

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2012 **11**
Vol.17 No.

中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2012年11月 第17卷 第11期(总第199期)

目次

综述

数字图像合成技术综述 吴昊, 徐丹(1333)

图像处理和编码

基于群稀疏的结构化字典学习 郭景峰, 李贤(1347)

SSIM 度量虚拟视点绘制失真的深度图帧内编码 喻莉, 张军涛, 邓慧萍, 向森, 周鹏, 左雯, 王宁(1353)

统计量移位的鲁棒无损图像信息隐藏 李晓博, 周诠(1359)

伪造图像典型篡改操作的检测 左菊仙, 刘本永(1367)

图像分析和识别

融合灰度和 SURF 特征的红外目标跟踪 范新南, 丁朋华, 刘俊定, 张学武(1376)

海面温度栅格图的锋面提取与矢量化 崔雪森, 周为峰, 王栋, 张胜茂(1384)

交通场景中车辆的运动检测与阴影消除 王彬, 冯远静, 郭海峰, 张贵军(1391)

基于随机点积图的图像标注改善算法 孙登第, 罗斌, 郭玉堂(1400)

图像理解和计算机视觉

有监督子空间建模和稀疏表示的场景分类 段菲, 章毓晋(1409)

对立色 LBP 模型的目标跟踪 张炯, 宁纪锋, 颜永丰, 于伟(1418)

计算机图形学

联合骨架与边界特征的平面形状分解…………… 蒋建国, 周丹凤, 郝世杰, 郭艳蓉, 詹曙(1425)

屏幕空间自适应的地形 Tessellation 绘制…………… 张兵强, 张立民, 艾祖亮, 张建廷(1431)

遥感图像处理

SAR 图像稀疏优化滤波…………… 杨萌, 张弓(1439)

分段线性动态矩匹配条带去除…………… 秦雁, 邓孺孺, 何颖清, 陈蕾, 陈启东(1444)

基于 Harris 角点和 SIFT 描述符的高分辨率遥感影像匹配算法…………… 陈梦婷, 闫冬梅, 王刚(1453)

第八届图像图形技术与应用学术会议征文通知…………… (1460)

中国图象图形学报

刊名题字: 宋 健

月刊(1996 年创刊)

第 17 卷 第 11 期

2012 年 11 月 16 日出版

主管单位 中国科学院

主 办 中国科学院遥感应用研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 李小文

编辑出版 《中国图象图形学报》编辑出版委员会

北京 9718 信箱 邮编 100101
电子信箱:jig@irsa.ac.cn
电话:010-64807995 010-82614429
网 址:www.cjig.cn

印刷装订 北京北林印刷厂

广告经营许可证 京朝工商广字第 0346 号

总 发 行 北京报刊发行局

订 购 全国各地邮局

国外发行 中国国际图书贸易总公司
(中国国际书店)
(北京 399 信箱 邮编 100044)

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing Application,
CAS China Society of Image and Graphics
Institute of Applied Physics and Computational
Mathematics

Chief editor LI Xiaowen

Editor, Publisher Editorial and Publishing Board
of Journal of Image and Graphics
(P. O. Box 9718, Beijing 100101, China)
E-mail:jig@irsa.ac.cn

Distributed by Beijing Bureau for Distribution of Newspapers
and Journals

Domestic All Local Post Offices in China

Foreign China International Book Trading Corporation
(P. O. Box 399, Beijing 100044, China)

Printed by Beijing Beilin Printing House

ISSN 1006-8961 CN11-3758/TB CODE ZTTXFZ 国内邮发代号: 82-831 国外发行代号: M1406 国内定价: 45.00 元

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2012)11-1400-09

论文引用格式: 孙登第, 罗斌, 郭玉堂. 基于随机点积图的图像标注改善算法[J]. 中国图象图形学报, 2012, 17(11): 1400-1408.

基于随机点积图的图像标注改善算法

孙登第^{1,2}, 罗斌^{1,2}, 郭玉堂³

1. 安徽大学计算机科学与技术学院, 合肥 230039; 2. 安徽省工业图像处理与分析重点实验室, 合肥 230039;
3. 合肥师范学院计算机科学与技术系, 合肥 230061

摘要: 针对自动图像标注中底层特征和高层语义之间的鸿沟问题, 提出一种基于随机点积图的图像标注改善算法。该算法首先采用图像底层特征对图像候选标注词建立语义关系图, 然后利用随机点积图对其进行随机重构, 从而挖掘出训练图像集中丢失的语义关系, 最后采用重启式随机游走算法, 实现图像标注改善。该算法结合了图像的底层特征与高层语义, 有效降低了图像集规模变小对标注的影响。在3种通用图像库上的实验证明了该算法能够有效改善图像标注, 宏F值与微平均F值最高分别达到0.784与0.743。

关键词: 图像标注改善; 随机点积图; 语义关系图; 重启式随机游走

Image annotation refinement based on a random dot product graph

Sun Dengdi^{1,2}, Luo Bin^{1,2}, Guo Yutang³

1. School of Computer Science and Technology, Anhui University, Hefei 230039, China;
2. Key Laboratory for Industrial Image Processing and Analysis of Anhui Province, Hefei 230039, China;
3. Department of Computer Science and Technology, Hefei Normal College, Hefei 230061, China

Abstract: In order to overcome the semantic gap between low-level features and high-level semantic concepts of imagery, a new image annotation refinement approach based on Random Dot Product Graph (RDPG) is proposed. In our approach, the visual features of images are used to construct a semantic graph of the candidate annotations. Then, we reconstruct the semantic graph with a RDPG, find the unobserved relevance in the incompletely observed semantic graph, and transform the random graph into the probabilities of state transition. Combined with Random Walk with Restart (RWR), the final annotations are chosen. This new method incorporates the visual and semantic information of images, and reduces the influence of the scale of database. Experiments conducted on three standard databases demonstrate that our approach outperforms the existing image annotation refinement techniques. The macro F-Score and micro average F-Score can reach 0.784 and 0.743 respectively.

