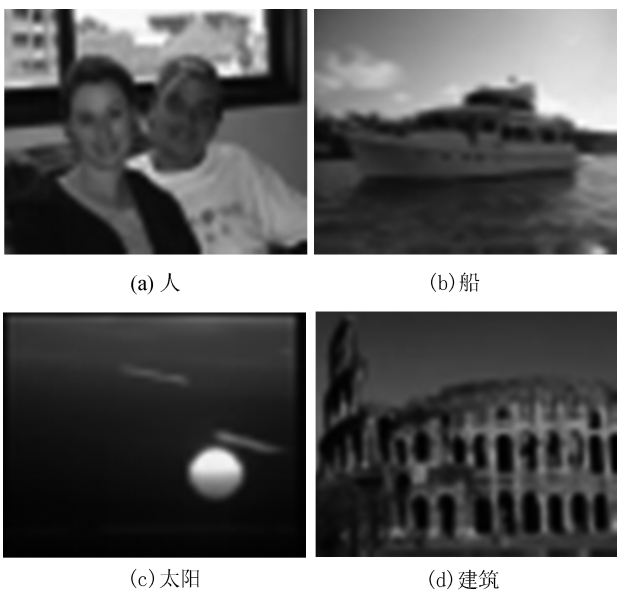


图 1 用 VAM 算法提取显著区域的例子

Fig.1 Some examples using our proposed VAM

在对图像的非显著区域进行标注时,采用了 k-NN 聚类算法。k-NN 聚类算法是最简单的机器学习算法之一,其中 k 值的选择对结果至关重要。实验测试了参数 k 取不同值时对标注结果的影响。图 2 展示的是用 k-NN 聚类算法在 3 个图像库上对非显著区域标注的性能。横坐标表示参数 k 取值的范围,纵坐标代表对应 k 值时 F_1 的变化。可以看到,当 k=100 时 F_1 达到最大值,也就是此时标注效果最好。所以,在下面的实验当中 k 取 100。

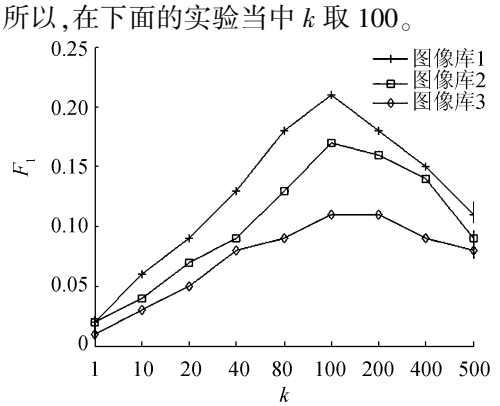
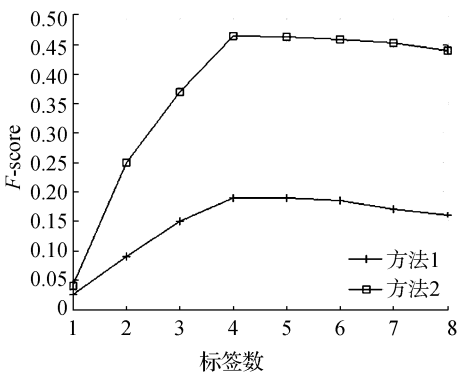


图 2 在 3 个图像库上 k 取不同值的标注结果

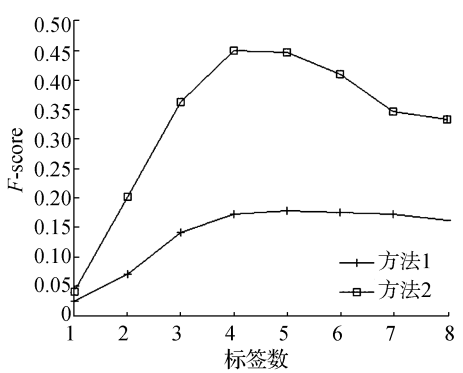
Fig.2 The results of k-NN with different k in the image datasets 1~3

