

互联网社群图像标签排序研究进展

吴焰樟¹ 刘宏哲¹ 冯松鹤² 袁家政¹ 张静怡¹

(北京联合大学北京市信息服务工程重点实验室 北京 100101)¹

(北京交通大学计算机与信息技术学院 北京 100044)²

摘要 互联网社群图像标签排序是目前计算机视觉、机器学习等领域最热门的课题之一。图像标签序列的合理性直接影响到图像检索等应用的效果。目前图像标签排序的方法多种多样,根据标签排序方法的不同将其划分为基于语义相关度与基于视觉显著性的标签排序,着重介绍了两类方法的典型标签排序方法,分析其各自的优缺点。最后就图像标签排序的评价方法以及发展趋势做了简单的论述。

关键词 标签排序,相关性,显著性

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2015.8.004

Advances in Tag Ranking for Internet Social Images

WU Yan-zhang¹ LIU Hong-zhe¹ FENG Song-he² YUAN Jia-zheng¹ ZHANG Jing-yi¹

(Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China)¹

(School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)²

Abstract Tag ranking for internet social images is one of the most popular topics in the computer vision and machine learning. The effect of image retrieval and other applications is directly affected by the reasonableness of the order of image tag. Currently, the existing methods on tag ranking are various. This paper divided them into relevance-based tag ranking and saliency-based tag ranking. This paper highlighted two typical image tag ranking ways and analyzes their advantages and disadvantages respectively. Finally, we discussed the evaluation methods and trends of image tag ranking simply.

Keywords Tag ranking, Relevance, Saliency

1 引言

随着科学技术的发展以及各类图像、视频采集设备的不断普及,近几年来社交媒体共享网站 Facebook、Google 的社交视频分享网站 YouTube 以及 Yahoo 的社交图像分享网站 Flickr,每天都会产生容量相当大的图像数据。在很多社交媒体网站中,网络用户可以为自已分享的图像信息添加关键字,这些关键字被称为图像标签(Tag)。有了这些丰富图像的标签信息,网络用户可以很方便地组织和分享使用图像信息^[1-4]。其中,图像标签是网络用户对图像的语义理解,这些标签的作用是为了让用户更快速地检索到它^[5]。但由于网络标注用户存在人为理解的主观性、偏差性以及用户的关注度不同等问题,即使对于同一幅图像,不同的网络用户标注的标签信息也有可能不一样,同时网络用户标注过程中往往忽略了图像内容与图像标签之间的相关度。其中,文献[6]研究指出网络用户对图像标注的标签不精确且仅有50%的标签与

图像内容相关,并且有50%的标签被用户遗漏或者与图像内容相关度甚小。文献[7]随机从 Flickr 图像中抽取 1200 幅图像,每一幅图像至少存在 10 个标签,同时对这些图像最相关的标签出现在标签列表中的位置进行统计,只有 10% 的图像将最相关的标签放置于图像标签列表顶部,因此得出大部分情况下标签随机出现在整个标签序列中的某个位置。

图像标注的标签与图像内容之间的相关度不大、标签精确性不高等问题,严重影响到社群图像处理方面的应用,例如图像检索^[8,9]、图像分类^[10]、图像显著性区域检测^[11]等应用。因此,如何利用社群图像的语义相关性对图像标签进行处理引起了很大关注^[12],目前已有很多研究机构或者个人对此进行研究。而本文主要针对图像标签排序问题进行综述,目前已有的图像多标签排序算法大致可以分为两类^[24]:基于语义相关度的多标签排序算法(Relevance-Based Tag Ranking)与基于视觉显著性的多标签排序算法(Saliency-Based Tag Ranking)。其中基于语义相关度的多标签排序算法主要考虑

到稿日期:2014-08-20 返修日期:2014-10-07 本文受国家自然科学基金(61372148,61271369,41101111),北京市教育委员会科技发展计划面上项目(SQKM201411417004),北京联合大学人才强校计划人才资助项目(BPHR2014A04,BPHR2014E02),北京市属高等学校创新团队建设与教师职业发展计划项目(CIT&TCD20130513,IDHT20140508)资助。

吴焰樟(1990-),男,硕士生,主要研究方向为数字图像处理,E-mail:wuyanzhang2013@buu.edu.cn;刘宏哲(1971-),女,副教授,主要研究方向为语义计算、数字图像处理、分布式系统、人工智能、数字博物馆等,E-mail:liuhongzhe@buu.edu.cn(通信作者);冯松鹤(1981-),男,博士,副教授,主要研究方向为图像语义检索及自动标注,E-mail:shfeng@bjtu.edu.cn;袁家政(1971-),男,博士,教授,主要研究方向为图形图像处理、文物遗迹的数字化处理、数字博物馆、导航定位等,E-mail:xxtjiazheng@buu.edu.cn。

图像标签与图像内容之间的相关度,通过相关度对图像标签进行排序。这类算法主要有两种代表性的工作,即基于统计模型的多排序算法与基于数据驱动的多排序算法。而基于视觉显著性的图像多标签排序算法主要利用视觉显著性提取算法求解图像的显著性区域,然后根据图像显著性区域的显著度对图像标签进行排序。

2 图像标签排序

图像标签排序问题在2009年首次被提出,在多媒体信息领域引起了广泛关注,国内外研究机构以及学者对此进行了大量的研究,并提出了一系列的图像标签排序方法。其中基于语义相关度的多标签排序算法为目前排序方法中较为主流的一类方法。其原理与思想较为简单:首先计算图像标签与图像内容之间的相关度,然后按照标签的相关度大小降序排序。相对于基于语义相关度的多标签排序算法,基于视觉显著性的多标签排序算法原理与思想相对比较复杂,首先需要计算图像中对象的显著值,然后根据对象的显著值获取对象相对应的标签信息,最后按照对象显著值的大小对标签进行降序排序。

