

基于社群隐含主题挖掘和多社群 信息融合的自动图像标注

陈 焯 邵 健 朱 科

(浙江大学计算机科学与技术学院, 杭州 310027)

摘 要 在 Flickr 图像共享网站上,大量无标签或者缺少标签的图像往往会因为标签信息的不完整,以致无法被有效地利用和检索。为了有效地进行图像检索,从 Flickr 用户经常会根据上传图像所隐含的主题而将其推荐到多个相关社群的特点出发,提出了一种新颖的基于社群隐含主题挖掘和多社群信息融合的自动图像标注算法。与传统的自动图像标注方法不同,该算法首先采用隐 Dirichlet 分配模型(latent Dirichlet allocation, LDA)对单个社群里的隐含主题(topic)进行挖掘,并利用隐含主题对由相似图像标签传播产生的初始“噪音”标签进行过滤;然后对同属于多个社群的图像,通过多社群信息融合来生成最终标注结果。实验结果显示了该新算法的有效性。

关键词 自动图像标注 社群 潜在主题挖掘 隐 Dirichlet 分配模型 多社群信息融合

中图法分类号: TP391.41 **文献标志码:** A **文章编号:** 1006-8961(2010)06-944-07

Automatic Image Annotation Using Social Group Latent Topic Mining and Multi-group Information Fusion

CHEN Ye, SHAO Jian, ZHU Ke

(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027)

Abstract At photo sharing websites like Flickr, a lot of images can not be effectively used and retrieved due to lack of tags. In order to retrieve images effectively, this paper presents a novel social group latent topic mining and multi-group information fusion based automatic image annotation algorithm by exploiting the property that users in Flickr often recommend their uploaded pictures to associated social groups according to the hidden topics in each picture. Different from traditional automatic image annotation methods, this algorithm first adopts the latent Dirichlet allocation model to mine the latent topics in single social group and makes use of the hidden topics to filter initial noisy tags generated by tag propagation among similar images, then utilizes multi-group information fusion to generate the final annotations for images simultaneously belonging to multiple social groups. Experimental results show the effectiveness of this algorithm.

Keywords automatic image annotation, social group, latent topic mining, latent Dirichlet allocation, multi-group information fusion

0 引 言

随着网络 and 多媒体技术的迅猛发展,互联网上的图像数量呈爆炸性增长。据统计,截至 2008 年,

Google 索引的 Web 网页规模达到 1 万亿,其中图像数据超过几十亿张。近年来,共享网络引起了互联网用户的特别关注,在提供数字图像共享的大众标注网站——Flickr^[1]上,其索引的图像已超过 30 亿张,并以每月几百万张的速度快速增长。

基金项目:国家自然科学基金项目(60833006);国家高技术研究发展计划(863)项目(2006AA010107);浙江省科技计划项目重点项目(2008C13G1410001)

收稿日期:2009-06-01; **改回日期:**2009-08-30

第一作者简介:陈 焯(1985—),男,硕士生。研究领域为多媒体分析与检索。E-mail: cyzeats@163.com

通讯作者:邵 健。E-mail: jshao@cs.zju.edu.cn

虽然互联网用户为 Flickr 图像手工添加的标签信息为图像的高效管理和检索带来了极大便利,但是, Sigurbjörnsson 等人对 Flickr 图像手工标注的结果进行了深入分析发现,64% 图像的标签均少于或等于 3 个^[2]。如何对大量无标签或者标签不足的图像自动添加标签或增强已有标签是当前研究的热点问题。

与普通图像不同,互联网共享图像具有如下几个特点:

1) 共享网络图像质量参差不齐,由不同用户通过不同相机在不同时间从不同的角度或使用不同拍照技巧拍摄得到;

2) 共享网络图像内容丰富,Flickr 图像的标签词条已超过了 1.3 亿条,它涵盖了 6 千多万个概念,囊括了风景、建筑、人物肖像、活动剪辑等多样的内容、事件或对象;

3) 共享网络图像语义复杂^[3],不仅有视觉信息,也有文本信息,而且一张图像往往同时包含多个主题信息,比如一张图像既包含了“天空”、“云彩”等主题信息,同时也包含了“水”、“河流”等主题信息;

