基于标签语义距离的图像多样化检索*

张震宇[1,2]，丁恒[1,2]，王瑞雪[1,2]，陆伟[1,2]
（1. 武汉大学信息管理学院，武汉 430072；2. 武汉大学信息资源中心，武汉 430072）

摘要：随着互联网图像资源的爆炸式增长，用户对图像多样化检索的需求愈发强烈。本文在对比图像视觉特征和图像文本内容算法的基础上，探讨不同标签语义距离算法在多样化检索中的效果，通过实验提供一种较好的基于标签语义距离的图像多样化检索算法。

关键词：标签语义距离；社会化标签；图像多样化检索；语义相似度
中图分类号：TP391.3
DOI：10.3772/j.issn.1673-2286.2017.2.006

1 引言

随着图片分享网站的兴起，互联网图像资源呈现爆炸式增长，用户图像检索的需求也随之增长。图像检索研究初期，大多数研究工作的目的是尽可能多地返回与检索词相关的图像[1-2]，这些研究忽略了图像间的关联，检索结果出现部分噪音图像[3]。近年来，图像多样化检索需求逐渐提上日程。多样化检索提供多种结果，并按检索结果的相关性反馈给用户[4]。

图像信息主要来源于图像视觉内容和文本内容。视觉内容指数字图像本身，如颜色、纹理、形状等；文本内容指与图像相关联的文字信息，如标题、描述、标签等[5]。研究表明图像底层视觉特征和高层语义概念间存在语义鸿沟（semantic gap）[6]，要满足用户图像多样化检索需求，关键是从图像中挖掘用户潜在需求。Yang等利用图像标签的差异性进行图像多样化检索[7]，但在标签语义差异性对图像多样化检索结果影响方面，未进行更深入的探讨。基于此，本文在NUS-WIDE图像检索数据集的基础上开展检索实验，探讨多种语义距离算法对图像多样化检索效果的影响。

2 图像多样化检索相关研究

2.1 社会化标签

随着Web 2.0技术，以及Flickr等社会化图像分享网站的出现[8]，用户可以自由上传、编辑和共享图像，同时也可以对图像设置个性化标签。截至2015年5月，Flickr累积拥有超过110亿幅图像[8]。这种由互联网用户生产、传播的图像，称为社会化图像（social image）；用户对社会化图像进行的标注行为，称为社会化标注（social tagging）；而由此产生的标签，则称为社会化标签（social tag）。社会化标签是人们在社会化环境下为实现资源共享和用户交互而使用的、描述资源的元数据，也是融入用户认知的高层次元数据[9]。社会化图像分享网站的流行，方便用户使用社会化标签对图像内容进行描述[10]。

2.2 图像多样化检索

目前，图像多样化检索包含多种算法。如Song等利

* 本研究得到海南省哲学社会科学规划课题“气候变化对海岛型旅游目的地游客流的影响及应对策略研究”（编号：HNSK（GJ）13-96）和中国科学技术信息研究所与武汉大学合作项目“科学文献的语义功能识别与深度利用研究”资助。

用图像自身特征计算相似度并以此作为连接概率[11]，借助Page等的思想计算最终排序值[12]，实现多样化排序，这种算法通过聚类算法将图像分成若干类，随后从每一类中选取最具代表性的图像作为多样化检索的结果[13]。Zhao等通过相似图像的检测和副本消除获得图像多样化检索结果[14]。朱胜平根据植物图像特点，利用形状和纹理等特征对植物图像检索领域开展研究，进行多样化检索[15]。

由于图像视觉特征与高层语义概念间的鸿沟，部分研究者以文本内容分析法展开研究。崔超然[16]和田枫[17]等指出图像标注的标签与图像内容高度相关，如果图像标注结果多样化，那么基于结果标签的图像检索必然也是多样化；Kim等提出图像标签多样化的概念，同时也给出计算图像标签多样化的算法[18]；Yang等运用图像标签中的语义差异，基于Google距离计算标签语义距离[7]。

