

Tag-Based Social Image Search: Toward Relevant and Diverse Results

Kuiyuan Yang, Meng Wang, Xian-Sheng Hua, and Hong-Jiang Zhang

Abstract Recent years have witnessed a great success of social media websites. Tag-based image search is an important approach to access the image content of interest on these websites. However, the existing ranking methods for tag-based image search frequently return results that are irrelevant or lack of diversity. This chapter presents a diverse relevance ranking scheme which simultaneously takes relevance and diversity into account by exploring the content of images and their associated tags. First, it estimates the relevance scores of images with respect to the query term based on both visual information of images and semantic information of associated tags. Then semantic similarities of social images are estimated based on their tags. Based on the relevance scores and the similarities, the ranking list is generated by a greedy ordering algorithm which optimizes Average Diverse Precision (ADP), a novel measure that is extended from the conventional Average Precision (AP). Comprehensive experiments and user studies demonstrate the effectiveness of the approach.

1 Introduction

There is an explosion of social media content available online, such as Flickr, Youtube and Zoomr. Such media repositories promote users to collaboratively cre-

K. Yang (✉)

Department of Automation, The University of Science and Technology of China, Hefei, Anhui
230027, China
e-mail: yky@ustc.edu

M. Wang

AKiiRA Media Systems Inc, Palo Alto, CA 94301, USA
e-mail: eric.mengwang@gmail.com

X.-S. Hua

Media Computing Group, Microsoft Research Asia, Beijing 100080, China
e-mail: xshua@microsoft.com

H.-J. Zhang

Microsoft Advanced Technology Center, Beijing 100080, China
e-mail: hjzhang@microsoft.com



Fig. 1 An example of a social image with its associated tags

ate, evaluate and distribute media information. They also allow users to annotate their uploaded media data with descriptive keywords called tags. As an example, Fig. 1 illustrates a social image and its associated user-provided tags. These valuable metadata can greatly facilitate the organization and search of the social media. By indexing the images with associated tags, images can be easily retrieved for a given query. However, since user-provided tags are usually noisy and incomplete, simply applying text-based retrieval approach may lead to unsatisfactory results. Therefore, a ranking approach that is able to explore both the tags and images' content is desired to provide users better social image search results.

Currently, Flickr provides two ranking options for tag-based image search. One is “most recent”, which orders images based on their uploading time, and the other is “most interesting”, which ranks the images by “interestingness”, a measure that integrates the information of click-through, comments, etc. In the following discussion, we name these two methods time-based ranking and interestingness-based ranking, respectively. They both rank images according to measures (interestingness or time) that are not related to relevance and it results in many irrelevant images in the top search results. As an example, Fig. 2 illustrates the top results of query “waterfall” with the two ranking options, in which we can see that many images are irrelevant to the query, such as those marked with red boxes. In addition to relevance, lack of diversity is also a problem. Many images from social media websites are actually close to each other. For example, several users get used to upload continuously captured images in batch, and many of them are visually and semantically close. When these images appear simultaneously as top results, users will get only limited information. From Fig. 2 we can also observe this fact, the images marked with blue or green boxes are very close to at least one of the other images.

Therefore, a ranking scheme that can generate relevant and diverse results is highly desired. This problem is closely related to a key scientific challenge that is recently released by Yahoo research: “*how do we combine both content-based retrieval with tags to do something better than either approach alone for multimedia retrieval*” [33].

The importance of relevance is clear. In fact, this is usually regarded as the bedrock of information retrieval: *if an IR system's response to each query is a ranking of documents in order of decreasing probability of relevance, the overall effectiveness of the system to its user will be maximized* [22]. The time-based and interestingness-based ranking options are of course useful. For example, users can easily browse the images that are recently uploaded via the time-based ranking. But

