ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.21, No.7, July 2010, pp.1561−1575
doi: 10.3724/SP.J.1001.2010.03704

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

图像-文本相关性挖掘的 Web 图像聚类方法*

吴 飞+, 韩亚洪, 庄越挺, 邵 健

(浙江大学 计算机科学与技术学院,浙江 杭州 310027)

Clustering Web Images by Correlation Mining of Image-Text

WU Fei+, HAN Ya-Hong, ZHUANG Yue-Ting, SHAO Jian

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

+ Corresponding author: E-mail: wufei@zju.edu.cn

Wu F, Han YH, Zhuang YT, Shao J. Clustering Web images by correlation mining of image-text. *Journal of Software*, 2010,21(7):1561−1575. http://www.jos.org.cn/1000-9825/3704.htm

Abstract: To cluster the retrieval results of Web image, a framework for the clustering is proposed in this paper. It explores the surrounding text to mine the correlations between words and images and therefore the correlations are used to improve clustering results. Two kinds of correlations, namely word to image and word to word correlations, are mainly considered. As a standard text process technique, tf-idf method cannot measure the correlation of word to image directly. Therefore, this paper proposes to combine tf-idf method with a feature of word, namely visibility, to infer the correlation of word to image. Through LDA model, it defines a topic relevance function to compute the weights of word to word correlations. Finally, complex graph clustering and spectral co-clustering algorithms are used to testify the effect of introducing visibility and topic relevance into image clustering. Encouraging experimental results are reported in this paper.

Key words: graph clustering; complex graph; visibility; latent Dirichlet allocation; spectral clustering

摘 要: 为了实现 Web 图像检索结果的聚类,提出了一种 Web 图像的图聚类方法.首先定义了两种类型关联:单词与图像结点之间的异构链接以及单词结点之间的同构链接.为了克服传统的 TF-IDF 方法不能直接反映单词与图像之间的语义关联局限性,提出并定义了单词可见度(visibility)这一属性,并将其集成到传统的 tf-idf 模型中以挖掘单词-图像之间关联的权重.根据 LDA(latent Dirichlet allocation)模型,单词-单词之间关联权重通过一个定义的主题相关度函数来计算.最后,应用复杂图聚类和二部图协同谱聚类等算法验证了在图模型上引入两种相关性关联的有效性,达到了改进了 Web 图像聚类性能的目的.

关键词: 图聚类;复杂图;可见度;LDA(latent Dirichlet allocation);谱聚类

中图法分类号: TP311 **文献标识码**: A

* Supported by the National Natural Science Foundation of China under Grant Nos.60603096, 60533090 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA010107 (国家高技术研究发展计划(863)); the Program for Changjiang Scholars and Innovative Research Team in University of China under Grant Nos.IRT0652, PCSIRT (长江学者和创新团队发展计划)

Received 2009-02-27; Revised 2009-05-21; Accepted 2009-07-23; Published online 2010-04-12

在Web上,使用关键字搜索图像仍然是有效的常用检索手段.在Web图像检索中,用户提交的查询关键字往往是视觉多义词(visually polysemous word)[1],这类单词包含多个不同视觉含义.例如,单词 mouse 可表示“computer mouse”,“mouse animal”和“Mickey mouse”等多个主题.因此,用这些视觉多义词查询图像,所返回的图像检索结果会包含多个主题,并且不同主题的图像混合在一起.这就需要在得到初始检索结果后对表达不同主题的图像进行归类等处理.近年来,研究者提出了若干 Web 图像聚类方法[2−8]来解决这个问题.由于图像的底层特征和高层语义之间存在“语义鸿沟”,这些聚类方法往往同时利用了被聚类图像集合所包含的视觉、文本和链接等多模态信息.属于不同特征空间的多模态信息是相互关联的,在模态信息融合学习中挖掘和利用这些相关性关联以正确理解多媒体隐含语义是近期机器学习研究的重点课题,代表性工作有多视角学习(multi-view learning)[9]和迁移学习(transfer learning)[10,11].前者同时利用同一数据的多种特征空间表示进行学习,而后者研究训练数据和测试数据有不同分布或属于不同特征空间的学习问题.本文挖掘文本与图像两种模态信息之间相关性关联,通过图模型对该关联关系进行建模,并利用图聚类算法对 Web 图像进行聚类.

Web 图像通常与其伴随文本共存于 HTML 页面之中,伴随文本以及一些文本标签(textual tags)描述了图像的语义内容.在 Web 图像检索和标注领域,很多研究利用了图像和文本之间的相关性关联.文献[12]中的实验表明:利用图像伴随文本和一些 HTML 标记中的文本对图像检索结果进行重排序可以提高图像检索性能.在图像自动标注领域,文献[13]研究从图像伴随文本中提取关键短语(salient term)来对图像进行标注,文献[14,15]研究如何将文本中的单词关联到图像某个区域,从而对图像进行更准确标注.但是,伴随文本中不同单词对图像语义描述所做贡献不同.对于文本中多个单词,有的单词能够找到合适的图像来形象地描述该单词的含义,例如 chairs;有的单词比较抽象,则很难找到一个合适图像来形象地描述该单词的含义,例如 statistics.从形象思维的角度,这种差异反映了单词和其指代图像之间存在或弱或强的关联,也反映单词具有可见度(visibility)属性.可见度定义为某个单词可以被视觉感知的概率.文献[16]中提出了一个单词可见度模型,并将之应用到文本-图片合成系统(TTP).另外一些研究者[17,18]则将单词可见度属性引入到图像标注领域,以改善图像标注结果.本文提出一种新的更合理的单词可见度模型,并将该模型与传统的 tf-idf 方法结合来定义单词-图像相关性关联.

另一方面,对于包含多个主题的 Web 图像集合,其伴随文本中的隐含主题(latent topic)信息间接地反映了图像间的主题相关性.文献[1,19]通过 LDA(latent Dirichlet allocation)学习发现图像伴随文本中的隐含主题,并依据单词上的隐含主题分布对图像进行聚类和排序.本文引入 LDA 模型挖掘图像伴随文本中的隐含主题,并通过其在各个单词上的边缘概率分布定义单词-单词的主题相关性关联.

因此,本文考虑两种关联关系:单词-图像相关性关联和单词-单词主题相关性关联.这种交叉关联可用图模型进行建模.近来,研究者已开始将图模型引入到 Web 图像聚类研究中.文献[2]利用三部图(tripartite graph)对图像-图像同构底层特征以及图像-文本异构结点间异构链接(heterogeneous links)关系进行建模,并基于图分割(graph cut)方法对 Web 图像进行聚类.与文献[2]模型不同,本文定义的关联关系不仅包含图像-文本间异构链接,也包含单词-单词间同构链接(homogeneous links).因此,本文用更一般的复杂图模型[20,21]对其进行建模,并应用复杂图聚类算法[20]对 Web 图像进行聚类.

基于图像-文本相关性挖掘,本文提出一种新的 Web 图像聚类框架.如图 1 所示,图模型中包含两类结点:图像结点和单词结点.本例中图像是将关键字 jaguar 作为查询提交给 Google 的图像搜索引擎返回结果,这些返回图像主要包含 3 个语义主题:jaguar animal,jaguar car 和 jaguar plane.图中单词来自图像所在网页中伴随文本.实线和虚线分别代表单词-图像以及单词-单词相关性关联.通过单词可见度模型与 tf-idf 方法的结合,高可见度单词与图像间的链接得到加强.在聚类过程中,高可见度单词结点将向与之关联的图像结点传递更多的主题相关性信息,从而提高 Web 图像的聚类性能.

本文第 1 节介绍图像-文本相关性挖掘.第 2 节介绍图聚类方法.第 3 节给出实验,对可见度计算模型进行分析,并对将单词可见度和主题相关度引入 Web 图像聚类的有效性进行验证.第 4 节总结全文.