基于语义相关度的多标签排序算法主要根据图像标签与图像内容之间的相关度高低进行排序。其中基于统计模型的标签排序方法主要根据核密度估计(Kernel Density Estimation, KDE)、贝叶斯框架(Bayesian Framework)等思想估计图像中每个标签与图像内容之间的语义相关度。Liu等^[7,13]对标签排序问题做出了先驱性的工作,通过融合核密度估计(KDE)和基于标签图的随机游走算法(Random Walk)实现图像标签排序。该算法首先利用核密度估计算法估计图像标签的相关度得分,然后利用随机游走算法进一步修正相关度得分,最后根据进一步修正后的相关度得分大小对图像原始标签进行排序。Sun等^[15]提出了一种自动标签排序算法,其考虑了图像视觉内容的相关性以及图像与图像之间的关联性。在基于数据驱动的多标签排序算法中,Li等^[16]利用基于整幅图像近邻投票(Entire Image based K Nearest Neighbors Voting, EKNN)的思想对图像标签进行重排序,即利用测试图像搜索得到的 k 个最近邻图像的标签对测试图像的标签进行投票,然后根据测试图像标签票数的大小进行降序排序。针对文献^[16]中只考虑了图像单一视觉特征的现象,Li等^[17]则通过学习图像的多种视觉特征得到图像标签的语义相关度实现图像的标签排序;Zhuang等^[18]提出了一个二维(文本与视觉)的学习方法计算图像标签与图像视觉内容之间的相关性;Li等^[19]则通过标签图表(Tag Graphs)与图像图表(Image Graphs)相关性传播算法实现图像标签排序。

基于视觉显著性的多标签排序算法主要通过使用与改进已有的图像显著性计算方法计算图像中对象的显著值,根据显著值的不同实现图像的多标签排序。李旻先等^[20]利用视觉注意力机制计算模型(Visual Attention Model)提取图像的初步显著性区域,然后利用外接矩形的方法提取最终的图像显著性区域,最后利用显著性区域近邻投票的方法对图像标签进行排序;Tang等^[21]提出了一种新颖的基于显著性区域的多标签排序方法,此方法不仅考虑了整幅图像特征,而且还考虑了图像的显著性区域特征;Feng等^[22]通过融合视觉显著性分析及多示例学习算法实现图像标签排序;经过进一步的研究,Feng等^[23,24]提出了一种自适应标签排序算法,该

算法首先利用机器学习技术将图像自动分为场景类与对象类,然后分别利用基于语义相关度与视觉显著性的多排序方法实现图像标签的自动排序。

2.1 基于语义相关度的多标签排序算法

2.1.1 基于统计模型的多标签排序算法

基于概率统计模型^[25-29]的多标签排序算法,其核心在于建立图像视觉特征与图像标签之间的概率共生关系。文献^[7,13]与文献^[15]分别利用核密度估计和贝叶斯估计算法估计图像中每个标签与图像内容的语义相关度,最后根据相关度求解合理的图像标签序列。其中基于核密度估计的图像标签排序算法^[7,13]是图像标签排序算法中一种典型的标签排序算法,为图像标签排序做了先驱性的研究。该算法主要从以下两个方面进行处理:(1)利用核密度估计标签的初始概率;(2)随机游走对图像标签相关度进行修正。给定任一幅图像及其对应的标签信息,首先利用核密度估计算法对图像的标签进行概率估计,此时得到的初始概率并没有考虑此幅图像标签列表中不同标签之间的相关性,随后利用随机游走算法对得到的初始标签概率进行进一步修正,最后根据修正加强后的标签概率对标签进行重排序。图1给出了该算法的框架。

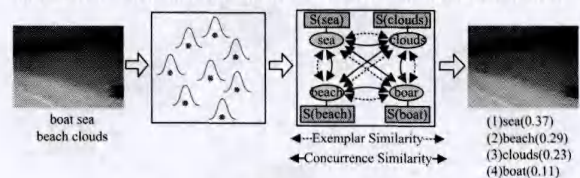


图1 基于核密度估计的标签排序算法

受Liu等^[7,13]提出的标签排序算法思路的启发,同时结合Zhao等^[14]提出的利用多个搜索引擎来计算标签与标签之间的关系,Sun等^[15]提出了一种自动标签排序框架,其不仅考虑了图像自身视觉内容的相关性,而且还考虑了图像与图像之间的关联性。首先,对于一个查询样例标签,利用多种搜索引擎工具搜索含有此标签的图像,例如百度、Google、Bing等搜索引擎工具,以确保覆盖查询样例标签的语义空间;其次,结合图像的局部特征与全局特征相似性,利用贝叶斯网络估计算法计算标签的初始概率;然后,构建标签相似度图表来缩小标签之间的关系;最后,利用随机游走算法对标签进行修正,并以相关度得分降序输出最终排序结果。其基本过程如图2所示。

