Signal Processing 110 (2015) 232–243

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 *

CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System, 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
Available online 13 August 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.

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

* 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/

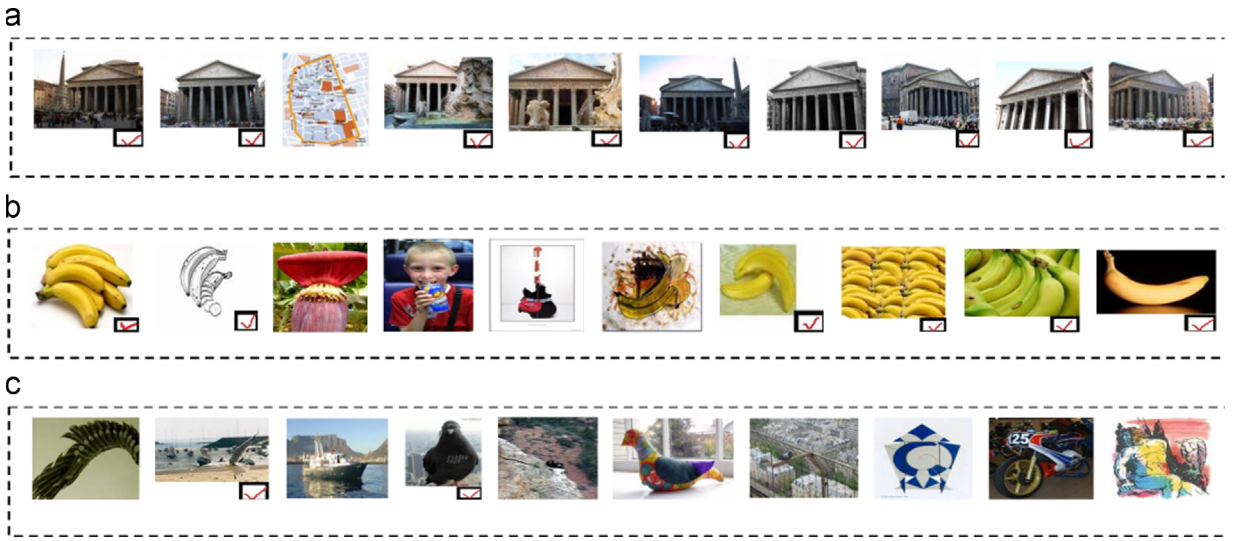

Fig. 1. Top-10 ranked images for three queries ("Pantheon Rome", "banana", "bird") returned by a text-based image search engine, ordered left to right. Query-relevant images are marked by red "☑". It illustrates that this image search engine suffers from a radical variance in search performance over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

been done in query difficulty estimation. For image retrieval, the query and the returned images are in two different 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 query length, the average inverse collection term frequency (AvICTF), query scope, simplified clarity score (SCS) and so on. Their experimental results showed AvICTF and SCS were the two best-performing predictors. Imran and Sharan [5] proposed two pre-retrieval predictors based on the co-occurrence information among query terms. It was assumed that higher co-occurrence of query terms means more information conveyed, which leads to lower query difficulty degree. Since pre-retrieval methods do not need to perform retrieval, they are much efficient.

In contrast, post-retrieval QDE methods additionally analyze the search results. Post-retrieval predictors mainly investigate the characteristics of the top ranked retrieval documents and can be divided into three categories: clarity-based methods, robustness-based methods and score distribution-based methods. Clarity-based predictors explore the distribution difference between the top ranked documents and the whole collection [1,23]. Clarity Score (CS) [1] measures the KL-divergence between the query language model of the top ranked documents and the model of the collection. Hauff et al. [23] improved the clarity score to solve the parameter sensitivity problem in CS. Robustness-based methods quantify the query difficulty by measuring the robust degree of the ranking list [3,17]. Yom-Tov et al. [17] estimated the query difficulty by measuring the agreement between the original ranking list and the ranking list of the query's each constituent terms. The idea behind this method is that for an easy query, the result list may not change considerably if only a subset of the query terms is used. Query Feedback [3] constructs a new query from the top ranked documents of the initial returned list. The overlap of documents in the initial list and the new list generated by searching the new query is used for query difficulty estimation. Score distribution-based predictors analyze the retrieval scores of documents in the result list [4]. Diaz et al. [4] measured the extent to which similar documents receive similar retrieval scores to indicate query performance. In general, post-retrieval methods are usually more expensive as the search results should be analyzed after retrieval, but are preferable to pre-retrieval methods.

For QDE in image retrieval, little research has been conducted. Xing et al. [6] adopted the textual information (surrounding text, image URL, *etc.*) to predict whether a query is difficult or easy. This method leverages the noisy textual information and neglects the rich content of the returned images. Li et al. [7] proposed a query difficulty predictor, by linearly integrating the language model-based clarity score, the spatial consistency of local descriptors and the appearance consistency of global features. Rudinac et al. [18] exploited the coherence of the top ranked video to predict the query performance, for selecting the best video search result. Tian et al. [8] proposed a set of features, including visual clarity score, coherence score, representativeness score and visual similarity distribution feature, by analyzing the search results, and then learnt a support vector regression model to predict the query difficulty. Those three methods belong to the QDE approach that investigates the statistical characteristics of the returned images. The method QAPE (Query-Adaptive Performance Estimation) in [9] utilizes the visual hyperlinks among the returned images to estimate the images' relevance probabilities to the given query and then calculates the mathematical expectation of average precision. This method is the most related work to ours. Different from this method, we estimate the relevance probability of each image via the improved pseudo relevance feedback and a simple but efficient voting scheme. The experimental results demonstrate the effectiveness of our method.

3. Query difficulty estimation via relevance prediction

The proposed approach automatically estimates the relevance probability of each image in response to the query and then derives the average precision. Fig. 2 illustrates the framework of our approach. Given a query, a list of ranked images is returned by the image search engine. We first adaptively choose a set of pseudo positive (query relevant) or pseudo negative (query irrelevant) images from the image list. Then, we estimate the relevance probability for each image in the list. With each image's relevance probability is estimated, we can derive the average precision of the image search results for each query.

In the proposed framework, there are two key components: the pseudo positive and negative image selection, and the relevance probability estimation for each image. We will detail them in Sections 3.1 and 3.2 respectively. Section 3.3 describes the average precision estimation. Some important notations are presented in Table 1.