4) 共享网络图像多数是弱标注的,其原因在于共享网络图像标签具有稀疏性以及含有很多噪音标签。

由于共享网络图像具有上述特点,因此难以使用传统算法对其进行有效标注。深入分析 Flickr 网

络上的共享图像可发现一个显著特征:当用户根据时间、地点或者事件将图像上传到个人相册后,会进一步根据图像主题将其推荐到相应社群中去。Flickr 网络中的社群是指包含某一特定主题的图像集合,当用户上传不符合社群主题的图像时,管理员会删除这些无关图像,这就保证了社群图像主题上的一致性。因此,可利用图像所在社群的主题信息对图像进行标注。同时,针对某一社群主题又可进一步细分为多个子主题(sub-topic)的事实,可先对社群进行隐含主题挖掘,再考虑结合图像视觉相似性生成更加精细的标注结果。

笔者在 Flickr 网络上随机选取 100 个用户,并选取每个用户所上传图像中的 10 张图像进行分析发现,平均每张图像被推荐到 5 个社群,这说明利用社群信息来提高图像标注结果是可行的。

综上,本文提出了一种新颖的基于社群隐含主题挖掘和多社群信息融合的算法,用于自动图像标注。该算法的实现框架如图 1 所示,具体可分两步。对于同属于多个社群的待标注图像,首先采用隐 Dirichlet 分配模型(latent Dirichlet allocation, LDA)^[4-5]对单个社群中的隐含主题(latent topic)进行挖掘,并根据候选单词与社群隐含主题之间的相关性对由相似图像标签传播产生的初始候选标注单词进行过滤;然后通过多社群的主题信息融合来产生最终标注结果。

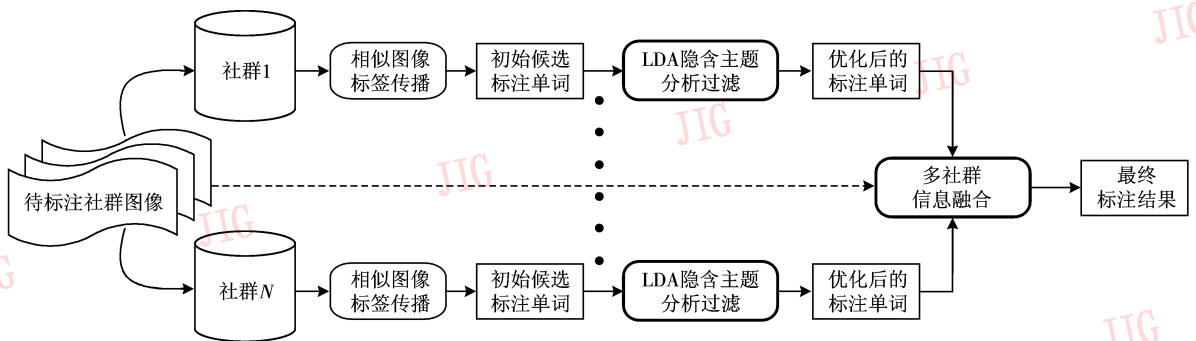


图 1 基于社群隐含主题挖掘和多社群信息融合的自动图像标注框架

Fig. 1 Social group latent topic mining and multi-group information fusion based automatic image annotation framework

1 相关工作

图像自动标注方法可分为翻译模型(translation model)、分类模型(classification model)、相关模型(relevance model)以及生成模型(generative model)

等几种。Duygulu 等人提出的翻译模型^[6]是通过寻找标注单词与图像特征之间的对应关系,使每一个单词都被翻译成对应的图像子块(Blob)来实现图像标注,其难点在于如何定义图像子块和单词的对应关系。而分类模型则是将每个单词都当作一个独立类别用来训练相应分类器,主要方法有支持向量