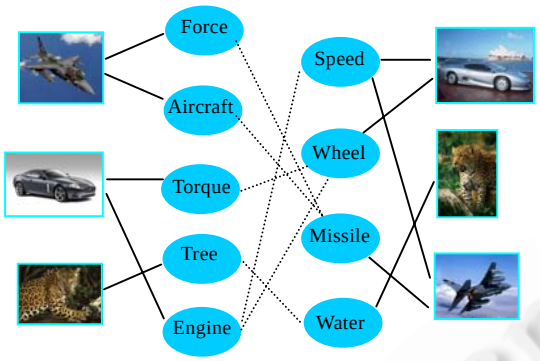


Fig.1 Graph model (two types of nodes: Image nodes & word nodes; two types of links: Heterogeneous links between image and word; homogeneous links (dashed) between words)

图 1 图模型(两种结点:图像结点和单词结点;两种链接:同构链接(虚线)和异构连接(实线))

1 图像-文本相关性挖掘

如图 1 所示,本聚类框架的核心是定义单词-图像和单词-单词这两种相关性关联,前者通过将单词可见度与 tf-idf 方法结合予以实现,后者通过 LDA 学习获得.

1.1 单词的可见度

本节讨论如何挖掘图像伴随文本中的单词和图像之间的相关性,即得到单词和图像两种不同类型结点之间的链接权重.不同的单词对图像语义描述所起到的贡献不同,本文用单词和图像的相关性来衡量这种差别.传统的通过 tf-idf 方法衡量单词对图像的重要性,在一定程度上忽略了图像本身具有的视觉特性,本节所定义的可见度模型弥补了这一不足.

1.1.1 传统方法:tf-idf

基于文本的图像检索方法[12]用查询单词与图像伴随文本之间的相关性代替单词与图像之间的相关性.根据向量空间模型(vector space model),单词在伴随文本中的重要性用该单词的 tf-idf 值来度量.假设单词 w 来自第 i 个图像的伴随文本 d_i 中,单词 w 的 tf-idf 值 $tfidf(w)$计算如下:

$$tfidf(w)=\sum_{i=1}^{N} freq(w,d_i)\cdot\log\frac{N}{num(w)} \tag{1}$$

其中,$freq(w,d_i)$是单词 w 在文档 d_i 中的词频,N 是图像的总数,$num(w)$是伴随文本中含有单词 w 的图像的总数.由于 tf-idf 方法来源于文本处理领域,$tfidf(w)$并不能直接地度量单词和图像之间的相关性.因此,需要进一步挖掘单词和图像之间语义联系.

1.1.2 可见度计算模型

单词的可见度(visibility)体现了单词(尤其是名词)所蕴含语义可用图像来描述的程度.从认知心理学和形象思维的角度,高可见度的单词(如 banana)要比低可见度的单词(如 Bayesian)更易在人脑中形成直接视觉形象.近期一些研究[16−18]已将可见度作为单词一种新的属性,用来表达单词与图像之间的语义关联.在 Web 页面中,图像周围每个单词具有不同程度的可见度,高可见度单词对图像的语义有更强的描述能力.

文献[16–18]分别提出了不同单词的可见度模型,但文献[16]中的可见度模型更加直接和简单,并运用到文本到图像(text-to-picture,简称 TTP)合成系统之中.TTP 系统研究如何自动地为文本寻找语义相关的图像,并将找到的图像自动合成一幅图画来形象地描述文本的语义内容.为了计算文本中单词可见度,文献[16]提出的可见度模型如下:

$$P(w)=\frac{1}{1+\exp(-(-2.78x+15.40))} \tag{2}$$

其中,$P(w)$表示单词 w 的可见度,$x=\log((C_1+10^{-9})/(C_2+10^{-9}))$,$C_1$ 是将单词 w 作为查询提交给 Google 图像搜索引擎返回的搜索结果数目,C_2 是将单词 w 作为查询提交给 Google 文本搜索引擎返回的搜索结果数目.根据式(2),$P(w)$是 x 的增函数.所以对于单词 w,如果在 Google 上得到的 C_1/C_2 值比较大,则 w 可见度就比较高.

通过实验发现,公式(2)给出的可见度计算模型对于某些主题宽泛的词会失效.如图 2 所示,该图像是用关键字 bass 作为查询,由 Google 图像搜索引擎返回的前 5 个结果中的一个.伴随文本中名词的 C_1 和 C_2 值见表 1.如图 3 所示,legend,record,scale 等词 C_1/C_2 值大于 largemouth 和 fishermen.但是根据可见度定义,因为 largemouth 和 fishermen 是这幅图像中两个主要对象,它们应有更高可见度.造成这种结果的原因是,record 等主题较宽泛单词大量地出现在 Web 页面上,同时也大量地出现在图像伴随文本中,从而提高了它们 C_1/C_2 值.因此,有必要对式(2)给出的可见度模型进行改进.由于主题宽泛单词的 C_2 值往往很大,因此可以用逆文档词频来对其可见度进行抑制.

Could this be it? The new *number* 1 for *largemouth* bass *fishermen*? Mike Winn is holding a 25.1-pound bass caught on Dixon Lake in southern California Monday by his *friend* Mac Weakley. By all accounts it smashes the standing *record*, but *questions* (as always) loom. Jed Dickerson, a bass fishing *legend* and *friend* of Weakley's, weighed the big'un with a digital *scale* and then released it.

Fig.2 Image and surrounding text (the italic words are nouns)

图 2 图像和伴随文本(斜体字是名词)

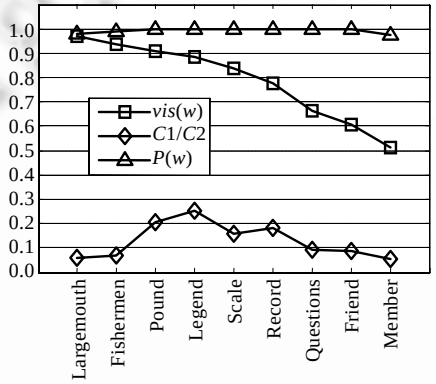


Fig.3 Visibility scores comparison for nouns in Fig.2

图 3 图 2 的伴随文本中名词的可见度对比

Table 1 Google image hit counts C_1 and Google Web hit C_2 counts for the nouns in the surrounding text in Fig.1 (retrieved from Google on Apr. 24, 2009)

表 1 图 1 的伴随文本中名词的 C_1 和 C_2 值(从 Google 上检索得到,2009 年 4 月 24 日)

Word	C_1	C_2
Largemouth	88 800	1 560 000
Fishermen	690 000	9 900 000
Pound	16 000 000	77 300 000
Legend	50 000 000	198 000 000
Scale	35 100 000	223 000 000
Record	110 000 000	606 000 000
Questions	79 400 000	869 000 000
Friend	114 000 000	1 310 000 000
Number	88 700 000	1 660 000 000

基于上述分析,本文提出了式(3)所示的一个可见度模型:

$$vis(w)=\left(\frac{C_1+10^{-9}}{C_2+10^{-9}}\right)^{-IDF_{\text{Google}}(w)} \tag{3}$$

其中,$IDF_{\text{Google}}(w)$定义如下:

$$IDF_{\text{Google}}(w)=\log\frac{|\boldsymbol{D}|}{C_2} \tag{4}$$

其中,$\boldsymbol{D}$ 是 Google 索引的所有 Web 页面集合.图 2 中名词的 $vis(w)$值如图 3 所示.largemouth 和 fishermen 的 $vis(w)$值最大,且不同单词可见度存在明显区分度.可见,改进的可见度模型更加合理.为了进一步验证本模型的合理性和性能,本文将在第 3.1 节对该模型进行详细定量分析.

我们提出将可见度模型集成到传统的 tf-idf 方法中来定义单词-图像的相关性,即单词 w 与图像之间的关联度为 $tfidf(w)\cdot vis(w)$.这样,高可见度单词与图像的相关性得到加强,低可见度单词与图像的相关性减弱.

1.2 单词间的主题相关度

本节介绍如何挖掘单词间的主题相关性,得到单词-单词之间的关联.这里提出通过 LDA 对图像伴随文本进行学习,发现分布在各个单词上的隐含主题,并利用这个概率分布对图像伴随文本中任意两个单词之间的主题相关性进行度量.