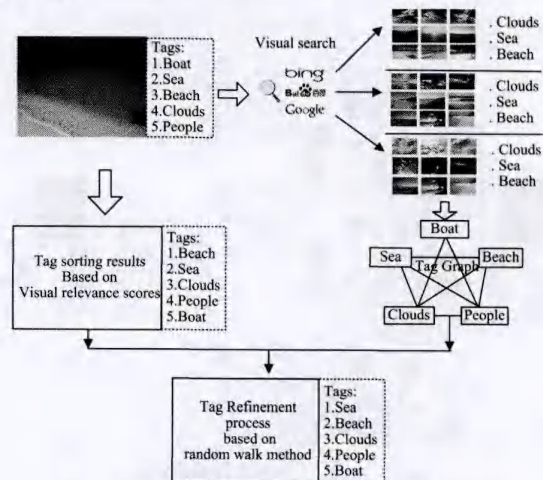


图2 基于贝叶斯估计的多标签排序算法框架

2.1.2 基于数据驱动的多标签排序算法

基于数据驱动的多标签排序算法,通过简单的图像底层视觉特征匹配得到给定图像的 K 近邻图像子集,通过近邻投票策略统计其测试图像中每个标签出现的频度,最后按照频度实现图像的标签排序。其中基于整幅图像的近邻投票多标签排序算法是基于语义相关度的多标签排序算法中另一类典型的算法。Li 等^[16]通过计算图像标签与图像内容的相关性,将社群图像的原始标签序列进行重排序。该算法首先提取图像的全局底层视觉特征;然后用图像相似度比较算法找到给定图像的 k 个视觉最近邻图像;其次根据 k 个视觉近邻图像的标签对给定图像的标签进行投票学习每个标签的相关度;最后按照每个标签相关度得分降序排列得到图像标签的排序结果。图 3 给出了该算法的大致过程,对于给定的种子图像(Seed Image),根据全局视觉底层特征找到种子图像的 8 个近邻图像,利用近邻图像的标签对种子图像标签进行投票,得到“tiger”的票数为 4,“cat”的票数为 2,“wildlife”的票数为 1,其他标签票数为 0,其中“tiger”标签票数高于其他标签的票数,因此可以将标签“tiger”放在标签列表的首位。

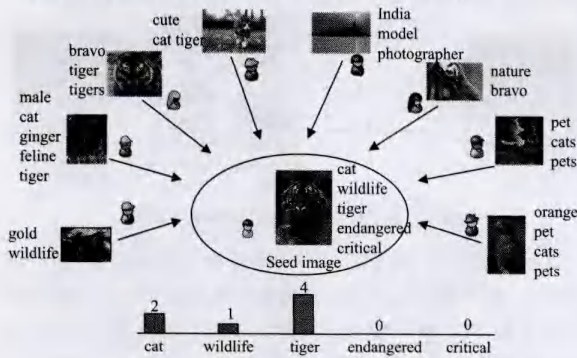


图 3 基于整幅图像的近邻投票标签排序算法

Li 等^[16]提出的基于整幅图像的近邻投票多标签排序算法中,在计算标签相关度时只使用了图像单一底层视觉特征。单一视觉特征并不能覆盖图像的整个语义空间,因此为了尽可能更全面地覆盖图像的语义空间,进一步提高图像标签相关度的精确度,Li 等^[17]对文献^[16]中的方法进行了改进,结合图像多种视觉特征以及模型参数,获取图像标签不同视觉特征下的相关度,并结合不同特征的相关度得分得到图像标签的最终相关度得分。具体相关度学习过程如图 4 所示。

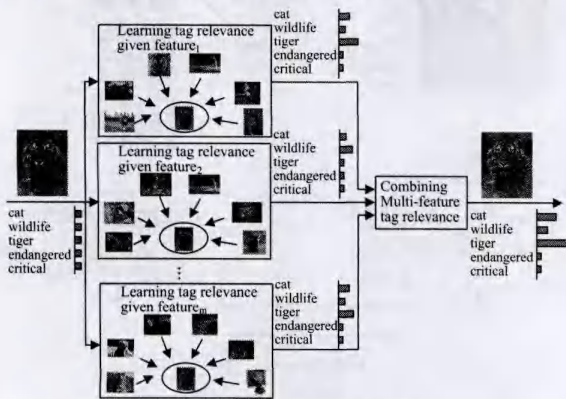


图 4 多特征标签相关度学习算法

除了上述基于数据驱动的多标签排序算法,Zhuang

等^[18]提出了一个二维(文本与视觉)的学习方法计算图像标签与图像视觉内容之间的相关性,该算法不同于传统学习算法需要设置大量的模型参数,属于纯数据驱动,不需要训练或者假设大量的隐藏模型参数。传统的排序算法一般采用启发式对图像标签进行处理,并没有考虑参数的最优化选择,因此较难得到一个理想的排序效果。Zhuang 等^[18]将本方法看作是最优化选择问题,计算与选择一个合理的标签与图像的相关性矩阵实现图像的标签排序,避免了上述存在的部分问题。Li 等^[19]提出了一种纯数据驱动的多标签排序方法,该算法首先对于图像库中的每一幅图像都建立相对应的标签图表(Tag Graphs),某一幅图像的标签图表中的标签初始概率通过随机游走算法计算得到;然后融入 K 近邻稀疏表示算法(KNN-Sparse)建立图像图表(Image Graphs)传播图像标签之间的相关性;最后通过标签的相关度得分对图像的标签实现降序排序。

无论是基于概率统计模型,还是基于数据驱动的图像多标签排序,其本质都是计算标签与图像内容之间的相关性,都采用了图像的全局底层视觉特征,并没有考虑图像的局部区域特征,因此在图像相似度匹配上,无法达到较为理想的效果;同时这类算法对于大规模的图像数据准确率较高,而对于小规模的数据其排序的结果并不理想。为了克服这一类算法的缺点,国内外研究机构以及学者通过图像分割、显著性区域提取等思路对图像进行处理,并分析各区域的显著性高低,实现图像标签的自动排序。

