

基于互 K 近邻图的自动图像标注与快速求解算法

郭玉堂

(合肥师范学院计算机科学与技术系 合肥 230061)

摘要 图像语义具有模糊性、复杂性、抽象性等特点,在提取图像语义时仅用低层特征进行描述是不够的,需要结合图像相关内容,以便提高图像标注的精确度。为此,提出了基于互 K 近邻图的图像标注方法,该方法用一个互 K 近邻图融合了图像的低层特征之间、标注词之间以及图像与标注词间的相互关系。利用互 K 近邻图实现了根据两个节点间的相互关系来提取语义信息,弥补了基于 K 近邻图的方法中单方向挖掘节点信息的不足,有效地提高了图像标注的性能。在对互 K 近邻图结构分析的基础上,结合重启随机游走,提出了一种快速求解算法,该算法在不明显降低图像标注精度下,实现了快速求解。在 Corel 图像数据集上进行了实验,结果验证了所提方法的有效性。

关键词 图像标注,互 K 近邻图,重启随机游走,快速求解

中图分类号 TP391.41 **文献标识码** A

Automatic Image Annotation Method and Fast Solution Based on the Mutual K Nearest Neighbor Graph

GUO Yu-tang

(Department of Computer Science and Technology, Hefei Normal College, Hefei 230061, China)

Abstract Image semantics has the characters of vague, complex, and abstractive, therefore only low-level features are not enough for describing image semantics, and require a combination of image-related content in order to improve the accuracy of the image annotation. In this paper, an image annotation method based on mutual K nearest neighbor graph (MKNNNG) was proposed, which builds the relationship between the low-level features, annotation words and images by a mutual K nearest neighbor graph. Mutual K nearest neighbor graph is to extract semantic information from paired-nodes, which overcomes the limitation of unilateral mining of the traditional K nearest neighbor graph, and effectively improves the image annotation performance. Based on the analysis on the structure of mutual K nearest neighbor graph, Combined with Random Walk with Restart (RWR), a fast algorithm was proposed without apparent reducing the precision of the image annotation. Experimental results on the Corel image datasets show the effectiveness of the proposed approach in terms of quality of the image annotation.

Keywords Image annotation, Mutual K adjacency graph, Random walk with restart, Fast solution

自动图像标注是基于内容图像检索中重要且具有挑战性的工作,它可以在一定程度上解决基于内容图像检索中存在的语义鸿沟问题。如果能实现自动图像标注,那么现有的图像检索问题就可以转化成技术已相当成熟的文本检索问题。自动图像标注已经引起了人们的关注,国内外已经有许多学者提出了各种不同的方法,并取得了相当多的成果。如 Duygulu 等人提出的翻译模型^[1],其利用传统语言统计翻译模型将语义概念翻译为由图像区域聚类产生的 blobs。Jeon 等人介绍了一种交叉媒体相关模型^[2],其将图像标注问题看作跨语言检索问题,Carneiro 和 Vasconcelos 等采用基于最小错误率的优化准则和统计分类的思想,提出了一种监督多标注方法(Supervised multiclass Labeling, SML)^[3,4]。国内学者在自动图像标注方面也进行了相关的研究。文献^[5]采用基于支持向量机(SVM)的多分类器为空间映射方法,将图像的低层特征映射为具有一定高层语义的模型特征,从而实现

基于概念索引的图像标注方法。

把图理论应用到自动图像标注与检索中是当今研究的热点。在文献^[6]中, Pan 等人提出了一种基于 Gcap 自动图像标注的方法,该方法对训练集中的各图像运用图像分割算法分割成一系列区域,用一个关联图表示图像与区域、图像与已标注词、区域与区域的关系。待标注的图像同样也分割成若干区域,并附加到训练图中,然后以待标注的图像节点为起点,运行随机游走算法估算待标注的图像的标注词的稳态概率。

Gcap 为自动图像标注提出了一个基于图的学习方法。但是它有一定局限性。在 Gcap 中,仅表达了图像区域的视觉特征相似性。

图像语义具有模糊性、复杂性、抽象性等特点,仅用低层特征进行描述是不够的,在提取图像语义时需要结合图像相关内容来提高图像标注的精确度。为此,本文提出了基于互

到稿日期:2010-03-19 返修日期:2010-08-08 本文受安徽省自然科学基金项目(11040606M134),安徽省高校自然科学基金重点项目(KJ2009A150)资助。

郭玉堂(1962-),男,博士,教授,主要研究方向为模式识别与图像处理,E-mail: aieyt@ah.edu.cn.