3.1. Adaptive positive and negative image selection

Automatically judging whether an image in the search results is relevant to the query is very challenging. In order to select positive or negative images, pseudo relevance feedback (PRF) method [39] is often utilized. PRF is a concept introduced from text retrieval. Its hypothesis is that a fraction of the top-ranked documents in the search results are pseudo-positive. Likewise, the pseudo-negative images are usually selected from either the bottom-ranked images or the database with the assumption that few samples in the database are relevant.

The pseudo positive and pseudo negative image selection plays an important role in our method, and will seriously affect the relevance probability estimation step. PRF directly treats the top ranked or bottom ranked images of the initial list as pseudo positive or pseudo negative ones. Those selected images may be very noisy and it is hard to determine how many images should be selected. In order to solve those problems, we improve the pseudo relevance feedback from the following three aspects.

3.1.1. Initial search results reranking

Since the initial image search results are not perfect, the images selected via PRF are noisy. In general, the better the image search results are, the less the noise in selected images will be. Thus, instead of directly applying PRF on initial image search results, we rerank it to get better search results. Here, any image search reranking methods

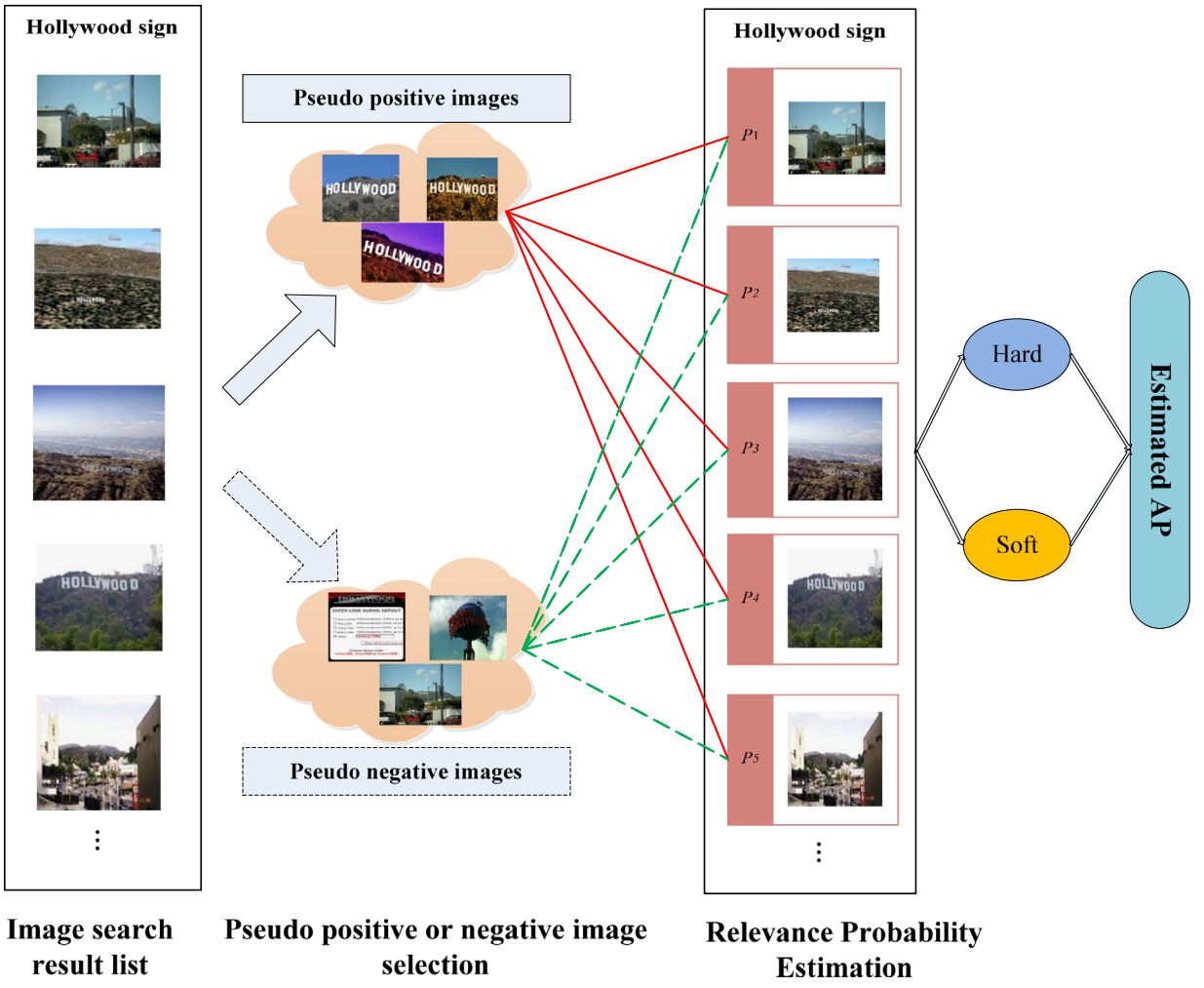

Fig. 2. The framework of the proposed approach. It contains three components: (1) adaptive pseudo positive and pseudo negative image selection: select positive or negative images via improved pseudo relevance feedback; (2) relevance probability estimation: evaluating each image's relevance probability in the search results via a voting scheme and two estimation methods; and (3) AP estimation: calculating the average precision of the search results for the given query based on hard estimation or soft estimation respectively. The dashed line indicates that it is optional.

Table 1
Important notations and their descriptions.

Notations	Descriptions
$I=\{I_i\}_{i=1}^{N}$	N images returned by the search engine for a given query Q, where$I_i$indicates the ith-ranked image
$V=\{w_i\}_{i=1}^{1000}$	The vocabulary with 1000 visual words
K	Number of positive or negative images selected via PRF
$P_0=\{P_k\}_{k=1}^{K}$	Positive (query relevant) image set
$N_0=\{N_k\}_{k=1}^{K}$	Negative (query irrelevant) image set
$x_i=[x_{i1},x_{i2},...,x_{i1000}]^T$	The vector representation of image I_i
$s(I_i,I_j)$	The visual similarity between $I_i$andI_i
$p=\{p_i\}_{i=1}^{N}$	The relevant probability set with N images returned for a given query Q

can be applied [13,31–33,37,38]. Visual reranking [13], which has been proven effective to refine the image search results, is adopted in this paper. This initial search results preprocessing step can efficiently reduce the noise in the selected images. Pseudo positive or pseudo negative images selected from the re-ranked search results contain less noise and therefore can generate much better performance.