机^[7]和贝叶斯分类器^[8]等。在翻译模型和分类模型中,都需要一个训练集用来为每个单词学习相应的图像子块或者分类器。但是,由于社会共享网络图像的弱标注性,使它不合适作为一个好的训练集用来训练分类器或者建立相应的翻译模型,因此翻译模型和分类模型不适合用于大规模社会共享网络图像的自动标注。而相关模型则是利用相似图像共同出现的标签单词来标注图像,主要方法有双向跨媒体相关模型(dual cross media relevance model)^[9]、多伯努利相关模型(multiple Bernoulli relevance model)^[10]等。多伯努利相关模型是首先将图像切分为规则的矩形用来取代复杂的区域分割算法,同时通过引入多伯努利分布取代多项式分布来刻画单词的概率分布。基于相关模型标注的效果往往容易受训练图像中的“噪音”标签影响。生成模型则假设在可观测数据之外,还存在一些不可观测的隐含变量,这些隐含变量往往对应一定的语义主题,是表达一定主题的图像和文本等数据“聚类”在一起的原因,生成模型的主要方法有隐含语义分析(latent semantic analysis, LSA), 概率隐含语义分析(probabilistic latent semantic analysis, pLSA)^[11]以及隐 Dirichlet 分配模型(LDA)。

本文提出的基于社群隐含主题挖掘和多社群信息融合的图像自动标注算法,主要是先利用相关模型的标签传播来产生初始候选标注单词,然后使用 LDA 生成模型挖掘图像隐含主题来对候选标注单词进行“去噪”,并进一步通过社会网络中多社群的主题信息融合来产生最终的标注结果。

随着 Web 2.0 的应用,社会网络信息被越来越多地应用到图像标注中。Stone 等人利用用户在 Facebook 中的好友信息对 Facebook 上的图像进行人名标注^[12],取得了较好的结果,但其只局限于人名标注。Lindstaedt 等人利用单个社群信息来对社群中的图像进行标注^[13],不过文献[13]的方法先人为地将图像分为若干类,且只能对有限类别图像进行标注,不适合大规模互联网共享图像的自动标注。Cao 等人提出了利用个人相片集进行图像标注的算法^[14],其出发点在于假设同一个相片集里的图像属于同一上下文(Context),因此可以利用这些上下文信息来实现图像所蕴含事件和场景的标注。该方法的局限性在于其标注信息只限于人为制定的几个事件和场景,且未能充分利用互联网共享图像中社群等社会网络信息。近年来,基于搜索技术的

图像标注^[15-16],即通过原始标注词查询 Google 等图像搜索引擎来获取与原始标注词相关的图像,并从其伴随文本中抽取新词作为扩展标注词集,最后结合原始标注词集及原始图像、扩展标注词集来进行图像标注的技术已成为一个热点,但是该方法仍然将图像标注看成是与用户无关的过程,未能有效利用社会网络的属性。

2 基于 LDA 的社群隐含主题挖掘

2.1 LDA 模型

LDA 模型主要用来对文本进行主题分析,相比于隐含语义分析和概率隐含语义分析等模型,LDA^[4-5]具有如下优点:

- 1) LDA 模型是全概率生成模型,具有清晰内在结构,可通过概率推断来进行高效计算;
- 2) LDA 模型的参数空间规模固定,与训练文档数目无关。

由于 Flickr 网络上的社群图像都是基于一个或者几个相近主题而成组显示的,因此通过 LDA 模型对社群图像的标注单词(标签)进行分析,即可得到社群图像的相关隐含主题信息以及每个隐含主题的相关单词信息。当某个单词被用来进行图像标注时,与图像隐含主题相关的单词则更有可能准确地标注图像信息,而与图像隐含主题无关的单词则更有可能是“噪音”单词。因此,挖掘图像标注单词的隐含主题的相关信息可以有效地去除图像标注中的“噪音”单词,以便改进图像标注的结果。如图 2 所示,LDA 模型中图像 D 、隐含主题 Z 、单词 W 之间的关系主