K近邻图的图像标注方法,该方法用一个互K近邻图融合了图像的低层特征之间、标注词之间以及图像与标注词间的相互关系。利用互K近邻图实现了根据两个节点间的相互关系来提取语义信息,弥补了基于K近邻图的方法中单方向挖掘节点信息的不足。并引入逆向文档频率(IDF)^[7]修正图像节点与其标注词节点之间边的权值,克服了传统方法中因高频词引起的偏差,有效地提高了图像标注的性能。在对互K近邻图结构分析的基础上,提出了一种快速求解算法,该算法在不明显降低图像标注精度下,实现了快速求解。

1 基于互K近邻图的图像标注

1.1 互K近邻图的构建

基于区域的自动图像标注,首先将训练集图像和测试图像(待标的图像)分割成若干区域,每个区域提取其形状、空间位置、颜色和纹理等视觉特征,构成区域的特征矢量 f_i 。设 I_q 表示待标注的图像, $T=\{i_1, i_2, \dots, i_n\}$ 表示训练图像,每一个训练图像 $i \in T$ 被分割为若干个区域,每个区域用其视觉特征矢量 f_i 表示。标注词列表记作 $w=\{w_1, w_2, \dots, w_l\}$ 。其中 $w_i(i=1, 2, \dots, l)$ 是标注词表中第 i 个标注词。我们用一个图 $G=\langle V, E \rangle$ 表示图像、区域和标注词三者间的邻接关系。每个图像与它的区域用一加权边相连,其权值取1。每个图像与它的标注词用一加权边相连,边的权值记为 $\text{sim}(i, w_j)$,每个区域根据相似测度与它的 K 个互近邻的区域用边相连,其权值记为 $\text{sim}(f_i, f_j)$ 。同理,每标注词根据相似测度与它有 K 个互近邻的标注词用边相连,其权值记为 $\text{sim}(w_i, w_j)$,如图1所示。

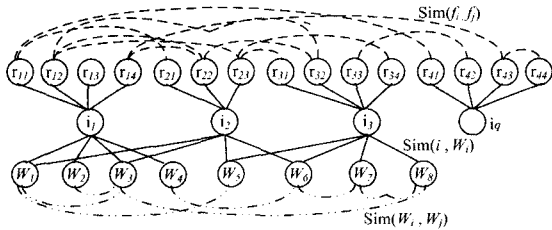


图1 MKNNG示意图

1.2 相似测度的计算

1.2.1 区域相似测度计算

在标注图像前,先采用图像分割方法对已标注图像和待标注图像进行分割,抽取分割后的每个区域的低层特征,包括形状、空间位置、颜色和纹理等,形成24维特征矢量,用 f 表示。

区域间的边权值 $\text{sim}(f_i, f_j)$ 通过式(1)计算:

$$\text{sim}(f_i, f_j) = \begin{cases} \exp(-\|f_i - f_j\|^2 / \sigma^2), & \text{如果 } f_i, f_j \text{ 是 } K\text{-互邻居} \\ 0, & \text{否则} \end{cases} \quad (1)$$

1.2.2 标注词相似测度

标注词之间的语义相似度量采用了基于数据集的概率统计的Co-occurrence方法^[8]。

基于数据集的概率统计的共现相似测度用 $\text{sim}(w_i, w_j)$ 表示,则:

$$\text{sim}(w_i, w_j) = \frac{N(w_i, w_j)}{\min(N(w_i), N(w_j))} \quad (2)$$

式中, $N(w_i, w_j)$ 表示标注词 w_i 和标注词 w_j 共同出现的次

数, $N(w_i)$ 和 $N(w_j)$ 分别表示标注词 w_i 和标注词 w_j 在训练集中出现的次数。

根据标注词相似测度 $\text{sim}(w_i, w_j)$,两标注词节点间有边连接,当且仅当这两节点是K-互邻居。

1.2.3 图像与标注词间关联度计算

在训练集词汇表中,各词的出现频率相差很大。如在Corel图像库中,词Water, Sky, tree出现的次数远大于Race, Canoe。这些高频词在相似矩阵中占有主导地位,不管待标注图像如何,这些高频词出现在最后结果中的概率都是比较大的,影响了最后的标注结果。