Key words: image annotation refinement; random dot product graph (RDPG); semantic graph; random walk with restart (RWR)

0 引言

为了有效管理与查询日益增多的图像资源, 图像检索技术应运而生。在基于内容的图像检索中,

自动图像标注是一项关键技术, 已成为该领域研究热点, 受到了广泛关注^[1-3]。目前的自动图像标注技术主要依赖于图像的纹理、颜色、形状等底层特征, 然后通过相似性度量实现标注。然而, 从图像底层特征到高级语义词汇之间存在着难以跨越的“语义

收稿日期: 2011-12-23; 修回日期: 2012-05-08

基金项目: 国家自然科学基金项目(61073116, 61003131); 安徽省自然科学基金项目(11040606M134)

第一作者简介: 孙登第(1983—), 男, 安徽大学计算机科学与技术学院博士研究生, 主要研究方向为随机图论、模式识别和数字图像处理。E-mail: sundengdi@163.com

鸿沟”,底层特征的相似往往并不能完全反映和匹配用户的查询意图。这使得标注结果必然会产生许多噪声词汇。因此对已有的标注候选词进行进一步筛选,获得更准确的标注结果,是弥补“语义鸿沟”的一种合理策略,这就是图像标注改善技术。

事实上,在图像理解过程中,图像内容之间的相关性远比图像底层特征间的相似性要重要得多,而内容之间的相关性表现为标注词汇之间的语义关系。所谓语义关系在图像标注中是指两条标注词出现在同一场景或同一图像中的可能性。因而如何合理描述图像集中的语义关系成为图像标注改善的关键问题,现有的图像标注改善方法也都建立在对待标注词间语义关系分析的基础上,并将图像的底层特征或文本信息融入到标注改善的过程中。

Jin 等人最早提出利用基于知识的 WordNet 来进行图像标注改善^[4]。首先对一幅待查询图像标注出一组候选词汇,依据这些词汇在 WordNet 之中的相关性,将相关词汇中数目最多的一种组合确定为正确的标注。由于 WordNet 是词汇间的一般关系,并不反映图像信息,且包含词汇有限,因而在剔除噪声词同时,该方法也可能会把正确标注去掉。

此后,Jin 等人又提出了一种基于知识和图算法的图像标注改善方法^[5]。该方法整合了多种度量方式来计算词汇间语义关系,并将图像标注改善问题转化为图割问题,对候选标注词构建加权语义关系图,再用图割方法将图划分为相关与不相关两部分,从而得到最终改善结果。

Wang 等人采用了另一种方式构建词汇语义关系,即计算样本图像库中两种语义共同出现在一幅图像上的频率,称为语义“共现性”^[6-7]。利用这一关系图,Wang 等人提出了基于内容的图像标注改善方法,一定程度上解决了前述问题。但是这类方法需要大规模的图像库,当库中图像较少或是事先标注不准确时,语义“共现性”的计算会产生巨大误差,从而影响最终改善结果。

在对语义关系图的研究基础上,提出一种基于随机点积图的图像标注改善方法,利用随机点积图重构了由训练图像集得到的图像语义关系图,通过对图的随机化处理,挖掘、补充和修正了语义相关性,并形成标注词转移概率,最终实现图像标注改善。实验结果表明,该方法可以有效提高图像标注精度,尤其在图像库较小时,具有明显优势。

1 随机点积图理论

1.1 随机点积图的概念

随机点积图(RDPG)是近年提出的一种新型随机图模型^[8],目前已有大量论文证明了其统计性质^[8-9],在社会网络关系挖掘中得到了广泛应用^[10-11]。随机点积图的基本思想是:

对节点集 V 中每个节点 $v_i, i=1, \dots, n$, 从 d 维单位空间中独立、均匀地随机选取一个 d 维向量 $\mathbf{x}_i \in \mathbf{R}^d$ 作为 v_i 的赋值,即 $\mathbf{x}_i \sim U[0, 1]^d$ 。每对节点 v_i, v_j 之间的边以概率

$$p_{ij} = f(\mathbf{x}_i \cdot \mathbf{x}_j) \quad (1)$$

随机独立出现,生成在赋值 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]_{d \times n}$ 下的随机点积图 $G_X(\mathbf{V}, \mathbf{P})$ 。

式(1)中 f 为一单调上升函数,一般简化为直接用向量的点积来表示边出现的概率 $p_{ij} = \mathbf{x}_i \cdot \mathbf{x}_j$ 。当两节点赋值接近时, p_{ij} 较大,在随机点积图理论中这意味着节点 v_i, v_j 所表示的对象具有相似的特征或兴趣,因而 v_i, v_j 之间存在某种关系的可能性较大。

1.2 随机点积图的求解

随机图中每条边都是随机独立出现,依据贝努利分布:随机点积图 $G_X(\mathbf{V}, \mathbf{P})$ 按概率 p_{ij} 生成边集 E ,得到某个图 $G(\mathbf{V}, E)$ 的似然概率为

$$P_X(\mathbf{G}) = \prod_{i < j, ij \in E} (\mathbf{x}_i \cdot \mathbf{x}_j) \prod_{i < j, ij \notin E} (1 - \mathbf{x}_i \cdot \mathbf{x}_j) \quad (2)$$

按照随机图的观点,在现实中我们所观测到的图只是随机图在某一时刻的状态,因此需要获得能产生这一状态的边概率,也就是找到最优的配置 \mathbf{X} ,使得产生的随机图在概率意义下最有可能生成当前已观测到的图 $G(\mathbf{V}, E)$,即

$$\mathbf{X} = \underset{\mathbf{X}}{\operatorname{argmax}} P_X(\mathbf{G}) \quad (3)$$