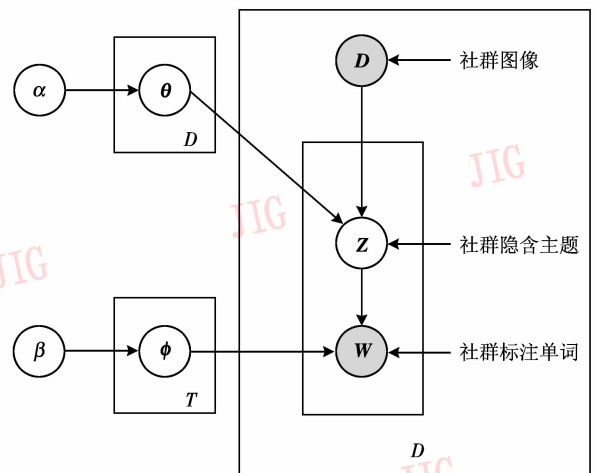


图 2 社群隐含主题 LDA 模型

Fig. 2 Social group latent topic LDA model

要由隐含变量 θ 和 ϕ 决定,其中 θ 表示图像 D 的主题分布, ϕ 为主题 Z 的单词分布, α, β 为隐含变量 θ, ϕ 的先验概率。先验概率 α, β 服从 Dirichlet 分布, T 为社群主题总数, D 为社群图像的总数。

2.2 Gibbs 采样

由于直接计算图像集中的隐含主题 Z 与图像 D 和单词 W 之间的概率比较复杂,所以 Griffiths 等人采用了 Gibbs 采样来简化 LDA 模型计算^[5]。对于第 i 个标注单词 token,该 token 所对应单词为 w_i ,token 所对应的图像为 d_i ,Gibbs 采样需轮流地考虑每一个单词 token,并通过计算其他 token 被赋给每一个主题的次数来估计当前 token 被赋给哪一个主题。在这个过程中,主题被循环采样,主题条件概率为

$$P(z_i = j | z_{-i}, w_i, d_i, \cdot) \propto \frac{c_{w_i, j}^{WT} + \beta}{\sum_{w=1}^W c_{w, j}^{WT} + W\beta} \cdot \frac{c_{d_i, j}^{DT} + \alpha}{\sum_{t=1}^T c_{d_i, t}^{DT} + T\alpha} \quad (1)$$

其中, $z_i = j$ 表示主题 j 被赋值给 token i , z_{-i} 表示除了 token i 之外其他单词 token 的主题分配,“ \cdot ”表示其他的所有已知信息,比如所有其他单词 w_{-i} ,其他图像 d_{-i} 以及先验参数 α, β 。 C^{WT} 与 C^{DT} 分别是大小为 $W \times T, D \times T$ 维的矩阵。 $c_{w_i, j}^{WT}$ 表示第 i 个单词 w_i 被赋给主题 j 的次数, $c_{d_i, j}^{DT}$ 表示第 i 个图像 d_i 里面的单词被赋给主题 j 的次数(不包括当前单词 token i)。

在式(1)里,影响当前单词 token 赋予某个主题的因素主要有以下两方面:第 i 个单词 w_i 属于主题 j 的概率,主题 j 在当前图像主题分布中的概率。若单词 w_i 被分配给主题 j 的次数越多以及单词所对应图像被分配给主题 j 的次数越多,则单词 token 被分配给主题 j 的概率也越大。

在每次 Gibbs 采样中,图像集里所有单词都被赋给某一个主题。当 Gibbs 采样被迭代足够次数后,主题概率就逼近于先验 Dirichlet 分布。Gibbs 采样结束之后,就得到了需要求解的单词-主题分布 ϕ 以及主题-图像 θ 分布,其求解迭代过程如下:

$$\phi_i^j = \frac{c_{w_i, j}^{WT} + \beta}{\sum_{t=1}^W c_{w_i, t}^{WT} + W\beta} \quad \theta_j^d = \frac{c_{d_i, j}^{DT} + \alpha}{\sum_{t=1}^T c_{d_i, t}^{DT} + T\alpha} \quad (2)$$

当得到单词-主题的概率分布 ϕ 之后,即可通过 ϕ 来计算以下单词 w_k 和 w_l 之间的主题相关度:

$$P(w_k | w_l) = \sum_{j=1}^T P(w_k | z_i = j) P(z_i = j | w_l) = \sum_{j=1}^T \phi_k^j \phi_l^j \quad (3)$$