3.1.2. The selection of pseudo positive and negative

In PRF, both pseudo positive and pseudo negative images are selected usually. In image retrieval, query relevant images are alike while each irrelevant image is irrelevant in its own way [24]. In other words, the number of irrelevant images are vast and they scatter in the whole space. The selected pseudo negative images are only a very small subset of all irrelevant images. Meanwhile, pseudo negative images may introduce some noise. By further considering the role of pseudo negative images in the following relevance probability estimation step, it is doubtful whether pseudo negative images should be involved. To verify its effect, we propose two selection strategies:

- Select both pseudo positive and pseudo negative images for the following relevance probability estimation. Here we directly treat the top-K ranked images and the bottom-K ranked images in the re-ranked list as pseudo positive (relevant) and pseudo negative (irrelevant) images respectively.
- Select the top-K ranked images as pseudo positive (relevant) images only. There is no pseudo negative image used for the following relevance probability estimation since it may nearly have no effect on it.

3.1.3. Adaptive pseudo positive image selection

As aforementioned, the pseudo positive image selection plays a crucial role in the whole method since all the following steps are based on the selected images. One important problem in pseudo positive image selection is that how many images we should select. We want to get as many as possible relevant images with the least noise. However, when K is too large, too much noise will be involved. When K is too small, the selected pseudo positive images are too few to well interpret the query, and thus lead to unsatisfactory performance.

To solve this problem, we propose an adaptive pseudo positive image selection method. This method can automatically determine how many images should be selected for different queries. Considering that the number of relevant images in the top ranked search list for different queries is varying, we had better adaptively set K for each query. For those queries which have good search performance with many relevant images returned, we can assign a high value to K. While for others, a low value of K should be given.

In order to achieve this goal, we adopt the CoS method [18] to estimate the K value for each query. The experimental results in [8] demonstrated that there is a positive correlation between a large CoS value and a high average precision for a given query. Meanwhile, a search list that has a higher average precision usually contains more relevant images in the top ranked search list. Thus, we assume that a higher CoS value for a given query indicates the top ranked returned results have more relevant images. Then we derive the optimum K^* for each query with a maximum CoS value, as shown in Eq. (1),

$$K^* = \underset{K \in [L,M]}{\operatorname{argmax}} CoS(K) \tag{1}$$

where L and M are the minimum and maximum of K value respectively. $CoS(K)$ is defined as the ratio of coherent pairs to all image pairs in the top-K-ranked images. This adaptive pseudo positive image selection can effectively avoid introducing noise (irrelevant images) to some extent.

3.2. Relevance probability estimation

With the selected pseudo positive or negative images, our approach intends to estimate the relevance probability of each image in the initial search results. Previous works in reranking [25–27] view the relevance probability estimation as a classification problem. Various classifiers, such as SVM [28], can be trained with the selected images to categorize each image in the initial search results. Different from these methods, we propose an efficient and effective voting scheme, and two estimation methods to estimate images' relevance probabilities in this paper.

3.2.1. Visual similarity measure

Before detailing the voting scheme and estimation method, we first introduce the visual similarity measurement. Given a query Q, let $I = \{I_i\}_{i=1}^{N}$ denotes the N images returned by the image search engine, where I_i is the ith-ranked image. For the image's visual representation, we adopt the popular bag-of-visual-words (BOVWs) model [11]. The SIFT local descriptors [12] are first extracted from each image in the collection. Then a vocabulary with 1000 visual words is built by clustering all the local descriptors. After quantizing local descriptors into visual words, each image can be viewed as a visual document consisting of a set of visual words. The $V = \{w_i\}_{i=1}^{1000}$ denotes the visual vocabulary with size 1000. We use the TF-IDF weighting scheme [19] to measure each visual word's importance and adopt the Vector Space Model [14] to represent each image. The vector representation of imageI_i is defined as $x_i = [x_{i1}, x_{i2}, ..., x_{i1000}]^T$ with

$$x_{ij} = tf_{ij} * idf_{ij}, \quad i = 1, 2, ...N, \quad j = 1, 2, ..., 1000 \tag{2}$$

where tf_{ij} is the frequency of visual word w_j in image I_i and idf_{ij} is the inverse document frequency that quantifies the importance of visual word w_j over the image collection.

Then we utilize $s(I_i, I_j)$ to denote the similarity between images I_i and I_j. In this paper, we adopt the cosine similarity:

$$s(I_i, I_j) = \frac{x_i \cdot x_j}{\|x_i\| \|x_j\|}. \tag{3}$$

3.2.2. Voting scheme

It is widely assumed that if an image is query relevant, it should be similar to other query relevant images and be different from query irrelevant images [24]. Based on this assumption, we propose a simple but very effective voting method.

For a given query Q, we have selected a set of pseudo positive or negative images in Section 3.1. Let $P_0 = \{P_k\}_{k=1}^{K}$ and $N_0 = \{N_k\}_{k=1}^{K}$ denote the pseudo positive

image set and pseudo negative image set respectively. For the ith-ranked image I_i for Q, we compare it with the images in P_0 andN_0. If the visual similarity between I_i and P_k is larger than certain threshold μ, then I_i obtains a positive vote. Likewise, a negative vote is given to I_i if the visual similarity between I_i and N_k is larger than certain threshold μ. The vote function can be written as:

$$vote_{ik}^{+} = \left\{ \begin{array}{l} 1, s(I_i, P_k) > \mu \\ 0, else \end{array} \right\}, \quad i = 1, 2, ..., N, \quad k = 1, 2, ..., K \tag{4}$$

$$vote_{ik}^{-} = \left\{ \begin{array}{l} 1, s(I_i, N_k) > \mu \\ 0, else \end{array} \right\}, \quad i = 1, 2, ..., N, \quad k = 1, 2, ..., K \tag{5}$$

where $vote_{ik}^{+}$ and $vote_{ik}^{-}$ respectively indicate the positive vote and negative vote from the selected images. The threshold μ is defined as that 70% of image pairs in the whole image collection have smaller visual similarity than this value.

Since the returned images are complex and the images selected via PRF may be noisy, some votes from the selected images casted to each image in the search results are invalid. In order to obtain an effective voting scheme, we add a decision mechanism. The decision mechanism is described as: if a given image owns the majority of votes from the pseudo positive or negative images, it obtains a true positive or negative vote. Then we form the decision function as:

$$r_i^{+} = \left\{ \begin{array}{l} 1, \sum_{k=1}^{K} vote_{ik}^{+} \geq K/2 \\ 0, else \end{array} \right\}, \quad i = 1, 2, ..., N \tag{6}$$

$$r_i^{-} = \left\{ \begin{array}{l} 1, \sum_{k=1}^{K} vote_{ik}^{-} \geq K/2 \\ 0, else \end{array} \right\}, \quad i = 1, 2, ..., N \tag{7}$$

where r_i^{+} and r_i^{-} denote the final vote to an image by pseudo positive and negative images respectively. The decision mechanism has the advantage of reducing noise (wrong votes) via the average of votes from the selected images.

