Multimedia Systems
DOI 10.1007/s00530-014-0379-8

SPECIAL ISSUE PAPER

Click-boosting multi-modality graph-based reranking for image search

Xiaopeng Yang · Yongdong Zhang · Ting Yao · Chong-Wah Ngo · Tao Mei

Abstract Image reranking is an effective way for improving the retrieval performance of keyword-based image search engines. A fundamental issue underlying the success of existing image reranking approaches is the ability in identifying potentially useful recurrent patterns from the initial search results. Ideally, these patterns can be leveraged to upgrade the ranks of visually similar images, which are also likely to be relevant. The challenge, nevertheless, originates from the fact that keyword-based queries are used to be ambiguous, resulting in difficulty in predicting the search intention. Mining useful patterns without understanding query is risky, and may lead to incorrect judgment in reranking. This paper explores the use of click-through data, which can be viewed as the footprints of user searching behavior, as an effective means of understanding query, for providing the basis on identifying the recurrent patterns that are potentially helpful for reranking. A new reranking algorithm, named click-boosting multi-modality graph-based reranking, is proposed. The algorithm leverages clicked images to locate similar images that are not clicked, and reranks them in a multi-modality graph-based learning scheme. Encouraging results are reported for image reranking on a real-world image dataset collected from a commercial search engine with click-through data.

Keywords Image search · Search reranking · Click-boosting · Multi-modality graph-based learning

National Natural Science Foundation of China (61172153, 61272290), National Key Technology Research and Development Program of China (2012BAH39B02), National High Technology Research and Development Program of China (2014AA015202).

X. Yang · Y. Zhang (✉)
Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, People's Republic of China
e-mail: zhyd@ict.ac.cn

X. Yang
e-mail: xiaopeng.ict@gmail.com

T. Yao · C.-W. Ngo
City University of Hong Kong, Kowloon, Hong Kong
e-mail: tingyao.ustc@gmail.com

C.-W. Ngo
e-mail: cscwngo@cityu.edu.hk

T. Mei
Microsoft Research, Beijing 100080, People's Republic of China
e-mail: tmei@microsoft.com

1 Introduction

The emergence of Web 2.0 has brought a new era of information production. A great number of community-contributed media contents, such as images and videos, are generated and shared on social media communities such as Flickr and YouTube. Many techniques have been developed for multimedia search [26, 38], and due to the success of information retrieval, most search engines employ text-based search techniques for multimedia search by leveraging surrounding textual information. However, as textual information is sometimes noisy and even unavailable, it cannot always accurately and comprehensively describe the media content.

To improve the text-based search performance, visual search reranking has attracted extensive attention in both academia and industry in recent years [20]. The research on visual search reranking has proceeded along two directions: recurrent pattern mining [7, 8, 18, 30] and multi-modality fusion [14, 15, 22, 25, 34, 36]. The former assumes the

Published online: 09 May 2014

existence of common patterns among relevant images for reranking. The later focuses on predicting or learning contributions of different modalities in search reranking.

Specifically, most reranking methods of recurrent pattern mining firstly detect dominant visual patterns and then perform reranking based on the following three assumptions: (1) images with dominant visual patterns should be ranked higher; (2) visually (or semantically) similar images should be ranked closely; (3) although unsatisfying, the initial ranks are important. While these approaches focus on the mining of recurrent patterns from different means, different modalities are treated independently. In contrast, multi-modality fusion explores effects of different visual features (modalities), such as color, edge, and texture. By predicting or learning a proper combination, the utility of each modality can be reflected in a principle way. Even though the aforementioned approaches have been proved effective, these methods ignore the significant effects of user feedback which is an explicit indication of relevance.

However, it is not easy to obtain sufficient and explicit user feedback as users are often reluctant to provide enough feedback to search engines. On the other hand, search engines may have a large quantity of user click-through data, e.g., the queries issued by users and the corresponding clicked images, which represent a kind of "implicit" user feedback. Although the clicked images, along with their corresponding queries, cannot reflect the explicit user preference on the relevance of particular query-image pairs, they statistically indicate the implicit relationship between individual images in the ranked list and the given query. Therefore, we can regard click-through data as implicit user feedback based on the assumption that most of the clicked images are relevant to the given query. Compared with explicit user feedback obtained by relevance feedback mechanism [40] which requests users to provide relevance scores for images, click-through data is more largely available and freely accessible by search engines.

Beyond the fact that click-through data have been widely used in the information retrieval area [2, 4–6, 13], in image search, users browse image thumbnails before selecting the images to click and the decision to click is likely dependent on the relevance of an image. Thus, intuitively click-through data can serve as a reliable feedback potentially useful for search reranking. Nevertheless, to the best of our knowledge, there are very few attempts leveraging click-through data for image reranking [10, 33]. It is inspiring to bring click-through data for image search reranking. However, it is worth noticing that clicked data are likely, but not absolutely, relevant. Some uncertain factors may cause users to click images which are irrelevant to the given query. This so-called user bias issue is mostly generated by users' prejudice, preference, interest or carelessness. Furthermore, an unclicked instance is not necessarily irrelevant. Thus, solely using clicked data may over (under) estimate the importance of clicked (unclicked) data.

The preliminary version of this work, which performs image search reranking using click-boosting random walk (CBRW), is published in [33]. Compared with [33] which is confined to leverage single modality to perform reranking, we generalize this work to multiple modalities. Furthermore, we address several issues arisen from this extension. These issues involve how to adaptively fuse different modalities and how to leverage click-through data and multiple modalities simultaneously. Therefore, in this paper, we focus on image search reranking in the way of multi-modality fusion using click-through data. Accordingly, a novel reranking approach, named click-boosting multi-modality graph-based reranking (CBMGR), is proposed. Figure 1 shows an overview of our proposed approach. Firstly, it boosts the initial ranked results by reordering images according to their click number. Then, it conducts multi-modality graph-based learning on several image graphs based on the click-boosted ranked results, and adaptively integrates multiple modalities to find visual recurrent patterns so as to rank the unclicked relevant images higher and the clicked irrelevant ones lower. Thus, through click-boosting multi-modality graph-based reranking, the reranking results are boosted by click-through data and improved by visual recurrent patterns learnt from appropriate combination of multiple visual features. We apply the proposed approach to perform image search reranking and conduct experiments over 40 image queries collected from a commercial image search engine. Experimental results show that the proposed reranking approach outperforms CBRW and several existing ones.

This paper makes three major contributions:

- We leverage the click-through data and image visual recurrent patterns mined from multiple modalities simultaneously to perform image search reranking.
- We propose an effective novel image search reranking method, named click-boosting multi-modality graph-based reranking, which conducts multi-modality graph-based learning on several image graphs constructed using the ranked list boosted by click-through data.
- The algorithm is demonstrated on a real-world data consisting of 115,792,480 image URLs collected from a commercial search engine.