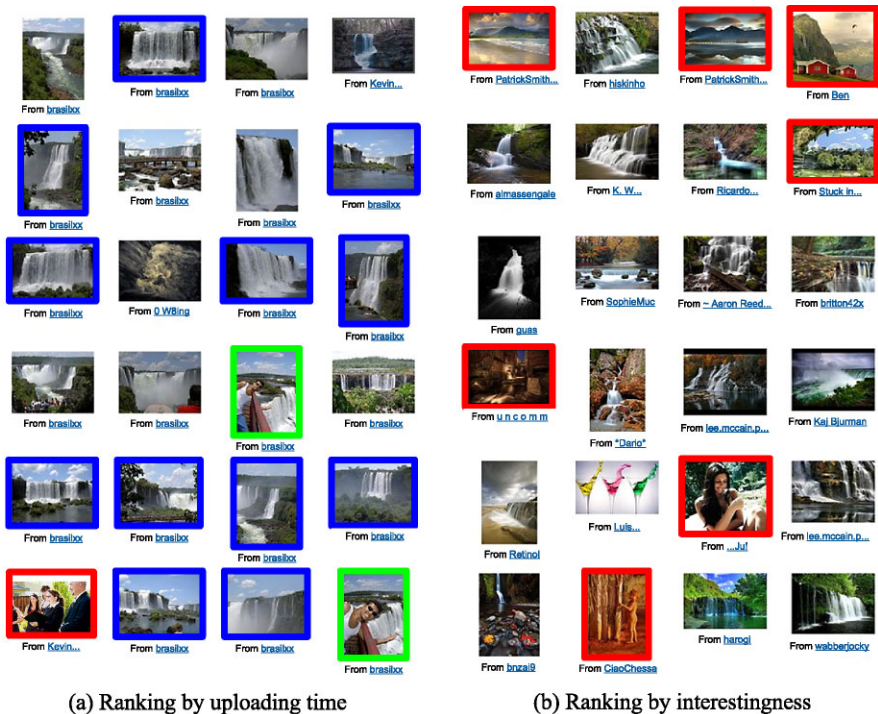


Fig. 2 (Color online) The top 20 search results of the query “waterfall” with the two ranking options. (a) Time-based ranking. (b) Interestingness-based ranking. We can see that many images are irrelevant to the query (marked with *red* border) or close to others (marked with *blue* or *green* border)

when users perform search with the intention of finding specific images, relevance will be more important than time and interestingness.

The necessity of diversity may seem less intuitive than relevance, but its importance has also been long acknowledged in information retrieval [9, 28]. One explanation is that the relevance of a document (can be a web page, image or video) with respect to the query should depend on not only the document itself but also its difference with the documents appearing before it. Now we observe this issue from another perspective. In many cases users cannot accurately and exhaustively describe their requests, and thus keeping diversity of the search results will provide users more chances to find the desired content quickly. For example, we can consider the following cases in image search:

- (1) The users only provide an ambiguous query [1]. For example, the query “apple” may refer to different topics, such as fruit, computer and mobile. Thus, it is better to provide diverse results to cover multiple topics.
- (2) The users cannot fully describe their requests by simple words. For example, although a user only provides a simple query “car”, he/she may actually want to

find a picture of a red car on grass. In this case, the hit probability of a diverse image set should be greater than a set of images that are quite close.

Therefore, diversity of results is also important for users [25, 28]. This fact can also be explained in the information theoretic point of view. If the returned images are all identical for a query, the information gained by the user is actually equivalent to only returning one image.

In this chapter, we will first review recent some research efforts related to the relevance and diversity problems of image search, and then introduce the Diverse Relevance Ranking (DRR) scheme. The organization of the rest of this chapter is as follows. In Sect. 2, we provide a review on the related work. In Sect. 3, we present DRR as a general ranking algorithm. In Sect. 4, we detail the relevance and semantic similarity estimation of social images. Empirical results is presented in Sect. 5. Finally, we conclude the chapter in Sect. 6.

2 Related Work Review

2.1 Social Image Search

The last decade has witnessed a great advance of image search technology [10, 17, 23, 34]. Different from general web images, social images are usually associated with a set of user-provided descriptors called tags, and thus tag-based image search can be easily accomplished by using these descriptors as index terms. Since user-provided tags are usually very noisy [14, 18] and it frequently results in unsatisfactory search results. In comparison with the extensive studies on how to help users better perform tagging or mining tags for other applications, the literature regarding tag-based image search is still very sparse. Most of such efforts focus on how to refine the image's tags or measure their relevance levels. Li et al. proposed a tag relevance learning method which is able to assign each tag a relevance score, and they have shown its application in tag-based image search [18]. Kennedy et al. [15] proposed a method to establish reliable tags by investigating highly similar images that are annotated by different photographers. Liu et al. [19] proposed an optimization scheme for tag refinement based on the visual and semantic connection between images. Sun and Bhowmick [27] proposed a method to measure the tag clarity score based on the query language model and the collection language model. These methods can help tag-based image search by improving the tags' quality, but they cannot deal with the aforementioned lack-of-diversity problem.