根据这一观点,为避免混淆,在下文中称这些非随机的图为观测图。

若观测图 $G(\mathbf{V}, E)$ 为一加权无向图,其邻接矩阵为 $\mathbf{A} = (a_{ij})_{n \times n}, a_{ij} \in [0, 1]$ 为边的权重,则

$$P_X(\mathbf{G}) = \prod_{i \neq j} (\mathbf{x}_i \cdot \mathbf{x}_j)^{a_{ij}} (1 - \mathbf{x}_i \cdot \mathbf{x}_j)^{1-a_{ij}}$$

其对数似然函数为

$$L_X(\mathbf{G}) = \sum_{i \neq j} a_{ij} \ln(\mathbf{x}_i \cdot \mathbf{x}_j) + (1 - a_{ij}) \ln(1 - \mathbf{x}_i \cdot \mathbf{x}_j) \quad (4)$$

在观测图中,边的权重反映了节点之间的相关

性。式(4)中当边的概率与权重成正比例时似然函数 $L_X(\mathbf{G})$ 变大。所以要求式(4)最大,就要求边的概率与权重大小尽可能对应。根据对偶原则可得

$$\max L_X(\mathbf{G}) = \min f_A(\mathbf{X})$$

式中

$$f_A(\mathbf{X}) = \sum_{i \neq j} (\mathbf{x}_i \cdot \mathbf{x}_j - a_{ij})^2 \quad (5)$$

因此求解随机点积图的目标函数转化为

$$\min f_A(\mathbf{X}) = \min \sum_{i \neq j} (\mathbf{x}_i \cdot \mathbf{x}_j - a_{ij})^2 \quad (6)$$

文献[8]对式(6)给出了求解方法。将观测图作为随机点积图在某时刻的状态,并求解出最大似然时生成该观测图的节点赋值 \mathbf{X} ,这一过程也称为对观测图 $\mathbf{G}(\mathbf{V}, \mathbf{E})$ 的随机点积化或随机点积重构。

1.3 随机点积图的传导性

定理 1 对随机点积图 $\mathbf{G}_X(\mathbf{V}, \mathbf{P})$ 中任意节点 u, v, w , “ \leftrightarrow ”表示两节点相连,存在如下概率关系

$$P(u \leftrightarrow w | u \leftrightarrow v \leftrightarrow w) > P(u \leftrightarrow w) \quad (7)$$

证明:由于 $\mathbf{x}_i \sim U[0, 1]^d$, 则 u, w 相连的事件关于随机变量 x_i 的概率为

$$P(u \leftrightarrow w) = \int_{[0,1]^d} \int_{[0,1]^d} (\mathbf{x}_u \cdot \mathbf{x}_w) d\mathbf{x}_u d\mathbf{x}_w = \frac{1}{(1+d)^2}$$

根据条件概率

$$P(u \leftrightarrow w | u \leftrightarrow v \leftrightarrow w) = \frac{P(u \leftrightarrow w, u \leftrightarrow v \leftrightarrow w)}{P(u \leftrightarrow v \leftrightarrow w)}$$

分子部分为 u, v, w 相互连通构成三角形的概率。

所以 $P(u \leftrightarrow w | u \leftrightarrow v \leftrightarrow w) =$

$$\frac{\int_{[0,1]^d} \int_{[0,1]^d} \int_{[0,1]^d} (\mathbf{x}_u \cdot \mathbf{x}_v)(\mathbf{x}_u \cdot \mathbf{x}_w)(\mathbf{x}_v \cdot \mathbf{x}_w) d\mathbf{x}_u d\mathbf{x}_v d\mathbf{x}_w}{\int_{[0,1]^d} \int_{[0,1]^d} \int_{[0,1]^d} (\mathbf{x}_u \cdot \mathbf{x}_v)(\mathbf{x}_v \cdot \mathbf{x}_w) d\mathbf{x}_u d\mathbf{x}_v d\mathbf{x}_w} = \frac{(1+d)^2(1+2d)}{(1+2d)^3} > \frac{1}{(1+d)^2} = P(u \leftrightarrow w)$$

证毕

由定理 1 可知,通过点积构造边概率,两节点 u, w 在有第 3 点 v 与其分别相连时,该两点连接的概率大于没有任何其他连接时的概率。也就是说,通过其他节点的“传导作用”,可以挖掘出隐藏的相互关系,这是传统基于欧氏距离的图结构所不具备的^[9]。

2 随机点积图对语义关系图的重构

2.1 语义关系图的构成

WordNet 并不适合反映图像中的语义相关性,

当前许多方法使用标注词在图像库中的“共现”情况构造语义关系图。传统的语义共现一般只计算两词汇同时出现于库中同一幅图像的频率,这样就丢失了图像的底层特征,因此提出一种与文献[6]类似的融合图像底层特征的语义关系计算方法。

对一幅待标注图像 \mathbf{I} 与其 k 个候选标注 w_1, \dots, w_k , 在该幅图像中语义 w_i, w_j 语义相关性为

$$\text{sim}(w_i, w_j) = \frac{\sum_{\mathbf{J} \in \mathbf{T}_{ij}} \text{sim}(\mathbf{J}, \mathbf{I})}{|\mathbf{T}_{ij}|} \quad (8)$$

这里 \mathbf{T}_{ij} 为图像库中同时被标注有 w_i, w_j 两个词的图像集, $|\mathbf{T}_{ij}|$ 为该图像集中图像个数。 $\text{sim}(\mathbf{J}, \mathbf{I})$ 为已标注图像与待标注图像的底层特征相似度

$$\text{sim}(\mathbf{J}, \mathbf{I}) = \frac{1}{\sqrt{2^p \pi^p |\Sigma|}} \exp\{- (\mathbf{g}_J - \mathbf{g}_I)^\top \Sigma^{-1} (\mathbf{g}_J - \mathbf{g}_I)\} \quad (9)$$