2.2 基于视觉显著性的多标签排序算法

2.2.1 视觉显著性区域检测

在图像检索、图像分割等领域中,网络用户通常只对图像中部分区域感兴趣,而感兴趣区域恰恰其视觉显著性也更高,代表了用户的查询、分割意图。基于视觉显著性的多标签排序算法的核心主要是找到最能吸引用户兴趣的区域,此类算法一般使用与改进现有的 Itti 视觉注意力计算模型或者灰度分布直方图(HC-Map)的方法进行图像的显著性分析。

根据对视觉注意力形成机制不同的研究,心理学家将人类的视觉注意力分为两种:一种是自底向上(Bottom-Up),另一种则是自上而下(Top-Down)的注意力机制。前者为纯数据驱动,通过分析图像的底层视觉特征,不借助于任何先验知识获取图像的显著性区域;后者则需要通过添加人类的先验知识进行学习,是有知识和任务驱动的。在现有的自底向上注意力机制的显著性提取模型中,最具有代表性的是 Itti 等^[30]提出的基于视觉注意力机制的计算模型方法。首先,提取图像的底层视觉特征,包括图像的亮度、颜色和方向等底层视觉特征图;其次,对特征图进行中央周边差分降采样,得到亮度、颜色和方向的特征图集合图像;然后,将特征图归一化线性组合得到初步的显著性区域图像;最后,采用胜者为王(Winner-Take-All, WTA)的策略产生图像最终的显著性区域图像。针对上述计算量较大、时间较长等问题,Frintrop 等^[31]使用了平方滤波器计算中央周边差,并且利用积分图像加速过程。Feng 等^[32]提出了一种基于三层高斯金字塔和对数理论的注意力模型,算法既考虑到了显著图生成的有效性,又节省了系统的开销,并将此算法应用到了图像标签排序算法当中。Cheng 等^[33]通过直方图对比度(Histogram-based Con-

trast, HC-Map)的方法实现图像的显著性检测,该算法简单高效,同时生成了具有全分辨率的显著度图。另一种则是自上而下的注意力机制不同形成的。该类算法通过学习获取的先验知识进行显著性分析,属任务驱动式,较数据驱动方式提取显著图速度慢,因此很少用于图像标签排序领域。

2.2.2 基于视觉显著性近邻投票的标签排序算法

目前大部分存在的标签排序算法根据整幅图像之间的关系进行标签处理,并没有考虑图像中最能够表达语义含义的显著性区域。鉴于Itti等^[30]提出的视觉注意力计算模型更注重模拟人眼的视觉注意力转移轨迹,Li等^[20]通过利用现有的Itti视觉注意力计算模型得到图像的初步显著性区域图像,然而得到的显著性区域一般都比较分散,而且被提取出来的显著性区域的形状也比较随意。为了更好地提取图像的显著性区域,Li等^[20]使用外接矩形提取图像的显著性区域,并在提取的过程中使用了若干次迭代。显著性区域提取过程如图5所示。



图5 显著性区域提取过程

受文献^[16,17]提出的基于整幅图像的K近邻图像投票方法的启发,Li等^[20]指出利用整幅图像的K近邻图像对图像标签进行排序是不够准确的,故利用显著性区域图像的K近邻显著性区域图像的方法对图像标签进行重排序。投票大致过程可描述为:首先,读取测试图像及其标签序列,并提取测试图像的显著性区域图像;然后利用测试图像的显著性区域图像获取k个近邻显著性区域图像;最后,利用K近邻显著性区域图像的标签序列对测试图像的标签序列进行投票,并对投票结果进行处理后得到最终的标签相关性。图像标签序列的顺序则按照图像最终的标签的相关性进行降序排列得到图像的排序后的标签序列。具体的标签投票过程如图6所示。

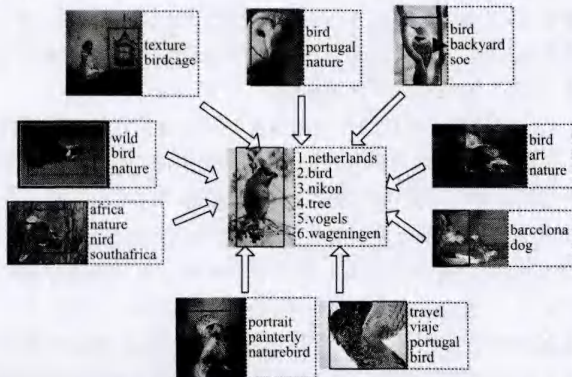


图6 基于显著性的近邻标签投票示意图

在此基础上,Tang等^[21]提出了一种新颖的基于显著性区域的多标签排序方法,此种方法不仅考虑了整幅图像特征,而且还考虑了图像的显著性区域特征。相比于文献^[18]中提

出的一种通过传播标签与图像图表的标签排序方法,Tang等^[21]融入了视觉注意力机制计算模型,不仅如此,还结合了Zha等^[34-36]与Yang等^[37]提出的基于图表的学习算法实现图像的标签排序。该算法同时考虑了图像局部特征以及全局特征来解决图像标签排序问题,其大致过程主要包括:首先,利用Itti等^[30]提出的视觉注意力机制计算模型获取测试图像的显著性区域图像;然后分别建立显著性区域级(Region-level Graph)、图像级(Image-level Graph)两个图表,通过两张图表来估算图像标签的相关度;最后根据图像标签的相关度得分对标签进行排序。其大致算法过程如图7所示。