The rest of this paper is organized as follows. We review the related work in Sect. 2. Section 3 gives the detailed introduction to the proposed algorithm, click-boosting multi-modality graph-based reranking. In Sect. 4, we analyze the click-through data collected from large-scale query logs. Experimental results are reported in Sect. 5. Finally, Sect. 6 concludes this paper.

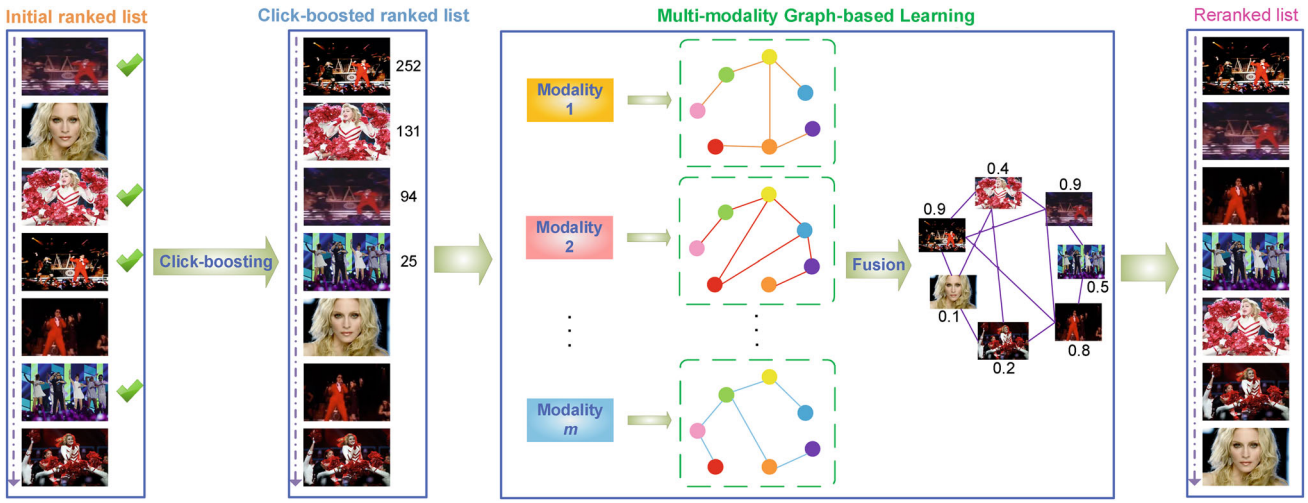


Fig. 1 The overview of the proposed click-boosting multi-modality graph-based reranking approach for image search

2 Related work

We next review related research on visual search reranking and search using click-through data, which are closely correlative with our work.

2.1 Visual search reranking

There are rich related work on visual search reranking in recent years [20]. Along the two research directions of visual search reranking, i.e., recurrent pattern mining and multi-modality fusion, we discuss some typical work in these topics respectively.

Recurrent pattern mining seeks to mine recurrent patterns from relevant images to improve the reranking performance. According to how external knowledge is exploited, the existing methods of recurrent pattern mining can be grouped into three categories: self-reranking [7, 8], example-based reranking [19, 30], and crowd-reranking [18, 37]. The first category focuses on detecting recurrent patterns in the initial search results, and then uses the recurrent patterns to perform reranking. Hsu et al. [7] propose an Information Bottleneck (IB) reranking method, which finds the optimal clustering of images as the recurrent patterns that preserves the maximal mutual information between the search relevance and visual features. In [8], they further formulate reranking as a random walk problem along the context graph, where video stories are represented as nodes and the edges between them are weighted by contextual similarities. Compared with the first category which is purely based on the initial ranked list, example-based reranking mainly relies on the query examples provided by users. For example, Yan et al. [30] propose to train a reranking classifier learnt with the pseudo-relevance feedback (PRF). They treat the query examples as pseudo-positives and choose the bottom-ranked initial results as pseudo-negatives. Similar to self-reranking, the objective of crowd-reranking is to find relevant visual patterns through crowd sourced knowledge, e.g., multiple initial ranked results from various search engines [18, 28] and the suggested queries augmented from the image collection on the Web [37].

Multi-modality fusion aims to fuse different modalities in a unified way and make them function well accordingly. To deal with this problem, one natural way is to concentrate multiple features into a long feature vector and then use this joint modality to perform specific task. Alternatively, we can fuse the reranking results produced by applying modalities separately in a reranking algorithm. These two methods are the well-known "early fusion" and "late fusion," respectively [23]. The early and late fusion approaches have been proved effective in boosting retrieval performance [9, 23, 31], yet, they still have their own limitations. Specifically, early fusion suffers from "curse of dimensionality" and late fusion cannot determine the proper fusion weights for different modalities. Thus, in order to learn appropriate fusion weights of multiple modalities, Snoek et al. [22] propose to assign weights heuristically and manually based on the type of query, such as text-, concept- and visual-oriented queries. Although this "rule-based fusion" is simple, it may degrade the retrieval performance due to the wrong weights assigned by users. To overcome this situation, Yan et al. [32] and Kennedy et al. [14] recommend to use "query-class-dependent fusion." They first classify each user query into categories which are predefined [32] or learnt by cluster algorithms [14], and then aggregate retrieval results with the help of query-class associated weights. Nevertheless, it is difficult to categorize a user query into a specific class accurately due to its complicated semantic meanings. Therefore, "adaptive fusion" is introduced to learn query-dependent fusion weights for multiple modalities using machine learning algorithms [12, 16, 17] or multimodal graph-based learning [25, 29].

To sum up, even though all the above reranking approaches have been proved effective, they ignore the significant effects of user feedback which is an explicit indication of relevance.

2.2 Search using click-through data

Click-through data have been widely used in the information retrieval area [2, 4–6, 13, 21, 35]. Joachims et al. [13] use eye tracking to analyze the relationship between the click-through data and the relevance of query web pages in web search. They prove that click-through data can be used to obtain relative relevance judgments. Since click-through data are informative but biased, many researchers devote to building models for predicting unbiased click-through data for web search ranking. Dupret et al. [5] propose a model based on user browsing behavior, which can estimate the probability that a document is seen, and thereby provide an unbiased estimate of document relevance. They further use the document relevance as a feature for a "learning to rank" machine learning algorithm [4]. Chapelle et al. [3] introduce the notion of satisfaction to separately model the relevance of the landing page and perceived relevance at the search result page, and build a Dynamic Bayesian Network (DBN) to provide the relevance from the click logs. Pan et al. [21] present a graph-based label propagation method by aggregating the nearest neighbors' clicked information to get the labels of a new image. In another work by Yao et al. [35], by combining click-through data and video document features for deriving a latent space, the dot product of the mappings in the latent space is taken as the similarity between videos and the similarity is further applied for video tagging tasks.