1.2.1 LDA 基本理论

给定一个文档集合,LDA 模型[22,23]假设每个文档是多个隐含主题 $z\in\{1,\ldots,K\}$的混合.假设文档集合中有 M 个文档,每个文档包含 N_d 个词,$d\in\{1,\ldots,M\}$.根据 LDA 模型,文档中多个单词由以下生成过程产生[22]:

1) 对于每一个隐含主题 $j=1,\ldots,K$,根据一个狄利克雷(Dirichlet)先验分布 $Dir(\beta)$选择一个分布在单词上的参数为 ξ^j 的多项分布(multinomial distribution);
2) 对于每一个文档 d,根据一个 Dirichlet 先验分布 $Dir(\alpha)$选择一个分布在隐含主题上的参数为 ψ的 Multinomial 分布;
3) 对于产生第 i 个单词,首先根据 Multinomial 分布 ψ确定一个主题 $z=j$,然后根据 Multinomial 分布 ξ^j 产生单词 w_i.

根据以上生成模型,生成一个文档 $d=w_1,\ldots,w_{N_d}$ 的概率定义为

$$P(w_1,\ldots,w_{N_d}\mid\xi,\psi)=\prod_{i=1}^{N_d}\sum_{z=1}^{K}P(w_i\mid z,\xi)P(z\mid\psi) \tag{5}$$

根据贝叶斯(Bayes)定理,隐含主题 $z=j$ 分布在文档 $d=w_1,\ldots,w_{N_d}$ 上的概率 $P(z=j|d)$可计算如下:

$$P(z=j\mid d)=\sum_{i=1}^{N_d}P(z=j\mid w_i)=\sum_{i=1}^{N_d}\frac{P(z=j\mid w_i)P(z=j)}{P(w_i)} \tag{6}$$

1.2.2 主题相关度函数

在式(6)中,$P(z=j|w_i)$是隐含主题 $z=j$ 在文档 $d=w_1,\ldots,w_{N_d}$ 中各个单词 w_i 上边缘分布,可以看作该隐含主题在各个单词上分配的份额.如果某一个单词对于某个隐含主题比较重要,那么该主题在这个单词上的边缘分布就相对比较大.因此,我们可以根据 LDA 模型定义主题相关度函数来计算文档集合中任意两个单词同时属于某一个主题的概率,这个概率值可以衡量两个单词的主题相关度.

定义 1(单词的主题相关度函数). 文档集合中任意两个单词 w_s 和 w_t 之间的主题相关度函数 $Topic_r(w_s,w_t)$定义为

$$\begin{aligned}Topic\_r(w_s,w_t)&=\max_j P(z=j\mid w_s)P(z=j\mid w_t)\\&=\max_j\frac{p(w_s\mid z=j)P(z=j)}{P(w_s)}\cdot\frac{p(w_t\mid z=j)P(z=j)}{P(w_t)}\\&=\max_j\frac{p(w_s\mid z=j)p(w_t\mid z=j)P(z=j)}{\sigma}\end{aligned} \tag{7}$$

在定义 1 中,由于乘积 $P(w_s)\cdot P(w_t)$与某个隐含主题 $z=j$ 无关,我们选择常数σ作为归一化(normalized)常数.

2 图聚类方法

传统的图聚类方法大多处理仅包含单一类型结点的同构结点图(homogeneous-node graph).但是,反映现实世界中对象之间复杂关系的图模型一般是异构结点图(heterogeneous-node graph),即图中包含多种类型结点,例

如科技著作网中包含作者和论文两种类型结点.异构结点图同时包含两种类型的链接关系:同类型结点之间的同构链接(homogeneous links)和不同类型结点之间的异构链接(heterogeneous links).例如,作者之间的合作关系以及论文之间的引用关系是同构链接;作者和论文之间的创作关系是异构链接.显然,同构结点图是异构结点图的一种特例.

如图 1 所示,本聚类框架中的图模型包含两种不同类型结点:图像结点和单词结点以及两种类型链接关系:单词-图像间的异构链接和单词-单词间的同构链接.文献[20]将这种更具一般性的异构结点图定义为复杂图(complex graph).利用复杂图聚类(complex graph clustering)[20]算法可对图像结点进行聚类.如果忽略单词-单词间的同构链接(如图 1 中虚线所示),该图模型简化为二部图(bipartite graph),可利用二部图协同谱聚类(spectral co-clustering)算法[24]对图像进行聚类.

2.1 复杂图聚类

定义 1(复杂图(complex graph)). 给定一个(无向的)复杂图 $G=(\{V_i\}_{i=1}^m,E)$,其中,$\{V_i\}_{i=1}^m$ 代表 m 个不同类型结点集合,E 代表图中边的集合,G 可以表示为一个相关矩阵的集合[20]:

$$\{\{S^{(i)}\in R_+^{n_i\times n_i}\}_{i=1}^m,\{A^{(ij)}\in R_+^{n_i\times n_j}\}_{i,j=1}^m\} \tag{8}$$

其中,$S^{(i)}$表示结点集合 V_i 中同构结点之间的同构链接(边)的权重矩阵,$A^{(ij)}$表示结点集合 V_i 和 V_j 之间的异构链接(边)的权重矩阵.结点集合 V_i 中的结点个数$|V_i|=n_i$.

2.1.1 复杂图聚类框架

文献[20]提出了一个复杂图聚类框架,该框架可以利用复杂图中的同构和异构链接信息对多种类型的结点分别聚类.在聚类过程中,不同类型结点之间的聚类结构可互相传递最终完成聚类.

假定 k_i 表示结点集合 V_i 聚类个数,$C^{(i)}\in R_+^{n_i\times k_i}$ 表示结点集合 V_i 中结点的类属关系指示矩阵,矩阵元素 $C_{pq}^{(i)}$ 表示 V_i 中的第 p 个结点隶属于第 q 类的类属权重,$D^{(i)}\in R_+^{k_i\times k_i}$ 代表 V_i 内结点的聚类模式,矩阵元素 $D_{pq}^{(i)}$ 表示 V_i 中第 p 类和第 q 类之间的链接强度,$B^{(ij)}\in R_+^{k_i\times k_j}$ 代表 V_i 和 V_j 内结点之间的聚类模式,矩阵元素 $B_{pq}^{(ij)}$ 表示 V_i 中第 p 类和 V_j 中的第 q 类之间的链接强度.复杂图聚类框架可以描述为如下优化问题[20]:

$$\left[\begin{array}{l}\underset{C^{(i)},C^{(j)}}{\arg\min}\,L\\ L=\sum_{i=1}^m\omega^{(i)}\boldsymbol{D}(S^{(i)},C^{(i)}D^{(i)}(C^{(i)})^T)+\sum_{ij=1}^m\omega^{(ij)}\boldsymbol{D}(A^{(ij)},C^{(i)}B^{(ij)}(C^{(j)})^T)\\ s.t.\ C^{(i)}\in\{0,1\}^{n_i\times k_i},C^{(i)}\mathbf{1}=\mathbf{1}\end{array}\right. \tag{9}$$

其中,向量 $\mathbf{1}$ 的每个分量都为 1,$\boldsymbol{D}$ 代表给定的距离函数,$\omega^{(i)}$和$\omega^{(ij)}$分别代表同构链接和异构链接的权重系数($\omega^{(i)}$和$\omega^{(ij)}$由用户根据具体问题求取).式(9)中并未指定距离函数 $\boldsymbol{D}$,因此,这个复杂图聚类框架有很好的推广性.从不同的现实系统中抽象得到的复杂图模型有不同的统计特性,要根据复杂图边权重的统计分布特性选择合适的距离函数.文献[20]中讨论了两种边权重分布及其对应的距离函数:

1) 正态分布(normal distribution).这时,距离函数 $\boldsymbol{D}$ 应选用欧氏距离(Euclidean distance);
2) 指数分布族(exponential family distribution),这是 1)的推广.欧氏距离、扩展的 I-散度(generalized I-divergence)距离、KL 散度(KL divergence)距离都可以推广为 Bregman 散度(Bregman divergences)距离,与之对应的是一系列指数分布.