3 基于社群隐含主题挖掘和多社群信息融合的图像标注算法

3.1 单个社群下基于相似图像标签传播的图像标注

给定社群中一张待标注图像 I_u ,则 I_u 和第 i 个标注单词 w_i 之间的联合概率可通过最大化如下公式得到:

$$w_i^* = \arg \max_{w_i \in V} \{P(w_i | I_u)\} = \arg \max_{w_i \in V} \left\{ \sum_{J \in T} P(w_i, I_u | J) \right\} = \arg \max_{w_i \in V} \left\{ \sum_{J \in T} P(w_i | J) P(I_u | J) \right\} \quad (4)$$

其中, V 表示社群的单词集, T 表示社群的训练数据集, $P(w_i | J)$ 表示图像 J 中单词 w_i 出现的次数, $P(I_u | J)$ 表示待标注图像 I_u 和已知图像 J 之间的视觉相似性。本文算法先选取训练集 T 中与待标注图像 I_u 视觉相似性最高的 10 幅图像中的标签作为 I_u 的候选标注单词,然后将通过相似图像标签传播得到的 $P(w_i | I_u)$ 值最大的 10 个单词作为候选标注单词^[17]。

对于每幅图像,提取它的 423 维视觉特征分别是 256 维的颜色直方图(color histogram)、6 维颜色矩(color moments)、128 维颜色相关图(color coherence)、15 维 MSRSAR 纹理、10 维 Tamura 粗糙度纹理和 8 维 Tamura 方向度纹理,然后通过余弦距离来计算图像之间的相似性。

3.2 单个社群下基于 LDA 隐含主题挖掘的图像标注

由于训练集中存在“噪音”,从而导致候选标注单词中存在一些相互之间主题无关的单词。本文认为,主题相关的单词被标注到同一张图像的概率会更大。因此,对于待标注图像所在的社群,可对它进行 LDA 隐含主题分析,先得到单词 w_i 和 w_j 之间的主题相似性 $P(w_j | w_i)$ (见式(3));接着,在计算了候选单词 w_i 和待标注图像 I_u 的标注概率 $P(w_i | I_u)$ (见式(4))之后,再按式(5)通过结合 LDA 分析得到的单词 w_i 与 w_j 之间的主题相似性 $P(w_j | w_i)$ 和候选单词 w_i 与待标注图像 I_u 的标注概率 $P(w_i | I_u)$ 来对候选单词 w_i 与待标注图像 I_u 的相关度(relevance score)进行计算:

$$RS(w_i, I_u) = P(w_i | I_u) \times e^{\sum_{j \neq i} P(w_j | w_i)} \quad (5)$$

其中, $\sum_{j \neq i} P(w_j | w_i)$ 表示单词 w_i 与其他所有候选标注单词之间主题相似性之和,即与待标注图像 I_u 的

主题相似性。 $\sum_{j \neq i} P(w_j | w_i)$ 值越大, 单词 w_i 作与待标注图像 I_u 的相关度也越大。结合 LDA 主题分析后得到的候选单词 w_i 和待标注图像 I_u 的相关度 $RS(w_i, I_u)$ 能有效地降低待标注图像与主题无关的候选标注单词的相关度, 进而可达到对标注单词“噪音”进行过滤的目的。候选单词 w_i 和待标注图像 I_u 的相关度越大, 候选单词被用来标注待标注图像的可能性就越大。

3.3 基于多社群信息融合的图像标注

由于图像可能包含不同的主题信息, 而用户则会把包含多个主题的图像同时推荐到不同的社群中去。不同社群包含了不同的主题信息, 这些社群的主题信息可能相互包含, 也可能相互独立, 比如“Beautiful Scenery”社群的主题信息包含了“Sky & Clouds”社群和“Water, Oceans, Lakes, Rivers, Creeks”社群的主题信息, 而“Sky & Clouds”、“Water, Oceans, Lakes, Rivers, Creeks”这两个社群之间的主题则相互独立。通过融合多社群的不同主题的信息进行图像标注, 图像标注结果的信息就会更加广泛和准确。假设待标注图像 I_u 同时属于 K 个不同社群集, 第 k 个社群的候选单词 w_i 和待标注图像 I_u 的相关度为 $RS(w_i, I_u)$, 则多社群的单词候选单词 w_i 和待标注图像 I_u 的相关度为