综上所述，针对图像多样化检索，科研人员已展开相关研究，而对于不同语义距离算法在图像多样化检索效果比较方面则较少。本文期望通过比较不同语义距离算法在图像多样化检索上的运用情况，为信息资源服务者提供更完善的多样化检索方案。

3 研究方法

3.1 标签语义距离

图像由若干语义概念构成，图像的任一视觉主题可以用语义概念的组合表示。社会化标签作为图像描述的元数据，在一定程度上表征图像的语义概念内容。图像语义概念越相似，图像间的视觉主题越相似；相反，图像语义概念相似度越低，视觉主题差异越大。

在一组图像集合$\{d_1, d_2, d_3, \cdots, d_N\}$中，对于第$i$幅图像$d_i$，$t(d_i)$表示该图像的所有标签集合$t(d_i)=\{t_{1_{di}}, t_{2_{di}}, t_{3_{di}}, \cdots, t_{m_{di}}\}$，$m$为图像$d_i$标签集合中的标签数。图像$d_i$和图像$d_j$的主题差异性值DiffScore$(d_i, d_j)$可用公式（1）计算。

$$\text{DiffScore}(d_i, d_j)=\text{dis}(t(d_i), t(d_j))=\frac{1}{\text{sim}(t(d_i), t(d_j))} \quad (1)$$

其中，dis$(t(d_i), t(d_j))$表示图像d_i和图像d_j的标签语义距离，sim$(t(d_i), t(d_j))$表示图像d_i和图像d_j的标签语义相似度。

标签语义距离主要依赖于标签语义相似度。为比较不同算法的效果，本文选取3种主流的语义相似度算法，分别是基于WordNet本体词典的规则化算法、基于显式语义分析算法（Explicit Semantic Analysis，ESA）和基于Google Distance的算法。

3.1.1 基于WordNet的标签语义距离

基于WordNet的算法本质上是基于本体词典的规则化算法。本体词典，即描述词语概念及关系的词典。WordNet是英文的通用本体词典，收录的名词、动词、形容词和副词数量超过14万个，分别建立同义词集合（synsets）、概念间同义关系（synonymy）、反义关系（antonymy）、上下位关系（hynernymy and hyponymy）以及部分关系（meronymy）等[19]。通过词间语义关系计算语义相似度，可获得标签语义距离。

3.1.2 基于ESA的标签语义距离

ESA是一种基于知识库的规则化算法。由于人工构建的语义词典存在语义关系不足、实际应用效果不显著等问题，基于知识库的规则算法则可以弥补这方面的不足，如维基百科知识库。本文通过维基百科知识库来计算语义相似度，结合公式（1），获得标签语义距离。

3.1.3 基于Google Distance的标签语义距离

Google Distance是一种基于搜索引擎的语义相似度算法。该算法源于两个相关概念在同一篇文章中的共现频次。通过搜索引擎输入待比较词汇，分别得到单个词汇和两个待比较词汇的结果返回页面个数，进而分析待比较词汇的语义相似度[20]。在此基础上利用公式（1）计算标签语义距离。

3.2 基于标签语义距离的多样化检索算法

对于图像集合$\{d_1, d_2, d_3, \cdots, d_N\}$，N表示集合中包含的图像总数，$D=\{D_1, D_2, \cdots, D_N\}$表示一个有序图像集合，按图像与检索词Q相关度得分由高到低排序。公式（2）中，simScore(D_i)表示图像D_i与检索词Q的相关度得分，DiffScore(D_i, D_{i-1})表示图像D_i与图像D_{i-1}间的主题差异性得分，即D_i的DiffScore。

$$\text{DivScore}(D_i)=\frac{N-(i-1)}{N}\times\text{simScore}(D_i)+\frac{i-1}{N}\times\text{DiffScore}(D_i, D_{i-1}) \quad (2)$$

