

# 词间相关性的 CMRM 图像标注方法

刘咏梅,代丽洁

(哈尔滨工程大学 计算机科学与技术学院,黑龙江 哈尔滨 150001)

**摘要:**自动图像标注因其对图像理解和网络图像检索的重要意义,近年来已成为新的热点研究课题.在图像标注的 CMRM 模型基础上,提出了一种基于词间相关性的 CMRM 标注方法.该方法提取了标注字之间的词间相关关系,并利用图学习算法,通过将词间相关性矩阵叠加到初始标注矩阵的方法对标注结果进行了改善.利用 Corel5k 标注图像库中的自然场景图像进行实验.实验结果表明,该方法很好地完成了对测试集图像的自动标注,在查全率与查准率上较 CMRM 模型有所提高.

**关键词:**图像标注;CMRM 模型;相关关系;词间相关性矩阵

**中图分类号:**TP391 **文献标识码:**A **文章编号:**1673-4785(2011)04-0350-05

## A method of CMRM image annotation based on inter-word correlation

LIU Yongmei, DAI Lijie

(School of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

**Abstract:** Automatic image annotation is significant for image understanding and retrieval of web images. As a result, it has become a hot research topic in recent years. On the basis of a CMRM image annotation model, an efficient image annotation method was proposed based on inter-word correlations. The correlations between the annotated words were picked up, and the inter-word correlation matrix was added to the initial labeling matrix by a graph learning algorithm. The proposed annotation approach was tested by a Corel5k database with natural scene images. The experimental result shows that the method can automatically label the images in the test set very well. The recall and precision are increased compared to CMRM.

**Keywords:** image annotation; CMRM; correlation; word-correlation matrix

有效的图像描述方法<sup>[1]</sup>是图像数据管理的基础.目前图像描述方法主要采用人工标注的方式进行,但是由于人工方式耗时费力而且缺乏客观性,因此需要对图像进行自动标注.

此外,对图像进行自动的语义标注是图像检索中重要且非常具有挑战性的工作<sup>[2]</sup>.对于提高检索系统的效率有着重要的意义.因此如何快速有效地进行自动的图像标注就变得异常重要.

通过对 CMRM 算法的研究发现,该模型只考虑了视觉特征与标注关键字之间的对应关系,却忽略了关键字本身所具有的相关性,因此,提出了一种基于词间相关性的 CMRM 图像标注方法.首先利用

CMRM 模型对图像进行标注,然后从训练集中提出关键字之间的相关关系,并利用图学习算法,将这种相关关系在各个关键字之间进行传播,从而得到新的图像标注关键字.

### 1 CMRM 图像标注方法

联合媒体相关模型(cross-media relevance models, CMRM)将相关语言模型(relevance-based language model)<sup>[3-4]</sup>应用到图像标注中,通过对训练图像集的学习获取图像的视觉词元与关键字的联合概率分布.利用概率统计方法,获得每个关键字作为图像标注的概率,并以此作为选择标注的依据.该模型在图像标注领域取得了较好的成绩,现在已经成为各种新图像标注方法进行性能对比的标准方法.

在利用 CMRM 算法对图像进行标注过程中常

用的图像描述方法如下:

1) 关键字记为  $w_i, i=1, 2, \dots, n$ .

2) 视觉词元记为  $b_i, i=1, 2, \dots, m$ . 并用它对图像进行表示.

3) 训练图像记为  $J_i = \{b_1, b_2, \dots, b_m, w_1, w_2, \dots, w_n\}$ , 其中  $i$  为图像的编号.

4) 测试图像记为  $I = \{b_1, b_2, \dots, b_m\}$ .

### 1.1 CMRM 标注的基本原理

CMRM 模型采用一种生成式的语言建模方法<sup>[4-5]</sup>, 该方法认为每幅图像的视觉特征与标注关键字之间都有一种潜在概率分布  $P(\cdot | I)$ , 并且将这个分布看作是  $I$  的相关模型<sup>[4-6]</sup>. 可以将图像的视觉词元表示  $\{b_1, b_2, \dots, b_m\}$  看作是从  $P(\cdot | I)$  中进行  $m$  次随机采样得到的. 同理, 很自然地想到对图像  $I$  进行标注可以看作是从它的相关模型  $P(\cdot | I)$  中随机抽样  $n$  次得到了  $n$  个关键字. 为了能对相关模型进行抽样需要对标注集中每个关键字都估计概率.