2.2 Diversifying Image Search Result

It has been long acknowledged that diversity plays an important role in information retrieval. In 1964, Goffman recognized that the relevance of a document must be determined with respect to the documents appearing before it [9]. Carbonell et al.

propose a ranking method named Maximal Marginal Relevance, which attempts to maximize relevance while minimizing similarity to higher ranked documents [4]. Zhai et al. propose a subtopic search method, which aims to return results that cover more subtopics [35, 36]. Santos et al. propose an approach to enhance diversity by explicitly modeling the query aspects and then actively seeking to maximize the coverage of the selected documents with respect to these aspects [24].

The diversity problem is actually more challenging in an image search, as it involves not only the semantic ambiguity of queries but also the visual similarity of search results [28]. Currently there are two typical approaches to enhancing the diversity in image search: search results clustering and duplicates removing. When performing search results clustering, a representative image can be selected from each cluster. Then we can either only present these representatives or put other images behind them in the ranking list. In [3], Cai et al. propose a method to cluster web image search results into different semantic clusters to facilitate users' browsing. Jing et al. [13] have proposed an IGroup system for clustering image search results. Song et al. have studied the topic coverage of image search diversification method [25]. Recently, Leuken et al. have investigated different clustering methods for visual diversification of image search results [28]. Different from clustering, the duplicates removing approach directly eliminates the duplicates or near-duplicates detected in image search results. Many different duplicate detection methods have been proposed, such as pair-wise image comparison [11], approximate search [29], and fingerprint-based algorithms [26]. Recent progress of image duplicate detection can be found in [37, 38].

Although encouraging results have been demonstrated, the clustering and duplicates removing techniques have their limitations due to the involved heuristics. For clustering, how to establish the number of clusters is a problem. If too many clusters are generated, then the diversity of search results cannot be guaranteed, and contrarily if the clusters are too few, then the search relevance may degrade. In addition, how to take images' relevance levels into the clustering process is also a problem. For duplicates removing, if we set a low threshold for near-duplicate detection, then the diversity of search results cannot be guaranteed, and contrarily if we set a high threshold for near-duplicate detection, many informative images will be removed.

2.3 Performance Evaluation Metric

To quantitatively evaluate different ranking schemes, many performance evaluation metrics are proposed in literature. The classical IR metrics such as AP [2] and NDCG [12] are widely used for measuring search quality. However, they only care about relevance but do not take diversity into account. Several metrics have been proposed for evaluating the diversity of search results, including α -NDCG [7], k -call metric [5] and Intent-Aware measures (NDCG-IA, MAP-IA, MRR-IA) [1]. In this chapter, we present a new performance metric which takes both relevance and diversity into account, and thus the images for a query can be ordered by directly optimizing the performance metric.

3 Diverse Relevance Ranking

We introduce the Diverse Relevance Ranking (DRR) approach in this section. Here we present it as a general ranking algorithm and leave the two flexible components, i.e., relevance score and similarity estimation of images, to the next section. We first prove that ranking by relevance scores can be viewed as a process of optimizing the mathematical expectation of the conventional Average Precision (AP) measure. Then we analyze the limitation of AP and generalize it to an Average Diverse Precision (ADP) measure to integrate diversity. The DRR algorithm is then derived by greedily optimizing the mathematical expectation of ADP measurement.

3.1 Average Precision

AP is a widely-applied performance evaluation measure in information retrieval. Given a collection of images $\mathcal{D} = \{x_1, x_2, \dots, x_n\}$, we denote the binary relevance label of x_i with respect to the given query as $y(x_i)$, i.e., $y(x_i) = 1$ if x_i is relevant and otherwise $y(x_i) = 0$. Denote by τ an ordering of the images, and let $\tau(i)$ be the image at the position of rank i (a lower number indicates image with a higher rank). Let R be the number of true relevant images in the set \mathcal{D} . Then the non-interpolated AP is defined as

$$AP(\tau, \mathcal{D}) = \frac{1}{R} \sum_{j=1}^n y(\tau(j)) \frac{\sum_{k=1}^j y(\tau(k))}{j}. \quad (1)$$

Thereby, ranking images with their relevance scores in decreasing order is the most intuitive approach. Now we prove that the ranking list generated in this way actually maximizes the mathematical expectation of AP measurement.

