网络图片搜索的打包回复模型

摘要：

图片搜索排序近年来已经成为了一个热门的搜索主题，并改善了现有的基于图片文本数据的网络图片搜索引擎的搜索结果。大量方法用于为搜索结果中的图片重顺序，但很多人讨论这些方案对于与特殊领域的图片查询并不是必选方案，如对象搜索，因为重排算法对整个图片进行处理，而不是图片中的相关部分。在这篇论文中，我们致力于为提供实物搜索的图片搜索引擎提供一种对象打包的回复模型。首先，我们运用一个普通算法从搜索引擎返回的图片元数据结果中查找相关对象。然后，查询返回的图片集会通过派生的重排算法呈现为查询相关对象词汇表的语言模型。因为常规对象查询不可靠的且有很多额外干扰，我们的目的是通过一个线性模型将查找到的对象（例如长度，位置等等）合并到排序结果中，且控制对象属性的权重。包括对象查询的两个网络查询数据库结果集的实验演示了我们的方法优化已有的排序方法。

分类及对象描述：

常规条款：

允许在整个工作过程中为了个人和团体项目需要进行软件或硬件的拷贝而不支付费用是允许的，但不可违背首页的注意事项将拷贝用于经济利益。否则拷贝、出版和用于服务器需提供特殊凭证和费用。

关键词：

图片搜索重排序、监测排重序、对象打包模型、回发模型

1. 简介

现今主要的图片搜索引擎，索引和排序主要基于相关文本信息，这通常导致一些并不准确的结果返回给用户。这是图片内容和文本描述不匹配造成的。图片搜索重排序目标是重新定义基于图片常规内容的文本搜索结果，这在多媒体回复方面已经是一个热门搜索主题。

现有的重排方法几乎全部是基于假设集群和PRF假设。假设集群认为相关图片在视觉上有相似性而不相关图片在视觉上完全不同。这在多数基于相似图形的图片重排算法中被广泛接受。在PRF假设认为文字搜索结果排序在顶部的图片是伪相关的，这可以用来发现一个或多个分类。尽管这些假设被证明在前期工作中是有效的，我们可以看出他们并不是十分合适实物搜索，因为用户可能搜索到包含一个实物（例如汽车）的图片。对于这部分搜索，用户感兴趣的图片可能只与搜索结果部分关联。换句话说，这些图片只有某一部分和搜索有关联而图片的其他区域则是不相关的。并且，簇假设和PRF假设都作用于整张图片，不能有效地从最初的文本搜索结果中检索出有效信息。

场景1展示了“埃菲尔铁塔”的搜索实力。上一行是文本搜索引擎的结果，下一行是PRF假设的排序结果，第一张图片是正确示例。在这个示例中，两个不相关图片被提至顶部，因为图片F和图片D相似于图片A。

我们假设问题主要源于基于以上假设的现有排序方法使用了整张图片视觉上的相似处，如数字3和37的直方图。但是，作为对象搜索，用户倾向于的图片结果包含了搜索对象，包括人脸，标志，动物，建筑物和工业产品，如果图片只有一部分相关于搜索对象就可以说是相关的。所以，基于整张图片的假设太过严格。居于以上示例，我们发现如果排序算法基于对象级别可能会获得更好的结果，所以图片B、C、E包含了埃菲尔铁塔但是任然有不同的观点可供选择。

因此，我们的目的是开发一套对象打包回发模型来代表查询对象并且返回的图片结果作为基于包含对象外贸的语言模型。为了使模型更多地关注与查询相关对象而忽略其他干扰如背景，我们以查询相关对象的单词代表图片和查询语言模型。

为了建立这样一个查询相关对象词汇表，我们扩展了迭代链接分析方法尝试将一组图片与常用对象关联。因为文本搜索的初始排序可更有效得以提供图片关联对象的信息，我们利用这个特性来提升算法，考虑文本搜索排序作为页面排序的先行版本以此区分不同排序位置的图片。尽管具体查询对象这很有效，但负责的情景下无法有效提供信息并可能导致不相关。

为了预估发现查询对象的关联性和证据，我们建立了一组属性，包括位置、尺寸和通常包含在图片中的对象密度的。这些属性被集成到回发模型中以建立线性权重排序算法。权重可以从人类标记的训练集RankSVM中获取。

目标方法以两个公共的网络请求数据集来评估。一个查询包含已命名的人查询，另一个则包含其他对象查询，结果表明打包回发对象模型在所有重排序方法中表现最为出色。在MVP查询方面提升了搜索引擎39.52%的准确性，国家艺术重排方法提升了6.84%。