由于  $P(\cdot | I)$  本身是未知的, 因此可以考虑利用条件概率  $P(w | b_1, b_2, \dots, b_m)$  近似  $P(w | I)$ <sup>[6]</sup>, 即

$$p(w | I) \approx P(w | b_1, b_2, \dots, b_m). \quad (1)$$

对于式(1)等号右边的概率, 可以首先利用已标注的训练图像集来估计在一幅图像中同时观察到关键字  $w$  和视觉词元  $b_1, b_2, \dots, b_m$  的联合分布, 然后对该分布按  $w$  进行边缘化得到. 而联合分布可以利用训练集中的图像  $J$  的期望得到

$$P(w, b_1, b_2, \dots, b_m) = \sum_{J \in T} P(J) P(w, b_1, b_2, \dots, b_m | J). \quad (2)$$

假设某一幅训练图像  $J$  被选定后, 关键字  $w$  和视觉词  $b_1, b_2, \dots, b_m$  是否出现是相互独立的<sup>[4-6]</sup>. 因此可以将式(2)重写为

$$P(w, b_1, \dots, b_m) = \sum_{J \in T} P(J) P(w | J) \prod_{i=1}^m P(b_i | J). \quad (3)$$

式中:  $P(J)$  可以对训练集中的所有图像保持一致. 由于  $J$  中同时包含了全部视觉词元与关键字, 因此可以利用平滑的期望最大化算法对式(3)中的概率进行估计, 当给定一幅训练图像  $J$  时, 可以同时观察到关键字  $w$  和视觉词元  $b$  的概率由式(4)、(5)给出.

$$P(w | J) = (1 - \alpha_J) \frac{\#(w, J)}{|J|} + \alpha_J \frac{\#(w, T)}{|T|}, \quad (4)$$

$$P(b | J) = (1 - \beta_J) \frac{\#(b, J)}{|J|} + \beta_J \frac{\#(b, T)}{|T|}. \quad (5)$$

式中:  $\#(w, J)$  表示  $w$  出现在图像  $J$  的标注中的次数, 因为同一个词很少会多次出现在同一个标题中,

所以该值一般取为 0 或 1;  $\#(w, T)$  表示  $w$  出现在训练集  $T$  中的总次数, 它等于用该关键字标注的训练图像的总数; 同理,  $\#(b, J)$  和  $\#(b, T)$  也表示相应的次数. 利用式(1)~(5)<sup>[4-7]</sup> 估计出概率  $P(w | I)$  后, 就可以直接利用这个概率对图像进行标注. 本文中将  $\alpha_J$  和  $\beta_J$  这 2 个参数经评估后分别取为  $\alpha_J = 0.1$ ,  $\beta_J = 0.9$ .

### 1.2 CMRM 的图像标注算法

利用联合媒体相关模型对图像进行标注的主要步骤总结如下:

1) 从数据集的文件中统计出的基本信息.

$\#(w, J)$ : 每个关键字出现在每幅图像的标注中的次数, 一般为 0 或 1.

$\#(w, T)$ : 关键字出现在数据集总次数.

$\#(b, J)$ : 视觉词元  $b$  出现在每幅图像中的次数.

$\#(b, T)$ : 视觉词元  $b$  出现在全部训练图像中的次数.

2) 利用统计信息计算关键字和视觉词元的条件概率  $P(w | J)$  和  $P(b | J)$ .

3) 利用条件概率对关键字和视觉词元的联合概率进行估计.

4) 利用  $p(w, b_1, b_2, \dots, b_m)$  来近似  $p(w | b_1, b_2, \dots, b_m)$ .

## 2 基于词间相关性的 CMRM 图像标注

CMRM 算法是根据图像的底层特征等对图像进行标注的, 由于“语义鸿沟”的存在, 使得该算法的标注性能受到影响. 在利用 CMRM 对图像进行初始的标注之后, 利用图学习方法, 将从训练集中得到的词间相关关系矩阵作用到初始标注矩阵上, 使得关键字之间的相关性在图的各个顶点之间进行传播, 从而实现了标注结果的改善.

### 2.1 词间相关性描述

词间相关性是指 2 个关键字同时标注一幅图像的特性. 相关程度用词间相关性矩阵表示. 为了获得关键字之间的相关性矩阵, 可以对训练集中关键字出现的次数进行统计, 文中将 2 个关键字作为同一幅图像的标注出现的次数称为它们的共生次数.