为了克服训练集中高频词对最终标注结果的偏导影响,我们采用了逆向文档频率(IDF)修正图像节点与其标注词节点之间边的权值,以改善因高频词引起的偏差。

在图像标注中,一幅图像相当于文档,图像的标注词相当于文档中的关键词。设 $df(w_i)$ 表示标注词 w_i 在训练集中出现的次数,我们用式(3)计算图像结点与连接它的标注词边的权值:

$$\text{sim}(i, w_j) = ((1-\lambda) + \lambda * \log((1 + |w|) / df(w_j))) * \delta(i, w_j) \quad (3)$$

式中, $|w|$ 是标注词汇总表的大小。当 w_j 是图像 i 的标注词时,函数 $\delta(i, w_j)$ 值取1,否则其值取0。 λ 是平滑系数,用来调节训练集中高频词与低频词的权重。

1.3 算法描述

算法1 生成互K近邻图

输入:已标注的图像集合 $T=\{i_1, i_2, \dots, i_n\}$ 和待标注的图像 I_q 。

输出:互K近邻图邻接矩阵 A 。

Step 1 将 I_q 分割成区域。

Step 2 为 T 中的每个图像 i 、区域 f_i 、标注词 w_i 创建顶点,为 I_q 创建图像顶点和区域顶点。

Step 3 对区域顶点 f_i ,与之K-互邻居区域顶点 f_j 添加一条边,并用式(1)计算权值 $\text{sim}(f_i, f_j)$ 。

Step 4 对每个图像顶点 i ,与之标注词顶点 w_i 添加一条边。根据式(3)计算边权值。

Step 5 对标注词顶点 w_i ,搜寻与之K-互邻居区域顶点 w_j 并添加一条边,用式(2)计算权值 $\text{sim}(w_i, w_j)$ 。

2 图像标注过程

待标注的图像首先分割成若干个区域,然后每个区域与训练集中互K近邻的区域分别用一条边相连。以待标注的图像节点为起点,运行随机游走算法^[9]确定与待标注的图像最相关的若干词作为待标注的图像的标注。RWR可以被认为基于起始节点的马尔可夫链的转换过程。它可用迭代式(4)表示:

$$R^{n+1} = cAR^n + (1-c)Y \quad (4)$$

式中, R 是代表节点的马尔可夫链转换状态的 N 维矢量, N 是图的节点数。 A 是图的近邻矩阵。 $(1-c)$ 表示游走过程中从各节点返回起点的概率。 Y 是 N 维矢量,除了对应起始点的分量为1外,其它分量均为0。为了确保方程(4)能收敛,要求 A 进行列规一化, R 的所有分量之和为1。当方程(4)收敛时,将对应于词节点的 R 分量按降序排列,选取前 n (如 $n=5$)个分量对应的词作为待标注图像的标注。

3 快速求解算法

在实现标注过程中,通常是通过迭代方程(4)直到其收

敛。在实际情况中迭代时间较长,尤其是对图像标注系统而言,训练数据集比较大,RWR 运行时间会更长,这是因为 RWR 迭代时间的复杂度正比于迭代次数和图的规模。如何快速求解 RWR 已成为一个有挑战性的工作。我们在对互 K 近邻图进行结构分析的基础上,提出一种快速求解的算法,该算法在不明显降图像低标注精度情况下,可避免迭代运算,实现快速求解。

方程(4)可改写为:

$$R^{n+1} = (1-c)(I-cA)^{-1}Y \quad (5)$$

令 $Q = I - cA$, 则式(5)可写为:

$$R^{n+1} = (1-c)Q^{-1}Y \quad (6)$$

式(6)是 RWR 算法的封闭解析式,但是直接用它求解在实际应用中还是不太现实的,因为计算并存储 Q^{-1} 的空间复杂度和时间复杂度分别是 $O(n^2)$ 和 $O(n^3)$, 对于像图像标注系统这样大的数据集中,直接计算 Q^{-1} 是不太现实的。