论文剩余部分按下列方式组织。在第二章中回顾图片搜索重排的相关工作之后，基于对象的图片回发和普通对象。第三章以一个说明性的示例我们描述目标对象打包回发模型。第四章展示了实验成果和方法分析。第五章中，我们以卓越的结论和一系列工作计划结束论文。

1. 相关工作

在这个章节，我们将回顾图片搜索重排的相关工作，基于对象的图片回发和普通对象

索并并对我们论文的成果做出定位。

图片搜索重排。已有的图片重排方法可以大致分为监督类和无监督类。物件都重排方法最多的是居于原始文本搜索结果结构的假设，其中最终名的是假设簇和PRF假设。假设簇认为相关图片在视觉上是相关的而非相关图片在视觉上是非相关的。因为假设可以很自然地在图片结构中展示，大部分基于图片的重排方法都致力于以不同观点阐释这个观点。特殊的是在这些方法中，图片首先作为图形的节点建立，边缘是相似的。图片搜索重排可以通过排序点达到。这些方法的最主要弊端就是视觉相似度是很难建立的。而都以整张图片为图片相似特定为基石，这不能很好地处理对象查询，因为图片相似度应当与相关对象进行比较。第二种PRF假设，严格假设文本搜索排序顶部的图片关联于查询。基于此假设的大量重排模型查找排序在顶部的图片的分类，用这些分类去为图片排序。因为分类基本都是居于整张图片而不是图片中的对象，这些方法在对象查询中的表现差强人意。

我们的排序算法用人类标记来搭建重排模型以便更好的关联用户知觉。不同方法见的最大区别是如何趋势文本结果和不同模型的图片之间的特点关联。重排特点通常是根据作者在图片搜索问题的领域知识手动设计的。我们的重排序方法的特点是通过接受大龄分类不同文本排序的图片重要性自动执行的。我们的目标方法遵循监督排序方法，但两个主要贡献是扩展了已有的工作。首先，我们不像已有方对整张图片进行操作。我们建立了对象模型以便排序过程可悲查询相关对象捕获。第二，在我们的工作中，我们并未尝试合并不同排序特点来创建方法。而是集中于学习打包回发对象模型的参数，而参数表明了对象的无用属性。

基于对象的图片回发。基于对象的图片回发是一个很好的研究点，因为用户通常只提供一个查询图片，而兴趣点则通过植入一个黑盒分辨。在过去不同的方法运用于不同的目的。数据库中的图片对象被标记他憨厚用图片、文本和图形标记。图片被分为小的范围，查询对象基于查询范围用LSA算法建模。图片呈现为打包的常用词汇，用户获取信息回发的语言模型对图片进行排序。作者认为对象上下文也是重要的，常用词汇定位的对象范围外部区域也应在回发过程中被考虑到。我们论文的烟酒店不同于基于对象的图片回发，基于对象的图片回发以关键字区分而不是以含搜索兴趣点的图片区分。因此，我们的问题更具有挑战性因为我们需要基于用户提供的查询推断兴趣对象的图片。

常用对象查找。常用对象查找是现在一个非常重要的主题，用于在一组图片中场找常用对象。不同的常用对象查找可以分为两类，基于分段的方法和基于边界的方法。基于分段的方法，也就是熟知的共同分段，同时为每张图片的常用对象分段。将问题最小化为分割图片为前景色和背景色。在作者的第一段每张图片放到一族像素中，然后通过算法对常用前景色分段。这些方法的最大缺陷有两个，我们的研究中限制它：首先，大多数的共同分段方法是让图片无背景，这对复杂背景的网络图片可能无效。其次，共同分段方法都非常浪费时间，通常需要花费数小时对几百张图片进行处理。

边界划分法目的是以边界形式找到常用对象。这种方法建立一系列假设的ROI。在某种情况下任意范围被划分出来，常用对象查找转变为在CRF中查找最佳的格局。连接查询算法运用于全部ROI中来查找常用对象的中心。对图片搜索排序而言，因为文本搜索引擎提供的初始的图片序号能在图片与查询对象不相关时提供有用信息，我们倾向于结合出事排序信息来建立查询相关对象的关键字。

3.方法

我们的方法始于建立查询相关对象关键字，最小化文本搜索引擎返回的图片中的查询相关对象。然后对象打包回发模型根据语言模型方法为回发信息定制排序问题。为了比较对象打包回发模型中每个语言模型的排序值，我们根据相关对象查询的结果预估了图片和查询模型。最后，我们展示了在回发模型中取得参数的方法。