However, there are very few attempts leveraging click-through data for image reranking, and the basic hypothesis of these few attempts is that images clicked in response to a given query are mostly relevant to the query. For instance, dealing with long-tail queries, Jain et al. [10] employ Gaussian Process regression to predict the normalized click count for each image, and combine it with the original ranking score for reranking. Yang et al. [33] leverage click-through data and detect recurrent visual patterns of images simultaneously to boost the performance of image retrieval. The proposed click-boosting random walk reranking (CBRW) is first formulated as a ranking problem according to the click number of images, and then formulated as a random walk problem on an image graph. When considering multiple features or modalities for reranking, nevertheless, CBRW cannot deal with the problem of multi-modality fusion, which plays an important role on improving reranking performance if the fusion weights could be learnt properly.

3 Click-boosting multi-modality graph-based reranking

3.1 Overview

There are two major steps in click-boosting multi-modality graph-based reranking. First, it boosts the initial ranked results according to the click number of images since click-through data can implicitly reflect the relevance feedback of users to images. Second, it conducts multi-modality graph-based learning on several image graphs based on the click-boosted ranked results. Each modality corresponds to an image graph where nodes in the graph represent images and edges between them are weighted by visual similarities accordingly [24, 26]. Through multi-modality graph-based learning, the fusion weights of different modalities can be adaptively modulated, and then these modalities can be optimally integrated to find visual recurrent patterns for reranking. Then the unclicked relevant images will be promoted if they are in close proximity with the clicked relevant images, and the clicked irrelevant images will be ranked lower due to the dissimilarity of relevant ones. The overview of the proposed click-boosting multi-modality graph-based reranking for image search is shown in Fig. 1.

Given a query, an initial ranked list of images is obtained by the search engine based on the text-based search technique. For instance, the initial ranked list of query "madonna gangnam style" is shown as the first column of Fig. 1 where images with checkmarks represent that they have been clicked by users according to the query log.

As mentioned above, we first rank the initial list in descending order according to the click number of each image as shown in the second column of Fig. 1. The numbers on the right side of the top four images are the click number of the corresponding images. Based on the rule of click-boosting, the most clicked image with 252 clicks is reranked to the top and the unclicked images are reranked to the bottom of the list. However, we can clearly see that the 2nd image with 131 clicks and the 4th image with 25 clicks are just partially relevant to the query "madonna gangnam style." In addition, the 6th image is quite relevant to the given query, but gets a low position in the click-boosted ranked list. This situation is just the so-called biased case produced by using click-through data solely. To improve the performance of ranked list based on click-boosting, we conduct multi-modality graph-based learning over image graphs, in which each modality corresponds to an image graph where images are represented as nodes and the edges between them are weighted by visual similarities. During multi-modality graph-based learning, on one hand, weights of multiple modalities can be learnt adaptively and query-dependently in a unified

way; on the other hand, with the aid of appropriate multi-modality fusion and iterative propagation of visual edge weights on the graphs, similar recurrent patterns will get closer to each other. When learning process converges, each image will get a static probability as its reranked score (shown in Fig. 1), and then a new reranked list can be generated by re-ordering the images in descending order according to their visual scores. As the last column on the right in Fig. 1 displays, after multi-modality graph-based learning, images with the same visual recurrent patterns are ranked closer to each other. Consequently, the unclicked relevant images can be promoted to higher rank; meanwhile the clicked irrelevant images may drop down to the bottom of the reranked list.

3.2 Formulation

For a given query, suppose we have an image set $\mathcal{X}$ with N images to be reranked from the search engine where $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ and $x_i\,(i \in \{1, 2, ..., N\})$ denotes the ith image of the initial ranked list. The ranked list can be recorded as $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$ where $r_i\,(i \in \{1, 2, ..., N\})$ is the ranking of x_i and the initial ranked list can be represented as $\mathcal{R} = \{r_i | r_i = i, i = 1, 2, ..., N\}$.

Because the search engines can record the click number of each image, then we have a click-through data set $\mathcal{C}$ of the image set $\mathcal{X}$ where $\mathcal{C} = \{c_1, c_2, ..., c_N\}$ and $c_i\,(i \in \{1, 2, ..., N\})$ denotes the click number of x_i.

Click-boosting multi-modality graph-based reranking is first formulated as a ranking problem according to the click number of images. Thus, we use quick-sort algorithm to rank the click-through data set $\mathcal{C}$ in descending order, and record r_{c_i} as the new ranking index of c_i. Then, the ranking r_i of image x_i is updated as r_{c_i}.

After click-boosting the initial image results, we get a new image ranked list $\mathcal{R}' = \{r_i | r_i = r_{c_i}, i = 1, 2, ..., N\}$. And then click-boosting multi-modality graph-based reranking is formulated as a multi-modality graph-based learning problem [25, 27, 29], in which each modality corresponds to an image graph where images are represented as nodes and the edges between them are weighted by visual similarities accordingly.

Suppose we have M modalities in total. We use y_i to represent the relevance score of image x_i which we intend to find, and a_i to indicate the relevance score calculated based on the updated ranking r_{c_i} of x_i after the step of click-boosting, i.e.,

$$a_i = 1 - r_{c_i}/N. \tag{1}$$

$\boldsymbol{Y}$ and $\boldsymbol{A}$ are the corresponding column vectors of y_i and a_i, i.e., $\boldsymbol{Y} \equiv [y_i]_{N\times 1}$ and $\boldsymbol{A} \equiv [a_i]_{N\times 1}$, respectively. $\boldsymbol{W}_m$ $(m = 1, 2, ..., M)$ is the affinity matrix of modality m, and $W_{m,ij}$ denotes the similarity between image x_i and image x_j based on modality m which is computed by cosine distance. Note that $\boldsymbol{W}_m$ is a symmetric matrix. α_m $(m = 1, 2, ..., M)$ is the fusion weight of modality m, and α is the weight vector.

Then, the regularization framework of multi-modality graph-based learning is formulated as following optimization problem

$$\begin{aligned} \min_{\alpha,\boldsymbol{Y}} f(\chi, \boldsymbol{W}, \alpha, \boldsymbol{Y}) = & \sum_{m=1}^{M} \sum_{i,j} \alpha_m W_{m,ij} \left(\frac{y_i}{\sqrt{D_{m,ii}}} - \frac{y_j}{\sqrt{D_{m,jj}}} \right)^2 \\ & + \lambda \|\boldsymbol{Y} - \boldsymbol{A}\|^2 + \gamma \|\alpha\|^2 \\ \text{s.t.} \quad & 0 \le \alpha_m \le 1, \sum_{m=1}^{M} \alpha_m = 1, \end{aligned} \tag{2}$$

where $\boldsymbol{D}_m$ is the diagonal matrix for $\boldsymbol{W}_m$, i.e., $D_{m,ij} = \sum_j W_{m,ij}$, and λ, γ are two positive weighting parameters which modulate the effects of the above three terms. Given a query, by solving this optimization problem, the relevance score of each image can be propagated effectively from a non-zero score image to another non-zero score image with a suitable combination of multiple modalities. The first term of Eq. (2) is based on the visual consistency assumption that if image x_i and image x_j have a high value in W_{ij}, then their relevance score y_i and y_j should be close. The second term ensures that image score $\boldsymbol{Y}$ does not change too much from the click-boosting image score $\boldsymbol{A}$. The third term $\|\alpha\|^2$ is a penalty parameter preventing too much reliance on one smooth graph resulting in not exploring the effects of other modalities.