为保证研究效果，本文中的simScore（D_i）利用Sun等提出的Tagir算法[21]。

在标签语义距离基础上，本文进一步提出基于标签语义距离的多样化检索算法，该算法借鉴Kharazmi的文本多样化检索思想[22]，在文本相似度排序的基础上，加入文档主题差异性要素，使检索结果更符合多样化需求。具体做法：首先，根据检索词Q，计算相关图像集合D；其次，利用公式（1）得出由D_2开始所有图像的DiffScore，根据得分对结果重新排序。实验结果表明simScore和DiffScore的线性加权计算可以获得较好的效果，基于标签语义距离的图像多样化检索算法DivScoreDiff的具体运行程序如下。

for $1<i\leqslant|D|$ do

$$\text{DivScore}(D_i)=\frac{N-(i-1)}{N}\times\text{simScore}(D_i)+\frac{i-1}{N}\times\text{DiffScore}(D_i,D_{i-1})$$

end for

SortD_i on DivScore（D_i）

4 数据集构建

实验采用NUS-WIDE图像检索数据集，该数据集来自Flickr约5 000名用户提供的近26万幅图像和40多万个不同的标签，反映网络中海量图像的真实情况。同时，该数据集提供部分检索词及其对应的相关性人工标注评测基准（ground-truth），并未提供对应检索词子主题和图像子主题的标注。为进行图像多样化检索实验，本文对NUS-WIDE数据集进行如下处理。

（1）确定检索词。为便于计算，筛选相关图像数不少于1 000个的词汇作为研究对象，得到29个词。同时，由于社会化图像中包含大量视觉对象（object）和场景（scene）描述[23]，为更好地进行分析，本文选取的检索词既包含视觉对象，也包含场景描述。

（2）子主题确定及图像标注。由于NUS-WIDE数据集不包含检索词的子主题，因此，本文依据Nov等[10]提出的子主题标注法，通过人工方式对照片进行归类，并选择合适的关键词对数据集进行子主题标注。子主题标注主要依据图像的视觉外貌差异（如天气、季节等），反映图像的不同主题。具体而言，本次调研共邀请5名志愿者参与，第1名和第2名志愿者分别对每个检索词的相关图像进行子主题划分，第3名志愿者对相似主题进行归并得到该检索词的子主题集合，并由剩下2名志愿者对子主题划分结果的合理性分别进行独立判断。实验结果显示，第1名和第2名志愿者的子主题标注数据通过了一致性检验，证明本文提出的算法对于数据集的构建是较合理的。

（3）数据获取。采用Tagir算法获取每一个检索词的前100幅相关图像，并标记子主题，同时对相邻图像之间子主题变化情况进行统计。由于该评测基准结果不包含子主题，因此对于检索词Q，通过Tagir算法得到相关结果集D；对照检索评测基准，对D中的相关结果进行子主题标记，对于不相关结果，给予“no-relevant”的标记；从D中的图像D_2开始，根据公式（1）计算每一幅图像的DiffScore得分，即DiffScore（D_i，D_{i-1}），并判断图像D_i和图像D_{i-1}是否存在子主题变化的情况；若子主题发生变化，该图像标记为1，否则标记为0。

5 实验及结果评测

5.1 实验设置

（1）预实验：验证各标签语义距离算法在图像多样化检索中的可行性。选取检索词Q，通过计算获取与该检索词相关的前100幅图像作为结果数据集，根据结果数据集中图像子主题是否发生变化，将主题差异得分DiffScore（D_i，D_{i-1}）分为两组，即子主题不变和子主题变化。若两组数据间存在显著差异，则表明标签语义距离能反映图像子主题差异。预实验对象为15个具有代表性的检索词。

（2）多样化重排序实验。借助Tagir算法计算每个检索词中每幅图像的相关度得分simScore（D_i），并按从大到小进行排序，得到图像相关度排序结果集D和图像多样化得分。然后，对每一个检索词的相关度结果集进行重新排序，得到最终的检索结果列表集。

5.2 预实验

针对不同标签语义距离算法，对其图像主题差异性与标签语义距离的相关性分别进行验证。计算相同主题和不同主题的标签语义距离，并将两组数据对比结果绘制成箱型图（见图1、图2和图3）。