3.1建立查询相关对象词汇

查询相关对象是不同的对象实例或相关连对象的不同外貌。举例来说，查询“埃菲尔铁塔”，查询相关对象可能包括不同角度或亮度的埃菲尔铁塔。查询“汽车”可能包括奥迪A6和宝马Q5这种不同的实例。

3.1.1算法回顾

我们先用显著对象检测发检测30个ROI中每张图片的显著区域作为查询相关对象的假设。

我们的方法由两部分组成。第一步，我们呢选择高度代表查询对象的假设。第二步则是聚集挑选的ROI并以簇作为查询对象词汇。

在ROI的选择步骤中，算法反复更新县官的ROI直至它趋于稳定。在每一次改变中，算法首先用一部分具有代表性的ROI作为最相关的查询。这些ROI成为“枢纽”。他憨厚，ROI重定义过程被运用于所有图片，相似度很高的ROI被选作查询相关对象。第二部挑选的查询相关ROI被用于下一次转换的参数。

4.实验

为了演示我们的研究成果，我们在两个公共数据库上做了实验研究，包括实物对象搜索和人物搜索。大部分基础方法用来自搜索引擎的方法，已有的重排方法对比来展示我们方法的优越性。

4.1实验步骤

4.1.1数据集

为了使我们的实验尽可能可充用，我们租用了一个公共的网络查询数据集来评估我们的方法并与基础方法做对比。数据集一共包括353个有代表性的图片搜索结果，范围覆盖了产品、名人和动物等。然后这些擦汗寻被发送到一个图片搜索引擎以获取顶部的图片。最终71478张图片被包含其中。每个查询文本和人们提供的检索图片间的二进制关联作为基础。

因为目标方法设计用于实物搜索，我们通过选择这些实物对象创建两个数据集。数据集比较查询对象包括地表，产品，标志。为了更综合地评估我们的方法，数据集以这些命名人的查询建立。最终，网络查询实物对象数据集以101个查询的19586个图片组成，人物数据集包括103个查询和20398张图片。

5.结论和未来工作计划

图片搜索研究已经被研究了几年，大量方法也已被开发出来，促进了基于文本的图片搜索引擎的搜索。在这片论文中我们发现没有单独的方法能很好地适应所有搜索，图片重拍的搜索要求在查询范围内查询开发特别的模型。

这篇论文是在这个方向上的首次尝试。我们发现已有的对整张图片操作的排序方法并不能有效地进行实物搜索，搜索出的图片只有部分是相关的。这种现状激励我们为实物查询创建一个打包回发模型。回发模型是基于语言模型技术创建的。通过查找文本搜索引擎返回结果中的查询相关对象，我们为每个查询建立对象语言模型和图片的相关对象的词汇。已有对象的属性被用于表现对象与查询之间的密切度，以权重来整合排序算法以避免查询相关对象的干扰。最终，创建排序方法的过程被转变为获得对象属性的权重。两个网络数据集的结果告诉我们最终的方法与现有重排方法相比可以提高6.84%的转准确率。

我们相信这是未来图片搜索重排序的正确方向。考虑到这个大方向和我们论文中的研究，我们继续下面的工作。首先，我们会对不同方向的图片查询分类。其次，我们会结合查询到的对象词汇和对象集来查询更具代表性的图片和查询，第三我们希望找出并解决系统方面的挑战是算法能真正头日到真实世界的图片搜索引擎中。

6.引用

[1] A. Bosch, A. Zisserman, and X. Muoz. Image

classification using random forests and ferns. In

CVPR, pages 1–8, 2007.

[2] S. Brin and L. Page. The anatomy of a large-scale

hypertextual web search engine. Computer networks

and ISDN systems, 30(1-7):107–117, 1998.

[3] G. Csurka, C. Dance, L. Fan, J. Willamowski, and

C. Bray. Visual categorization with bags of keypoints.

In Workshop on statistical learning in computer

vision, ECCV, volume 1, page 22, 2004.

[4] T. Deselaers, B. Alexe, and V. Ferrari. Localizing

objects while learning their appearance. ECCV, pages

452–466, 2010.

[5] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun. Salient

object detection by composition. In ICCV, pages

1028–1035, 2011.

[6] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman.

Learning object categories from google’s image search.

In ICCV, volume 2, pages 1816–1823, 2005.