2.1.2 复杂图聚类算法

常用的复杂图模型构建在两种类型的结点集合之上.如图 1 所示.在复杂图 $G_1=(V_1,V_2,E)$中,同构链接来自 V_1 中的结点之间,异构链接来自 V_1 和 V_2 中的结点之间.根据定义 1,G_1 可表示为 $\{S\in R_+^{n_1\times n_1},A\in R_+^{n_1\times n_2}\}$.如果忽略权重系数$\omega$,式(9)中目标函数 L 简化为

$$L=\boldsymbol{D}(S,C^{(1)}D(C^{(1)})^T)+\boldsymbol{D}(A,C^{(1)}B(C^{(2)})^T) \tag{10}$$

如果 $\boldsymbol{D}$ 选用欧氏距离,则复杂图 G_1 的聚类任务可以表示为如下的优化问题[20]:

$$\begin{bmatrix} \min\limits_{C^{(1)},C^{(2)},D,B} \| S-C^{(1)}D(C^{(1)})^T \|^2+\| A-C^{(1)}B(C^{(2)})^T \|^2 \\ \text{s.t. } C^{(1)}\in\{0,1\}^{n_1\times k_1}, C^{(2)}\in\{0,1\}^{n_2\times k_2}, C^{(1)}\mathbf{1}=\mathbf{1}, C^{(2)}\mathbf{1}=\mathbf{1} \end{bmatrix} \tag{11}$$

定理 1. 如果 $C^{(1)}\in\{0,1\}^{n_1\times k_1}, C^{(2)}\in\{0,1\}^{n_2\times k_2}, D\in R_+^{k_1\times k_1}, B\in R_+^{k_1\times k_2}$ 是优化问题(11)的最优解,那么有

$$D=((C^{(1)})^TC^{(1)})^{-1}(C^{(1)})^TSC^{(1)}(C^{(1)})^TC^{(1)})^{-1} \tag{12}$$

$$B=((C^{(1)})^TC^{(1)})^{-1}(C^{(1)})^TAC^{(2)}(C^{(2)})^TC^{(2)})^{-1} \tag{13}$$

文献[20]给出了定理 1 的证明.根据定理 1,复杂图 G_1 的聚类算法如下:

算法 1. 复杂图 G_1 的聚类算法 CGC.

输入:矩阵 S 和 A,结点集合 V_1 的聚类个数 k_1,V_2 的聚类个数 k_2;

输出:类属关系指示矩阵 $C^{(1)}$和 $C^{(2)}$.

步骤 1. 重复迭代步骤 2~步骤 5 直到收敛;

步骤 2. 根据式(12)计算 D;

步骤 3. 根据式(13)计算 B;

步骤 4. 固定 D,B 和 $C^{(2)}$,逐行更新 $C^{(1)}$,使得最小化 $L=\| S-C^{(1)}D(C^{(1)})^T \|^2+\| A-C^{(1)}B(C^{(2)})^T \|^2$;

步骤 5. 固定 D,B 和 $C^{(1)}$,逐行更新 $C^{(2)}$,使得最小化 $L=\| S-C^{(1)}D(C^{(1)})^T \|^2+\| A-C^{(1)}B(C^{(2)})^T \|^2$.

本文中,对称矩阵 $S\in R_+^{n_i\times n_i}$ 为单词-单词相关性矩阵,n_i 是词典(vocabulary)中单词的个数,矩阵元素 $S_{ij}(i\neq j)$表示单词 w_i 和 w_j 之间的主题相关度,$S_{ij}=Topic\_r(w_i,w_j)$.矩阵 $A\in R_+^{n_i\times n_j}$ 为单词-图像相关性矩阵,n_j 是图像集合中图像的个数,矩阵元素 A_{ij} 表示单词 w_i 和第 j 个图像之间的相关度,$A_{ij}=tfidf(w_i)\cdot vis(w_i)$.

2.2 基于二部图的协同谱聚类

定义 2(二部图(bipartite graph)). 给定一个(无向的)二部图 $G=(V_1,V_2,E)$,其中,$V_1=\{d_1,\ldots,d_n\}$和 $V_2=\{w_1,\ldots,w_m\}$代表两个不同类型结点集合,$E=\{\{d_i,w_j\}:d_i\in V_1,w_j\in V_2\}$ 代表图中边的集合.

定义 3(图 G 的分割(cut)). 给定图 G 的邻接矩阵 M,图 G 中边(d_i,w_j)的权重为 E_{ij},则 M 定义为

$$M_{ij}=\begin{cases} E_{ij}, & \text{如果存在边}\{d_i,w_j\} \\ 0, & \text{否则} \end{cases} \tag{14}$$

如果图 G 的一个分割将结点集 V 分成两个子集 V_1 和 V_2,则图 G 的分割定义为

$$cut(V_1,V_2)=\sum_{d_i\in V_1,w_j\in V_2}M_{ij} \tag{15}$$

如果推广到 k 个子集,则图 G 的 k-分割(k-cut)定义为

$$k\text{-}cut(V_1,V_2,\ldots,V_k)=\sum_{i<j}cut(V_i,V_j) \tag{16}$$

根据定义 3,二部图聚类框架可以描述为式(17)所示优化问题:

$$k\text{-}cut(W_1\cup D_1,\ldots,W_k\cup D_k)=\min_{V_1,\ldots,V_k}k\text{-}cut(V_1,\ldots,V_k) \tag{17}$$

文献[24]中的协同谱聚类(spectral co-clustering)算法 Multipartition(k)给出了以上优化问题的解.将单词-图像相关性矩阵 $A\in R_+^{n_i\times n_j}$ 作为算法 Multipartition(k)的输入可对图像进行聚类.

3 实 验

3.1 可见度计算模型分析

本节将在 5 个单词集合上对本文提出的单词可见度计算模型进行定量分析,以验证其合理性和性能.第 3.1.2 节通过对可见度计算模型在 5 个单词集合上的总体性能进行分析,表明该模型的合理性.第 3.1.3 节进一步验证了本模型对主题宽泛单词可见度有抑制能力.

3.1.1 单词集合

可见度用来度量单词语义可被视觉感知的能力,为了验证本文提出的可见度计算模型的合理性,我们选取了下述 5 个单词集合进行定量分析和对比实验:

(1) Columbia374[25].该集合选自 LSCOM 本体[26],包含 374 个单词,分别对应 374 个视觉概念(visual concepts),它们由研究者定义并用来对视频语义内容进行标注;

(2) Flickr Hot Tags.该集合选自 Flickr(http://www.flickr.com/photos/tags),包含 144 个单词,是 Flickr 上用户最常用的图像标签;

(3) IAPR Annotations.该集合选自开放数据集 IAPR TC-12 Benchmark[27]中的图像标注数据.我们选取了 IAPR TC-12 的一个子集,包含 2 000 个图像,并提取了其标注文本中的名词,共包含约 500 个单词;

(4) Surrounding Words.该集合由我们构建,首先将视觉多义词 apples 作为查询提交给 Google 的图像搜索引擎,对返回结果中的每个图像,我们下载了图像文件以及该图像所在的 Web 页面,并提取了图像伴随文本中的名词构成单词集合,共包含约 1 200 个单词;

(5) NIPS BOWs.该集合选自 Topic Modeling Toolbox[28]中的数据集“NIPS proceedings papers (bag of words)”,该数据集包含 9 244 个 NIPS 会议论文集中的单词,我们随机选取了其中的 500 个名词构成单词集合.