$$\widehat{RS}(w_i, I_u) = \sum_{k=1}^K RS_k(w_i, I_u) \quad (6)$$

4 实验与分析

4.1 数据集与评测指标

为了测试本文算法, 本文通过图片共享网站 Flickr 提供的 API 进行用户共享图片的下载。本文选择了 3 个社群, 并将 2009 年 5 月之前每月 1 到 3 日期间上传的所有图像作为实验数据集: 社群 1——“Beautiful Scenery” (<http://www.flickr.com/groups/78249294@N00>, 34 585 张)、社群 2——“Sky & Clouds” (<http://www.flickr.com/groups/89594630@N00>, 29 434 张)、社群 3——“Water, Oceans, Lakes, Rivers, Creeks” (<http://www.flickr.com/groups/92767609@N00>, 29 462 张)。3 个社群的图像总和为 93 481 张, 其中 252 张图像同属于其中的 3 个社群。本文先选取了同时属于 3 个社群的 252 张图像中的 100 张图像作为实验的测试集, 然后选取了除了同时属于 3 个社群的 252 张图像以外的

表 1 主题词

Tab.1 Topic words

Sea	Lake	Beach	River	Ocean
Water	Flowers	Grass	Rocks	Trees
Snow	White	Black	Green	Coast
Sand	Island	Boat	Sunset	Sky
Sun	Sunrise	Mountains	Forest	Tree
Blue	Clouds	Landscape	Nature	Park

93 229 张图像作为实验的训练集。

为了评测实验结果, 本文从 3 个社群中选出表 1 所列的 30 个主题词对测试集中的 100 张图像进行了手工标注, 并以手工标注结果为参考计算查准率 (precision) 与查全率 (recall):

$$R_{\text{precision}}(w) = \frac{N_c}{N_s} \quad (7)$$

$$R_{\text{recall}}(w) = \frac{N_c}{N_r} \quad (8)$$

其中, 给定一个查询单词 w 下, N_r 表示人工标注下测试集中查询返回的图像数目, N_s 表示自动标注下查询返回的图像数目, N_c 表示其中正确返回的图像数目 (即 N_s 与 N_r 重叠的部分)。图像检索的查准率和查全率越高, 图像标注的结果也越好。

4.2 单个社群下基于 LDA 隐含主题挖掘的实验

首先, 在单个社群下进行了基于 LDA 隐含主题挖掘的实验。实验中 LDA 参数分别为: $T=100, \alpha=0.1, \beta=0.5$ 。在实验中, 选取与待标注图像相关度最大的 5 个标注单词作为最终的标注结果。

表 2 和图 3 给出了在社群 1、社群 2、社群 3 上单独进行 LDA 隐含主题挖掘的实验结果。

表 2 基于 LDA 隐含主题挖掘的实验结果

Tab.2 Experiment results on LDA based latent topic mining

社群	基线查准率/%	基线查全率/%	LDA 查准率/%	LDA 查全率/%
1	13.3	6.1	12.4	6.0
2	12.6	0.7	15.3	0.7
3	18.5	7.2	19.9	7.2
平均	14.8	4.6	15.9	4.6

从表 2 的实验结果可以看出, 相比于传统的基于相似图像标签传播的相关模型产生候选标注词的基线 (Baseline) 系统, 通过 LDA 隐含主题的分析, 3 个社群的实验结果在查全率保持基本不变的情况