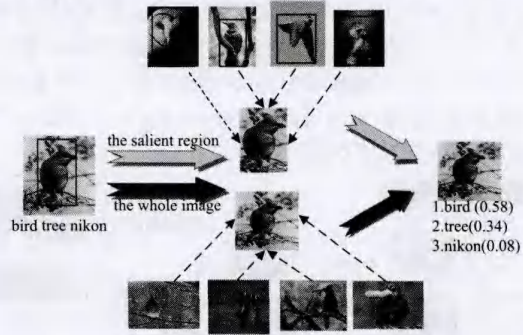


图7 基于显著性区域传播算法的图像标签排序

2.2.3 融合视觉显著性及多示例学习的多标签排序算法

文献^[20,21]提出的基于显著性区域的近邻投票多标签排序算法中只考虑了图像中某个感兴趣区域问题,并没有考虑图像中除去显著性区域的那一部分子区域,同时网络用户标注的标签往往标注整幅图像,而非图像中的每个对象上。鉴于此,Feng等^[22]通过将多示例学习融入到显著性标签排序算法中,以进一步改善上述存在的问题。如图8所示,对于给定任一幅图像,首先采用图像分割算法对测试图像进行分割,同时利用视觉注意力模型提取图像的显著性区域图;然后利用多示例学习的经典算法即DD算法将图像标签从图像级向区域级传播^[38,39],完成区域标注(Label-to-Region)的工作;最后,利用显著性区域图,将标签列表按照标签对应的视觉显著度高低进行降序排列,从而得到最终的标签排序结果。

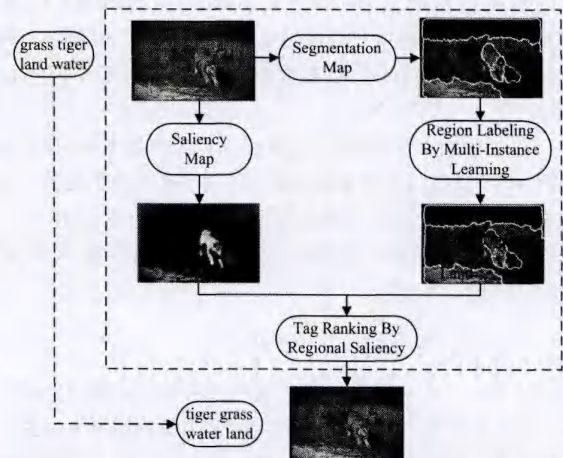


图8 融合视觉显著性及多示例学习的多标签排序

2.2.4 自适应的标签排序方法

Feng等人经过进一步的研究,发现可以将图像分为场景

类图像与对象类图像,然后分别利用基于语义相关度与基于视觉显著性的方法进行标签排序,其中具有显著性对象的图像称作对象类图像,没有显著性对象的或者前景与背景比较接近的图像称为场景类图像。Feng 等^[23,24]通过改进现有的 Itti 视觉注意力机制计算模型,提出一种基于三层高斯金字塔和对比理论的注意力模型算法计算图像的显著性区域。由于场景类与对象类图像的显著性区域图的显著性分布不一样,通过提取显著性图的灰度直方图特征进行训练分类器。对于给定的任意一幅测试图像,首先利用分类器对图像进行自动分类,场景类图像则使用基于语义相关度的多标签排序算法,反之使用基于视觉显著性的多标签排序算法实现图像标签排序;不仅如此,还在利用基于视觉显著性的多标签排序算法中融入了多示例学习学法(Multi-instance Learning, MIL)以及目前比较热门的稀疏表示算法(Sparse Representation)。该算法的大致过程如图 9 所示。

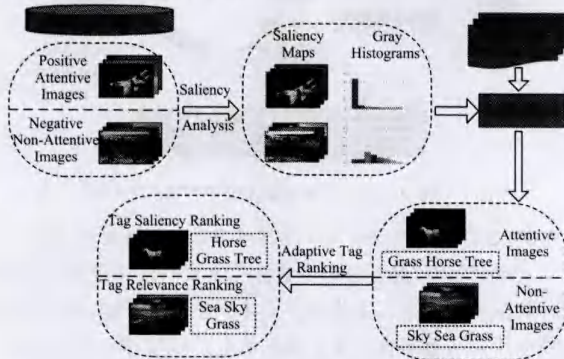


图 9 自适应的图像标签排序

Feng 等人提出的图像标签自适应排序算法综合考虑了基于语义相关度与视觉显著性的多标签排序算法的优点,但是图像文本信息与图像内容之间的“内容鸿沟”、图像底层视觉特征与图像高层语义特征之间的“语义鸿沟”还有进一步研究的空间。

2.3 评价方法

目前已有很多图像标签排序算法,而如何评价图像标签排序算法的好坏成为实验中最重要的一部分。一般采用归一化的折扣累积增益(Normalize Discount Cumulative Gain, NDCG)^[40]、平均精度(Mean Average Precision, MAP)^[41]指标对实验结果进行评价,实验数据集一般有 Corel5K、MSRC、NUS-WIDE、ECCV5K。

平均精度(MAP)指标主要用于文档信息检索中,单个主题的 MAP 是每篇相关文档检索出后的准确率的平均值。在文献[19]中, π^* 表示文档的标准序列(Ground-Truth Ranking), π 表示利用相关度指标 r 检索到的排序结果,则 MAP 指标可以定义为:

$$MAP(\pi^*, \pi) = \frac{1}{rel} \sum_{j: \pi_j^* = 1} Prec@j \quad (1)$$

其中, $rel = \{i: \pi_i^* = 1\}$ 代表检索到相关文档的个数, $Prec@j$ 表示第 j 个相关文档位置与本文档在所有返回结果 π 中的位置百分比。观察式(1)可以看到其合理性:不仅考虑了相关文档的位置信息,同时考虑了返回相关文档的数目。