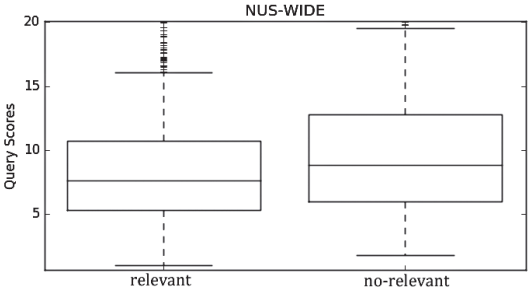

图 1 基于WordNet的标签语义距离

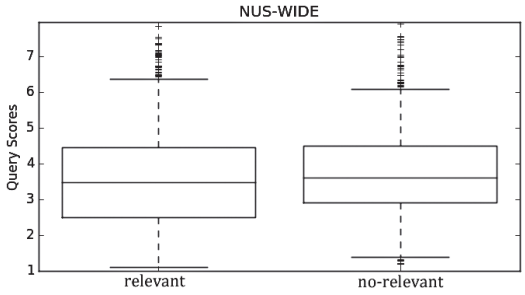

图 2 基于ESA的标签语义距离

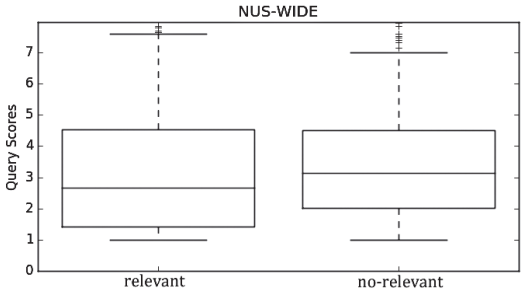

图 3 基于Google Distance的标签语义距离

图1、图2、图3箱型图中的横线表示两组数据对应的中位数，子主题变化时的语义距离中位数均高于子主题不变时的语义距离中位数，这与假设相符，即图像主题差异越大，图像标签的语义距离越大。同时，对3种算法进行曼-惠特尼U检验，其中基于WordNet的算法检验结果为0.000 000 2，基于维基百科的ESA算法检验结果为0.002，基于Google Distance的算法检验结果为0.000 6，均表明两种主题的得分在统计学意义上差异显著。由此得出标签语义距离能够反映图像主题差异。

5.3 实验结果与讨论

为评价图像多样化检索的性能，采用子主题覆盖率CR@N来衡量图像多样化检索的质量，表示结果中与图像所覆盖的检索词相关的子主题比例。

$$CR@N=\frac{|\cup_{i=1}^{K} topic(D_i)|}{N_t} \quad (3)$$

公式（3）中K表示结果集中相关图像的个数，D_i表示第i个相关图像，$topic(D_i)$表示D_i对应的子主题，N_t表示与检索词相关的子主题的总数；N表示检索结果总数。

基于WordNet、ESA和Google Distance算法得出的标签语义距离总体结果对比，如表1所示。3种算法对于图像多样化检索的效果都有提升。其中，基于WordNet标签语义距离的多样化检索结果在CR@10这一指标下提升最大（18.59%），在CR@5指标下提升最少（8.72%），在该算法下所有指标平均提升15.10%；基于ESA标签语义距离的多样化检索结果在CR@10这一指标下提升最大（8.91%），这与基于WordNet标签语义距离的最小提升效果（CR@5）几乎持平，在CR@20指标下提升最少（3.60%），该算法所有指标平均提升6.48%；基于Google Distance标签语义距离的多样化检索结果在CR@5这一指标下提升最多（10.22%），在CR@20指标下提升最少（0.43%），该算法平均提升6.23%。

表 1 基于标签语义距离的图像多样化检索结果对比

指 标	Tagir	WordNet	ESA	Google Distance
CR@5	0.262	0.285 （8.72%）	0.279 （6.34%）	0.289 （10.22%）
CR@10	0.354	0.420 （18.59%）	0.385 （8.91%）	0.373 （5.36%）
CR@20	0.449	0.517 （15.18%）	0.465 （3.60%）	0.451 （0.43%）
CR@30	0.503	0.580 （15.15%）	0.541 （7.54%）	0.538 （6.84%）
CR@40	0.537	0.633 （17.88%）	0.569 （6.00%）	0.580 （8.29%）