3.1.2 可见度模型总体性能

应用公式(3),我们计算了 5 个数据集中每个单词的可见度.见表 2,对每个数据集,从可见度最高的单词开始,每个可见度数值区间选取一个单词,列出了 20 个代表单词.每个数据集的平均可见度如图 4 所示.实验结果表明:

(1) 数据集 Colimbia374,Surrounding Words,Flickr Hot Tags 和 IAPR Annotations 中单词是图像标注所用标签或视觉概念,见表 2 中的 football 和 oceans,其语义可被视觉感知的能力普遍较强,应用可见度计算模型其单词有较高 $vis(w)$值;

(2) 数据集 NIPS BOWs 中单词多数属于机器学习领域的抽象概念,见表 2 中的 statistics 和 probability,其语义可被视觉感知的能力较弱,因此平均 $vis(w)$值较低;

(3) 在图像标签或视觉概念中,描述图像中特定对象的单词,见表 2 中的 pavilions 和 railings,往往获得更高 $vis(w)$值,这和图 3 中的单词 fishermen 和 largemouth 的高可见度一致.

Table 2 20 representative words and corresponding $vis(w)$ values from each of the 5 word sets (values of C_1 and C_2 are retrieved from Google on May 27, 2009)

表 2 5 个数据集中 20 个代表单词及其 $vis(w)$值(C_1 和 C_2 值来自 Google,2009 年 5 月 27 日)

NIPS BOWs		Columbia374		Surr. words		Flickr hot tags		IAPR annotations	
Fields	0.930 6	Pavilions	0.986 5	Fuji	0.952 1	Museum	0.925 5	Football	0.991 2
Dimension	0.889 2	Harbors	0.977 0	Statue	0.947 3	Urban	0.919 9	Foreground	0.987 4
Feature	0.874 3	Oceans	0.965 0	Oranges	0.934 2	Autumn	0.906 6	Railings	0.978 3
Pattern	0.855 7	Microphones	0.959 6	Bananas	0.927 1	Portrait	0.897 3	Pavement	0.969 9
Kernel	0.845 3	Suburban	0.944 8	Apples	0.913 1	Mountains	0.890 7	Pillows	0.955 6
Distance	0.830 3	Supermarket	0.935 0	Spices	0.904 9	Landscape	0.878 8	Chairs	0.946 9
Output	0.808 4	Farms	0.923 1	Fruits	0.897 9	Festival	0.868 5	Cliff	0.938 1
Detection	0.796 5	Flood	0.917 0	Orchard	0.880 2	Ireland	0.856 5	Fountain	0.923 0
Properties	0.792 1	Clouds	0.909 6	Downtown	0.879 6	Holiday	0.845 5	Sunrise	0.914 6
Abstract	0.784 8	Flags	0.896 4	Bush	0.862 4	Florida	0.836 6	Lamp	0.900 5
Weight	0.773 9	Laundry	0.886 1	Ridge	0.858 2	Island	0.823 6	Desk	0.899 1
Statistics	0.752 9	Desert	0.875 2	Wine	0.844 6	Christmas	0.813 4	Peak	0.886 1
Variable	0.740 3	Politics	0.867 3	Temperature	0.833 2	Wedding	0.800 4	Bathroom	0.870 9
System	0.727 4	Traffic	0.858 0	Wood	0.829 3	Food	0.796 1	Mirror	0.864 4
Information	0.622 3	Block	0.847 7	Storage	0.814 6	Fashion	0.784 0	Stone	0.856 5
Probability	0.601 9	Oil	0.830 4	Ground	0.801 4	Birthday	0.774 2	Areas	0.841 8
Size	0.585 9	Hand	0.820 3	Settings	0.794 2	Family	0.765 8	Plant	0.836 4
Results	0.551 2	Frame	0.813 8	Format	0.785 4	Green	0.737 3	Town	0.825 0
Type	0.512 4	Party	0.804 1	Firm	0.777 0	Art	0.717 0	Night	0.818 0
Search	0.352 8	Driver	0.797 0	Point	0.768 7	Black	0.700 7	Border	0.802 2

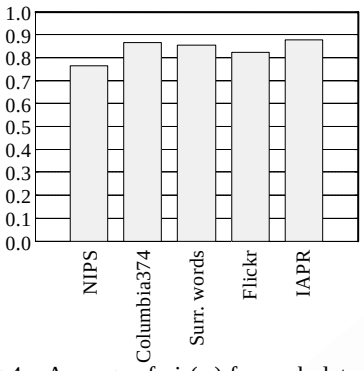


Fig.4 Average of *vis*(*w*) for each datasets

图 4 5 个数据集的平均可见度

3.1.3 可见度与 NOS

文献[1]利用 WordNet(http://wordnet.princeton.edu/)得到视觉多义词的不同含义(sense),然后通过伴随文本检索某个特定含义的图像.基于此方法,我们提出利用 NOS(number of senses in WordNet),即单词在 WordNet 中不同含义(sense)的个数这一指标来度量单词的主题宽泛程度.例如,名词 place 和 desk 的 NOS 值分别为 16 和 1,则单词 place 比 desk 更宽泛.

为了考察式(3)对主题宽泛单词可见度的抑制效果,本实验对比不同可见度单词的 NOS 值.如图 5 所示,分 3 个 *vis*(*w*)值区间对比相应单词的平均 NOS 值.高可见度单词(*vis*(*w*)∈[0.9,1])的平均 NOS 值约为 2.5,而低可见度单词(*vis*(*w*)≤0.8)的平均 NOS 值约为 6.

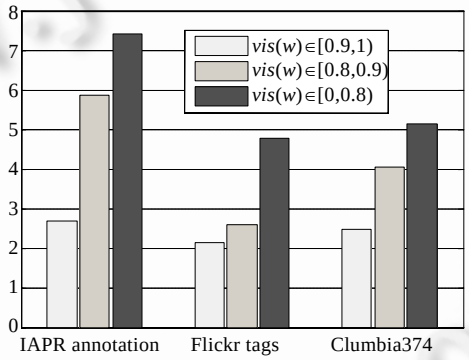


Fig.5 Average of NOS for each datasets

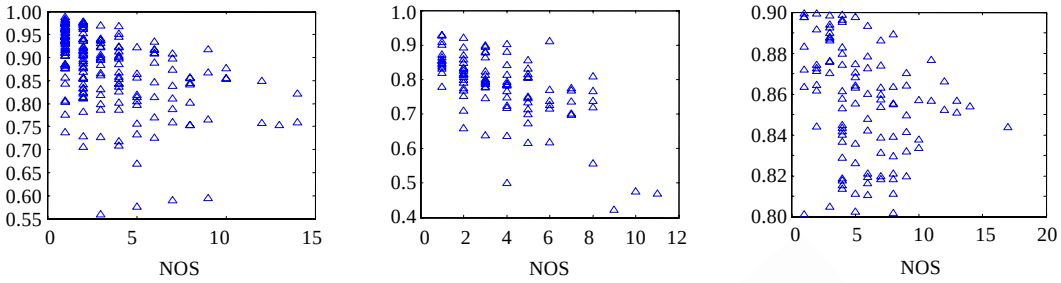
图 5 不同 *vis*(*w*)值区间的平均 NOS 值

为了进一步考察可见度与 NOS 的关系,我们将单词可见度 *vis*(*w*)与其 NOS 值进行了二维可视化表示.如图 6 所示,大部分高 *vis*(*w*)值单词其 NOS 值很小,而随着 *vis*(*w*)值的减小,相应单词的 NOS 值逐渐增加.

以上实验结果表明:

(1) 整体上,式(3)对主题宽泛单词可见度的抑制效果是明显的.

(2) 通过对 5 个数据集中单词 NOS 值分析发现,描述图像中特定对象的标签或视觉概念,其单词的 NOS 较小,如 desk,pavilion,railings,pillows,largemouth 和 fishermen 等单词的 NOS 值为 1,而这些单词通过单词可见度模型计算得到的 *vis*(*w*)值较高.因此,本文所提出的单词可见度计算模型在抑制主题宽泛单词可见度的同时,进一步提升了描述图像中特定对象单词的可见度,从而使得这些单词和图像的链接权重在本文提出的单词-图像相关性定义下得到增强.