Denote by $r(x_i)$ the relevance score of x_i (how to estimate it will be introduced in the next section), and it is reasonable for us to assume that $r(x_i) = P(y(x_i) = 1)$, i.e., we regard the relevance score $r(x_i)$ as the probability that x_i is relevant. Since R can be regarded as a constant, we do not take it into account in the expectation estimation. We also assume that the relevance of an image is independent with other images, and hence the expected value of $AP(\tau, \mathcal{D})$ can be computed as follows

$$\begin{aligned} E\{AP(\tau, \mathcal{D})\} &= \frac{1}{R} \sum_{j=1}^n \sum_{k=1}^j \frac{E\{y(\tau(k))y(\tau(j))\}}{j} \\ &= \frac{1}{R} \sum_{j=1}^n \frac{1}{j} \left(r(\tau(j)) + \sum_{k=1}^{j-1} r(\tau(k))r(\tau(j)) \right). \end{aligned} \quad (2)$$

Then we have the following theorem:

Theorem 1 *Ranking the images in \mathcal{D} with relevance scores $r(x_i)$ in non-increasing order maximizes $E\{AP(\tau, \mathcal{D})\}$.*

Proof Denote by τ^* the ranking of images in \mathcal{D} with relevance scores in non-increasing order, i.e., $r(\tau^*(i)) \geq r(\tau^*(i+1))$. Then we only need to prove $E\{AP(\tau^*, \mathcal{D})\} \geq E\{AP(\tau, \mathcal{D})\}$ for every possible τ .

Without loss of generality, we consider an ordering τ' that has exchange the documents at the positions of rank i and $i+1$ in τ^* , i.e., $\tau'(i) = \tau^*(i+1)$ and $\tau'(i+1) = \tau^*(i)$. Actually it is not difficult to find that any change on the τ^* can be decomposed into a series of such adjacent exchanges. So, our task is simplified to prove $E\{AP(\tau^*, \mathcal{D})\} \geq E\{AP(\tau', \mathcal{D})\}$.

For simplicity, we denote $r_i = r(\tau^*(i))$ and $r'_i = r(\tau'(i))$. Since $r'_i = r_{i+1}$, $r'_{i+1} = r_i$, and $r'_k = r_k$ when $k \neq i$ and $i+1$, we have

$$\begin{aligned}
 \Delta &= E\{AP(\tau^*, \mathcal{D})\} - E\{AP(\tau', \mathcal{D})\} \\
 &= \frac{1}{R} \left(\sum_{1 \leq j \leq n, j \neq i, j \neq i+1} \frac{r_j + \sum_{k=1}^{j-1} r_k r_j}{j} + \frac{r_i + \sum_{k=1}^{i-1} r_k r_i}{i} \right. \\
 &\quad \left. + \frac{r_{i+1} + \sum_{k=1}^i r_k r_{i+1}}{i+1} \right) \\
 &\quad - \frac{1}{R} \left(\sum_{1 \leq j \leq n, j \neq i, j \neq i+1} \frac{r'_j + \sum_{k=1}^{j-1} r'_k r'_j}{j} + \frac{r'_i + \sum_{k=1}^{i-1} r'_k r'_i}{i} \right. \\
 &\quad \left. + \frac{r'_{i+1} + \sum_{k=1}^i r'_k r'_{i+1}}{i+1} \right) \\
 &= \frac{r_i - r_{i+1} + \sum_{k=1}^{i-1} r_k (r_i - r_{i+1})}{i} - \frac{r_i - r_{i+1} + \sum_{k=1}^{i-1} r_k (r_i - r_{i+1})}{i+1} \\
 &= \left(1 + \sum_{k=1}^{i-1} r_k \right) (r_i - r_{i+1}) \left(\frac{1}{i} - \frac{1}{i+1} \right). \tag{3}
 \end{aligned}$$

Since $r_i \geq r_{i+1}$, we have $\Delta \geq 0$, i.e., $E\{AP(\tau^*, \mathcal{D})\} \geq E\{AP(\tau', \mathcal{D})\}$, which completes the proof. \square

This demonstrates that the AP performance evaluation measure encourages prioritizing images with high relevance. However, the measure may not be consistent with users' experience due to the neglect of diversity. Figure 3 illustrates an example to demonstrate this fact. In Fig. 3(a), all images are relevant and several images in (b) are irrelevant. Therefore, most probably illustrating images in (a) on the top of the ranking list will introduce higher AP measurement than (b), but clearly it provides little information for users because the images are just duplicates. Therefore, the conventional AP measure can be improved to be more consistent with user experience by integrating diversity.