[7] M. Fritz and B. Schiele. Decomposition, discovery and

detection of visual categories using topic models. In

CVPR, pages 1–8, 2008.

[8] D. Hochbaum and V. Singh. An efficient algorithm for

co-segmentation. In ICCV, pages 269–276, 2009.

[9] L. Hohl, F. Souvannavong, B. Merialdo, and B. Huet.

Enhancing latent semantic analysis video object

retrieval with structural information. In ICIP,

volume 3, pages 1609–1612, 2004.

[10] W. Hsu, L. Kennedy, and S. Chang. Video search

reranking through random walk over document-level

context graph. In ACM Multimedia, pages 971–980,

2007.

[11] Y. Jing and S. Baluja. Visualrank: Applying pagerank

to large-scale image search. IEEE Trans. on PAMI,

30(11):1877–1890, 2008.

[12] T. Joachims. Making large-scale svm learning

practical. Advances in Kernel Methods Support Vector

Learning, pages 169–184, 1999.

[13] T. Joachims. Training linear svms in linear time. In

ACM SIGKDD, pages 217–226, 2006.

[14] A. Joulin, F. Bach, and J. Ponce. Discriminative

clustering for image co-segmentation. In CVPR, pages

1943–1950, 2010.

[15] G. Kim and A. Torralba. Unsupervised Detection of

Regions of Interest using Iterative Link Analysis. In

NIPS, 2009.

[16] G. Kim, E. Xing, L. Fei-Fei, and T. Kanade.

Distributed cosegmentation via submodular

optimization on anisotropic diffusion. In Computer

Vision (ICCV), 2011 IEEE International Conference

on, pages 169–176, 2011.

[17] J. Krapac, M. Allan, J. Verbeek, and F. Juried.

Improving web image search results using

query-relative classifiers. In CVPR, pages 1094–1101,

2010.

[18] J. Lafferty and C. Zhai. Document language models,

query models, and risk minimization for information

retrieval. In ACM SIGIR, pages 111–119, 2001.

[19] Y. Lee and K. Grauman. Object-graphs for

context-aware category discovery. In CVPR, pages

1–8, 2010.

[20] L. Li, H. Su, E. Xing, and L. Fei-Fei. Object bank: A

high-level image representation for scene classification

and semantic feature sparsification. NIPS, 2010.

[21] Y. Liu, T. Mei, X. Hua, J. Tang, X. Wu, and S. Li.

Learning to video search rerank via pseudo preference

feedback. In ICME, pages 297–300, 2008.

[22] D. Lowe. Distinctive image features from

scale-invariant keypoints. IJCV, 60(2):91–110, 2004.

[23] E. Oomoto and K. Tanaka. Ovid: Design and

implementation of a video-object database system.

IEEE Trans. on KDE, 5(4):629–643, 1993.

[24] E. Parzen. On estimation of a probability density

function and mode. The annals of mathematical

statistics, 33(3):1065–1076, 1962.

[25] C. Rother, T. Minka, A. Blake, and V. Kolmogorov.

Cosegmentation of image pairs by histogram

matching-incorporating a global constraint into mrfs.

In CVPR, volume 1, pages 993–1000, 2006.

[26] S. Sav, G. Jones, H. Lee, N. O’Connor, and

A. Smeaton. Interactive experiments in object-based

retrieval. Image and Video Retrieval, pages 1–10, 2006.

[27] S. Sav, H. Lee, A. Smeaton, N. O’Connor, and

N. Murphy. Using video objects and relevance

feedback in video retrieval. 6015:353–364, 2005.

[28] F. Schroff, A. Criminisi, and A. Zisserman. Harvesting

image databases from the web. In ICCV, pages 1–8,

2007.

[29] X. Tian, Y. Lu, L. Yang, and Q. Tian. Learning to

judge image search results. In ACM Multimedia, pages

363–372, 2011.

[30] X. Tian, L. Yang, J. Wang, Y. Yang, X. Wu, and

X. Hua. Bayesian video search reranking. In ACM

Multimedia, pages 131–140, 2008.

[31] X. Tian, L. Yang, X. Wu, and X. Hua. Visual

reranking with local learning consistency. Advances in

Multimedia Modeling, pages 163–173, 2010.

[32] S. Vicente, V. Kolmogorov, and C. Rother.

Cosegmentation revisited: Models and optimization.

ECCV, pages 465–479, 2010.

[33] R. Yan, A. Hauptmann, and R. Jin. Multimedia