MAP 指标虽然考虑了文档的位置信息,但没有考虑相关度的高低,在很多图像标注、图像标签排序体系中,常常关心

的是相关度的高低,有另外一部分研究者使用归一化折扣累积增益(NDCG)指标对图像标签排序结果进行评价,主要步骤为:首先,从图像数据库中随机选择 N 张图像作为测试图像,并请多人对这 N 张图像的标签与图像的内容进行相关度标记。在文献[7,13]中,每个标签与图像的相关度被分为 5 个等级:最相关(分数 5)、相关(分数 4)、部分相关(分数 3)、一点点相关(分数 2)、相关(分数 1),相关度等级也可以设置为其他的,文献[42]则将相关度划分为 4 个等级。其次,利用标签排序算法对选取的 N 张图像的原始标签序列进行排序。最后,利用 NDCG 评价指标对排序结果进行评价。例如,给定一张图像排序之后的标签序列为: t_1, t_2, \dots, t_n , 其 NDCG 值计算过程如下:

$$N_n = Z_n \sum_{i=1}^n \frac{2^{r(i)} - 1}{\log(1+i)} \quad (2)$$

其中, i 为第 i 个标签的次序, $r(i)$ 为第 i 个标签与当前图像内容的相关度, Z_n 为 NDCG 归一化的常数, N_n 的取值范围为 $[0, 1]$, 当 N_n 为 1 时为最佳的排序结果。

结束语 图像标签排序问题一直是计算机视觉、图像检索、机器学习等领域的研究热点与难点,近年来国内外研究机构或者学者对此进行了大量的实验研究,已经获得了一定程度的发展,并提出了一些有意义的算法:Liu Dong 等人利用了核密度估计(KDE)和随机游走算法(Random Walk)对图像标签进行了重排序;Li Xirong 等人通过近邻投票(KNN)的思想实现图像的标签排序;鉴于 Li Xirong 等人提出的算法只考虑了图像的全局特征,并没有关注图像的局部特征, Li Minxian 等人提出了基于显著性区域的近邻投票的策略实现图像标签排序;Feng 等人通过进一步研究,结合基于语义相关度与基于视觉显著性的多标签排序算法各自的优点,提出了自适应的图像多标签排序算法。上述这些方法都从一定程度上解决了图像标签乱序的问题,但是离理想效果还有一定的差距,其中基于语义相关性的图像标签排序方法在很大程度上依赖于图像数据的规模。对于大规模的图像数据,此类算法的实现效果比较好,但对于小规模图像数据,其得到的排序结果并不理想。而基于显著性区域标签排序算法通过分析图像的显著性区域实现图像的标签排序,图像数据的规模大小对于此类算法影响较小,但对于图像中并没有显著性区域的图像,排序的结果显然不能很好地解决图像标签乱序的问题。而 Feng 等人通过结合基于相关性与显著性的排序算法的优点,实现了图像的标签自动排序。

众多计算机领域的专家和学者们对图像标签排序问题做出了很大的贡献,他们从不同的视角利用不同的方式提出了自己的见解。本文通过阅读有关图像标签排序问题的重要文献,分析了现有的社群图像标签排序方法:大部分基于语义相关性、视觉显著性的图像标签排序算法针对的是图像标签语义特征、图像底层视觉特征分别进行研究,而很少有算法将图像高层语义特征融入到基于图像底层视觉特征、图像标签语义特征的图像标签排序算法中对图像标签进行处理,同时图像文本信息与图像内容之间的“内容鸿沟”、图像底层视觉特征与高层语义特征之间的“语义鸿沟”并没有得到彻底的解决。因此基于图像底层视觉特征以及高层语义特征的社群图像标签排序还有进一步研究的空间及发展前景。