在实验中我们发现互 K 近邻图具有两个显著的特点:一是矩阵 Q 中,行与列存在线性相关,二是图 G 中存在着疏密不同的区域,整个图的结构呈块状分布。针对互 K 近邻图的这些特点,权衡标注的精度与速度的关系,我们提出了一种快速、近似的求解 RWR 算法。

首先,根据第一个特性,我们可以通过降维,用一低秩矩阵来低阶近似 Q^{-1} ,再根据第二个特性,我们可以把图 G 分解为 k 个子图,各子图内的邻接矩阵用 $A_{1,i} (i=1,2,\dots,k)$ 表示,子图之间的邻接矩阵用 A_2 表示。这样,把原来求高阶矩阵的逆转化为求若干个低阶矩阵的逆。算法描述如下。

设有互 K 近邻图 $G = \langle V, E \rangle$, 其中 V 是 n 个节点集,其邻接矩阵为 A 。对矩阵 A 进行规一化:

$$L = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \quad (7)$$

式中, $D_{ii} = \sum_{k=1}^n A_{ik}$, $D_{ij} = 0, i \neq j$ 。利用 K-way 规一化割算法^[10]将图 G 剖分成 k 个子图,并分解 L 为两个矩阵之和:

$$L = L_1 + L_2 \quad (8)$$

式中, L_2 是各子图间的邻接矩阵, L_1 可表示为:

$$L_1 = \begin{bmatrix} L_{1,1} & 0 & \cdots & 0 \\ 0 & L_{1,2} & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & L_{1,k} \end{bmatrix} \quad (9)$$

$L_{1,i} (i=1,2,\dots,k)$ 表示各子图内邻接矩阵。对 L_2 进行特征值分解 $L_2 = USV$, 则:

$$L = L_1 + USV \quad (10)$$

所以

$$(I-cL)^{-1} = (I-cL_1 - cUSV)^{-1} \quad (11)$$

由 Sherman-Morrison 引理^[11], 结合式(10)可得:

$$\begin{aligned} (I-cL)^{-1} &= (I-cL_1 - cUSV)^{-1} \\ &= Q_i^{-1} + cQ_i^{-1}U(S^{-1} - cVQ_i^{-1}U)^{-1}VQ_i^{-1} \\ &= Q^{-1} \end{aligned}$$

由式(6)得

$$\begin{aligned} R^{n+1} &= (1-c)Q^{-1}Y \\ &= (1-c)(Q_i^{-1} + cQ_i^{-1}U(S^{-1} - cVQ_i^{-1}U)^{-1}VQ_i^{-1})Y \end{aligned} \quad (12)$$

又因 L_1 是对角矩阵, 对应地 Q_i^{-1} 改写为:

$$Q_i^{-1} = (I - cL_{1,i})^{-1} = \begin{bmatrix} Q_{i,1}^{-1} & 0 & \cdots & 0 \\ 0 & Q_{i,2}^{-1} & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & Q_{i,k}^{-1} \end{bmatrix} \quad (13)$$

式中, $Q_{i,i}^{-1} = (I - cL_{1,i})^{-1}$ 。

根据式(12)和式(13)可知求解高阶矩阵 Q 的逆 Q^{-1} , 现可转化为求解规模较小的 $Q_{i,i}^{-1} (i=1,2,\dots,k)$ 和 $(S^{-1} - cVQ_i^{-1}U)^{-1}$ 。如果预先计算并存储 $Q_{i,i}^{-1} (i=1,2,\dots,k)$ 和 $(S^{-1} - cVQ_i^{-1}U)^{-1}$, 我们就可以快速计算出 Q^{-1} , 进而可实现快速计算 R 。算法总结如下。