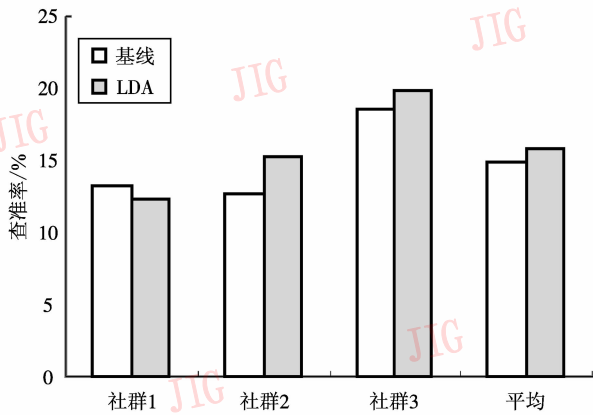


图 3 基于 LDA 隐含主题挖掘的实验结果

Fig. 3 LDA based latent topic mining experiment results

下,其查准率基本得到了改善,平均从 14.8% 提高到 15.9%,相对提高 7.4%。该实验结果表明, LDA 隐含主题分析可有效去除标注单词“噪音”和能提高图像的标注精度。需要说明的是,实验结果中的查准率和查全率相对较低,这主要是由于训练集图像的弱标注性以及采用的评测指标是以并不完整的手工标注结果作为参考的缘故。仔细分析发现,许多自动标注结果虽然与手工标注结果不一致,但事实上与被标注的图像是主题相关的。

4.3 基于多社群信息融合的实验

其次,本文在基于单个社群 LDA 隐含主题挖掘实验的基础上,还进行了多社群信息融合的实验,实验结果如表 3 所示。

表 3 多社群信息融合的实验结果

Tab. 3 Experiment results on multi-group fusion

社群	查准率/%	查全率/%	平均查准率/%	平均查全率/%
1,2	16.4	12		
1,3	15.7	9.6	17.1	10.2
2,3	19.3	9.1		
1,2,3	20.2	16.3	20.2	16.3

表 3 的第 2—4 列给出的是利用 2 个社群进行信息融合的实验结果。其平均查全率从单个社群下的平均 4.6% 提高到 10.2%,相对提高了 121.7%;其平均查准率从 15.9% 提高到 17.1%,相对提高了 7.5%。

最后一列给出的是利用 3 个社群进行信息融合的实验结果。其平均查全率达到 16.3%,比两个社群融合下的 10.2% 相对提高了 59.8%,比单个社群下的 4.6% 相对提高了 254.3%;其平均查准率达到 20.2%,

比两个社群融合下的 17.1% 相对提高了 18.1%,比单个社群下的 15.9% 相对提高了 27.0%。该实验结果表明,通过多社群信息的融合,在显著改进图像标注的覆盖性的同时,可显著提高标注的精度。

图 4 给出了用本文所提算法进行图像标注的一个具体例子。实验时,首先单独在社群 3 “Water, Oceans, Lakes, Rivers, Creeks” 上,通过直接采用相似图像标签传播来产生初始标注结果,更多的偏于 “River, Water” 等主题;接着,通过 LDA 隐含主题挖掘分析, “Antonio” 这个人名单词由于跟社群的主题相关性较小,因此其标注概率被降低了;而单词 “Blue” 则由于与社群的主题相关性更大,从而其标注概率被提高了,这一定程度上实现了 “噪声” 标注单词的消除。进一步地,通过融合社群 2 “Sky & Clouds” 的信息提升了 “Clouds” 和 “Sky” 等单词的概率,通过融合社群 1 “Beautiful Scenery” 的信息提升了 “Mountain” 和 “Landscape” 的概率,最终为一张复杂图像提供了多主题的标注信息。



社群 3	River San Water Antonio Bexar County Courthouse Blue Clouds Sea
社群 3 + LDA	River San Water Bexar Blue Courthouse Antonio County Clouds Sea
社群 3 + 社群 2 + LDA	Clouds River San Sky Water Bexar Blue Courthouse Antonio Count
社群 3 + 社群 2 + 社群 1 + LDA	Sky Blue Clouds River Water San Landscape Bexar Courthouse Mountains

图 4 图像标注例子(单词按照概率从大到小排列)

Fig. 4 Image annotation example

(words sorted by probability from big to small)

5 结 论

本文从 Flickr 网络用户经常会根据上传图像的隐含主题而将其推荐到多个主题相关的社群的特点出发,提出了一种新颖的基于社群隐含主题挖掘和多社群信息融合的自动图像标注算法。该算法首先

采用隐 Dirichlet 分配模型对单个社群里的隐含主题进行挖掘,并根据候选标注词与隐含主题之间相关性对由相似图像标签传播产生的初始“噪音”标签进行过滤;然后,对同属于多个社群的图像,通过多社群信息融合来实现最终标注。