\mathbf{g}_J 与 \mathbf{g}_I 分别为已标注图像 \mathbf{J} 与待查询图像 \mathbf{I} 的底层特征, p 为特征的维数, Σ 为特征向量的协方差矩阵, $\text{sim}(\mathbf{J}, \mathbf{I}) \in [0, 1]$ 。

式(8)计算已标注 w_i, w_j 的图像与待查询图像的底层特征相似度作为 w_i, w_j 的相关性,给出最符合此图像的语义关系,这同时考虑图像的底层特征与图像库中的高层语义。由此得到关于待测图像 \mathbf{I} 的候选词汇语义关系图 $\mathbf{G}(\mathbf{V}, \mathbf{E})$:

k 个候选词汇构成节点集 $\mathbf{V}\{w_1, \dots, w_k\}$,

$$\mathbf{A} = (a_{ij})_{k \times k}, \quad a_{ij} = \begin{cases} \text{sim}(w_i, w_j) & |\mathbf{T}_{ij}| > 0 \\ 0 & |\mathbf{T}_{ij}| = 0 \end{cases} \quad (10)$$

2.2 语义关系图的随机点积重构

式(8)中的语义关系本质上仍依赖于标注词在图像中的“共现”情况。当标注词 w_i, w_j 从未出现于库中同一幅图像时, $|\mathbf{T}_{ij}| = 0$, w_i 与 w_j 的语义关系依然为 0。常用图像库中基本都是描述单一目标的图像,这些图像识别性强,但不利于描述图像中的语义关系。借助下面的例子来说明这个问题。

Corel5k 是当前图像标注领域使用最多的图像库,规模较大,共包含 50 个不同的主题,同一主题下的标注词会在该主题图像中较频繁地“共现”。按 2.1 节的方法构造了自然主题的语义关系图。

图 1(a)是从自然主题的图像中得出的标注词语义关系图 $\mathbf{G}(\mathbf{V}, \mathbf{E})$ 的一部分,图 1(b)是对应该部

分的邻接矩阵,其中颜色由深到浅表示对应的语义相关性由大到小。

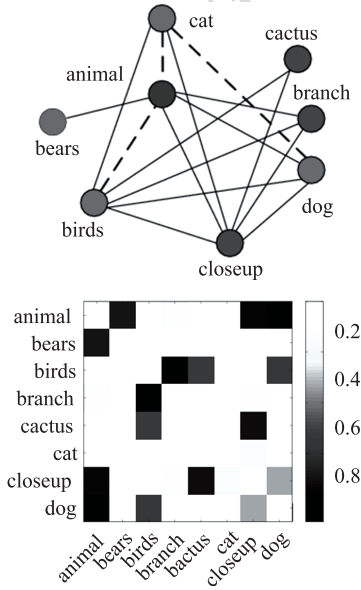


图 1 初始语义关系图与邻接矩阵的一部分

Fig. 1 A part of semantic graph and adjacent matrix

图 1 中每种词汇看做一个节点,实线表示两词汇在图像库中“共现”,虚线表示缺失的“共现”。图 1 充分反映了仅通过图像库得到的语义关系具有明显不足:“animal”与“birds”、“cat”与“dog”这些常见的语义关系在该图像库中都没有“共现”。

按照随机图的观点,所有标注词之间完整的语义关系可以用随机图 $G(V, P)$ 来描述,而当前语义关系图 $G(V, E)$ 是 $G(V, P)$ 在此时刻(该图像库中)的状态。通过式(6)可以解出最大似然时的节点赋值 X ,从而得到生成这个观测图的边概率,实现对 $G(V, E)$ 的随机点积化得到 $G_X(V, P)$ 。由定理 1 可知,在 $G_X(V, P)$ 中标注词 w_i, w_j 与 w_k 之间存在概率关系 $P(w_i \leftrightarrow w_k | w_i \leftrightarrow w_j \leftrightarrow w_k) > P(w_i \leftrightarrow w_k)$ 。

在图 1 的例子中,虽然“animal”与“birds”不相连,但“dog”、“branch”和“closeup”3 个词汇都分别与“animal”和“birds”相连,成为这两个词汇的中间连接词,符合 $w_i \leftrightarrow w_j \leftrightarrow w_k$ 的条件。因此在标注词语义关系的随机点积图 $G_X(V, P)$ 中,有

$$P(w_{\text{animal}} \leftrightarrow w_{\text{birds}} | w_{\text{animal}} \leftrightarrow w_{\text{dog}} \leftrightarrow w_{\text{birds}}) > P(w_{\text{animal}} \leftrightarrow w_{\text{birds}})$$

$$P(w_{\text{animal}} \leftrightarrow w_{\text{birds}} | w_{\text{animal}} \leftrightarrow w_{\text{branch}} \leftrightarrow w_{\text{birds}}) > P(w_{\text{animal}} \leftrightarrow w_{\text{birds}})$$

$$P(w_{\text{animal}} \leftrightarrow w_{\text{birds}} | w_{\text{animal}} \leftrightarrow w_{\text{closeup}} \leftrightarrow w_{\text{birds}}) > P(w_{\text{animal}} \leftrightarrow w_{\text{birds}})$$

上述关系说明,多个中间标注词的传导作用使得“animal”与“birds”之间也可能具有显著的语义关系。虽然“bears”与“dog”也通过“animal”相连,但二者间仅有这一个中间标注词,在 $G_X(V, P)$ 中的语义关系概率显然没有前一个例子显著。这与真实的语义关系是完全一致的:随机点积图可以获得任意标注词之间的关系,常见的语义关系概率较大,不常见的语义关系概率较小但不为 0,因为可能存在少量的语义不一致的异常图像;对于不同的图像库,语义“共现”情况并不相同,最常见的语义关系也可能不会在库中表现出来,但在概率意义下常见的语义关系总是会出现较多,从而可以在随机点积图中挖掘出丢失的关系。