4.3.2 标签数目对标注的影响

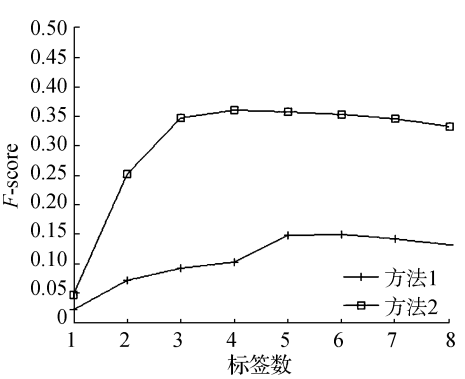
标注性能的好坏有很多影响因素,标签数目就是其中一种因素,为了验证标签数目对标注性能的影响,选取了不同的标签数进行实验。图 3 分别显示了在 3 个数据库上不同的标签数目对标注的影响。横坐标表示所取标签的个数从 1~8,纵坐标代表对应标签数时 F_1 的变化。这是在两种方法下所做的实验,方法 1 是使用视觉注意机制和 SVM 求得显著区域的标注词,然后利用 k-NN 求得非显著区域的标注词;方法 2 是在方法 1 的基础上利用条件随机场对所获得的标注词进行校正。



(a) Corel5k 数据库



(b) IAPR TC-12 数据库



(c) ESP Games 数据库

图 3 不同标签数对标注的影响

Fig.3 The effect of different tag numbers on annotation

从图 3 可以看出,当只给图像一个标签时,

标注结果够不好,随着使用标签数目的增加,标注的准确度都在增加。但是使用的标签数不易过多,如果过多反而会使标注准确度下降。方法 2 比方法 1 的效果更好些,说明标注词之间的共生关系对标注效果也是十分重要的。标签相关性的引入使得标注结果更符合实际的标签集,由此证明了本文算法的优势。

4.3.3　比较和结果分析

为了验证本文所提出算法的标注性能,在 Corel5k、IAPR TC-12 和 ESP Game 3 个图像库中的测试图像集进行了实验,并对 AR、AP 和 F_1 的值进行对比。在表 2~4 给出了比较结果。

表 2 显示了 VAMCRF 算法在 Corel5k 上得到的 AR、AP 和 F_1 的值。从表中数据可见,VAMCRF 算法取得了最高 AR 值 0.48,AP 最高值为 0.45,F_1 最大值为 0.464。与其他 6 种算法 F_1 最高值 0.439 比较,VAMCRF 的最大值 0.464 至少高出了 0.014。

表 2　在 Corel5k 数据库上和其他算法标注性能的比较

Table 2　Performance comparison with other algorithms on the Corel5k

算法	AR	AP	F_1
2PKNN+ML	0.46	0.44	0.450
CCD(SVRMKL+KPCA)	0.41	0.36	0.383
MBRM	0.25	0.24	0.245
JEC	0.32	0.27	0.293
TagProp+ML	0.37	0.31	0.337
TagProp+ σ ML	0.42	0.33	0.370
VAMCRF	0.48	0.45	0.464

表 3 显示了 VAMCRF 算法在 IAPR TC-12 上得到的 AR、AP 和 F_1 的值。从表中数据可见,2PKNN+ML 算法和 VAMCRF 算法取得了最高 AR 值 0.37,AP 最高值 0.56,F_1 最大值 0.445。与其他 6 种算法 F_1 最高值 0.450 比较,VAMCRF 的最大值 0.445 至少高出了 0.006。

表 4 显示了 VAMCRF 算法在 ESP Game 上得到的 AR、AP 和 F_1 的值。从表中数据可见,VAMCRF 算法取得了最高 AR 值 0.28,2PKNN+ML 最高 AP 值 0.53,F_1 最大值 0.358。与其他 6 种算法 F_1 最高值 0.357 比较,VAMCRF 的最大值 0.358 至少高出了 0.001。

表 4　在 ESP Game 数据库上和其他算法标注性能的比较

Table 4　Performance comparison with other algorithms on the ESP Game

算法	AR	AP	F_1
2PKNN+ML	0.27	0.53	0.357
CCD(SVRMKL+KPCA)	0.24	0.36	0.284
MBRM	0.19	0.18	0.185
JEC	0.25	0.22	0.234
TagProp+ML	0.20	0.49	0.284
TagProp+ σ ML	0.27	0.39	0.319
VAMCRF	0.28	0.50	0.358