注：括号中数值为新算法较Tagir算法提升的百分比。

实验显示，本文提出的基于标签语义距离的算法在图像多样化检索中具有较好的效果。总体而言，基于WordNet标签语义距离的多样化检索算法略优于基于ESA和Google Distance距离的标签语义距离多样化检索算法，产生这种效果差异的原因之一在于图像本身特征。图像多样化重点在于图像子主题间的多样性，WordNet包含上下位关系、同义关系、反义关系及部分关系等，这种语义网络构建方式更适用于表达子主题的多样化关系；而Google Distance侧重于共现关系，ESA中语义关系不够丰富。导致效果差异的另一个原因是信息噪声。基于WordNet的标签语义距离算法采用人工构建的本体词典，而ESA采用维基百科知识库，Google Distance采用互联网资源，相比较而言，WordNet的语义概念关系较准确，信息噪声较小。

6 结语

在图像多样化检索研究方面，基于图像视觉特征的算法受限于语义鸿沟，有一定的局限性，而基于图像上下文的算法也只是简单探讨基于Google Distance的标签语义距离算法在图像多样化检索中的可行性。本文在此基础上，对比分析基于WordNet、ESA和Google Distance语义距离算法在图像多样化检索中的不同表现及特点。实验表明，基于WordNet的语义距离算法语义概念关系较准确，更适用于多样化检索。此外，本文的图像标签处理较简单，今后将着重考虑将图像标签优化与标签语义距离结合，以进一步提升多样化检索效果。

参考文献

[1] JING Y,BALUJA S.Visualrank:applying pagerank to large-scale image search[J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2008,30(11):1877-1890.

[2] LIU Y, MEI T, HUA X S.Crowd Reranking:exploring multiple search engines for visual search reranking[C]//Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval.New York:ACM,2009:500-507.

[3] 田新梅.基于内容的图像搜索重排序研究[D].合肥:中国科学技术大学, 2010.

[4] 冯晓华,陆伟,张晓娟.检索结果多样化研究综述[J].情报学报,2015,34(7):776-784.

[5] ZWOL R V,MURDOCK V,PUEYO L,et al.Diversifying image search with user generated content[C]//Proceedings of the 1st ACM international conference on Multimedia information retrieval.New York:ACM,2008:67-74.

[6] SMEULDERS A W M,WORRING M,SANTINT S,et al.Content-based image retrieval at the end of the early years[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2000,22(12):1349-1380.

[7] YANG K,WANG M, HUA X S,et al.Tag-based social image search:Toward relevant and diverse results[M]//HOI S C H,LUO J,BOLL S,et al.Social Media Modeling and Computing. London:Springer London, 2011:25-45.

[8] MISLOVE A,KOPPULA H S,GUMMADI K P,et al.Growth of the flickr social network[C].Proceedings of the 1stWorkshop on Online Social Networks.[S.l.]:[s.n.],2008:25-30.

[9] 蒋盛益,陈东沂,王连喜,等.国内外社会化标签挖掘研究综述[J].图书情报工作,2014,58(21):136-145.

[10] NOV O,CHEN Y.Why do people tag?Motivations for photo tagging[J]. Communications of the ACM,2010,53(7):128-131.

[11] SONG K,TIAN Y,GAO W,et al.Diversifying the image retrieval results[C]// Proceedings of the 14th Annual ACM International Conference on Multimedia.New York:ACM,2006:707-710.

[12] PAGE L,BRIN S,MOTWANI R,et al.The PageRank citation ranking: bringing order to the web[J].Stanford Infolab,1999,9(1):1-14.

[13] VAN LEUKEN R H,GARCIA L,OLIVARES X,et al.Visual diversification of image search results[C]//Proceedings of the 18th International Conference on World Wide Web.New York:ACM,2009:341-350.

[14] ZHAO W L,NGOC W.Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection[J].IEEE Transactions on Image Processing,2009,18(2):412-423.

[15] 朱胜平.基于内容的多样性植物图像检索技术研究[D].厦门:华侨大学, 2014.