3.2.3. Estimating relevance probability

With the final positive or negative vote being casted to each image in the search results, we intend to estimate each image's relevance probability to the given query Q. In this paper we design two different methods to estimate the relevance probability, denoted as hard estimation and soft estimation respectively. Here we denote p_i as the relevance probability of the ith-ranked image I_i.

(A) Hard estimation

We follow the assumption: a relevant query should be visually similar to those positive images and be away from those negative images. Then the hard estimation is defined as:

$$p_i = \left\{ \begin{array}{l} 1, \ if \ r_i^{+} = 1 \ and \ r_i^{-} = 0 \\ 0, \ else \end{array} \right\}, \quad i = 1, 2, ..., N \tag{8}$$

From Eq. (8), it can be considered that we label an image which obtains only a final positive vote as relevance and others as irrelevance. Due to this characteristic, we can directly apply the existing formula in [8] to calculate the average precision.

(B) Soft estimation

As discussed in Section 3.1, here we only utilize the pseudo positive images to estimate the images' relevance probabilities. We employ the classification thought and view each image in P_0 as a classifier to produce relevance score $f_k(I_i)$ $(1 \leq k \leq K)$ for any images $I_i (1 \leq i \leq N)$ in response to the given query Q. $f_k(I_i)$ is set to 1 for positive vote and 0 for others from the pseudo positive images, as shown in Eq. (4). Then we adopt a logistic regression algorithm to combine the outputs of all the K classifiers. Thus the soft estimation is expressed as:

$$p_i = \frac{\exp\left(\sum_{k=1}^{K} f_k(I_i)/K - 0.5 \right)}{1 + \exp\left(\sum_{k=1}^{K} f_k(I_i)/K - 0.5 \right)}, \quad i = 1, 2, ..., N \tag{9}$$

In summary, Algorithm 1 in Fig. 3 outlines our relevance probability estimation method. It contains two steps. (1) In the voting step, by comparing the visual similarity between each image in the search list and the selected pseudo images, we cast pseudo positive or negative votes to each image. Then we determine whether those votes are valid and give the final positive or negative vote to each image. (2) In the estimating step, based on the final vote, we evaluate the relevance probability of each image in the initial search results via the hard estimation or soft estimation.

3.3. Average precision estimation

Our main idea is to automatically estimate the average precision of the image search results for each query. After the aforementioned steps, we have obtained the relevance probability of each image in response to the given query. Based on those two relevance probability estimation methods, we adopt two different average precision estimation strategies.

For the hard estimation method, we directly utilize the commonly used truncated average precision (AP) method [8] to estimate AP, as shown in Eq. (10).

$$EAP@T = \frac{1}{Z_T} \sum_{i=1}^{T} rel(i) \frac{1}{i} \sum_{j=1}^{i} rel(j) \tag{10}$$

where $rel(i)$ is the binary function on the relevance of the ith-ranked image with "1" for relevant and "0" for irrelevant, Z_T is a normalization constant that is chosen to guarantee that EAP@T = 1 for the perfect ranking result and T is a variable which indicates the truncation level. Here we replace $rel(i)$ in Eq. (10) withp_i estimated via hard estimation to measure AP.

For the soft estimation method, we tend to calculate the mathematical expectation of AP. Let $p(rel(i) = 1)$ denotes the relevance probability of the ith-ranked image

Algorithm 1 Relevance Probability Estimation

Input: $I = \{I_i\}_{i=1}^{N}$; Positive image set: $P_0 = \{P_k\}_{k=1}^{K}$; Negative image set: $N_0 = \{N_k\}_{k=1}^{K}$

Output: $p = \{p_i\}_{i=1}^{N}$

Step 1: Voting Step

For each image $I_i \in I$

Compute $vote_{ik}^{+}$ by Equation (4);

If $N_0 = \emptyset$

$vote_{ik}^{+} = 0$;

Else

Compute $vote_{ik}^{-}$ by Equation (5);

End

End for

For i=1:N

Compute r_i^{+} by Equation (6);

If $N_0 = \emptyset$

$r_i^{-} = 0$;

Else

Compute r_i^{-} by Equation (7);

End

End for

Step2: Estimating Step

For i=1:N

Estimate P_i by Equation (8) or (9);

End for

Return $p = \{p_i\}_{i=1}^{N}$

Fig. 3. Algorithm details of implementations for relevance probability estimation.

I_i. Assume that the relevance of two images that are in different positions of the returned list is completely independent, then we derive the mathematical expectation of AP as:

$$
\begin{aligned}
&\mathrm{E}(AP@T)\\
&= E\left[\frac{1}{Z_T}\sum_{i=1}^{T} rel(i)\frac{\sum_{j=1}^{i} rel(j)}{i}\right]\\
&= \frac{1}{Z_T}\sum_{i=1}^{T}\sum_{j=1}^{i}\frac{E[rel(i)rel(j)]}{i}\\
&= \frac{1}{Z_T}\sum_{i=1}^{T}\frac{1}{i}\left\{E[rel(i)^2] + \sum_{j=1}^{i-1} E[rel(i)rel(j)]\right\}\\
&= \frac{1}{Z_T}\sum_{i=1}^{T}\frac{1}{i}\left\{p(rel(i)=1) + \sum_{j=1}^{i-1} p(rel(i)=1, rel(j)=1)\right\}\\
&= \frac{1}{Z_T}\sum_{i=1}^{T}\frac{1}{i}\left\{p(rel(i)=1) + \sum_{j=1}^{i-1} p(rel(i)=1)p(rel(j)=1)\right\}
\end{aligned}
\tag{11}
$$

Here we replace $p(rel(i)=1)$ in Eq. (11) with p_i estimated via soft estimation to calculate the mathematical expectation of AP.

4. Experiments

4.1. Dataset

We conduct experiments on a large public Web image search dataset "Web353" available from [15]. This dataset contains 71478 images returned for 353 queries. The 353 queries are diverse in topics, including landmark, animal, plant, sports, flag, people and instruments, *etc*. For each query, the top ranked images are collected by the search engine [16]. And there are about 200 images on average for each query. Each image is manually labeled as relevant or irrelevant. Fig. 4 shows some examples in this dataset.