(a) Columbia374 words
(a) Columbia374 单词

(b) Flickr tags
(b) Flickr 标签

(c) IAPR annotations
(c) IAPR 标注单词

Fig.6 2D visualization of *vis*(*w*) and NOS for words

图 6 单词可见度 *vis*(*w*)与 NOS 值的二维可视化表示

3.2 Web图像聚类

本节通过 Web 图像聚类实验对所提出的聚类框架的有效性进行验证.

3.2.1 数据集

本实验采用两个数据集:(1) 自己构建的 Google 图像搜索结果数据集;(2) 开放的图像检索基准数据集 IAPR TC-12 Benchmark[27].

我们编写了爬虫程序,根据提交的关键字作为查询自动提取 Goolge Image Search™ 的返回结果.对返回结果中的每个图像,下载了图像文件以及该图像所在的 Web 页面.为了更好地验证聚类效果,根据文献[1]中实验,选择了 5 个视觉多义词[4]作为查询,它们是:apple,bass,jaguar,mouse 和 tower.由于 Google 限制了搜索实际返回的结果数量,通过爬虫获取的数据集共包含约 4 000 个数据项.为了提取图像的伴随文本,对图像所在的 Web 页面进行解析,提取了图像周围窗口为 300 个单词的文本作为该图像的伴随文本.所有的伴随文本通过词性标注(part of speech tagging)提取了其中名词.按照文献[29]中的实验,也提取了 Web 页面上 alt 和 title 标签中的名词.经过以上处理,每个查询返回结果的所有伴随文本中的名词词汇表的规模为 1 000~2 000 个单词.

IAPR TC-12 Benchmark 数据集包含 20 000 张照片和相应的文本标注(annotations)数据.文本标注数据以 XML 格式存放,内容包括照片的标题、创建时间、拍摄地点、照片内容的语义描述(description)和备注(notes),因此可以作为照片的伴随文本.由于本实验需要图像附带比较丰富的伴随文本,我们选取了 IAPR 中 Description 和 Notes 文本数量比较大的 2 000 张照片进行聚类实验.通过提取 Description 和 Notes 中的名词形成的词汇表包含约 550 个单词.

3.2.2 聚类性能评价标准

标准的聚类性能评价标准是归一化聚类互信息(normalized mutual information,简称 NMI[30]).NMI 定义为:

定义 4. 给定聚类个数 k,类属列表向量$\lambda=(\lambda_1,\ldots,\lambda_K)$中$\lambda_i$的取值范围为$\lambda_i=1,\ldots,k$,$\lambda_i=j$ 表示第 i 个数据项属于第 C_j 类.用$\lambda^{(a)}$和$\lambda^{(b)}$分别表示聚类结果和基准(ground truth)类属列表向量,则$\lambda^{(a)}$和$\lambda^{(b)}$的归一化聚类互信息$\phi^{(NMI)}$定义为[30]

$$\phi^{(NMI)}(\lambda^{(a)},\lambda^{(b)})=\frac{\sum_{h-1}^{k}\sum_{l-1}^{k}n_{hl}\log\left(\frac{n\cdot n_{hl}}{n_h^{(a)}n_l^{(b)}}\right)}{\sqrt{\left(\sum_{h-1}^{k}n_h^{(a)}\log\frac{n_h^{(a)}}{n}\right)\left(\sum_{l-1}^{k}n_l^{(b)}\log\frac{n_l^{(b)}}{n}\right)}} \tag{18}$$

其中,$n_h^{(a)}$是对应于$\lambda^{(a)}$的类C_h中的数据项个数,$n_l^{(b)}$是对应于$\lambda^{(b)}$的类C_l中的数据项个数,C_{hl}表示同时被聚在$\lambda^{(a)}$的类C_h中和$\lambda^{(b)}$的类C_l中的数据项的个数.

根据定义 4,某次聚类结果$\lambda^{(a)}$和基准类属$\lambda^{(b)}$之间的互信息值$\phi^{(NMI)}(\lambda^{(a)},\lambda^{(b)})$越大,表示本次聚类效果越好.理

想的聚类是$\phi^{(NMI)}(\lambda^{(a)},\lambda^{(b)})$=1.为了获得基准(ground truth)类属列表向量,我们对数据集进行了手工标注.

3.2.3 引入可见度的有效性实验

为了验证引入单词可见度的有效性,本节对 $tfidf(w)\cdot vis(w)$方法和 $tfidf(w)$方法在 Google 图像搜索结果数据集上进行实验对比.在这个实验中,我们忽略图 1 所示的图模型中单词-单词之间的同构链接,只考虑单词-图像的异构链接.因此,每个查询输入是单词-图像相关性矩阵 A.矩阵元素 $A_{ij}(i\neq j)$表示单词 w_i 和第 j 个图像之间的链接权重.这里分别采用 $A_{ij}=tfidf(w_j)$和 $A_{ij}=tfidf(w_j)\cdot vis(w_j)$进行对比.图像-单词矩阵 A 对应的图模型是二部图,因此采用基于二部图的协同谱聚类算法.为了进一步研究图聚类算法,将基于二部图的协同谱聚类算法与 k-means 聚类算法进行了对比.实验结果如图 7 所示.k-means 和 Spectral Co.表示单词-图像链接权重采用 $A_{ij}=tfidf(w_j)$时的 k-means 聚类和协同谱聚类的聚类互信息;而+$vis(w)$表示单词-图像链接权重采用 $A_{ij}=tfidf(w_j)\cdot vis(w_j)$时的聚类互信息.实验结果表明,将本文提出的可见度检索模型集成到 tf-idf 方法中能够提高图像聚类的精确性.高可见度单词在聚类过程中能够向图像类传递更多的语义描述信息,从而使得协同谱聚类在图像类和单词类的交互下得到比较好的聚类结果.

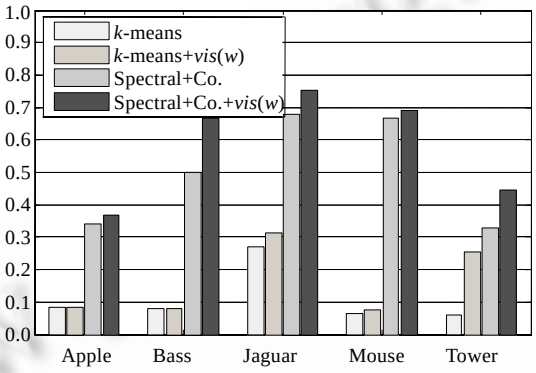


Fig.7 NMI of k-means and spectral co-clustering clustering results

图 7 k-means 和协同谱聚类的聚类互信息对比

3.2.4 引入主题相关度的有效性实验

在实验中,考虑单词之间的两种关联:主题相关性关联和共生相关性关联.共生相关性关联定义如下:

定义 5(单词的共生相关性关联(co-occurring relevance)). 图像的伴随文本中,任意两个单词 w_s 和 w_t 的共生相关性定义为它们同时出现在某个图像的伴随文本中的概率 $P(w_s,w_t)$:

$$P(w_s,w_t)=\frac{num(w_s,w_t)}{N} \tag{19}$$

其中,N 是图像的总数,$num(w_s,w_t)$是其伴随文本中同时包含单词 w_s 和 w_t 的图像的个数.

在通过主题相关度函数 $Topic_r(w_s,w_t)$计算两个单词间主题相关性时,本文采用 LDA Topic Modeling Toolbox[28].基于吉布斯采样(Gibbs sampling)方法,使用对称 Dirichlet 先验,参数设置为:α=50/K 和β=0.01.其中,K 是给定的每个查询所包含的主题个数.吉布斯采样次数为 500 次迭代.通过 LDA 学习得到主题分布 $P(w|z)$以及 $P(z)$,由式(7)可计算 $Topic\_r(w_s,w_t)$.