表 3　在 IAPR TC-12 数据库上与其他算法标注性能的比较

Table 3　Performance comparison with other algorithms on the IAPR TC-12

算法	AR	AP	F_1
2PKNN+ML	0.37	0.54	0.439
CCD(SVRMKL+KPCA)	0.29	0.44	0.350
MBRM	0.23	0.24	0.235
JEC	0.29	0.28	0.285
TagProp+ML	0.25	0.48	0.329
TagProp+ σ ML	0.35	0.46	0.398
VAMCRF	0.37	0.56	0.445

5　结论

本文提出了一种基于视觉注意机制和条件随机场的算法进行图像的标注,并在 Corel5k, IAPR TC-12 和 ESP Game 图像库上进行实验。首先,用视觉注意机制提取图像的显著区域,然后利用 SVM 进行标注,之后使用 k-NN 聚类算法对图像的非显著区域进行标注,最后利用条件随机场对图像的标注词向量进行校正。实验结果表明,与传统方法相比,本文所提出的算法在标注性能上取得了很好的效果,但是从时间复杂度方面来看还需要很多的改进工作,在未来的研究中可以对算法进行进一步改进以期降低时间复杂度。

参考文献:

[1] WANG Meng, NI Bingbing, HUA Xiansheng, et al. Assistive tagging: a survey of multimedia tagging with human-computer joint exploration[J]. ACM computing surveys, 2012, 44(4): 25.

[2] JIN Cong, JIN Shuwei. Image distance metric learning based on neighborhood sets for automatic image annotation[J].

Journal of visual communication and image representation, 2016, 34: 167-175.

[3] 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]//Proceedings of the 7th European Conference on Computer Vision. Berlin Heidelberg: Springer-Verlag, 2002: 97-112.

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

[5] LOOG M. Semi-supervised linear discriminant analysis through moment-constraint parameter estimation[J]. Pattern recognition letters, 2014, 37: 24-31.

[6] FU Hong, CHI Zheru, FENG Dagan. Recognition of attentive objects with a concept association network for image annotation[J]. Pattern recognition, 2010, 43(10): 3539-3547.

[7] FAREED M M S, AHMED G, CHUN Qi. Salient region detection through sparse reconstruction and graph-based ranking[J]. Journal of visual communication and image representation, 2015, 32: 144-155.

[8] JIA Cong, QI Jinqing, LI Xiaohui, et al. Saliency detection via a unified generative and discriminative model[J]. Neurocomputing, 2016, 173: 406-417.

[9] KHANDOKER A H, PALANISWAMI M, KARMAKAR C K. Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings[J]. IEEE transactions on information technology in biomedicine, 2009, 13(1): 37-48.

[10] PRUTEANU-MALINICI I, MAJOROS W H, OHLER U. Automated annotation of gene expression image sequences via non-parametric factor analysis and conditional random fields[J]. Bioinformatics, 2013, 29(13): i27-i35.

[11] VERMA Y, JAWAHAR C V. Image annotation using metric learning in semantic neighbourhoods[C]//Proceedings of the 12th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012: 836-849.

[12] NAKAYAMA H. Linear distance metric learning for large-scale generic image recognition[D]. Tokyo, Japan: The University of Tokyo, 2011.

[13] FENG S L, MANMATHA R, LAVRENKO V. Multiple Bernoulli relevance models for image and video annotation [C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE, 2004, 2: II-1002-II-1009.

[14] MAKADIA A, PAVLOVIC V, KUMAR S. A new baseline for image annotation[C]//Proceedings of the European Conference on Computer Vision. Berlin Heidelberg: Springer-Verlag, 2008: 316-329.

[15] GUILLAUMIN M, MENSINK T, VERBEEK J, et al. TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation[C]//Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 309-316.

作者简介:

孙庆美 ,女,1989 年生,硕士研究生,主要研究方向为数字图像处理

金聪,女,1960 年生,教授,博士。主要研究方向为数字图像处理