4.2. Ground-truth performance and correlation measurements

To measure the effectiveness of the proposed method, we calculate the correlation coefficients between the estimated performance and the ground-truth performance. In this paper, the ground-truth performance for each query is measured based on manual relevance labels via the commonly used truncated average precision (AP) [8] in Web search. Fig. 5 illustrates the ground-truth search performance in terms of AP@10 for each of the 353 queries in Web353. For better view, in this figure the queries are sorted in ascending order of AP. It shows the image search system suffers from a radical variance in performance over various queries.

As to the correlation measurements, there exist three commonly used measurements, including the Pearson's r liner correlation [20], non-parametric rank correlation Kendall's τ [21] and Spearman's ρ [22]. All the above three correlation coefficients vary from -1 to 1, where -1 means perfect reverse and 1 means perfect agreement. In our experiment, all three correlation measurements are adopted.

4.3. Baselines

To demonstrate the effectiveness of the proposed method, we compare it with four QDE methods for image search, including Visual Clarity Score (VCS) [8], Coherence Score (CoS) [18], Representativeness Score (RS) [8], and Query-Adaptive Performance Estimation (QAPE) [9]. VCS is a variant of clarity score [1] applied to the image retrieval query difficulty estimation. It measures query difficulty via the KL-divergence between the language model of the returned images and the language model of the whole image collection, where a high divergence suggests an "easy" query. CoS measures the portion of coherent image pairs in the top ranked image results. A pair of images is coherent if their visual similarity exceeds certain threshold which is empirically set. This method assumes the tightness of the top ranked images can indicate the search performance. RS is defined as the mean of the density of the top ranked images in the returned results. The density of each image is estimated via kernel density estimation. In general, a large RS corresponds to a well-performing query. QAPE estimates the mathematical expectations of average precision for each query via the relevance

Fig. 4. Example images in Web353 dataset.

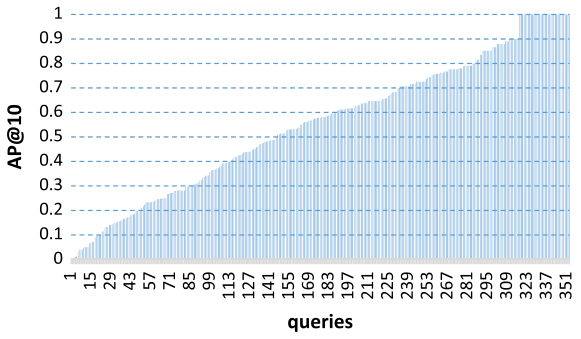

Fig. 5. The AP@10 on each of the 353 queries. For better view, here queries are sorted in ascending order according to their AP@10. This figure shows that the performance of the image search engine varies largely over different queries.

probability of each image to the given query. The image's relevance probability is estimated by utilizing the visual hyperlinks among the returned images.

4.4. Experimental results

In this paper, we show the correlation coefficients at several truncation levels for T, i.e. {AP@T, T=10, 20, 40, 60}. The choice of T depends on the need in real application.

4.4.1. Analysis of our approach

In this subsection, we intend to analyze the performance of the proposed method. We first compare the correlation coefficients of our method with and without pseudo negative images, denoted as EAP_PN and EAP_P respectively. Then we analyze our method with fixed and adaptive K value for each query respectively.

4.4.1.1. Comparisons between EAP_PN and EAP_P. For the proposed approach, the only parameter is K. We vary K from 10 to 45 with interval 1, and observe that there is little variance in our predicted performance. Thus, in this experiment, we empirically set K=15 for all queries.

Table 2 demonstrates the correlation coefficients of EAP_PN and EAP_P. It shows that EAP_P achieves much better performance than EAP_PN. The reasons are that: (1) the images in the top T results might hardly bear resemblance to the pseudo negative images and then the pseudo negative images have almost no effect on the voting scheme; (2) the pseudo negative images might contain some relevant images which will result in wrong votes.

From Table 2, we can also observe that soft estimation outperforms hard estimation in most cases. This is because the soft estimation method preserves more relevance information of the returned results.

4.4.1.2. Comparisons between fixed and adaptive K value for each query. As described in Section 3.1, we investigate the effect of the number of pseudo positive images on query difficulty estimation. For the fixed method, we set

Table 2
The correlation coefficient comparison of our approach with pseudo negative images (EAP_PN) and without pseudo negative images (EAP_P). Hard and Soft respectively indicates the hard estimation and soft estimation. It shows that EAP_P achieves much better performance than EAP_PN. Meanwhile, we can also observe that soft estimation outperforms hard estimation in most cases. The best performances are highlighted in bold.

	T	10	20	40	60
Kendall's τ (*P*-value)	EAP_PN (Hard)	0.300 (7.8e-17)	0.287 (9.0e-16)	0.266 (8.4e-14)	0.253 (1.2e-12)
	EAP_P (Hard)	0.304 (2.5e-17)	0.302 (2.6e-17)	0.274 (2.6e-14)	0.255 (8.7e-13)
	EAP_P (Soft)	**0.320** (**5.4e-19**)	**0.305** (**1.3e-17**)	**0.274** (**1.7e-14**)	**0.256** (**7.1e-13**)
Pearson's *r* (*P*-value)	EAP_PN (Hard)	0.448 (7.6e-19)	0.445 (1.4e-18)	0.408 (1.5e-15)	0.390 (3.1e-14)
	EAP_P (Hard)	0.450 (5.1e-19)	0.456 (1.4e-19)	0.417 (3.0e-16)	**0.391** (**2.6e-14**)
	EAP_P (Soft)	**0.468** (**1.4e-20**)	**0.459** (**9.5e-20**)	**0.420** (**2.1e-16**)	0.390 (2.9e-14)
Spearman's ρ (*P*-value)	EAP_PN (Hard)	0.432 (1.6e-17)	0.415 (4.1e-16)	0.383 (8.7e-14)	0.367 (1.1e-12)
	EAP_P (Hard)	0.438 (5.2e-18)	0.435 (9.1e-18)	0.395 (1.2e-14)	0.369 (8.2e-13)
	EAP_P (Soft)	**0.459** (**8.4e-20**)	**0.444** (**1.8e-18**)	**0.396** (**1.0e-14**)	**0.369** (**7.5e-13**)

Table 3
The correlation coefficient comparison of our approach with fixed and adaptive *K* value. Hard and Soft respectively indicates the hard estimation and soft estimation. From this Table we can see that the approach with adaptive *K* value and the soft estimation achieves the best results. Meanwhile, the method with adaptive *K* value outperforms that with fixed *K* value. The best performances are highlighted in bold.