参考文献

- [1] Tang J, Yan S, Hong R, et al. Inferring semantic concepts from community-contributed images and noisy tags[C]//Proceedings of the 17th ACM international conference on Multimedia. ACM, 2009:223-232
- [2] Xiang Y, Zhou X, Chua T S, et al. A revisit of generative model for automatic image annotation using markov random fields [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). IEEE, 2009:1153-1160
- [3] Zhang S, Tian Q, Hua G, et al. Descriptive visual words and visual phrases for image applications[C]//Proceedings of the 17th ACM international conference on Multimedia. ACM, 2009:75-84
- [4] Xiang Y, Zhou X, Liu Z, et al. Semantic context modeling with maximal margin conditional random fields for automatic image annotation[C]//2010 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2010:3368-3375
- [5] Ames M, Naaman M. Why we tag: motivations for annotation in mobile and online media[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2007: 971-980
- [6] Kennedy L S, Chang S F, Kozintsev I V. To search or to label?: predicting the performance of search-based automatic image classifiers[C]//Proceedings of 8th ACM International Workshop on Multimedia Information Retrieval. ACM, 2006:249-258
- [7] Liu D, Hua X S, Yang L, et al. Tag ranking [C]//Proceedings of 18th International Conference on World Wide Web. ACM, 2009: 351-360
- [8] Li X, Snoek C G M, Worring M. Learning tag relevance by neighbor voting for social image retrieval[C]//Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval. ACM, 2008:180-187
- [9] Feng S, Xu D, Yang X. Attention-driven salient edge(s) and region(s) extraction with application to CBIR[J]. Signal Processing, 2010, 90(1):1-15
- [10] Guillaumin M, Verbeek J, Schmid C. Multimodal semi-supervised learning for image classification[C]//2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2010:902-909
- [11] Wang W, Lang C, Feng S. Contextualizing Tag Ranking and Saliency Detection for Social Images[M]//Advances in Multimedia Modeling. Springer Berlin Heidelberg, 2013:428-435
- [12] Liu D, Hua X S, Zhang H J. Content-based tag processing for internet social images[J]. Multimedia Tools and Applications, 2011, 51(2):723-738
- [13] 刘东. 基于内容的互联网社交图像标签处理与分析[D]. 哈尔滨:哈尔滨工业大学, 2012
Liu D. Content-based tag processing and analysis for internet social images [D]. Harbin: Harbin Institute of Technology, 2012
- [14] Zhao Y, Zha Z J, Li S, et al. Which tags are related to visual content? [M]//Advances in Multimedia Modeling. Springer Berlin Heidelberg, 2010:669-675
- [15] Sun F, Li H, Zhao Y, et al. Towards tags ranking for social images[J]. Neurocomputing, 2013, 120:434-440
- [16] Li X, Snoek C G M, Worring M. Learning social tag relevance by neighbor voting[J]. IEEE Transactions on Multimedia, 2009, 11(7):1310-1322
- [17] Li X, Snoek C G M, Worring M. Unsupervised multi-feature tag relevance learning for social image retrieval[C]//Proceedings of the ACM International Conference on Image and Video Retrieval. ACM, 2010:10-17
- [18] Zhuang J, Hoi S C H. A two-view learning approach for image tag ranking[C]//Proceedings of fourth ACM International Conference on Web Search and Data Mining. ACM, 2011:625-634
- [19] Li M, Tang J, Li H, et al. Tag ranking by propagating relevance over tag and image graphs[C]//Proceedings of 4th International Conference on Internet Multimedia Computing and Service. ACM, 2012:153-156
- [20] 李曼先, 赵春霞, 唐金辉, 等. 基于视觉显著性近邻投票的标签排序方法[J]. 南京理工大学学报(自然科学版), 2012, 36(4):561-566
Li M X, Zhao C X, Tang J H, et al. Visual saliency neighbor voting based tag ranking[J]. Journal of Nanjing University of Science and Technology(Natural Science), 2012, 36(4):561-566
- [21] Tang J, Li M, Li Z, et al. Tag ranking based on salient region graph propagation[J]. Multimedia Systems, 2014:1-9
- [22] Feng S, Lang C, Xu D. Beyond tag relevance: integrating visual attention model and multi-instance learning for tag saliency ranking[C]//Proceedings of ACM International Conference on Image and Video Retrieval. ACM, 2010:288-295
- [23] Feng S, Lang C, Li B. Towards relevance and saliency ranking of image tags [C]//Proceedings of the 20th ACM international conference on Multimedia. ACM, 2012:917-920
- [24] Feng S, Lang C, Liu H, et al. Adaptive all-season image tag ranking by saliency-driven image pre-classification[J]. Journal of Visual Communication and Image Representation, 2013, 24(7): 1031-1039
- [25] Barnard K, Duygulu P, Forsyth D, et al. Matching words and pictures[J]. The Journal of Machine Learning Research 3, 2003: 1107-1135
- [26] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using crossmedia relevance models[Z]. New York: ACM Press, 2003
- [27] Lavrenko V, Manmatha R, Jeon J. A Model for Learning the Semantics of Pictures[C]//NIPS. 2003, 1:2
- [28] Feng S L, Manmatha R, Lavrenko V. Multiple bernoulli relevance models for image and video annotation[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004
- [29] Monay F, Gatica-Perez D. Modeling semantic aspects for cross-media image indexing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(10):1802-1817
- [30] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259
- [31] Frintrop S, Klodt M, Rome E. A real-time visual attention system using integral images[C]//Proceedings of 5th International Conference on Computer Vision Systems. 2007