由于关键字的共生次数不能全面地反映词间的相关性, 因此可以采用文本检索领域中的 TF-IDF (term frequency-inverse document frequency) 思想对关键字的相关性进行度量, 得到词间相关性矩阵. 其定义为

$$K_{wc}(w_1, w_2) = K(w_1, w_2) \times \log\left(\frac{N_T}{n_1}\right). \quad (6)$$

式中:  $K(w_1, w_2)$  代表关键字  $w_1, w_2$  在训练集中共同作为一幅图像的标注而出现的次数, 可以将它作为一个矩阵存储记为  $K$ , 不难看出,  $K$  是一个对称矩阵;  $N_T$  为训练集的大小, 它代表了训练图像集中所包含的全部图像的数目;  $n_1$  则表示标注中出现了关键字  $w_1$  的全部图像的数量, 它就决定了  $K_{wc}(w_1, w_2)$  与  $K_{wc}(w_2, w_1)$  并不相等, 因为常见的关键字, 在图像集的标注中出现的次数相对于一般的关键字要多, 也就是它的  $n_1$  值大, 相应的  $\log(\frac{N_T}{n_1})$  就小。

要准确地获取关键字之间的词间相关性矩阵, 需要对训练集数据进行统计和整理, 并进行相应的计算。下面将详细说明计算该矩阵的算法。算法的主要步骤如下:

- 1) 读入训练图像集中的标注信息, 即训练图像本身自带的标注关键字;
- 2) 从训练集的标注信息中, 统计出任意 2 个关键字的共生次数, 即  $K(w_1, w_2)$ ;
- 3) 统计每个关键字在训练图像集中作为标注出现的次数  $n_1$ ;
- 4) 按照式(6)计算出词间相关性矩阵;
- 5) 对矩阵进行归一化处理。

## 2.2 相关关系传播算法

本文利用图学习算法来实现词间相关性在各个关键字之间的传播。文献[8]将图学习算法应用于图像标注中, 所提出的基于图学习的标注框架首先进行初始标注, 即以图像为节点, 以图间相似性为边建立图, 通过图学习算法将标注信息从已标注图像传递到未标注图像, 然后对标注结果进行改善。这时, 利用词间相关性建立以词为节点的图, 以初始标注结果设置初始状态向量, 通过学习算法得到图像的最终标注结果。本文方法借鉴了这一框架。

设  $G = (V, E)$  表示图, 其中  $V = \{x_1, x_2, \dots, x_n\}$  是图的顶点, 对应标注的关键字;  $E$  代表图的边, 边上的权值对应它所连接的 2 个关键字之间的相关强度。同时, 设 CMRM 标注结果用矩阵  $Y$  表示, 它的元素  $y_{ij}$  表示第  $i$  幅图像被标注为第  $j$  个关键字的概率。结果矩阵  $R$  的元素表示每幅图像被标注为每个关键字的最终概率结果, 由下面的迭代过程产生<sup>[8]</sup>:

$$R(t+1) = \beta \cdot S \cdot R(t) + (1 - \beta)Y, \quad (7)$$

$$R^* = (1 - \beta)(1 - \beta S)^{-1}Y. \quad (8)$$

式中:  $t$  表示迭代次数, 其初值为 0,  $R(0) = Y\beta$  是传播系数, 它决定了相关矩阵对标注结果的影响程度, 本文中  $t$  的终值取 500,  $\beta$  取 0.25。

从上面的描述可以看出, 迭代过程中有 2 个最

重要的矩阵, 相似矩阵  $S$  和初始标注矩阵  $Y$ , 前者对应着上一节中得到的共生矩阵, 而后者由 CMRM 算法对测试图像进行标注而得到。

## 2.3 基于词间相关性标注算法

下面给出算法的具体步骤如下。

- 1) 从数据集的文件中读入训练数据信息, 进而完成对基本信息  $\#(w, J)$ 、 $\#(w, T)$ 、 $\#(b, J)$ 、 $\#(b, T)$  的统计, 并转 2)。
- 2) 估计关键字  $w$  和视觉词元  $b$  的条件概率, 并计算关键字与视觉词元之间的联合概率。
- 3) 提取词间相关性矩阵  $S$ 。
- 4) 将 2) 中得到标注矩阵与式(7)中的初始状态矩阵  $Y$  对应, 进行迭代计算, 直到收敛, 得到的新矩阵就是最终标注结果。
- 5) 对矩阵每行的元素按概率值排序, 取概率最大的  $N$  个作为最终的标注关键字, 在本文方法中  $N=5$ 。

## 3 实验结果与分析

### 3.1 实验数据

本文采用带有标注字的 Corel5k 图像集<sup>[9]</sup> 实验。整个图像集中共包含 50 个文件夹, 每个文件夹 100 张图像, 共 5 000 张图像。从每个文件夹中选取 80 张图像作为训练图像, 再先取 10 张作为评估图像集, 用于对参数进行评估, 其余 10 张图像作为测试图像<sup>[4]</sup>。