基于这一原理,提出如下语义关系图随机点积重构算法,通过多次重构得到完整的语义关系。

算法 1 语义关系图的随机点积重构算法

输入:待标注图像 I , 候选标注词汇 w_1, \dots, w_k 。

输出:候选词语义关系的随机点积图 $G_X(V, P)$ 。

- 1) 用式(10)构造 w_1, \dots, w_k 的语义关系图 $G(V, E)$;
- 2) 初始化 $p_{ij} = \text{sim}(w_i, w_j), i, j = 1, \dots, k$, 阈值 t ;
- 3) 取 $p_{ij} \geq t$ 的边与所有节点 V 构造 $G_t(V, E)$;
- 4) 对 $G_t(V, E)$ 随机点积重构,用式(6)求解赋值 X ;
- 5) 计算 $P = X^T X$,用式(4)计算当前对数似然函数 $L_X(G_t)$,返回步骤 2),直到 $L_X(G_t)$ 收敛(达到最大)。

迭代收敛时的 $P = (p_{ij})_{k \times k}$ 就是重构后的随机点积图的边概率矩阵。设置阈值 t 为

$$t = \min_{ij \in E, |T_{ij}| > 0} (\text{sim}(w_i, w_j)) \quad (11)$$

即训练得到的语义关系图中语义相似度的最小值,这使得在初始阶段 $G_t(V, E)$ 与 $G(V, E)$ 相同。事实上,由于常见语义关系概率较大,只需要迭代 3~5 次就可以达到收敛。

图 2 是随机点积图重构之后的部分语义关系。该结果成功挖掘出了图 1 中丢失的语义关系;显示出高层语义“animal”与这些自然词汇的强相关性。这些隐藏信息从有限的图像库中是无法获得的。

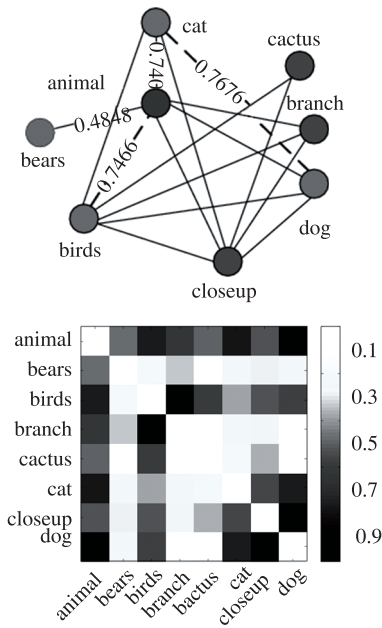


图 2 随机点积图与邻接矩阵的一部分

Fig. 2 A part of random dot product graph and adjacent matrix

3 基于随机点积图的图像标注改善

随机游走算法是实现自动图像标注的有效方法之一,采用带重启的随机游走算法(RWR)^[6]对候选标注词进行最终筛选。对 k 个候选标注 w_1, \dots, w_k 的语义关系图 $G(V, E)$ 随机点积化,可以获得完整的语义关系随机点积图 $G_X(V, P)$ 。在此基础上,将 $G_X(V, P)$ 的边概率矩阵 P 进行归一化处理就可以得到随机游走中的转移概率矩阵

$$TP_{ij} = p(w_i | w_j, I) = \frac{P_{ij}}{\sum_{l=1}^k P_{il}} \quad (12)$$

此时 $TP_{ij} > 0$, $\sum_{j=1}^k TP_{ij} = 1$ 。 $p(w_i | w_j, I)$ 表示对于图像 I ,若已标注有词汇 w_j ,则还可能具有标注词 w_i 的概率。带重启的随机游走算法迭代公式为

$$\pi^t = (1 - c)TP\pi^{t-1} + ce \quad (13)$$

式中, e 是在标注阶段所确定的候选标注置信分数,对其进行归一化操作,使得总和为 1,将此组归一化的数据作为重启向量。 c 为随机游走的重启概率,取为 0.3。 π^t 为 t 时刻图像 I 被标注为 w_i 的概率,即 $\pi_i^t = p^{(t)}(w_i | I)$ 。通过迭代最终求得游走停止时图像 I 具有各标注词的概率 π ,对 π 按从大到小排序,取前 m 个大的为最后标注。具体算法

如下:

算法 2 基于随机点积图的标注改善算法(RD-PG)

输入:待标注图像 I 。

输出:改善后的图像标注 w_1, \dots, w_m 。

1) 利用自动图像标注技术产生候选语义;

2) 保留前 k 个标注词及其置信分数并对其归一化,记为 $p_0(w_i)$;

3) 利用算法 1 求得描述 k 个候选标注词语义关系的随机点积图 $G_X(V, P)$

4) 用式(12)得到转移概率矩阵 TP ;采用重启式随机游走对候选语义进行改善:在 t 时刻得到的概率为

$$p^{(t+1)}(w_i | I) = (1 - c) \sum_{j=1}^k p(w_i | w_j, I)p^{(t)}(w_j | I) + cp_0(w_i)$$

迭代直到收敛

$$p(w_i | I) = \lim_{t \rightarrow \infty} p^{(t)}(w_i | I) \quad (14)$$

5) 对 $p(w_i | I)$ 按降序排列,取前 m 个标注词作为最终标注。

4 实验结果与分析

4.1 实验数据集和评价标准

为充分验证本文方法在图像标注改善中的有效性,采用 3 个常用图像库进行实验,具体情况见表 1,其中 Corel5k 与 MSRC 图像库均包含有大量高分辨率图像,而 TRECVID 图像库则主要是新闻视频中截取的镜头图片^[13]。