	T	10	20	40	60
Kendall's τ (*P*-value)	Fixed (Hard)	0.304 (2.5e-17)	0.302 (2.6e-17)	0.274 (2.6e-14)	0.255 (8.7e-13)
	Adaptive (Hard)	0.313 (2.7e-18)	0.304 (1.7e-17)	0.277 (9.0e-15)	0.259 (4.0e-13)
	Fixed (Soft)	0.320 (5.4e-19)	0.305 (1.3e-17)	0.274 (1.7e-14)	0.256 (7.1e-13)
	Adaptive (Soft)	**0.333** (**1.6e-20**)	**0.324** (**1.1e-19**)	**0.286** (**1.1e-15**)	**0.263** (**1.6e-13**)
Pearson's *r* (*P*-value)	Fixed (Hard)	0.450 (5.1e-19)	0.456 (1.4e-19)	0.417 (3.0e-16)	0.391 (2.6e-14)
	Adaptive (Hard)	0.457 (1.2e-19)	0.458 (9.3e-20)	0.417 (2.9e-16)	0.391 (2.4e-14)
	Fixed (Soft)	0.468 (1.4e-20)	0.459 (9.5e-20)	0.420 (2.1e-16)	0.390 (2.9e-14)
	Adaptive (Soft)	**0.482** (**6.0e-22**)	**0.481** (**8.0e-22**)	**0.432** (**1.6e-17**)	**0.399** (**5.6e-15**)
Spearman's ρ (*P*-value)	Fixed (Hard)	0.438 (5.2e-18)	0.435 (9.1e-18)	0.395 (1.2e-14)	0.369 (8.2e-13)
	Adaptive (Hard)	0.452 (3.2e-19)	0.444 (1.7e-18)	0.401 (4.7e-15)	0.374 (3.5e-13)
	Fixed (Soft)	0.459 (8.4e-20)	0.444 (1.8e-18)	0.396 (1.0e-14)	0.369 (7.5e-13)
	Adaptive (Soft)	**0.480** (**9.5e-22**)	**0.469** (**1.0e-20**)	**0.414** (**5.0e-16**)	**0.381** (**1.2e-13**)

$K=15$ for all the queries. For the adaptive method, in this experiment, we empirically set $L=20$ and $M=45$ respectively.

Table 3 demonstrates the correlation coefficients comparison of the proposed method with two different *K* value settings. From this Table we can see that the approach with adaptive *K* value and the soft estimation achieves the best results. Meanwhile, the method with adaptive *K* value outperforms that with fixed *K* value. This indicates the effectiveness of the proposed adaptive pseudo positive image selection method.

4.4.2. Correlation coefficient comparison between our approach and four baseline methods

The correlation coefficients and corresponding *P*-values are given in Table 4. For VCS, CoS, RS and QAPE, we have tried various parameter settings and we report their best results.

From this Table, we can see that our approach outperforms baseline methods over all *T*s consistently. The *P*-values are far less than 0.05, which indicates that the correlation between the predicting performance and the actual performance is statistically significant. We also observe that: with the increase of *T*, our method has a constantly high Pearson's *r* correlation coefficient, which implies our estimated average precision well matches the ground-truth performance.

Among those four baselines, CoS provides the best predicting performance. The correlation coefficients for CoS with the retrieval performance become continuously worse with the increase of *T*. The reason is that, when *T* increases, more and more irrelevant images occur in the top ranked list, making it more difficult to measure the coherence of the returned images. For QAPE, our approach has higher correlation coefficients than QAEP's. The reason why QAPE does not work well is that: since the returned images are complex and noise (mismatch, irrelevant links, *etc.*) can be continually propagated via visual hyperlinks, the image's estimated relevance probability is inaccurate. For VCS, the correlation coefficients are much worse than others. But the performance of VCS rises with the increase of *T*. [1] explains the clarity method needs a great number of documents to adequately measure the coherence of the ranked list. Therefore, the performance of VCS is poor at a small value of *T*.

4.4.3. Effectiveness of voting scheme

As discussed in Section 3.2, we can view the relevance probability estimation as a classification problem. With the selected pseudo images, different classifiers can be trained to categorize each image in the search results. In order to

Table 4
Correlation Coefficients and *P*-value of Query Difficulty Prediction on the Web353 dataset. From this Table, we can see that our approach outperforms baseline methods over all *T*s consistently. Among those four baselines, CoS provides the best predicting performance. The best performances are highlighted in bold.

	T	10	20	40	60
Kendall's τ (*P*-value)	VCS	0.119 (9.1e-04)	0.120 (7.6e-04)	0.132 (0.4e-05)	0.138 (1.1e-04)
	CoS	0.318 (1.7e-19)	0.306 (1.6e-19)	0.252 (1.6e-12)	0.232 (8.0e-11)
	RS	0.208 (7.4e-9)	0.200 (2.0e-08)	0.184 (2.4e-07)	0.188 (1.3e-07)
	QAPE	0.241 (1.8e-11)	0.223 (3.8e-10)	0.140 (8.6e-5)	0.124 (4.8e-4)
	Ours	**0.333** (**1.6e-20**)	**0.324** (**1.1e-19**)	**0.286** (**1.1e-15**)	**0.256** (**7.1e-13**)
Pearson's r (*P*-value)	VCS	0.138 (0.009)	0.155 (0.0035)	0.187 (4.1e-04)	0.216 (4.4e-05)
	CoS	0.453 (2.9e-19)	0.438 ($<$ 1e-50)	0.390 (3.0e-14)	0.353 (8.2e-12)
	RS	0.313 (1.8e-9)	0.295 (1.6e-08)	0.268 (3.3e-07)	0.268 (3.3e-07)
	QAPE	0.294 (1.7e-8)	0.263 (5.1e-7)	0.187 (4.3e-4)	0.203 (1.2e-4)
	Ours	**0.482** (**6.0e-22**)	**0.481** (**8.0e-22**)	**0.432** (**1.6e-17**)	**0.399** (**5.6e-15**)
Spearman's$s\rho$ (*P*-value)	VCS	0.173 (0.001)	0.182 (6.1e-04)	0.194 (2.4e-04)	0.202 (1.3e-04)
	CoS	0.456 (1.7e-19)	0.447 ($<$ 1e-50)	0.365 (1.4e-12)	0.335 (1.0e-10)
	RS	0.303 (6.2e-9)	0.294 (1.8e-08)	0.272 (2.2e-07)	0.276 (1.3e-07)
	QAPE	0.361 (2.9e-12)	0.339 (5.9e-11)	0.204 (1.2e-04)	0.180 (6.6e-04)
	Ours	**0.480** (**9.5e-22**)	**0.469** (**1.0e-20**)	**0.414** (**5.0e-16**)	**0.381** (**1.2e-13**)

Table 5
The correlation coefficients for our method using the voting scheme and SVM respectively. It shows our method has higher correlation coefficients than SVM's, which indicates the effectiveness of the proposed voting scheme in query difficulty estimation. The best performances are highlighted in bold.