- in computer science. Berlin, Germany: Springer Verlag, 1996, 1141:178-187
- [3] Pelikan M, Godberg D E, paz E C. Linkage problem, distribution estimation, and Bayesian networks[J]. *Evolutionary Computation*. 2000, 8(3): 311-340
- [4] Paul T K, Iba H. Linear and combinatorial optimizations by estimation of distribution algorithms[C]//9th MPS Symposium on Evolutionary Computation, IPSJ, Japan, 2002
- [5] 梁玉洁, 许峰. 自适应混合多目标分布估计进化算法[J]. *计算机工程与应用*, 2014, 50(5): 46-50, 207
Liang Yu-jie, Xu Feng. Adaptive hybrid multi-objective estimation of distribution evolutionary algorithm [J]. *Computer Engineering and Applications*, 2014, 50(5): 46-50, 207
- [6] 戚玉涛, 刘芳, 刘静乐, 等. 基于免疫算法和 EDA 的混合多目标优化算法[J]. *软件学报*, 2013, 24(10): 2251-2266
Qi Yu-tao, Liu Fang, Liu Jing-le, et al. Hybrid immune algorithm with EDA for multi-objective optimization[J]. *Journal of Software*, 2013, 24(10): 2251-2266
- [7] 丁有军, 钟声. 基于分布估计算法的连续函数全局优化问题研究[J]. *计算机科学*, 2012, 39(10): 218-219, 223
Ding You-jun, Zhong Sheng. Global Optimization Problem of Continuous Function Based on Distribution Estimation Algorithm[J]. *Computer Science*, 2012, 39(10): 218-219, 223
- [8] 周雅兰, 朱耀辉, 张军. 具有学习机制的离散差分演化算法[J]. *计算机科学*, 2011, 38(7): 225-227, 249
Zhou Ya-lan, Zhu Yao-hui, Zhang Jun. Discrete Differential Evolution with Learning Mechanism[J]. *Computer Science*, 2011, 38(7): 225-227, 249
- [9] 王凌. 智能优化算法及其应用[M]. 北京: 清华大学出版社, 2004
Wang Ling. Intelligent optimization algorithm with applications [M]. Beijing: Tsinghua University Press, 2004
- [10] 李盼池, 李士勇. 求解连续空间优化问题的混沌量子免疫算法[J]. *模式识别与人工智能*, 2007, 20(5): 654-660
Li Pan-chi, Li Shi-yong. A Chaos Quantum Immune Algorithm for Continuous Space Optimization[J]. *PI & AI*, 2007, 20(5): 654-660
- [11] 寇晓丽, 刘三阳, 张建科. 一种随机蚁群算法求解连续空间优化问题[J]. *系统工程与电子技术*, 2006, 28(12): 1909-1911
Kou Xiao-li, Liu San-yang, Zhang Jian-ke. Stochastic ant colony algorithm for continuous space optimization[J]. *Systems Engineering and Electronics*, 2006, 28(12): 1909-1911
- [12] 张锐, 高辉, 张涛. 求解连续空间优化问题的量子差分混合优化算法[J]. *系统工程与电子技术*, 2012, 34(6): 1288-1292
Zhang Rui, Gao Hui, Zhang Tao. Hybrid optimization algorithm based on quantum and differential evolution for continuous space optimization [J]. *Systems Engineering and Electronics*, 2012, 34(6): 1288-1292
- [13] 郭源源, 王谦, 梁峰. 基于粒子群优化算法的车间布局设计[J]. *计算机集成制造系统*, 2012, 18(11): 2476-2484
Guo Yuan-yuan, Wang Qian, Liang Feng. Facility layout design based on particle swarm optimization [J]. *Computer Integrated Manufacturing Systems*, 2012, 18(11): 2476-2484
- [14] 黄敏, 靳婷, 钟声, 等. 基于改进蚁群算法求解连续空间寻优问题[J]. *广西师范大学学报(自然科学版)*, 2013, 31(2): 34-38
Huang Min, Jin Ting, Zhong Sheng, et al. Ant Colony Algorithm for Solving Continuous Function Optimization Problem Based on Pheromone Distributive Function [J]. *Journal of Guangxi Normal University(Natural Science Edition)*, 2013, 31(2): 34-38
- [15] 马卫, 朱庆保. 求解函数优化问题的快速连续蚁群算法[J]. *电子学报*, 2008, 36(11): 2120-2124
Ma Wei, Zhu Qing-bao. Fast Continuous Colony Optimization Algorithm for Solving Function Optimization Problems[J]. *Acta Electronica Sinica*, 2008, 36(11): 2120-2124
- [16] 张腾飞, 王锡淮, 肖健梅. 基于微粒群优化的连续属性离散化算法[J]. *计算机工程*, 2006, 32(3): 44-46
Zhang Teng-fei, Wang Xi-huai, Xiao Jian-mei. Algorithm for Discretization of Continuous Attributes Based on Particle Swarm Optimization[J]. *Computer Engineering*, 2006, 32(3): 44-46
- [17] 高尚, 杨静宇. 群智能算法及其应用[M]. 北京: 中国水利水电出版社, 2006
Gao Shang, Yang Jing-yu. Swarm intelligence algorithm and its application[M]. Beijing: China Water and Power Press, 2006

(上接第 27 页)

- [32] 冯松鹤. 面向感知的图像检索及自动标注算法研究[D]. 北京: 北京交通大学, 2009
Feng S H. Research on perception oriented image retrieval and automatic image annotation [D]. Beijing: Beijing Jiaotong University, 2009
- [33] Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2011: 409-416
- [34] Zha Z J, Wang M, Zheng Y T, et al. Interactive video indexing with statistical active learning[J]. *IEEE Transactions on Multimedia*, 2012, 14(1): 17-27
- [35] Zha Z J, Yang L, Mei T, et al. Visual query suggestion[C]//Proceedings of 17th ACM International Conference on Multimedia. ACM, 2009: 15-24
- [36] Zha Z J, Hua X S, Mei T, et al. Joint multi-label multi-instance learning for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2008 (CVPR 2008). IEEE, 2008: 1-8
- [37] Yang Y, Yang Y, Shen H T. Effective transfer tagging from image to video[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 2013, 9(2): 14
- [38] Chen Y, Bi J, Wang J Z. MILES: Multiple-instance learning via embedded instance selection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 1931-1947
- [39] Yakhnenko O, Honavar V. Multi-Instance Multi-Label Learning for Image Classification with Large Vocabularies[C]//BMVC. 2011: 1-12
- [40] Järvelin K, Kekäläinen J. Cumulated gain-based evaluation of IR techniques [J]. *ACM Transactions on Information Systems (TOIS)*, 2002, 20(4): 422-446
- [41] Baeza-Yates R, Ribeiro-Neto B. Modern information retrieval [M]. New York: ACM press, 1999
- [42] Lan T, Mori G. A Max-Margin Riffled Independence Model for Image Tag Ranking[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE, 2013: 3103-3110