结合主题相关性和共生相关性关联,单词间的同构链接权重定义为($\lambda(0<\lambda<1)$是可调参数):

$$\lambda\cdot P(w_s,w_t)+(1-\lambda)Topic\_r(w_s,w_t) \tag{20}$$

由于同时考虑单词-图像的异构链接和单词-单词的同构链接,采用复杂图聚类算法.对于式(20)中的参数λ,考虑 3 种情况:

1) λ=1.式(20)转化为 $P(w_s,w_t)$.这时忽略了单词间的主题相关性关联;

2) λ=0.式(20)转化为 $Topic\_r(w_s,w_t)$.这时忽略了单词间的共生相关性关联;

3) λ=0.15.这时,同时考虑两种相关性关联.

在 Google 图像搜索数据集上的实验结果如图 8 所示.对于所有 5 个查询的复杂图聚类性能都在λ=0 时达到最好,因此可得到结论:首先,由 LDA 学习得到的主题相关性关联要比共生相关性关联重要;其次,在复杂图聚类过程中,单词间的主题相关性能够通过单词-图像异构链接进行传递,获得较好的聚类结果.

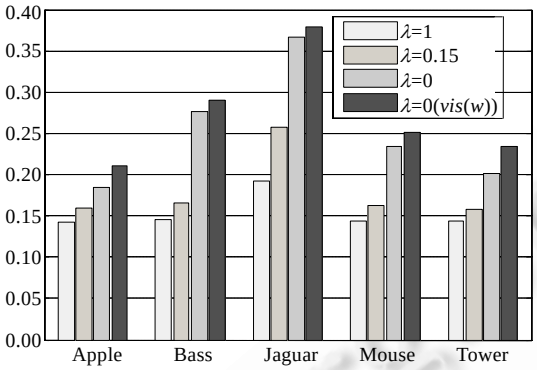


Fig.8 NMI of complex graph clustering results

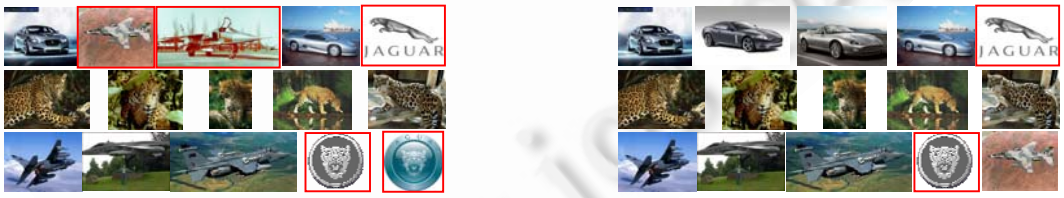
图 8 复杂图聚类的互信息对比

为了进一步考察可见度模型的有效性,我们在复杂图聚类中也区分了 $tfidf(w)\cdot vis(w)$方法和 $tfidf(w)$方法.如图 8 所示,$\lambda=0(vis(w))$表示单词-图像链接权重采用 $A_{ij}=tfidf(w_j)\cdot vis(w_j)$.由聚类互信息结果可以看到,在复杂图聚类中,将单词的可见度引入单词-图像链接权重,使得高可见度单词向与之关联的图像结点传递更多的主题相关性信息,从而进一步提高聚类性能.

以图 9 所示对查询 jaguar 检索图像聚类结果为例(查询 jaguar 的 3 个主题 jaguar car,jaguar animal 和 jaguar plane 分别列出前 5 个结果,加边框图像指示错误的聚类项),对比图 9(a)、图 9(b)可得以下结论:

(1) 对主题 jaguar animal 采用 $tfidf(w)$和 $tfidf(w_j)\cdot vis(w_j)$前后聚类效果均较好,这是由于该主题图像伴随文本中频繁出现单词,如 water,cat 和 tree,在其他主题图像伴随文本中出现频率很低,使得 LDA Gibbs 采样得到其伴随文本隐含主题分布较稳定,从而使得其伴随文本中单词间的主题相关度较高,获得比较好的复杂图聚类结果;

(2) 很多单词,如 car,speed 和 wheel,同时出现在主题 jaguar car,jaguar plane 和 jaguar logo 图像伴随文本中,使得这 3 个主题图像在图 9(a)的聚类结果中出现交叉.但是,在图 9(b)的 $tfidf(w_j)\cdot vis(w_j)$方法引入可见度增强某些描述图像特定对象的单词与图像链接权重情况下,聚类结果得到一定改善.



(a) The link weight between word-image by $tfidf(w)$ method

(a) 单词-图像链接权重采用 $tfidf(w)$方法

(b) The link weight between word-image by $tfidf(w_j)\times vis(w_j)$ method

(b) 单词-图像链接权重采用 $tfidf(w_j)\times vis(w_j)$方法

Fig.9 Examples of complex graph clustering results under condition 1 by $tfidf(w)$ and $tfidf(w_j)\cdot vis(w_j)$ method

图 9 分别采用 $tfidf(w)$和 $tfidf(w_j)\cdot vis(w_j)$方法的复杂图聚类结果

3.2.5 IAPR TC-12 Benchmark 数据集上实验结果

为了进一步验证引入可见度和主题相关度的有效性及其对整个聚类框架聚类性能的改进效果,我们在开放基准数据集 IAPR TC-12 Benchmark[27]上重复以上实验,实验结果如图 10 所示.

(1) 引入可见度对算法 *k*-means,Spectral Co.和 Complex Graph Clustering(CGC)在 IAPR 数据集上聚类互信息提高率分别为 3.6%,0.4%和 9.8%;

(2) 对比图 7、图 8 和图 10 中算法 *k*-means 和 Spectral Co-clustering 的聚类结果可以发现其聚类性能不稳定,这是由于算法 *k*-means 对初值的依赖性所致(spectral co-clustering 算法中对谱分解后特征向量的聚类采用的也是 *k*-means 算法[24]).但是,引入可见度后均提高了其 NMI;

(3) 复杂图聚类(CGC)性能有待提高.本文对于该算法距离函数 $\boldsymbol{D}$ 采用欧氏距离,见式(11).经实验发现,当 CGC 算法收敛后其损失函数值 L 仍然较大,表明距离函数 $\boldsymbol{D}$ 在具体聚类任务中需进一步研究和改进.但是,引入可见度后仍然提高了其 NMI,再次验证了所提出的聚类框架的有效性.

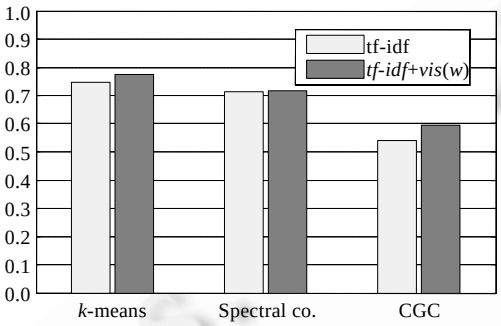


Fig.10 NMIs of clustering results for 3 algorithms by $tfidf(w)$ and $tfidf(w_j)\cdot vis(w_j)$ methods

图 10 3 个聚类算法分别采用 $tfidf(w)$和 $tfidf(w_j)\cdot vis(w_j)$方法的聚类互信息对比

4 结 论

本文介绍了一种 Web 图像图聚类方法,该方法对包含图像和单词两种类型结点的图进行聚类.在聚类过程中,定义了两种关联关系:单词-图像的相关性关联以及单词-单词主题相关性关联.前者由一种新的单词可见度模型与 tf-idf 方法的集成来定义;后者通过 LDA 模型对图像的伴随文本进行学习得到.实验结果表明,将可见度和单词间的主题相关性引入到图聚类算法中能够改善 Web 图像的聚类性能.

进一步挖掘图像和伴随文本之间的语义关联,例如动词和图像的语义关联,是今后我们的研究重点.同时,针对具体的图模型定义更合适的距离函数 $\boldsymbol{D}$ 并应用到复杂图聚类中,也是值得研究问题.

References:

[1] Saenko K, Darrell T. Unsupervised learning of visual sense models for polysemous words. Advances in Neural Information Processing Systems 21. Cambridge: MIT Press, 2009. 1393−1400.