对所有图像均先利用 NCut 算法进行图像分割, 取前 10 个面积较大的区域作为有效区域, 进行特征提取。区域的视觉特征为 36 维特征向量, 包括 18 种颜色特征、12 种纹理特征和 6 种形状特征。对训练图像, 利用  $K$ -均值聚类算法 ( $K=500$ ) 对所有有效区域进行聚类, 由聚类结果得到 500 个视觉词元。

### 3.2 实验结果

对 500 幅测试图像进行了自动的标注, 计算了每个关键字(标注字)的平均查全率和平均查准率。查全率(recall)度量出对单个词查询的完整性, 查准率(precision)反映出查询的精度。对于给定的标注字  $w$ , 若在测试图像集的手工标注结果中包含  $w$  的图像个数为  $N_m$ , 使用自动标注模型的标注结果中包含该词的图像个数为  $N_a$ , 其中参照手工标注结果有  $N_r$  个是正确的, 则单个标注字的查全率和查准率为

$$\text{recall} = \frac{N_r}{N_m},$$

$$\text{precision} = \frac{N_r}{N_a}.$$

采用标注字的平均查准率和平均查全率来考察

本文方法的标注性能. 与 CMRM 算法相比, 部分关键字的查全率与查准率有明显的提高, 如图 1 及图 2 所示.

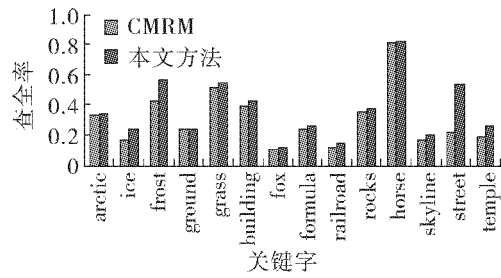


图1 本文方法与 CMRM 模型的部分关键字的查全率对比

Fig.1 Recall contrast results of partial words with CMRM model

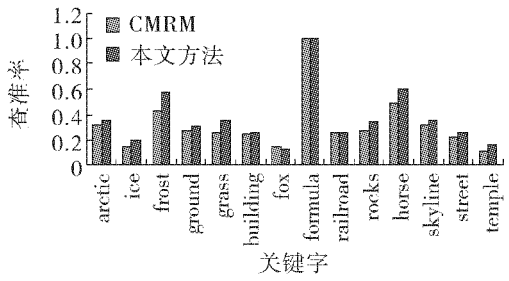


图2 本文方法与 CMRM 模型的部分关键字的查准率对比

Fig.2 Precision contrast results of partial words with CMRM model

对测试图像集中 500 幅图像的标注结果进行统计, 对标注性能最好的前 49 个标注字的平均查全率为 0.375, 平均查准率为 0.38. 本文方法与 CMRM 模型的标注性能进行了对比, 见表 1.

表1 与 CMRM 模型的性能对比

Table 1 Efficiency contrast results with CMRM model

方法名称	平均查全率	平均查准率
CMRM 模型	0.370	0.33
本文方法	0.375	0.38

本文方法充分考虑了标注关键字之间的相关性, 这种相关性反映了各个关键字之间的语义联系, 通过相关性将 2 个紧密联系的关键字的标注概率进行调整, 并对 CMRM 的标注结果进行调整, 使得标注结果在查准率和查全率上较 CMRM 有所提高. 表 2 给出了部分图像的标注结果.

本方法无论在查全率还是查准率方面, 都较 CMRM 模型有所提高. 但是由于图像的视觉词元对最终标注结果的影响是非常大的, 不可避免地出现了一些标注效果不理想的情况. 标注性能受以下几方面因素影响: 1) 颜色、形状特征在全部特征中所

占的比重, 各种特征直接影响聚类效果; 2) 对于词间关系的提取不够全面; 3) 所有图像的标注长度相同, 影响了最终的查全率与查准率.

表2 标注结果举例

Table 2 Examples of annotation results

测试图像	CMRM 标注结果	本文方法标注结果	人工标注结果
(a)	sun water cloud tree lake	sun sky cloud water tree	cloud sea sun tree
(b)	grass gores cars bear water	grass cars tracks bear people	grass cars tracks
(c)	sky plane jet clouds water	sky clouds plane jet eagle	jet plane sky

## 4 结 论

本文在 CMRM 模型基础上, 利用标注字之间的词间相关性对 CMRM 标注方法进行了改进, 取得了更好的标注效果. 下一步的研究工作可以从以下几个方面进行.