实验是在从 Flickr 网络上下载的 3 个社群中的 93 481 张图像中进行的。实验结果显示,与传统的基于标签传递的相关模型相比较,通过 LDA 对社群隐含主题进行挖掘,在保持查全率基本不变的情况下,查准率相对提高了 7.4%,充分体现了 LDA 在单词“去噪”上的有效性;在此基础上,通过多个社群信息的融合,其查全率和查准率分别相对提高了 254.3%,27.0%,在显著改进图像标注覆盖性的同时,显著提高了标注精度,这充分体现了多社群信息融合的有效性。

本文利用了 Flickr 图像共享网站中 3 个社群的信息来做图像标注,如何更好的有效利用 Flickr 图像共享网站中更多的社群信息来对更多的图像做标注将是下一步的研究重点之一。同时,Flickr 网络中还有用户和图像相册等信息,如何结合用户、相册和社群等信息来提高图像标注的结果也是将来的研究方向。

参考文献 (References)

- [1] Flickr[EB/OL]. <http://www.flickr.com>
- [2] Sigurbjörnsson B, Zhol R. Flickr tag recommendation based on collective knowledge [C]//Proceedings of the International Conference of World Wide Web, Beijing, China: ACM, 2008: 327-336.
- [3] Wu Fei, Liu Yanan, Zhuang Yueting. Transductive multi-modality video semantic concept detection with tensor representation[J]. Journal of Software, 2008, 19 (11): 2853-2868. [吴飞, 刘亚楠, 庄越挺. 基于张量表示的直推式多模态视频语义概念检测[J]. 软件学报, 2008, 9 (11): 2853-2868.]
- [4] Blei D, Ng A, Jordan M. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [5] Griffiths T, Steyvers M. Finding scientific topics [C]//Proceedings of the National Academy of Science, Humlebaek, Denmark: PNAS, 2004, 101: 5228-5235.
- [6] Duygulu P, Barnard K, Freitas J, 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, Copenhagen, Denmark: Springer, 2002: 97-112.
- [7] Claudio C, Gianluigi C, Raimondo S. Image annotation using SVM[C]//Proceeding of Internet Imaging, San Jose, CA, USA: SPIE, 2004, 4: 330-338.
- [8] Chang E, Goh K, Sychay G, et al. CBSA: content-base soft annotation for multimodal image retrieval using Bayes point machines[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(1): 26-38.
- [9] Liu J, Wang B. Dual cross-media relevance model for image annotation[C]//Proceedings of ACM International Conference on Multimedia, Augsburg, Germany: ACM, 2007: 605-614.
- [10] Feng S, Manmatha R, Laverenko V. Multiple Bernoulli relevance models for image and video annotation [C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA: IEEE, 2004: 1002-1009.
- [11] Monay F, GaticaPerez D. On image auto annotation with latent space models[C]//Proceedings of ACM International Conference on Multimedia, Berkeley, CA, USA: ACM, 2003: 275-278.
- [12] Stone Z, Zickler T, Darrell T. Autotagging facebook: Social network context improves photo annotation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA: IEEE, 2008: 23-28.
- [13] Lindstaedt S, Mörzinger R, Sorschag R. Automatic image annotation using visual content and folksonomies[J]. Multimed Tools and Application, 2009, 42(1): 97-113.
- [14] Cao L, Luo J B, Huang T. Annotating photo collections by label propagation[C]//Proceedings of ACM International Conference on Multimedia, Beijing, China: ACM, 2008: 121-129.
- [15] Rui X, Li M, Li Z, et al. Bipartite graph reinforcement model for web image annotation [C]//Proceedings of the 15th International Conference on Multimedia, Augsburg, Germany: ACM, 2007: 25-29.
- [16] Wang X, Zhang L, Li X, et al. Annotating images by mining image search results[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1919-1932.
- [17] Xia D Y, Wu F, Zhang X Q, et al. Local and global approaches of affinity propagation clustering for large scale data[J]. Journal of Zhejiang University-Science A, 2008, 19(10): 1373-1381.