3.3 Solution

As we can see from Eq. (2), we need to solve two variables in the multi-modality graph-based learning scheme, that is, the reranking relevance score vector of images $\boldsymbol{Y}$, and the multi-modality fusion weight vector α. We adopt an alternating optimization algorithm to solve Eq. (2).

First, fixing α, we can derive that

$$\boldsymbol{Y} = \left(\boldsymbol{I} + \frac{1}{\lambda} \sum_{m=1}^{M} \alpha_m \boldsymbol{L}_m \right)^{-1} \boldsymbol{A}, \tag{3}$$

where $\boldsymbol{I}$ is an identity matrix of which diagonal elements are all 1 and the others are 0, $\boldsymbol{L}$ is the so-called normalized graph Laplacian for $\boldsymbol{W}$, i.e., $\boldsymbol{L} = \boldsymbol{D}^{-1/2}(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{D}^{-1/2}$. Generally, an iterative process shown in Algorithm 1 can be used to solve Eq. (3), which is widely known as label

propagation or manifold ranking [39]. The detailed proof of the convergence of this iterative process can be found in [27].

Algorithm 1 Iterative Solution for Relevance Score Vector

1: Initialize $\boldsymbol{Y} = \boldsymbol{A}$ when $t = 0$.
2: Update $\boldsymbol{Y}$ by
$$\boldsymbol{Y}^{(t+1)} = \frac{1}{1+\lambda}(\boldsymbol{I} - \sum_{m=1}^{M} \alpha_m \boldsymbol{L}_m)\boldsymbol{Y}^t + \frac{\lambda}{1+\lambda}\boldsymbol{A}.$$
3: Let $t = t + 1$ until step 2 converges.

Next, considering $\boldsymbol{Y}$ is fixed, Eq. (2) transforms into a linear combination of two terms containing α shown as follows

$$\min_{\alpha} \sum_{m=1}^{M} \alpha_m g_m + \gamma \|\alpha\|^2 \quad s.t. \quad 0 \le \alpha_m \le 1, \sum_{m=1}^{M} \alpha_m = 1, \tag{4}$$

where $g_m = \sum_{i,j} W_{m,ij}(\frac{y_i}{\sqrt{D_{m,ii}}} - \frac{y_j}{\sqrt{D_{m,jj}}})^2 = \boldsymbol{Y}^T \boldsymbol{L}_m \boldsymbol{Y}$. Then we leverage a coordinate descent method to compute the optimal fusion weight vector α, and the updating rules are

$$\begin{cases} \alpha_i^* = 0, \alpha_j^* = \alpha_i + \alpha_j, & \text{if} \quad 2\gamma(g_i + g_j) + (\alpha_j - \alpha_i) \le 0 \\ \alpha_i^* = \alpha_i + \alpha_j, \alpha_j^* = 0, & \text{if} \quad 2\gamma(g_i + g_j) + (\alpha_i - \alpha_j) \le 0 \\ \alpha_i^* = \dfrac{2\gamma(g_i + g_j) + (\alpha_j - \alpha_i)}{4\gamma}, \alpha_j^* = \alpha_i + \alpha_j - \alpha_i^*, & \text{otherwise.} \end{cases} \tag{5}$$

In each iteration, we conduct a pair-wise updating policy, in other words, we select two elements of vector α to update and fix others. Since the value of Eq. (4) will not increase in each iteration, the above updating process is guaranteed to converge. Clearly, this process can integrate multiple modalities adaptively.

Finally, we illustrate our proposed reranking algorithm, click-boosting multi-modality graph-based reranking, in Algorithm 2.

Algorithm 2 Click-boosting Multi-modality Graph-based Reranking

Step 1: Click-boosting.
Rerank images $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ according to their click number $\mathcal{C} = \{c_1, c_2, \ldots, c_N\}$ in descending order.
Step 2: Multi-modality graph-based learning.
2.1 Initialize iteration $t = 0$.
2.2 Update relevance score vector $\boldsymbol{Y}$ by Algorithm 1.
2.3 Update weight vector α based on Eq. (5).
2.4 $t = t + 1$.
2.5 if $t > T$, end iteration and output the relevance scores; otherwise, go to step 2.2.

4 Click-through data analysis

We have collected query logs from a commercial image search engine in Nov. 2012. The query logs are represented as plain text files that contain a line for each HTTP request satisfied by the Web server. For each record, the following fields are used in our data collection:

$$< Query, ClickedURL, ClickCount, Thumbnail >$$

where the *ClickedURL* and *ClickCount* represent the URL and the number of clicks on this URL when users submit the *Query*, respectively. *Thumbnail* denotes the corresponding image information on the *ClickedURL*.

For analyzing the click-through bipartite graph, we used all the queries in the log with at least one click. There are 34,783,188 queries and 115,792,480 image URLs on the bipartite graph. Figure 2 shows the main characteristics of the query and URL distribution. The left plot of Fig. 2 shows the query click distribution. Each point represents the number of queries (y axis) with a given number of clicks (x axis). The plot on the right hand shows the clicked image URL distribution. Each point denotes the number of URLs with a given number of clicks. We can see that these two distributions clearly follow power laws. The observation is similar to [1], which also states that user search behavior follows a power law. Figure 2 shows the associated law. According to the statistics, each query has on average 11.59 clicked URLs and each URL was clicked an average of 3.48 times.

5 Experiments

To facilitate evaluation and compare our proposed method with other methods, we randomly select 40 queries[1] (as Fig. 3 shows) from the queries mentioned in Sect. 4 as test samples. Since the images after the top 500 results are typically irrelevant, we use the top 500 images from the initial search results to perform reranking.

[1] The queries include: (1) 15 party dresses, (2) baby shower, (3) backsplash ideas, (4) back tattoos, (5) billy beer, (6) boston fire hazmat, (7) bouncy castles, (8) cute offices, (9) deadmau5, (10) fall decorating ideas, (11) fb profile pics for girls, (12) fresh beat band logo, (13) funny spongebob pictures, (14) gingerbread house glue, (15) gingerbread man, (16) girl teen bedrooms, (17) graffiti drawings, (18) gray wolf, (19) guitar factory, (20) gymnastics pictures, (21) hawaii, (22) jesus married mary magdalene, (23) lady gaga, (24) ledge stone hearth, (25) lyndon b. johnson ranch, (26) madonna gangnam style, (27) man with the golden gun, (28) marianas trench, (29) mermaids, (30) michael jackson house, (31) monster high pictures, (32) murano glass, (33) nudity children, (34) ohio state backgrounds, (35) proletariat, (36) skull, (37) vintage christmas prints, (38) vote for me posters, (39) water fountains, (40) winter coloring pages.