[16] 崔超然,马军.一种结合相关性和多样性的图像标签推荐方法[J].计算机学报,2013,36(3):654-663.

[17] 田枫,沈旭昆,周凯.一种大规模图像多样化语义标注方法[J].系统仿真学报,2014,26(9):2085-2090.

[18] KIM E,YAMAMOTO T,TANAKA K.Computing Tag-Diversity for Social Image Search[M]//TUAMSUK K,JATOWT A,RASMUSSEN E.The Emergence of Digital Libraries—Research and Practices.New York:Springer International Publishing,2014:328-335.

[19] PEDERSEN T,KOLHATKAR V.WordNet:SenseRelate:AllWords:a broad coverage word sense tagger that maximizes semantic relatedness[C]// Proceedings of Human Language Technologies:The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics,Companion Volume:Demonstration Session. Madison:Association for Computational Linguistics,2009:17-20.

[20] BOLLEGALA D,ISHIZUKA M,MATSUO Y.Measuring semantic similarity between words using web search engines[C]//Proceedings of

World Wide Web.[S.l.]:[s.n.],2007:766.

[21] SUN A,BHOWMICK S S,NGUYEN K N,et al.Tag-based social image retrieval:an empirical evaluation[J].Journal of the Association for Information Science and Technology,2011,62(12):2364-2381.

[22] KHARAZMI S,SANDERSON M,SCHOLER F,et al.Using score differences for search result diversification[C]//Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval.New York:ACM,2014:1143-1146.

[23] 夏召强.面向互联网社会化图像的标签优化算法研究[D].西安:西北工业大学,2014.

作者简介

张震宇，男，1992年生，硕士研究生，研究方向：信息检索，E-mail：137263109@qq.com。
丁恒，男，1988年生，博士研究生，研究方向：知识挖掘，E-mail：hengding@whu.edu.cn。
王瑞雪，女，1991年生，博士研究生，研究方向：信息检索，E-mail：815247188@qq.com。
陆伟，男，1974年生，副院长，教授，博士生导师，研究方向：信息检索与数据挖掘，E-mail：weilu@whu.edu.cn。

Image Diversity Retrieval Based on Semantic Distance of Tags

ZHANG ZhenYu[1,2], DING Heng[1,2], WANG RuiXue[1,2], LU Wei[1,2]
(1. School of Information Management, Wuhan University, Wuhan 430072, China;
2. Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China)

Abstract: With the explosive growth of image resources on the internet, the need for image diversity retrieval is becoming stronger. Comparing the visual characteristics and context description using the image diversity retrieval, this paper discusses the effect of different semantic distance algorithms (WordNet、ESA and Google Distance) in image diversity retrieval. At the same time, an image diversity retrieval algorithm has been provided in this paper, which based on the best of the semantic distance algorithms.

Keywords: Semantic Distance of Tags; Social Tag; Image Diversity Retrieval; Semantic Similarity

（收稿日期：2016-12-22）

书 讯

《汉语主题词表》（工程技术卷）

《汉语主题词表》自1980年问世以后，经1991年进行自然科学版修订，在我国图书情报界发挥了应有的作用，曾经获得了国家科学技术进步二等奖。为了适应网络环境下知识组织与数据处理的需要，2009年由中国科学技术信息研究所主持，并联合全国图书情报界相关机构，完成《汉语主题词表（工程技术卷）》的重新编制工作。

全书共收录优选词19.6万条，非优选词16.4万条，等同率0.84。在体系结构、词汇术语、词间关系等方面进行改进创新。为了方便工程技术领域不同专业用户使用，《汉语主题词表》（工程技术卷）按专业分13个分册出版，同时建立《汉语主题词表》服务系统，提供在线概念检索和辅助标引服务，通过可视化技术展示各类概念关系，是图书馆、档案馆、出版社、期刊杂志社、文献信息中心等专业工作者及科研、教育及工程技术领域人员必备的参考书。

《汉语主题词表（工程技术卷）》已于2014年由科学技术文献出版社出版，全书2 300余万字，总定价3 880元，可分册购买。