Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Query difficulty estimation via relevance prediction for image retrieval

Qianghuai Jia, Xinmei Tian *

University of Science and Technology of China, Hefei, Anhui 230027, PR China

ARTICLE INFO

Article history:
Received 3 May 2014
Received in revised form
4 June 2014
Accepted 25 July 2014

Keywords:
Query difficulty estimation
Image retrieval
Average precision
Pseudo relevance feedback
Voting scheme

ABSTRACT

Query difficulty estimation (QDE) attempts to automatically predict the performance of the search results returned for a given query. QDE has long been of interest in text retrieval. However, few research works have been conducted in image retrieval. Existing QDE methods in image retrieval mainly explore the statistical characteristics (coherence, specificity, *etc.*) of the returned images to derive a value for indicating the query difficulty degree. To the best of our knowledge, little research has been done to directly estimate the real search performance, such as average precision. In this paper, we propose a novel query difficulty estimation approach which automatically estimates the average precision of the image search results. Specifically, we first adaptively select a set of query relevant and query irrelevant images for each query via modified pseudo relevance feedback. Then a simple but effective voting scheme and two estimation methods (hard estimation and soft estimation) are proposed to estimate the relevance probability of each image in the search results. Based on the images' relevance probabilities, the average precision for each query is derived. The experimental results on two benchmark image search datasets demonstrate the effectiveness of the proposed method.

© 2014 Published by Elsevier B.V.

1. Introduction

The state-of-the-art image search systems suffer from a radical variance in retrieval performance over various queries. For some queries, they are easy to be retrieved (the search engine can return very good search results). While for others, they are difficult (the search results are very unsatisfactory). For instance, Fig. 1 shows the top-10 ranked images of three queries returned by an image search engine [16]. It illustrates that this search system performs well on query "Pantheon Rome" with 9 out of 10 relevant images returned, but poor on query "bird" with only 2 out of 10 relevant images returned. Thus, it is desirable for the image search engines to identify "difficult" queries in order to handle them properly.

Query difficulty estimation (QDE) attempts to estimate the search difficulty level for a given query by predicting the retrieval performance of the search results returned for this query without relevance judgments or user feedback [17]. Such technique can allow users or search engines to provide better search experience. For users, they can rephrase the "difficult" queries to improve the search results if an instant feedback of query difficulty is provided. For search engines, they can adopt alternative retrieval strategies (reranking, query suggestion, *etc.*) for different queries via the estimated query difficulty.

QDE has been investigated in text retrieval for several years and many valuable approaches have been proposed [1–5,17,23]. However, in image retrieval, little research has been done in query difficulty estimation. For image retrieval, the query and the returned images are in two different

* Corresponding author. Tel.: +86 18355102690.
E-mail addresses: jqh218@mail.ustc.edu.cn, xinmei@ustc.edu.cn (X. Tian).

http://dx.doi.org/10.1016/j.sigpro.2014.07.018
0165-1684/© 2014 Published by Elsevier B.V.