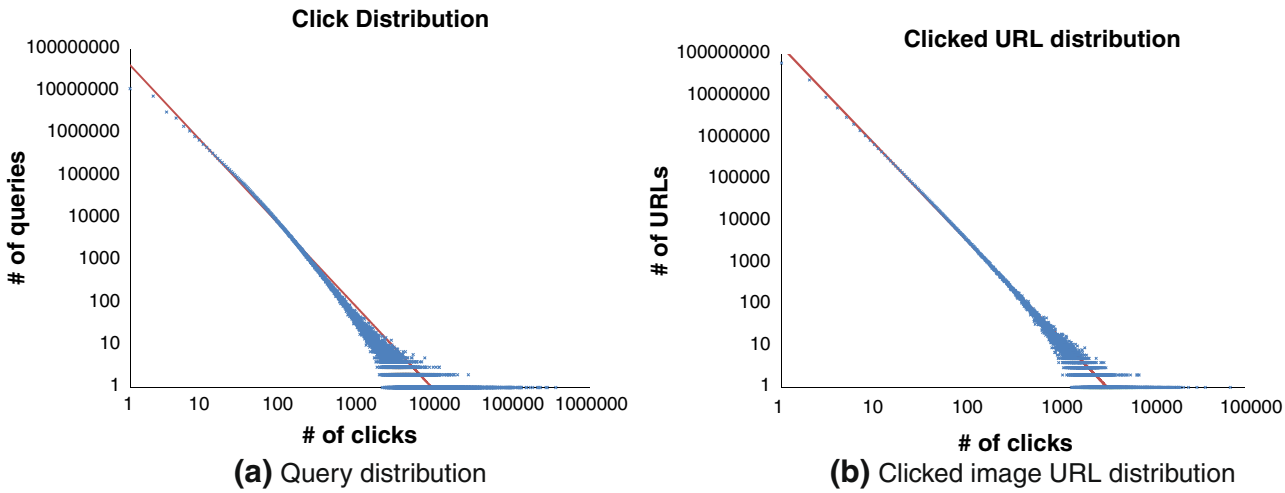


Fig. 2 Query and clicked image URL distribution for the click-through data. *Red lines* denote the fitting power law curves

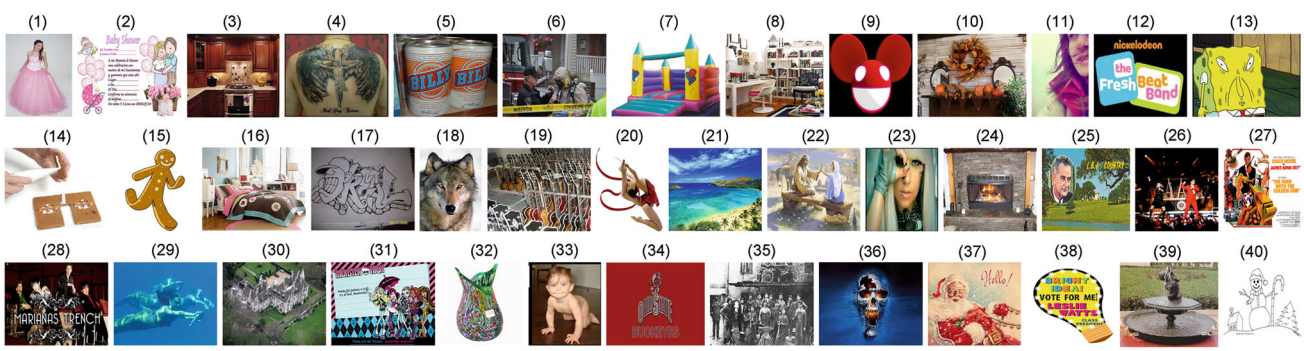


Fig. 3 The exemplary relevant image thumbnails for the 40 queries in our dataset

For query images, we extract six global visual features: (1) 225-dimensional block-wise color moments; (2) 64-dimensional HSV color histogram; (3) 144-dimensional color autocorrelogram; (4) 128-dimensional wavelet texture; (5) 75-dimensional edge distribution histogram; (6) 7-dimensional face features.

5.1 Methodologies

We adopt Normalized Discounted Cumulative Gain (NDCG) [11] to measure the performance, which is widely used in information retrieval when there are more than two relevance levels. Given a ranked list, the NDCG score at the depth d is defined as

$$\mathrm{NDCG@}d = Z_d \sum_{j=1}^{d} \frac{2^{r^j} - 1}{\log(1+j)}, \tag{6}$$

where r^j is relevance score of the jth image, and Z_d is a normalization constant to guarantee that a perfect ranking's NDCG@d is equal to 1. In our experiments, the relevance of each image was labeled manually with three levels: "irrelevant," "fair," and "relevant."

To demonstrate the effectiveness of our proposed approach click-boosting multi-modality graph-based reranking (CBMGR), we compare it with the following five reranking methods, where the parameters are optimized to achieve the best possible performance.

- Pseudo-relevance feedback (PRF) [30]. A typical example-based reranking method which assumes that top-ranked results are more relevant to the bottom-ranked results and performs reranking as a classification problem.
- Random walk (RW) [8]. A representative self-reranking method which conducts random walk on an image graph where nodes are images and edges are weighted by image visual similarities.
- Multimodal graph-based reranking (MGR) [29]. An effective reranking method which leverages the initial ranked list and explores the effects of multiple modalities in a graph-based learning scheme.

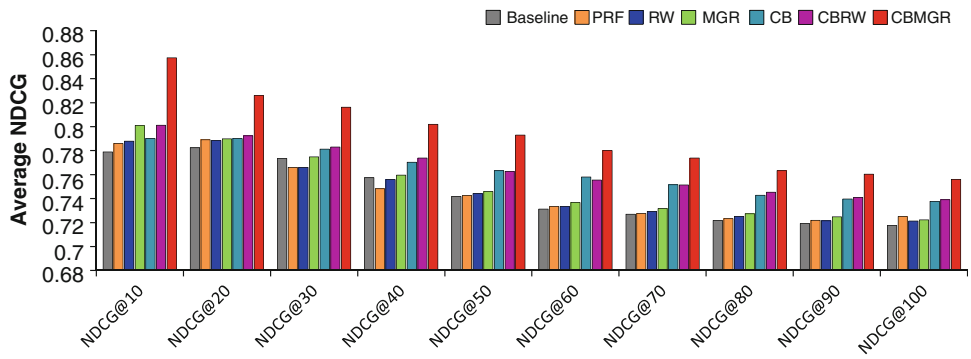


Fig. 4 Comparison of reranking approaches in terms of NDCG

- Click-boosting (CB). CB performs reranking by leveraging click-through data only, namely, CB reranks images according to their click number in descending order.
- Click-boosting random walk (CBRW) [33]. A two-step reranking method which is a combination of using click-through data and detecting visual recurrent patterns for image search reranking. Compared with CBMGR, CBRW can only handle single visual feature. Therefore, in our implementation, the six global visual features are concatenated as one feature vector to perform CBRW.