1) 优化图像的特征提取方法, 在不影响算法性能的前提下, 尽量全面地采用各种特征, 以消除某些特征对聚类的影响.

2) 对于共生关系矩阵的提取, 可以借助一些结构化的词典来进行, 普林斯顿大学开发的 WorldNet 就是一个不错的选择, 该词典收录了大量的单词, 且按语义对单词进行组织, 能够更加全面地衡量关键字之间的相关性.

3) 可以对图像进行长度不固定的标注, 例如对图像中每个区域进行单独的标注, 然后对于标注相同的区域进行融合, 这样可以更好地消除噪声关键字对查全率和查准率的影响.

## 参考文献:

[1] 陈世亮, 李战怀, 袁柳. 一种基于区域特征关联的图像语义标注方法[J]. 计算机工程与应用, 2007, 43(2): 53-57.  
CHEN Shiliang, LI Zhanhuai, YUAN Liu. Image semantic annotating method based on region features relevancy[J]. Computer Engineering and Applications, 2007, 43(2): 53-57.  
[2] YU Feiyang, HORACE H S. Semantic content analysis and annotation of histological images[J]. Computers in Biology

- and Medicine, 2008, 38: 635-649.
- [3] LAVRENKO V, CROFT W. Relevance-based language models[C]//Proceedings of the 24th Annual International ACM SIGIR Conference. New Orleans, USA, 2001: 120-127.
- [4] LAVRENKO V, CHOQUETTE M, CROFT W. Cross-lingual relevance models[C]//Proceedings of the 25th Annual International ACM SIGIR Conference. Tampere, Finland, 2002: 175-182.
- [5] 王斌. 图像检索中自动标注与快速相似搜索技术研究[D]. 长沙: 中国科学技术大学, 2007: 5-45.
- WANG Bin. Researches on automatic image annotation and fast similarity searching technology in CBIR[D]. Changsha: University of Science and Technology of China, 2007: 5-45.
- [6] 袁曾任. 人工神经网络及其应用[M]. 北京: 清华大学出版社, 1999: 40-50.
- [7] JEON J, LAVRENKO V, MANMATHA R. Automatic image annotation and retrieval using cross-media relevance models[C]//Proceedings of the 26th International ACM SIGIR Conference. Toronto, Canada, 2003: 119-126.
- [8] 卢汉清, 刘静. 基于图学习的自动图像标注[J]. 计算机学报, 2008, 31(9): 1629-1639.
- LU Hanqing, LIU Jing. Image annotation based on graph learning[J]. Chinese Journal of Computers, 2008, 31(9): 1629-1639.
- [9] [http://www.cs.arizona.edu/people/kobus/research/data/eccv\\_2002](http://www.cs.arizona.edu/people/kobus/research/data/eccv_2002)
- [10] 纪颖. 一种基于实例的图像自动语义标注方法[J]. 哈尔滨理工大学学报, 2009, 14(1): 38-42.
- Ji Ying. An instance-based method for automatic image semantics annotation[J]. Journal of Harbin University of Science and Technology, 2009, 14(1): 38-42.

#### 作者简介:



刘咏梅, 女, 1973 年生, 副教授, 硕士生导师, 博士. 主要研究方向为模式识别、图像理解和生物信息学.



代丽洁, 女, 1983 年生, 硕士, 主要研究方向为图像处理与模式识别、图像标注.

## 第九届全球智能控制与自动化大会(WCICA 2012)

### The 9th World Congress on Intelligent Control and Automation

The 9th World Congress on Intelligent Control and Automation (WCICA 2012) will be held in Beijing, China, from July 6 to 8, 2012.

WCICA 2012 is technically sponsored by Academy of Mathematics and Systems Science, Chinese Academy of Sciences, IEEE Robotics and Automation Society, IEEE Control System Society, National Natural Science Foundation of China, the Chinese Association of Automation, the Chinese Association of Artificial Intelligence, IEEE Hong Kong Robotics and Control Chapter, IEEE Beijing Control System Chapter, and IEEE Singapore Control System Chapter.

WCICA 2012 features plenary keynotes and plenary panel discussion sessions by world renowned researchers as well as awards to honor outstanding papers presented at this Congress. The awards include Best Paper on Theory, Best Paper on Applications, Best Student Paper, Best Paper on Biomedical & Biosystem Related Areas, Best Poster Paper, SUPCON Best Paper on Industrial Automation, and AIAG Best Paper on Supply Chain Related Topics Awards.

All accepted papers in the Proceedings of WCICA 2012 will be published by IEEE and indexed by EI.

**Web site:** <http://wcica12.amss.ac.cn>