	T	10	20	40	60
Kendall's τ (*P*-value)	SVM	0.268 (8.6e-14)	0.248 (3.4e-12)	0.179 (5.6e-07)	0.133 (2.0e-04)
	Ours (Hard)	0.313 (2.7e-18)	0.304 (1.7e-17)	0.277 (9.0e-15)	**0.259** (**4.0e-13**)
	Ours (Soft)	**0.333** (**1.6e-20**)	**0.324** (**1.1e-19**)	**0.286** (**1.1e-15**)	0.256 (7.1e-13)
Pearson's r (*P*-value)	SVM	0.384 (8.1e-14)	0.357 (4.3e-12)	0.260 (7.2e-07)	0.192 (2.8e-04)
	Ours (Hard)	0.457 (1.2e-19)	0.458 (9.3e-20)	0.417 (2.9e-16)	0.391 (2.4e-14)
	Ours (Soft)	**0.482** (**6.0e-22**)	**0.481** (**8.0e-22**)	**0.432** (**1.6e-17**)	**0.399** (**5.6e-15**)
Spearman's$s\rho$ (*P*-value)	SVM	0.390 (2.9e-14)	0.366 (1.2e-12)	0.264 (4.9e-07)	0.197 (1.9e-04)
	Ours (Hard)	0.452 (3.2e-19)	0.444 (1.7e-18)	0.401 (4.7e-15)	0.374 (3.5e-13)
	Ours (Soft)	**0.480** (**9.5e-22**)	**0.469** (**1.0e-20**)	**0.414** (**5.0e-16**)	**0.381** (**1.2e-13**)

Table 6
Correlation coefficients and *P*-value of Query Difficulty Prediction on the MSRA-MM_V1.0 dataset. From the Table we can see that our approach achieves the best results in almost all cases. Among those four baseline methods, CoS provides moderate performance, even better than ours at some *T*s. Since this dataset is much complex, the relevance probability estimated via QAPE is rather inaccurate, resulting in the poor predicting performance. The best performances are highlighted in bold.

	Method	NDCG@20	NDCG@40	NDCG@60
Kendall's τ (*P*-value)	VCS	0.037 (0.660)	0.129 (0.121)	0.088 (0.292)
	CoS	0.200 (0.018)	0.183 (0.028)	0.246 (0.003)
	RS	0.150 (0.071)	0.174 (0.036)	0.204 (0.014)
	QAPE	0.163 (0.050)	0.155 (0.063)	0.139 (0.095)
	Ours	**0.221** (**0.018**)	**0.310** (**5.0e-04**)	**0.316** (**2.6e-04**)
Pearson's r (*P*-value)	VCS	0.072 (0.562)	0.066 (0.591)	0.091 (0.459)
	CoS	0.226 (0.064)	0.299 (0.013)	**0.315** (**0.009**)
	RS	0.164 (0.182)	0.199 (0.104)	0.226 (0.064)
	QAPE	0.176 (0.152)	0.134 (0.277)	0.118 (0.338)
	Ours	**0.244** (**0.045**)	**0.328** (**0.006**)	0.313 (0.009)
Spearman's ρ (*P*-value)	VCS	0.061 (0.619)	0.171 (0.164)	0.127 (0.303)
	CoS	**0.292** (**0.016**)	0.285 (0.018)	0.355 (0.003)
	RS	0.207 (0.090)	0.261 (0.032)	0.284 (0.019)
	QAPE	0.238 (0.051)	0.205 (0.094)	0.178 (0.147)
	Ours	0.285 (0.017)	**0.412** (**4.8e-04**)	**0.438** (**1.9e-04**)

demonstrate the effectiveness of our voting scheme in query difficulty estimation for image retrieval, in this paper we compare our method with SVM based relevance probability estimation method. We have tried several parameter settings for SVM with RBF kernel and reported the best result. The experimental results are given in Table 5. Here we present the results of our method with hard estimation and soft estimation. It shows our method

has higher correlation coefficients than SVM's, which indicates the effectiveness of the proposed voting scheme in query difficulty estimation.

4.4.4. Performance on MSRA-MM_V1.0

We also test our approach on the dataset of MSRA-MM_V1.0 [29]. This dataset has 60257 images from Microsoft Live search for 68 representative queries. For each image, its relevance to the corresponding query is labeled with three levels: very relevant, relevant and irrelevant, which are indicated by scores 2, 1 and 0 respectively. Here we adopt the truncated normalized discounted cumulative gain (NDCG) [30], which is widely used for graded relevance judgments, to calculate the actual performance of each query in this dataset. The other experimental settings are the same as in Web353.

Table 6 demonstrates the correlation coefficient comparison between our approach and four baseline methods with different performance metrics. From the Table we can see that our approach achieves the best results in almost all cases. Among those four baseline methods, CoS provides moderate performance, even better than ours at some *T*s. Since this dataset is much complex, the relevance probability estimated via QAPE is rather inaccurate, resulting in the poor predicting performance.

5. Conclusion

In this paper, we propose a novel query difficulty estimation approach for Web image retrieval. Our method automatically estimates the average precision for each query via improved pseudo relevance feedback and a simple but effective voting scheme. Our method has the advantage of estimating the returned images' relevance labels (hard estimation) in response to the query, which can be well applied to image labeling and image annotation [34–36]. Experiments on two real Web image search datasets demonstrate the effectiveness of our proposed method.

Acknowledgment

This work is supported by the NSFC under the contract No. 61201413, the Fundamental Research Funds for the Central Universities No. WK2100060007 and No. WK2100060011, the Specialized Research Fund for the Doctoral Program of Higher Education No. WJ2100060003, to Dr. Xinmei Tian.