算法2 基于互 K 近邻图的图像标注快速求解

输入: 互 K 近邻图加权矩阵 A , 和初始矢量 Y

输出: 矢量 R

// 离线预计算

Step 1 计算 $L = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 。

Step 2 利用 K-way 规一化割算法将图 G 剖分成 k 个子图。

Step 3 分解 L 为两子矩阵之和: $L = L_1 + L_2$ 。

Step 4 根据式(9)计算 $L_{1,i}$ 。

Step 5 for $i=1$ to k

$$Q_{i,i}^{-1} \leftarrow (I - cL_{1,i})^{-1}$$

End for

Step 6 对 L_2 进行特征值分解: $L_2 = USV$ 。

Step 7 根据式(13), 计算并存储 $(S^{-1} - cVQ_i^{-1}U)^{-1}$ 。

// 在线计算

Step 8 根据式(12), 计算并输出 R 。

4 算法的复杂度分析

算法由离线预处理和在线计算 R 两部分组成。下面我们分别讨论两部分的时间与空间的复杂度。

(1) 在线计算 R 时间复杂度: 根据式(12)可知, 计算 R 可分解为 6 次矩阵乘积:

$$R_0 \leftarrow Q_i^{-1}Y$$

$$R \leftarrow VR_0$$

$$R \leftarrow (S^{-1} - cVQ_i^{-1}U)^{-1}R$$

$$R \leftarrow UR$$

$$R \leftarrow Q_i^{-1}R$$

$$R \leftarrow (1-c)(R_0 + cY)$$

由于在离线预处理阶段已计算并存储了 Q_i^{-1} , $(S^{-1} - cVQ_i^{-1}U)^{-1}$ 和 V , 因此在在线计算 R 时能达到实时输出的要求。

(2) 离线预处理复杂度: 离线预处理主要包含几个步骤:

a) 图 G 的剖分;

b) 计算 k 个 $Q_{i,i}^{-1} = (I - cL_{1,i})^{-1}, i=1,2,\dots,k$ 的逆矩阵 $Q_{i,i}^{-1}$;

c) 计算 L_2 的特征值分解;

d) 计算 $(S^{-1} - cVQ_i^{-1}U)^{-1}$ 。

计算 $(S^{-1} - cVQ_i^{-1}U)^{-1}$ 的时间复杂度与计算 $Q_{i,i}^{-1}$ 是同一数量级, 因此整个离线预处理时间复杂度决定于计算 $k+1$ 个小规模的矩阵逆运算、 L_2 的特征值分解所需的时间和图 G 的剖分。

在空间复杂度上, 离线预处理阶段只需存储 $k+1$ 个小规模的矩阵、一个 U 矩阵和一个 V 矩阵。另外, 在实验中我们发现矩阵 $Q_{i,i}^{-1}, U$ 和 V 中有许多元素几乎接近于零, 把这些元

素设置为零,并不影响输出精度。如果我们采用稀疏矩阵来表示矩阵 Q_i^j , U 和 V , 会有效地减少其存储空间。

另一方面,由于规一化加权矩阵 L 具有对称性,因而 k 个低阶矩阵 Q_i^j 也具有对称性,并且 $U=V^T$, 利用对称性质,我们能减少近 50% 的存储空间。

5 实验结果与分析

实验数据集:为了验证算法的有效性,实验数据集选用了 Corel 图像集。我们选取其中的 500 幅图像进行测试比较。其中 450 幅图像作为训练集,50 幅图像作为测试图像。每幅图像运用图像分割算法分割成 1-10 个区域,共形成 4460 个图像区域。用算法 1 构造互 K 近邻图 G , 图 G 有 4460 个节点。

参数的设定:为了方便描述,我们把第 2 节的算法简称为 MKNN 算法,本节算法简称为 Fast-MKNN 算法。在实验中两个算法中的参数设置相同,即:互近邻参数 $K=27$, 参数 $C=0.35$, 参数 $\mu=0.4$, $\lambda=0.3$ 。

实验 1 子图个数 k 对该算法的影响

下面,我们将通过实验来考察子图个数 k 对算法的性能影响,实验中分别选取 $k=1, 50, 100$ 和 4460, 考察预处理存储空间、预处理时间、在线计算 R 时间和输出的精度的影响。实验结果如表 1 所列。