5.2 Evaluations

5.2.1 Evaluation of reranking performance

For the proposed reranking method click-boosting multi-modality graph-based reranking (CBMGR), we conduct experiments to find out the most suitable weighting parameters λ and γ in Eq. (2) for the 40 queries. Based on the experiment results, we set the weighting positive parameters $\lambda = 4$ and $\gamma = 0.01$ since they achieve the best possible performance on average.

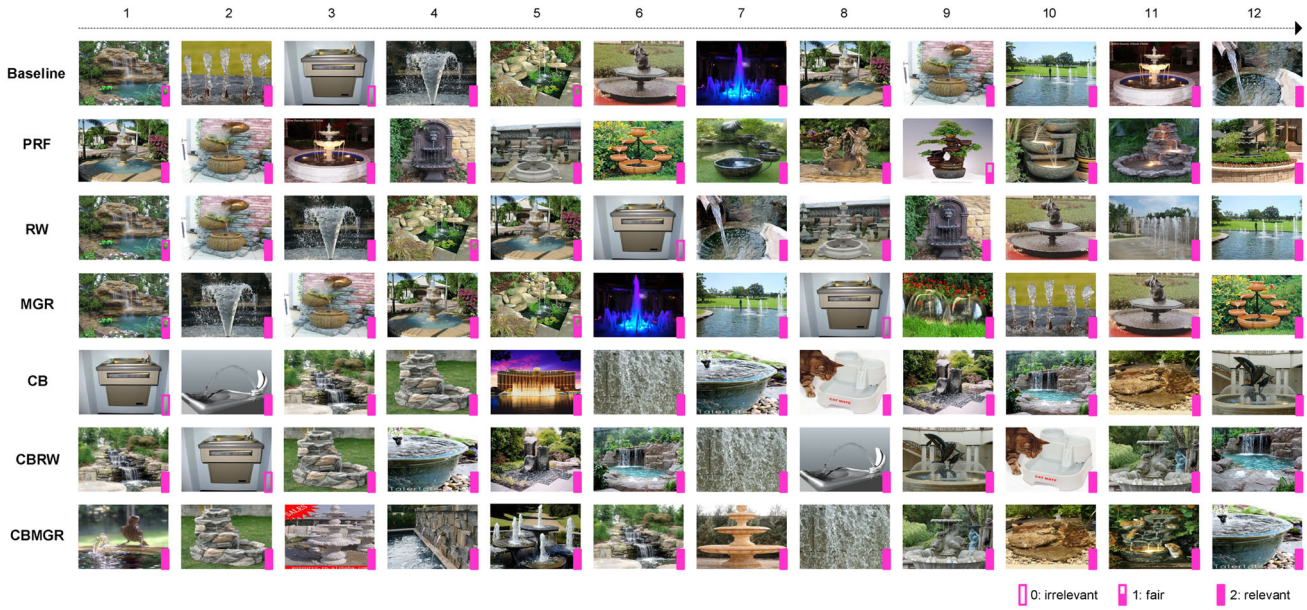
Figure 4 shows the overall performance of different methods. Overall, our proposed click-boosting multi-modality graph-based reranking (CBMGR) outperforms other methods, and the improvements are consistent and stable at different depths of NDCG. Especially, using CBMGR, the value of NDCG@10 shows an obvious improvement of 10.09 % over the baseline. The click-boosting reranking (CB) performs mostly better than the baseline, PRF, RW and MGR, which indicates that click-through data can provide much helpful information of user feedback for image reranking. Furthermore, the superiority of CBRW to CB demonstrates that by means of finding image visual recurrent patterns we can solve the bias problem of click-through data to some extent. Accordingly, the unclicked relevant images can be recommended to be ranked higher and the clicked irrelevant images can be ranked lower and even excluded from the top ranked list. Moreover, there is a significant performance gap between the two runs CBRW and CBMGR. Though both runs involve utilization of all the six visual modalities, they are fundamentally different in the way that the performance of CBRW is as a result of concatenating different modalities, while, CBMGR is by reflecting the utility of each modality in a multi-graph-based learning framework. As indicated by our results, our proposed reranking approach can not only leverage the useful information of "implicit" relevance feedback, i.e., click-through data, but also adjust the reranked list further making unclicked relevant (clicked irrelevant) images rank higher (lower) with appropriate query-dependent fusion weights of multiple modalities.

Figure 5 shows the top 12 images of different reranking approaches for the query "water fountains" and "gymnastics pictures." We can clearly find that the most satisfying results can be obtained using our proposed method.