References

[1] Cronen-Townsend, Steve, Yun Zhou, and W. Bruce Croft,Predicting query performance, in: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, ACM, 2002.
[2] Ben He, Iadh Ounis, Query performance prediction, Inf. Syst. 31 (7) (2006) 585–594.
[3] Yun Zhou and W.Bruce Croft, Query performance prediction in web search environments, in: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, pp. 543–550, 2007.
[4] Fernando Diaz, Performance prediction using spatial autocorrelation, in: Proceedings of the 30th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, pp. 583–590, 2007.
[5] Hazra Imran, Aditi Sharan, Co-occurrence based predictors for estimating query difficulty, 2010 IEEE International Conference on IEEE Data Mining Workshops (ICDMW), 2010.
[6] Yi Zhang Xing Xing, Mei Han, Query difficulty prediction for contextual image retrieval, Advances in Information Retrieval, 581-585, Springer, Berlin Heidelberg, 2010.
[7] Yong Luo Yangxi Li, Dacheng Tao, Chao Xu, Query difficulty guided image retrieval system, Advances in Multimedia Modeling, 479-482, Springer, Berlin Heidelberg, 2011.
[8] Yijuan Lu Xinmei Tian, Linjun Yang, Query Difficulty Prediction for Web Image Search, Multimed. IEEE Trans. 14 (4) (2012) 951–962.
[9] Meng Wang Liqiang Nie, Zheng-Jun Zha, Tat-Seng Chua, Oracle in image search: A content-based approach to performance prediction, ACM Trans. Inf. Syst. (TOIS) 30 (2) (2012) 13.
[10] Jaime G Carbonell, Yiming Yang, Robert E Frederking, Ralf D Brown, Yibing Geng, Danny Lee, Translingual Inf. Retr. Comp. Eval. (1997).
[11] Josef Sivic Andrew Zisserman, Video Google: A text retrieval approach to object matching in videos, in: Proceedings of the Ninth IEEE International Conferenceon Computer Vision 2003,IEEE, 2003, pp. 1470–1477.
[12] David G Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[13] Yushi Jing, Shumeet Baluja, Visualrank: Applying pagerank to large-scale image search, Pattern Anal. Mach. Intell. IEEE Trans. 30 (11) (2008) 1877–1890.
[14] Anita Wong Gerard Salton, Chung-Shu Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1975) 613–620.
[15] ⟨https://lear.inrialpes.fr/∼krapac/webqueries/webqueries.html⟩.
[16] ⟨http://www.exalead.com/search/image⟩.
[17] Shai Fine Elad Yom-Tov, David Carmel, Adam Darlow, Learning to estimate query difficulty: Including applications to missing content detection and distributed information retrieval, Proc. ACM SIGIR Spec. Interest Group Inf. Retr. (2005) 512–519.
[18] Stevan Rudinac, Martha Larson, Alan Hanjalic, Exploiting result consistency to select query expansions for spoken content retrieval, in European Conference on Information Retrieval, pp. 645–648, 2010.
[19] Gerard Salton, Christopher Buckley, Term-weighting approaches in automatic text retrieval, Inf. Process. Manage. 24 (5) (1988) 513–523.
[20] Erwin Kreyszig, Advanced Engineering Mathematics, Wiley. com, 2007.
[21] Maurice George Kendall, Rank Correlation Methods, 1948.
[22] Jean Dickinson Gibbons, Subhabrata Chakraborti, Nonparametric Statistical Inference, 168, CRC press, Boca Raton, FL, USA, 2003.
[23] Claudia Hauff, Vanessa Murdock, Ricardo Baeza-Yates, Improved query difficulty prediction for the web, in: Proceedings of the 17th ACM conference on Information and Knowledge Management, ACM, 2008.
[24] X.S. Zhou, T.S. Huang, Small sample learning during multimedia retrieval using biasmap, IEEE Int. Conf. Comput. Vis. Pattern Recogn. (2001) 11–17.
[25] Yuan Liu, Tao Mei, Xian-Sheng Hua, Jinhui Tang, Xiuqing Wu, and Shipeng Li, Learning to video search rerank via pseudo preference feedback, IEEE International Conference on Multimedia and Expo, 2008, IEEE, 2008.
[26] Rong Yan, Alexander G Hauptmann, Co-retrieval: A boosted reranking approach for video retrieval, Image and Video Retrieval, 60-69, Springer, Berlin Heidelberg, 2004.
[27] Alexander Hauptmann Rong Yan, Rong Jin, Multimedia search with pseudo-relevance feedback, Image and Video Retrieval, 238-247, Springer, Berlin Heidelberg, 2003.
[28] Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 27.
[29] Meng Wang, Linjun Yang, Xian-Sheng Hua, MSRA-MM: bridging research and industrial societies for multimedia information retrieval, Microsoft Res. Asia Technol. Rep. (2009).
[30] Kalervo Järvelin, Jaana Kekäläinen, Cumulated gain-based evaluation of IR techniques, ACM Trans. Inf. Syst. (TOIS) 20 (4) (2002) 422–446.
[31] Linjun Yang Xinmei Tian, Jingdong Wang, Xiuqing Wu, Xian-Sheng Hua, Bayesian visual reranking, IEEE Trans. Multimed. 13 (4) (2011) 639–652.
[32] Yong Rui Jun Yu, Bo Chen, Exploiting click constraints and multi-view features for image reranking, IEEE Trans. Multimed. 16 (1) (2014) 159–167.
[33] Yong Rui Jun Yu, Dacheng Tao, Click prediction for web image reranking using multimodal sparse coding, IEEE Trans. Image Process. 23 (5) (2014) 2019–2032.

[34] Jiwoon Jeon, Victor Lavrenko, Raghavan Manmatha, Automatic image annotation and retrieval using crossmedia relevance models, in Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2003.
[35] Weifeng Liu Dacheng Tao, Multiview hessian regularization for image annotation, IEEE Trans. Image Process 22 (7) (2013) 2676–2687.
[36] Dacheng Tao Weifeng Liu, Jun Cheng, Yuanyan Tang, Multiview Hessian discriminative sparse coding for image annotation, Comput. Vision Image Understand. 118 (2014) 50–60.
[37] Hao Li Meng Wang, Dacheng Tao, Ke Lu, Xindong Wu, Multimodal graph-based reranking for web image search, IEEE Trans. Image Process. 21 (11) (2012) 4649–4661.
[38] Chao Zhou Yangxi Li, Bo Geng, Chao Xu, Hong Liu, A comprehensive study on learning to rank for content-based image retrieval, Signal Process. 93 (6) (2013) 1426–1434.
[39] J Kalpana, R Krishnamoorthy, Generalized adaptive Bayesian relevance feedback for image retrieval in the orthogonal polynomials transform domain, Signal Process. 92 (12) (2012) 3062–3067.