[2] Rege M, Dong M, Hua J. Graph theoretical framework for simultaneously integrating visual and textual features for efficient Web image clustering. In: Proc. of the 17th Int'l Conf. on World Wide Web. New York: ACM Press, 2008. 317−326. http://www2008.org/papers/Proceedings.html

[3] Ding HY, Liu J, Lu HQ. Hierarchical clustering-based navigation of image search results. In: Proc. of the 16th Annual ACM Int'l Conf. on Multimedia. New York: ACM Press, 2008. 741−744. http://portal.acm.org/citation.cfm?id=1459359

[4] Cai D, He X, Li Z, Ma W, Wen J. Hierarchical clustering of WWW image search results using visual, textual and link information. In: Proc. of the 13th Annual ACM Int'l Conf. on Multimedia. New York: ACM Press, 2008. 952−959. http://portal.acm.org/citation.cfm?id=1027527

[5] Hu Y, Yu N, Li Z, Li M. Image search result clustering and re-ranking via partial grouping. In: Proc. of the 2007 IEEE Int'l Conf. on Multimedia and Expo. Washington: IEEE Press, 2007. 603−606. http://www.ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4284552

[6] Jing F, Wang CH, Yao YH, Deng KF, Zhang L, Ma WY. IGroup: A Web image search engine with semantic clustering of search results. In: Proc. of the 14th Annual ACM Int'l Conf. on Multimedia. New York: ACM Press, 2006. 23−27. http://portal.acm.org/citation.cfm?id=1180639

[7] Wu F, Liu YA, Zhuang YT. Tensor-Based transductive learning for multi-modality video semantic concept detection. IEEE Trans. on Multimedia, 2009,11(5):868−878. [doi: 10.1109/TMM.2009.2021724]

[8] Zhang H, Wu F, Zhuang YT, Chen JX. Cross-Media retrieval method based on content correlations. Chinese Journal of Computers, 2008,31(5):820−826 (in Chinese with English abstract).

[9] Long B, Yu PS, Zhang ZF. A general model for multiple view unsupervised learning. In: Proc. of the 2008 SIAM Int'l Conf. on Data Mining. SIAM Press, 2008. 822−833. http://www.siam.org/proceedings/datamining/2008/dm08.php

[10] Raina R, Battle A, Lee H, Packer B, Ng A. Self-Taught learning: Transfer learning from unlabeled data. In: Proc. of the 24th Int'l Conf. on Machine Learning. New York: ACM Press, 2007. 759−766.

[11] Pan SJ, Kwok JT, Yang Q. Transfer learning via dimensionality reduction. In: Proc. of the 23rd AAAI Conf. on Artificial Intelligence. AAAI Press, 2008. 677−682. http://www.aaai.org/Press/Proceedings/aaai08.phpc

[12] Coelho TAS, Calado PP, Souza LV, Ribeiro-Neto B, Muntz R. Image retrieval using multiple evidence ranking. IEEE Trans. on Knowledge and Data Engineering, 2004,16(4):408−417. [doi: 10.1109/TKDE.2004.1269666]

[13] Wang XJ, Zhang L, Li XR, Ma WY. Annotating images by mining image search results. IEEE Trans. on Pattern Anal Mach Intell, 2008,30(11):1919−32. [doi: 10.1109/TPAMI.2008.127]

[14] Barnard K, Duygulu P, Forsyth D, de Freitas N, Blei D, Jordan M. Matching words and pictures. Journal of Machine Learning Research, 2003,3:1107−1135. [doi: 10.1162/153244303322533214]

[15] Blei D, Jordan M. Modeling annotated data. In: Proc. of the Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 2003. 127−134. http://portal.acm.org/citation.cfm?id=860435

[16] Zhu XJ, Goldberg A, Eldawy M, Dyer C, Strock B. A text-to-picture synthesis system for augmenting communication. In: Proc. of the 22nd AAAI Conf. on Artificial Intelligence: Integrated Intelligence Track. AAAI Press, 2007. 1590−1595. http://www.aaai.org/Press/Proceedings/aaai07.php

[17] Xia DY, Wu F, Zhuang YT. Search-Based automatic Web image annotation using latent visual and semantic analysis. In: Proc. of the 9th Pacific Rim Conf. on Multimedia: Advances in Multimedia Information Processing. Berlin, Heidelberg: Springer-Verlag Press, 2008. 842−845. http://www.informatik.uni-trier.de/~ley/db/conf/pcm/pcm2008.html

[18] Deschacht K, Moens MF. Text analysis for automatic image annotation. In: Proc. of the 45th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics Press, 2007. 1000−1007. http://ufal.mff.cuni.cz/acl2007/

[19] Berg T, Forsyth D. Animals on the web. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society Press, 2006. 1463−1470. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=10924

[20] Long B, Zhang ZF, Xu TB. Clustering on complex graphs. In: Proc. of the 23rd Conf. on Artificial Intelligence. AAAI Press, 2008. 659−664. http://www.aaai.org/Press/Proceedings/aaai08.phpc

[21] Yang B, Liu DY, Liu JM, Jin D, Ma HB. Complex network clustering algorithms. Journal of Software, 2009,20(1):54−66 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3464.htm [doi: 10.3724/SPJ.1001.2009.03464]

[22] Blei D, Ng A, Jordan M. Latent dirichlet allocation. Journal of Machine Learning Research, 2003,3:993−1022. [doi: 10.1162/jmlr.2003.3.4-5.993]

[23] Li WB, Sun L, Zhang DK. Text classification based on labeled-LDA model. Chinese Journal of Computers, 2008,31(4):620−627 (In Chinese with English abstract).

[24] Dhillon IS. Co-Clustering documents and words using bipartite spectral graph partitioning. In: Proc. of the 7th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2001. 269−274. http://www.sigkdd.org/kdd2001/Papers/papers.html

[25] Naphade M, Smith JR, Tesic J, Chang SF, Hsu W, Kennedy L, Hauptmann A, Curtis J. Large-Scale concept ontology for multimedia. IEEE Multimedia, 2006,13(3):86−91. [doi: 10.1109/MMUL.2006.63]

[26] LSCOM lexicon definitions and annotations version 1.0. ADVENT Technical Report, #217-2006-3, Columbia University, 2006.

[27] Grubinger M., Clough P, Müller H, Deselaers T. The IAPR TC-12 benchmark: A new evaluation resource for visual information systems. Proc. of the Int'l Workshop on Image 2006 Language Resources for Content-Based Image Retrieval, Held in Conjunction with LREC 2006. 2006. 13–23.

[28] Steyvers M, Griffiths T. Matlab topic modeling toolbox. http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm

[29] Schroff F, Criminisi A, Zisserman A. Harvesting image databases from the Web. In: Proc. of the 11th Int'l Conf. on Computer Vision. Washington: IEEE Press, 2007. 1–8. http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4408818

[30] Strehl A, Ghosh J. Cluster ensembles—A knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 2002,3:583–617. [doi: 10.1162/153244303321897735]

附中文参考文献:

[8] 张鸿,吴飞,庄越挺,陈建勋.一种基于内容相关性的跨媒体检索方法.计算机学报,2008,31(5):820–826.

[21] 杨博,刘大有,Liu JM,金弟,马海宾.复杂网络聚类方法.软件学报,2009,20(1):54–66. http://www.jos.org.cn/1000-9825/3464.htm [doi: 10.3724/SPJ.1001.2009.03464]

[23] 李文波,孙乐,张大鲲.基于 Labeled-LDA 模型的文本分类新算法.计算机学报,2008,31(4):620–627.

吴飞(1973－),男,湖南冷水江人,博士,副教授,CCF 高级会员,主要研究领域为多媒体分析与检索,计算机动画,统计学习理论.

庄越挺(1965－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为多媒体检索,计算机动画,人工智能,计算机图形学,数字图书馆.

韩亚洪(1977－),男,讲师,主要研究领域为多媒体分析与检索,机器学习.

邵健(1982－),男,博士,讲师,主要研究领域为多媒体分析与检索.