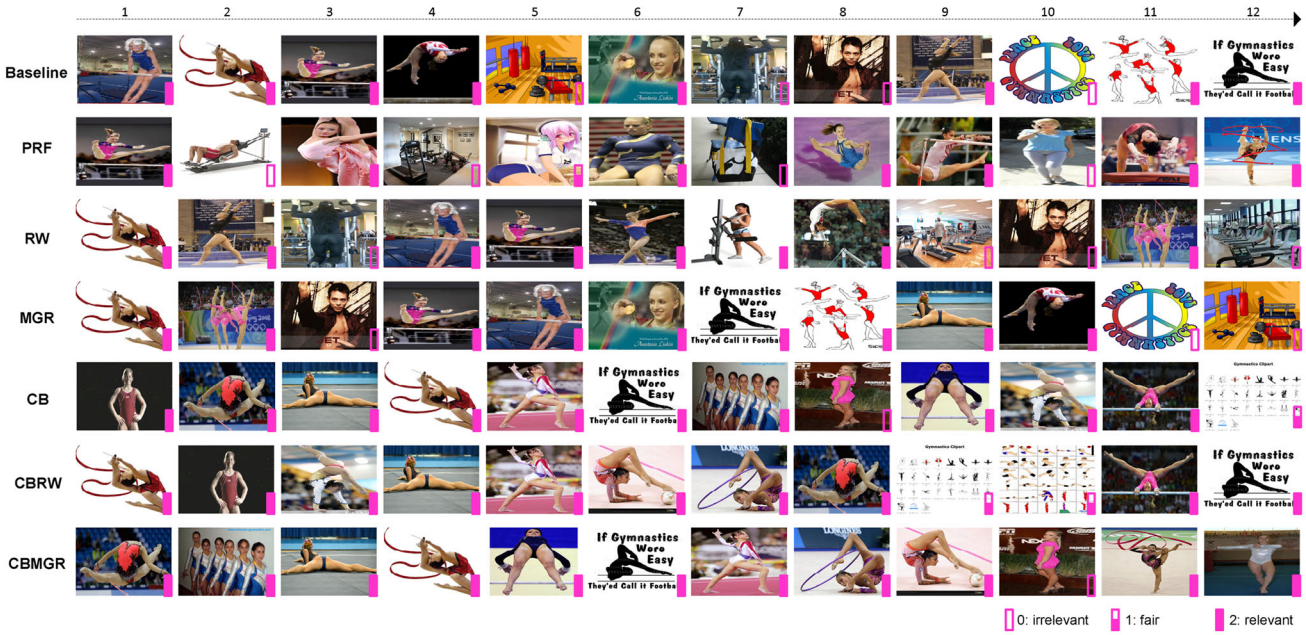
5.2.2 Evaluation of different queries

Figure 6 displays the NDCG performance at depth 50 across different queries. The NDCG values are normalized with respect to the maximum value and the minimum value. As we can see from Fig. 6, our proposed method achieves the best performance in 25 out of 40 queries. It is worth noting that some of these queries obtain the perfect NDCG@50, for example, NDCG values of query "monster high pictures" and "graffiti drawings" are all increased to 1. Moreover, among the 25 queries which exhibit better NDCG using click-boosting multi-modality graph-based reranking, there are two queries remaining the same NDCG value from using CB, CBRW to CBMGR. This is mainly because that the NDCG values of these queries have already been upgraded to 1 by CB, such as "backsplash ideas," and "deadmau5."

However, for the queries, such as "fb profile pics for girls," "girl teen bedrooms," and "murano glass," the performance of CBMGR is worse than CBRW'. These



(a) "water fountains"



(b) "gymnastics pictues"

Fig. 5 Reranked lists from different approaches of specific queries [best viewed in *color*]

cases are understandable because of the selection of parameters λ and γ in Eq. (2). Even though multi-modality fusion weights are assigned query-dependently, these two parameters still determine the effects of CBMGR to some extent. Thus, due to the fact that we set parameters $\lambda = 4$ and $\gamma = 0.01$ which obtain the best possible performance on average, it is unavoidable that some individual queries perform worse using CBMGR than CBRW.

6 Conclusions

In this paper, we demonstrate the effects of the combination of using click-through data and detecting visual recurrent patterns via multi-modality fusion for image search reranking. From our formulation, the initial ranked list is first boosted based on the corresponding click data of images, and then promoted through multi-modality graph-

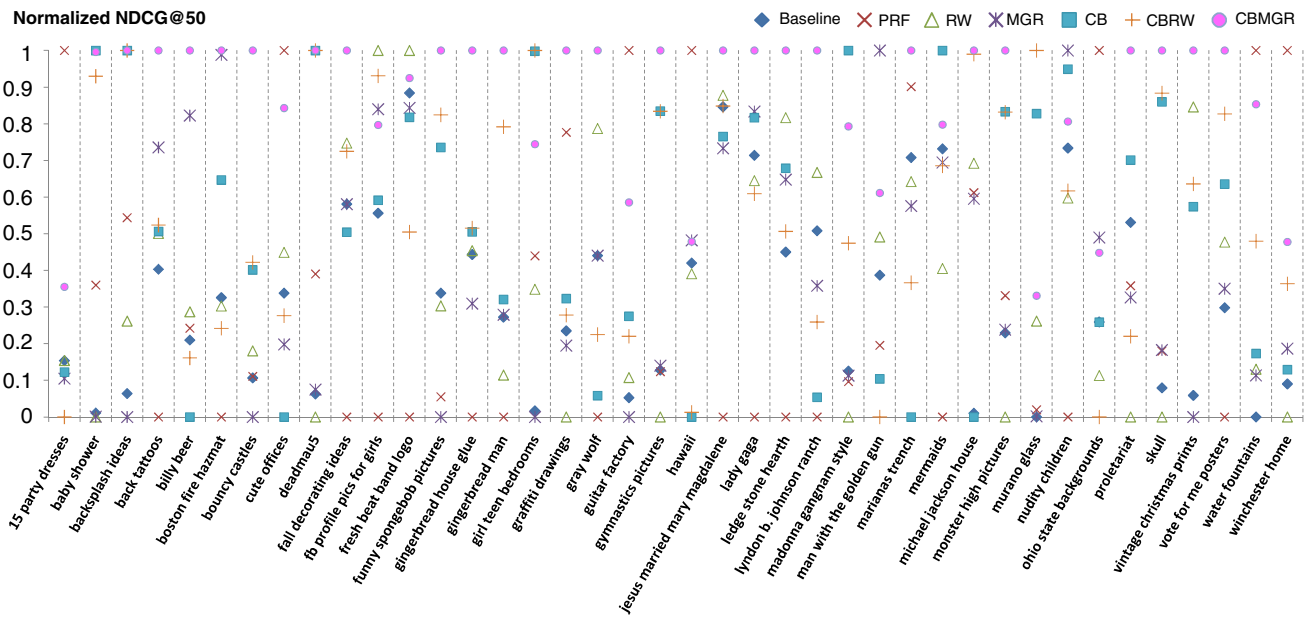


Fig. 6 Normalized NDCG@50 of different methods across 40 queries. Note that NDCG is scaled with *max-min* normalization

based learning which produces the reranked result iteratively based on the fusion of multiple image graphs and the click-boosted reranked list. This can not only promote the position of relevant images according to the number of clicks, but also rank the unclicked relevant images higher owing to multi-modality graph-based learning. Our experimental results demonstrate that the proposed click-boosting multi-modality graph-based reranking method outperforms several state-of-the-art approaches. Since the number of clicks is dependent on the query, we will focus on studying the prediction of click-through data for queries and the deeper influence of click-through data on image retrieval according to different kinds of queries in our future work.