表 1 实验中使用的 3 个图像集

Table 1 Three Data Sets Used in Our Experiments

	MSRC	TRECVID	Corel5k
图像数目	591	61 901	5 188
标注词数目	23	39	374

分别将每个图像库等分成 5 份,进行交叉检测实验,对任一待检测的图像,从同一图像库的标注词中随机挑选未标注到此图像的词汇作为噪声词,将噪声词与正确的人工标注都作为待检测图像的候选标注,其中噪声词与人工标注词数量相等。

为评价改善结果,统计所有测试图像中正确词汇被成功挑选出来的总和,计算出宏查准率(MP)

和宏查全率(MR),并获得总体的宏 F 值(MF):

$$MF = 2 \frac{MP \times MR}{MP + MR} \quad (15)$$

由于一幅图像会被标注为多个词汇,因而上述宏 F 值的计算中有许多重复记数^[14]。针对这一问题,对每一词汇 w_i 分别计算在测试图像集中的标注效果,即微查准率(mP_i)和微查全率(mR_i)。将微 F 值(mF_i)作为对该词汇标注情况的综合评价标准:

$$mF_i = 2 \frac{mP_i \times mR_i}{mP_i + mR_i}$$

对所有词汇求平均得到微平均 F 值(mAF):

$$mAF = \frac{\sum F_i}{K} \quad (16)$$

宏 F 值虽然包含了对正确标注的重复记数,但它反映了标注改善方法对测试图像库整体的改善效果。而微平均 F 值则更适用于评价改善方法对每个具体词汇的标注准确度。在实验中,宏 F 值往往大于微平均 F 值。本文将微平均 F 值与宏 F 值都作为评价标准,同时考虑算法对图像库整体与标注词个体的标注改善效果。

4.2 标注改善效果的对比实验与时间复杂度分析

比较 4 种图像标注改善算法,即基于 WordNet 的方法(WordNet)、基于内容的图像标注改善算法(CIAR)、基于知识和图算法的图像标注改善方法(KBIAR-MC)以及本文基于随机点积图的改善算法(RDPG)。在交叉检测的基础上,为每幅图像随机选择 3 组噪声词进行实验,再求平均,比较实验结果如表 2 所示。

表 2 3 种算法的标注改善结果比较

Table 2 Comparison of the 3 different algorithms

图像库	评判标准及时间	WordNet	CIAR	KBIAR-MC	RDPG
MSRC	MF	0.350	0.629	0.476	0.656
	mAF	0.341	0.606	0.421	0.621
	Time/s	0.74	3.32	8.83	3.19
TREC VID	MF	0.349	0.682	0.541	0.727
	mAF	0.341	0.539	0.443	0.571
	Time/s	1.42	4.07	10.66	4.25
Corel5k	MF	0.354	0.714	0.486	0.784
	mAF	0.347	0.674	0.462	0.743
	Time/s	1.05	3.24	9.49	3.57

从实验结果可以看出,本文算法 RDPG 在微平均 F 值与宏 F 值两种标准上都要优于其他 3 种算

法,这主要是因为:1) RDPG 算法在构造标注词语义关系图时考虑了待查询图像的底层特征相似度而非简单的词汇共现性,而 WordNet 描述词汇在语法句法中的相关性,并不能准确反映图像中的语义关系,且词汇量有限,因而标注改善结果最差;2) CIAR 方法依赖于训练集中的语义“共现”关系,当进行交叉检测时,训练图像集并不完整,随机划分往往使其丢失了很多反映词汇关系的关键图像。RDPG 算法利用随机点积图挖掘出了图像集中缺失的标注词语义关系,因而改善效果比 CIAR 方法要好。这在图像库较小(交叉检测分块较大)时尤其明显;3) KBIAR 方法用图的最大割算法将候选语义二分为相关与噪声。由于噪声词汇之间缺乏相关性,通过图的二分很难准确地将噪声词汇作为一个整体与正确词汇分割开。与之相比,本文在重构的随机点积图基础上采用随机游走获得置信分数的策略显然更加合理。

基于 WordNet 的方法本质上就是一种多数派投票方法,时间复杂度为 $O(n)$,耗时最少但精度最差。KBIAR 方法采用图的最大割对语义图进行划分,图的最大割是非确定性多项式时间难题(NP-hard),其近似解法的时间复杂度为 $O(2^n)$,而不精确的简化算法复杂度依然高达 $O(cn^{1/2} \times (\ln(n) + \ln(1/\epsilon))n^3)$, c 为迭代次数, ϵ 为参数。从表 2 的实验结果中也可看出 KBIAR 方法十分耗时。CIAR 构造的是 $n+1$ 阶转移矩阵进行重启随机游走算法,其复杂度为 $O(c(n+1)^3)$,而本文方法先构造随机点积图完善语义关系,然后再用 n 阶转移矩阵进行随机游走。每次对观测图随机点积化的复杂度也为 $O(n^3)$,所以最终的时间复杂度为 $O((c_1 + c_2)n^3)$ 。在算法 1 中迭代次数一般只有 3 次,所以基于随机点积图的方法在时间复杂度上与 CIAR 相差不多。

表 3 给出了实验的部分结果,从中可以看出,在施加噪声词的影响下,本文算法在 3 个图像库上的改善结果都很准确。

4.3 随机点积图对语义关系图重构效果

本文的最主要工作就是利用随机点积图对语义关系图的随机点积化重构,从中挖掘出丢失掉的语义关系。为验证随机点积重构的鲁棒性与准确性,设计了以下实验:

首先用式(10)对 Corel5k 中的 374 个词汇构造语义关系图 $G(V, E)$,然后从现有的共现关系(边

表 3 RDPG 方法图像标注改善结果举例

Table 3 Examples of image annotation refinement by RDPG

样图	人工标注	噪声词	改善结果
	Water Boat Sky Mount	Road Building Bird Grass	Water Boat Sky Mount
	Ocean Coral Water	People Bears Island	Water Ocean Coral Island
	Verkehr Highway Grass	Tree Car River	Verkehr Highway Grass Car

集 E 中随机选取 30%、40%、60%、80% 的边构成子图 $G_{sub}(V, E)$ 。利用随机点积图对这些子图进行重构:也就是假设某一部分的边未知,用剩下的边通过随机点积图的方法来进行挖掘。计算真实的相似度 $a_{ij} = sim(w_i, w_j)$ 与重构的边概率 $b_{ij} = x_i \cdot x_j$ 之差

$$\Delta sim = a_{ij} - b_{ij} \quad (17)$$

当 Δsim 小于某一阈值 τ 时,则认为该边在重构的随机点积图中被探测出。对每组比例各随机挑选 10 次进行实验,统计原图中存在但未被挑选的边被探测出的比例的平均值,结果如图 3。

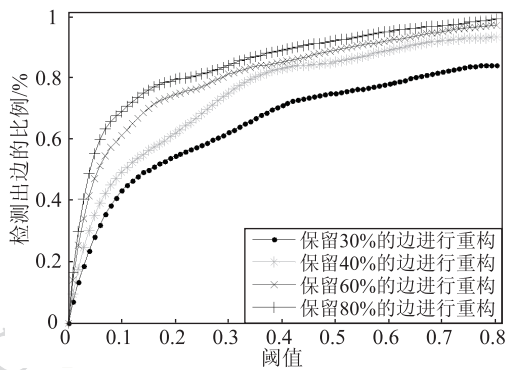


图 3 语义关系的随机点积重构中边的探测情况

Fig. 3 Proportion of edges found in the semantic graph reconstructed by random dot product graph

差值大于阈值的边是应该存在的,从图 3 中可知,即使仅有 30% 的边参与重构,随着阈值变大,挖掘出的正确边数目也不断增加;曲线初始时上升比较明显,说明在取较小差值情况下就可以发现大量丢失的边;整个重构效果随着参与重构的边数目增加而显著提高。

图 4 是对 Corel5k 数据库中所有 374 种词汇的语

义关系图随机点积化重构结果。图 4(a)是从图像库训练得到的语义关系图的邻接矩阵,图 4(b)是随机点积化重构后的结果。颜色由深到浅表示对应的相似度数据值由大到小。从图 4 中可以明显看出,重构之前有很多空白,说明存在着大量的语义关系没有在图像库中反映出来,而重构后的关系明显完整。

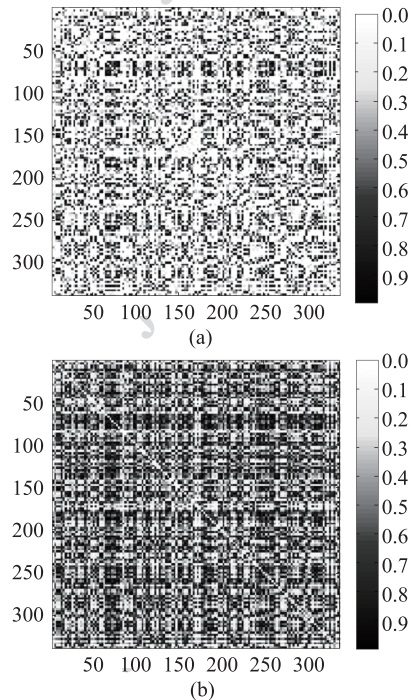


图 4 Corel5k 的语义关系图的随机点积化重构

Fig. 4 Reconstruction of semantic graph of Corel5k by RDPG

4.4 随机赋值参数 d 对结果的影响

参数 d 表示随机点积化过程中对节点赋值的向量维数。显然,不同维数的赋值向量直接影响到最终重构后边概率的计算。本实验将验证该参数对于

标注结果的影响。再次采用前述 3 种图像库,分别在 $d=2, d=5, d=8, d=10$ 这 4 种情况下进行 RD-PG 实验,实验结果如图 5 所示。

从图 5 中可见,当赋值 x 维数较低时,并不能完全拟合真实的语义关系,改善精度不高;随着维数 d 升高,改善精度逐渐升高;当维数 d 过高时,赋值 x 包含的噪声分量也会增多,又降低了改善精度。因此取 $d=5$,此时 F 值基本处于峰值。

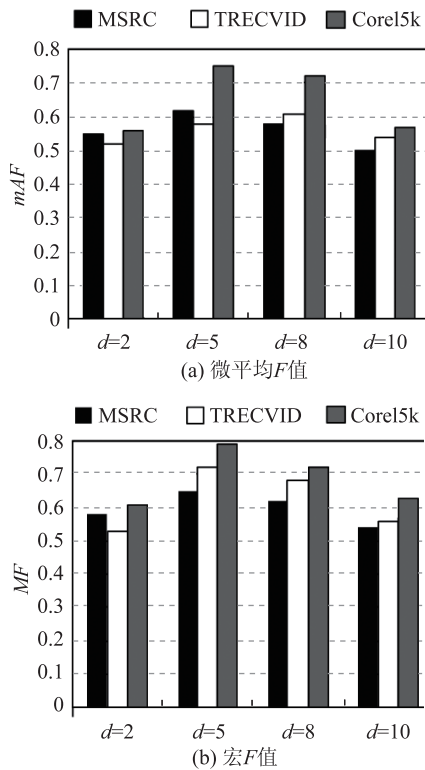


图 5 参数 d 对标注结果的影响

Fig. 5 The influence of parameter d to annotation results

4.5 图像库规模对结果的影响

基于随机点积图的图像标注改善方法可以从不完整的图像库中得到较好的标注结果,这是该方法最大的优势。为验证这一结论,设计在不同图像集规模下的标注改善实验。由于 Core5k 数据库相对较为完整,语义种类较丰富,因此本实验仅采用该图像库进行。从 Core5k 图像库的每种场景类中各抽取相同比例的部分图像构成训练图像库,将剩下的图像作为待查询图像,添加人工噪声,再用前述 4 种方法进行标注改善比较实验。实验结果如图 6 所示。