UNCORRECTED PROOF

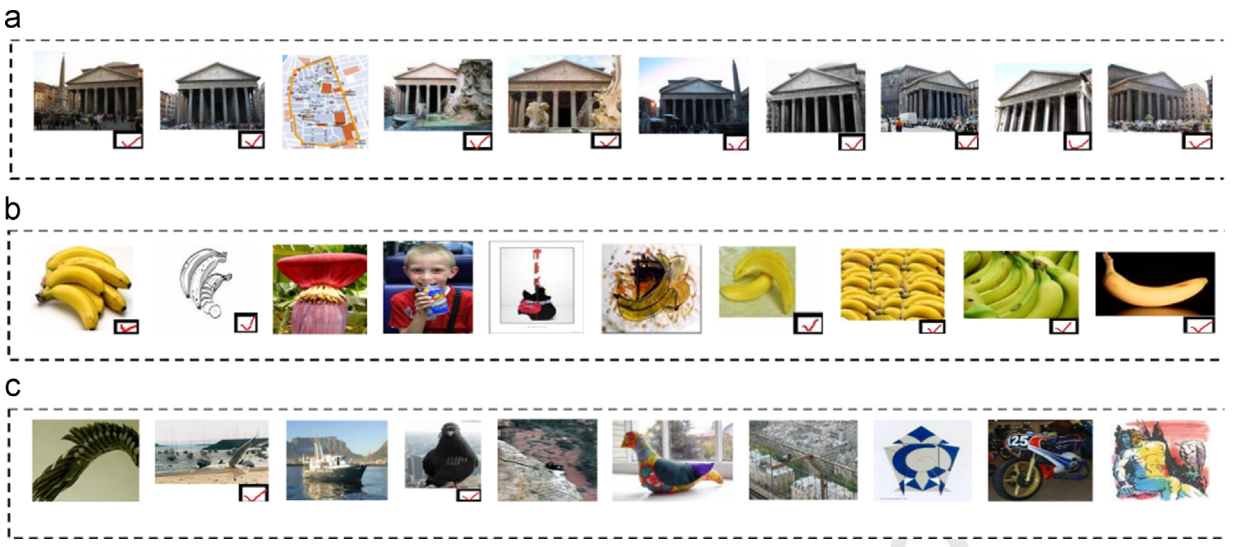

Fig. 1. Top-10 ranked images for three queries ("Pantheon Rome", "banana", "bird") returned by a text-based image search engine, ordered left to right. Query-relevant images are marked by red "☑". It illustrates that this image search engine suffers from a radical variance in search performance over different queries.

domains: textual and visual respectively. This domain gap makes it a challenge for image retrieval query performance prediction. Besides, the textual description (image URL, surrounding text, *etc.*) associated with the image is noisy and insufficient to represent the rich content of images completely. Thus it is hard to directly employ QDE methods in text retrieval to predict query performance for image retrieval.

Existing QDE methods in image retrieval mainly investigate the statistical characteristics (coherence, specificity, *etc.*) of the search images returned for a given query and then derive a value, such as clarity score [1], to indicate the query difficulty level [7,8]. Li et al. [7] proposed a query difficulty predictor by analyzing the prominence character of the top ranked returned images. Tian et al. [8] measured the tightness of the returned images to predict query performance.

To the best of our knowledge, little work has been done to directly estimate the real search performance, such as average precision, for a given query. One preliminary study was performed in [9]. Nie et al. [9] calculated the mathematical expectation of average precision by estimating the relevance probability of each image returned to a given query. In this method, each image's relevance probability is estimated via the visual hyperlinks among the returned images. There exists one major disadvantage in this work. Since the returned results are not perfect, there is noise (mismatch, irrelevant links, *etc.*) among images. With the noise being continually propagated via visual hyperlinks, the final estimated images' relevance probabilities are inaccurate.

In this paper, we propose a novel query difficulty estimation approach for image retrieval. This method can automatically estimate the average precision of the search results returned to a given query. First, by utilizing pseudo relevance feedback (PRF) [10] and adaptive selection, we select a set of query relevant and query irrelevant images for each query. Then a simple but effective voting scheme and two different estimation methods are proposed to estimate the relevance probability of each image in the search results in responds to this query. Finally, based on those images' relevance probabilities, the average precision of each query is derived.

The main contributions introduced in this paper are summarized as follows. (1) We have proposed a novel query difficulty estimation method to automatically predict the actual performance instead of only an indicator. (2) We have proposed an adaptive pseudo positive image selection method which solves the problem that how many images should be assigned to different queries. (3) We have proposed an efficient voting scheme to estimate the image's relevance probability. (4) Our work can be well applied to query difficulty estimation in interactive image retrieval systems. By replacing PRF with users' relevance feedback, the predicting performance can be further improved.

The rest of this paper is organized as follows. Section 2 briefly introduces the related work. Our query difficulty estimation approach is described in Section 3. Section 4 reports the experimental results, followed by the conclusion in Section 5.

2. Related work

QDE has been of interest in the information retrieval (IR) field for many years and its importance has been widely recognized in IR community. In this section, we will introduce the related work of QDE in text retrieval and image retrieval respectively.

QDE in text retrieval can be roughly categorized to pre-retrieval approaches and post-retrieval approaches. Pre-retrieval QDE methods estimate the search difficulty before the search takes place. He and Ounis [2] proposed and evaluated a range of pre-retrieval predictors, including

UNCORRECTED PROOF

Please cite this article as: Q. Jia, X. Tian, Query difficulty estimation via relevance prediction for image retrieval, Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.07.018

Download English Version:

https://daneshyari.com/en/article/6959455

Download Persian Version:

https://daneshyari.com/article/6959455

Daneshyari.com