References

1. Baeza-Yates, R., Tiberi, A.: Extracting semantic relations from query logs. In: Proceedings of ACM SIGKDD pp. 76–85 (2007)
2. Carterette, B., Jones, R.: Evaluating search engines by modeling the relationship between relevance and clicks. Adv. Neural Inf. Process. Syst. 217–224(2007)
3. Chapelle, O., Zhang, Y.: A dynamic bayesian network click model for web search ranking. In: Proceedings of ACM WWW pp. 1–10 (2009)
4. Dupret, G., Liao, C.: A model to estimate intrinsic document relevance from the clickthrough logs of a web search engine. In: Proceedings of ACM WSDM pp. 181–190 (2010)
5. Dupret, G., Piwowarski, B.: A user browsing model to predict search engine click data from past observations. In: Proceedings of ACM SIGIR pp. 331–338 (2008)
6. Gu, S., Yan, J., Ji, L., Yan, S., Huang, J., Liu, N., Chen, Y., Chen, Z.: Cross domain random walk for query intent pattern mining from search engine log. In: Proceedings of IEEE ICDM pp. 221–230 (2011)
7. Hsu, W.H., Kennedy, L.S., Chang, S.F.: Video search reranking via information bottleneck principle. In: Proceedings of ACM Multimedia pp. 35–44 (2006)
8. Hsu, W.H., Kennedy, L.S., Chang, S.F.: Video search reranking through random walk over document-level context graph. In: Proceedings of ACM Multimedia pp. 971–980 (2007)
9. Iyengar, G., Nock, H.J.: Discriminative model fusion for semantic concept detection and annotation in video. In: Proceedings of ACM Multimedia pp. 255–258 (2003)
10. Jain, V., Varma, M.: Learning to rerank: Query-dependent image reranking using click data. In: Proceedings of ACM WWW pp. 277–286 (2011)
11. Järvelin, K., Kekäläinen, J.: IR evaluation methods for retrieving highly relevant documents. In: Proceedings of ACM SIGIR pp. 41–48 (2000)
12. Jhuo, I.H., Lee, D.: Boosting-based multiple kernel learning for image re-ranking. In: Proceedings of ACM Multimedia pp. 1159–1162 (2010)
13. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., Gay, G.: Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. ACM Trans. Inf. Syst. **25**(2), Art no. 7 (2007)
14. Kennedy, L.S., Natsev, A.P., Chang, S.: Automatic discovery of query-class-dependent models for multimodal search. In: Proceedings of ACM Multimedia pp. 882–891 (2005)
15. Li, H., Wang, X., Tang, J., Zhao, C.: Combining global and local matching of multiple features for precise item image retrieval. Multimed. Syst. **19**, 37–49 (2013)
16. Li, Z., Liu, J., Zhu, X., Liu, T., Lu, H.: Image annotation using multi-correlation probabilistic matrix factorization. In: Proceedings of ACM Multimedia pp. 1187–1190 (2010)
17. Li, Z., Yang, Y., Liu, J., Zhou, X., Lu, H.: Unsupervised feature selection using nonnegative spectral analysis. In: Proceedings of AAAI pp. 1026–1032 (2012)
18. Liu, Y., Mei, T., Hua, X.S.: CrowdReranking: Exploring multiple search engines for visual search reranking. In: Proceedings of ACM SIGIR pp. 500–507 (2009)

19. Liu, Y., Mei, T., Hua, X.S., Tang, J., Wu, X., Li, S.: Learning to video search rerank via pseudo preference feedback. In: Proceedings of IEEE ICME pp. 297–300 (2008)
20. Mei, T., Rui, Y., Li, S., Tian, Q.: Multimedia search reranking: A literature survey. ACM Computing Surveys (2013)
21. Pan, Y., Yao, T., Yang, K., Li, H., Ngo, C.W., Wang, J., Mei, T.: Image search by graph-based label propagation with image representation from dnn. In: Proceedings of ACM Multimedia pp. 397–400 (2013)
22. Snoek, C., van de Sande, K., de Rooij, O., et al.: The mediamill trecvid 2008 semantic video search engine. NIST TRECVID Workshop (2008)
23. Snoek, C.G.M., Worring, M., Smeulders, A.W.M.: Early versus late fusion in semantic video analysis. In: Proceedings of ACM Multimedia pp. 399–402 (2005)
24. Song, Y., Zhang, Y., Cao, J., Xia, T., Liu, W., Li, J.: Web video geolocation by geotagged social resources. IEEE Trans. Multimed. **14**(2), 456–470 (2012)
25. Tan, H.K., Ngo, C.W.: Fusing heterogeneous modalities for video and image re-ranking. In: Proceedings of ACM ICMR (2011)
26. Tang, J., Yan, S., Hong, R., Qi, G.J., Chua, T.S.: Inferring semantic concepts from community-contributed images and noisy tags. In: Proceedings of ACM Multimedia pp. 223–232 (2009)
27. Wang, M., Hua, X.S., Yuan, X., Song, Y., Dai, L.R.: Optimizing multi-graph learning: Towards a unified video annotation scheme. In: Proceedings of ACM Multimedia pp. 862–871 (2007)
28. Wang, M., Li, G., Lu, Z., Gao, Y., Chua, T.S.: When amazon meets google: Product visualization by exploring multiple information sources. ACM Trans. Internet Technol. **12**(4), Art no. 12 (2013)
29. Wang, M., Li, H., Tao, D., Lu, K., Wu, X.: Multimodal graph-based reranking for web image search. IEEE Trans. Image Process. **21**(11), 4649–4661 (2012)
30. Yan, R., Hauptmann, A., Jin, R.: Multimedia search with pseudo-relevance feedback. In: Proceedings of ACM CIVR pp. 238–247 (2003)
31. Yan, R., Hauptmann, A.G.: The combination limit in multimedia retrieval. In: Proceedings of ACM Multimedia pp. 339–342 (2003)
32. Yan, R., Yang, J., Hauptmann, A.G.: Learning query-class dependent weights in automatic video retrieval. In: Proceedings of ACM Multimedia pp. 548–555 (2004)
33. Yang, X., Zhang, Y., Yao, T., Zha, Z.J., Ngo, C.W.: Click-boosting random walk for image search reranking. In: Proceedings of ACM ICIMCS pp. 1–6 (2013)
34. Yao, T., Mei, T., Ngo, C.W.: Co-reranking by mutual reinforcement for image search. In: Proceedings of ACM CIVR pp. 34–41 (2010)
35. Yao, T., Mei, T., Ngo, C.W., Li, S.: Annotation for free: Video tagging by mining user search behavior. In: Proceedings of ACM Multimedia pp. 977–986 (2013)
36. Yao, T., Ngo, C.W., Mei, T.: Circular reranking for visual search. IEEE Trans. Image Process. **22**(4), 1644–1655 (2013)
37. Zha, Z.J., Yang, L., Mei, T., Wang, M., Wang, Z.: Visual query suggestion. In: Proceedings of ACM Multimedia pp. 15–24 (2009)
38. Zhang, L., Zhang, Y., Tang, J., Gu, X., Li, J., Tian, Q.: Topology preserving hashing for similarity search. In: Proceedings of ACM Multimedia pp. 123–132 (2013)
39. Zhou, D., Bousquet, O., Lal, T.N., Weston, J., Schölkopf, B.: Learning with local and global consistency. Adv. Neural Inf. Process. (2004)
40. Zhou, X.S., Huang, T.S.: Relevance feedback in image retrieval: A comprehensive review. Multimed. Syst. **8**(6), 536–544 (2003)