表 1 子图个数 k 对性能的影响

子图个数 K	预处理存储空间	预处理时间(秒)	在线计算 R 时间(秒)	精度
$K=1$	180M	840	0.4	0.371
$K=50$	21M	90	0.7	0.353
$K=100$	20.2M	147	0.9	0.345
$K=4460$	17M	303	0.6	0.342

当 $K=1$ 时, $L_1=L, L_2=0$, 这种情况下相当于直接根据式(6)求解 R , 所以预处理时间较长(840 秒)。另一方面,由于没有近似处理,因此输出精度不受影响。当 $K=4460$ (相当于 $K=n$)时,由于子图的增加,预处理时间相应延长,同时增加了近似处理,其输出精度有所降低。

实验 2 相关算法的比较与分析

在本组实验中,我们通过与第 2 节中所讨论的迭代算法做比较,验证了 Fast-MKNN 方法的有效性,比较结果如表 2 所列。由此可能得出:Fast-MKNN 方法能显著提高运行速度。虽然其对图像标注的查准率和查全率有所降低,但影响不大。

表 2 实验结果对比

methods	Avg_precision	Avg_recall	accuracy
GCap	0.1308	0.1607	0.3431
MKNN	0.1578	0.1919	0.3707
Fast_MKNN	0.1501	0.1892	0.3687

结束语 本文研究了基于互 K 近邻图的自动图像标注,该算法用一个 MKNN 图融合了低层特征间、标注词间以及图像与标注词间的关联关系,并结合随机游走(RWR)算法,实现了图像语义自动标注。通过运用互 K 邻接关系和逆向

文档频率(IDF)技术,有效地提高了图像标注的性能。在 Corel 图像数据集上进行的仿真实验证明了所提算法的有效性,无论是查准率还是查全率都比 Pan 等人提出的 GCap 方法有明显的提高。同时,在对互 K 近邻图的结构分析的基础上,提出了一种随机游走(RWR)快速求解的算法。该算法在不明显降图像低标注精度情况下,可避免迭代运算,实现快速求解。实验结果表明本文所提新方法取得了令人满意的效果。

参考文献

- [1] Duygulu P, Barnard K, de Freitas J F G, et al. Object recognition as machine translation; learning a lexicon for a fixed image vocabulary[J]. Lecture Notes in Computer Science, Heidelberg: Springer, 2002, 23(53): 97-112
- [2] Jeon J, Lavrenko V, Mnmatha 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. Toronto, 2003, 119-126
- [3] Carneiro G, Chan A B, Moreno P J, et al. Supervised learning of semantic classes for image annotation and retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410
- [4] Vasconcelos N. Minimum probability of error image retrieval [J]. IEEE Transactions on signal Processing, 2004, 52(8): 2322-2336
- [5] Cusano C, Ciocca G, Schettini R. Image annotation using SVM [J]. Proceedings of SPIE, 2004, 53(41): 330-338
- [6] Pan Jia-yu, Yang Hyung-jeong, Faloutsos C, et al. GCap: Graph-based Automatic Image Captioning[C]// Proceedings of the 4th International Workshop on Multimedia Data and Document Engineering(MDDE 04), in conjunction with Computer Vision Pattern Recognition Conference(CVPR 04). 2004: 146-156
- [7] Witten I H, Moffat A, Bell T. Managing Gigabytes: Compressing and Indexing Documents and Images [M]. Morgan Kaufmann Publishers, 1999
- [8] Miller G, Beckwith R, Fellbaum C, et al. WordNet: An on-line lexical database [J]. International Journal of Lexicography, 1990, 3(4): 235-244
- [9] Bailloeuil T, Zhu C Z, Xu Y. Automatic image tagging as a random walk with priors on the canonical correlation subspace [C]// Proceedings of ACM Multimedia Information Retrieval. 2008: 75-82
- [10] Meila M, Xu L. Multiway cuts and spectral clustering [R]. Washington Tech Report. 2003
- [11] Tong H, Faloutsos C, Pan J Y. Random walk with restart: fast solutions and applications[J]. Knowledge and Information Systems, 2008, 14: 327-346
- [12] Piegorsch W, Casella G E. Inverting a sum of matrices [J]. SIAM Review, 1990, 32(3): 470-470