图 6 中左侧为微平均 F 值,右侧为宏 F 值。从图中可见,基于随机点积图的方法在训练图像规模在 40%~80% 的范围内,改善效果都明显高于其他

方法,当图像集的规模变大时,实验结果有显著提升;在仅有 50% 的训练图像时,RDPG 方法仍然可以达到 0.6 左右的 F 值,而 CIAR 方法完全依赖于从图像库训练得到的词汇“共现”关系,使得其仅在 80% 训练图像时才有 0.65 左右的 F 值,训练图像规模减少到 70% 时, F 值就迅速下降到了 0.4 左右。这说明 RDPG 方法可以有效降低训练图像集规模变小对标注的影响:即使仅有 50% 的训练图像,RDPG 方法依然可以达到较好的改善效果。

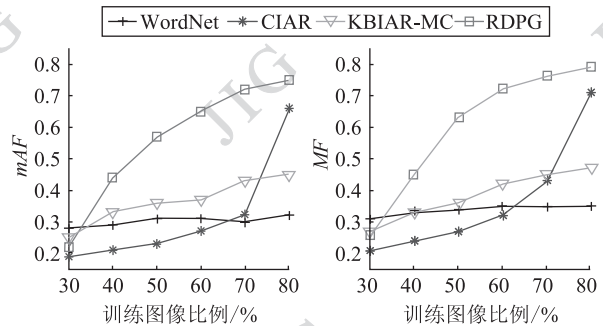


图 6 图像库规模对标注结果的影响

Fig. 6 The influence of the scale of database to annotation results

此外,基于 WordNet 的方法利用 WordNet 描述语义关系,在不同的训练图像规模下 F 值都只在 0.3 左右。KBIAR 方法除 WordNet 之外还整合了其他几种语义相关性度量,改善效果随训练图像规模变大而缓慢提高。

5 结 论

由于底层特征和高层语义间“鸿沟”的存在,现有自动图像标注方法的性能还不能令人满意,因此有必要对已有的标注结果进行改善,从而提高标注的准确性。提出一种新的图像标注改善算法:同时考虑图像的底层特征与高层语义建立针对待查询图像的语义关系图,在此基础上利用随机点积重构,挖掘出图像库中丢失的语义关系,最终结合重启式随机游走算法实现图像标注改善。在 3 种常用图像库中的多项实验结果表明,该算法使得标注结果的准确性得到显著提高,尤其在训练图像集规模较小时,具有明显优势。

参考文献 (References)

[1] Lu H Q, Liu J. Image annotation based on graph learning [J].

- Chinese Journal of Computers, 2008, 31(9): 1629-1639. [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.]
- [2] Rui X G, Yuan P B, He F, et al. A new approach for automatic image annotation based on semantic clustering and graph algorithm [J]. Journal of Image and Graphics, 2007, 12(2): 239-244. [芮晓光, 袁平波, 何芳, 等. 一种新的基于语义聚类 and 图算法的自动图像标注方法[J]. 中国图象图形学报, 2007, 12(2): 239-244.]
- [3] Guo Y T, Luo B. Image semantic annotation method based on multi-modal relational graph [J]. Journal of Computer Applications, 2010, 30(12): 3295-3397. [郭玉堂, 罗斌. 基于多模态关联图的图像语义标注方法[J]. 计算机应用, 2010, 30(12): 3295-3397.]
- [4] Jin Y, Khan L, Wang L, et al. Image annotations by combining multiple evidence & WordNet[C] //Proceedings of the 13th annual ACM International Conference on Multimedia. New York: ACM, 2005: 706-715.
- [5] Jin Y, Khan L, Prabhakaran B. Knowledge based image annotation refinement[J]. Journal of Signal Processing Systems, 2010, 58(3): 387-406.
- [6] Wang C H, Jing F, Zhang L, et al. Image annotation refinement using random walk with restarts[C] //Proceeding of the 14th Annual ACM International Conference on Multimedia. New York: ACM, 2006: 647-650.
- [7] Wang C H, Jing F, Zhang L, et al. Content-based image annotation refinement[C] //Proceeding of IEEE Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2007: 1-8.
- [8] Young S, Scheinerman E. Random dot product graph models for social networks[C] //Proceeding of the 5th Workshop on Algorithms and Models for The Web-Graph. San Diego: University of California, 2007: 138-149.
- [9] Young S. Random dot product graphs a flexible model for complex networks [D]. Baltimore, Maryland: Johns Hopkins University, 2008.
- [10] Zhang D M, Sun D D, Fu M S, et al. Extended dot product representations of graphs with application to radar image segmentation[J]. Optical Engineering, 2010, 49(11): 117201-117210.
- [11] Marchette D, Priebe C. Predicting unobserved links in incompletely observed networks[J]. Computational Statistics & Data Analysis, 2008, 52(3): 1373-1386.
- [12] Scheinerman E, Tucker K. Modeling graphs using dot product representations[J]. Computational Statistics, 2010, 25(1): 1-16.
- [13] Wang H, Ding C, Huang H. Multi-label linear discriminant analysis[C] //Proceeding of the 11th European Conference on Computer Vision. Berlin: Springer-Verlag, 2010: 126-139.
- [14] Olson D, David L. Advanced Data Mining Techniques[M]. Berlin: Springer -Verlag, 2008: 137-147.