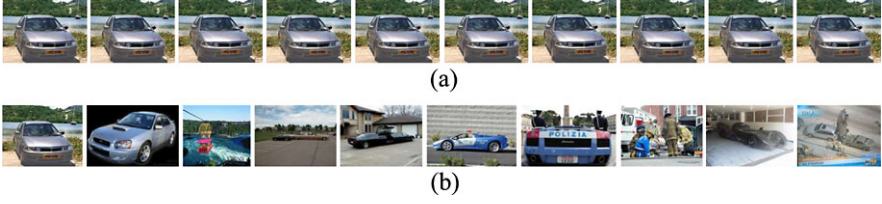


Fig. 3 An extreme example to illustrate the limitation of the conventional AP measure that only considers relevance. In (a), all the top 10 results are highly relevant to “car” and in (b) several images are irrelevant. Therefore, illustrating the images in (a) on top in the ranking list will most probably introduce higher AP measurement than (b), but clearly the images in (b) are more informative because (a) only illustrates duplicates

3.2 Average Diverse Precision

Here we generalize the existing AP measure to Average Diverse Precision (ADP) to take diversity into account, which is defined as

$$ADP(\tau, \mathcal{D}) \triangleq \frac{1}{R} \sum_{j=1}^n y(\tau(j)) Div(\tau(j)) \left(\frac{\sum_{k=1}^j y(\tau(k)) Div(\tau(k))}{j} \right), \quad (4)$$

where $Div(\tau(k))$ indicates the diversity score of $\tau(k)$. We define $Div(\tau(k))$ as its minimal difference with the images appearing before it, i.e.,

$$Div(\tau(k)) = \min_{1 \leq t < k} (1 - s(\tau(t), \tau(k))), \quad (5)$$

where $s(\cdot, \cdot)$ is a similarity measure between two images. Comparing the definition of AP and ADP (see Eq. (1) and Eq. (4)), we can see that the only difference is that we have changed $y(\tau(k))$ to $y(\tau(k))Div(\tau(k))$. For an image in the ranking list, its contribution to the ADP measure is not only determined by its relevance with respect to the query but also its difference with the images appearing before it. If an image is identical to one of the images that have previously appeared, it will add no contribution to the ADP measurement. Thus the ADP measure takes both relevance and diversity into account. Denote by τ^* the optimal ranking list under the ADP performance evaluation measure, i.e., the list that achieves the highest ADP measurement, we can prove that $y(\tau(i))Div(\tau(i)) \geq y(\tau(j))Div(\tau(j))$ for any $i \leq j$. This indicates that the top images will tend to be more relevant and diverse. Here we omit its proof since it is analogous to Theorem 1.

3.3 Diverse Relevance Ranking

The DRR algorithm is actually a greedy approach to optimizing the expected value of the ADP measurement. Analogous to AP, we can estimate the expected value of ADP as

$$\begin{aligned}
& E\{ADP(\tau, \mathcal{D})\} \\
&= \frac{1}{R} \sum_{j=1}^n \sum_{k=1}^j \frac{E\{y(\tau(k))y(\tau(j))Div(\tau(k))Div(\tau(j))\}}{j} \\
&= \frac{1}{R} \sum_{j=1}^n r(\tau(j))Div(\tau(j)) \left(\frac{Div(\tau(j)) + \sum_{k=1}^{j-1} r(\tau(k))Div(\tau(k))}{j} \right). \quad (6)
\end{aligned}$$

The direct optimization of $E\{ADP(\tau, \mathcal{D})\}$ is a permutation problem and the solution space scales is $O(n!)$. Thus here we propose a greedy method to solve it. Considering the top $i - 1$ documents have been established, based on Eq. (6) we can derive that the i th image should be decided as follows

$$\tau(i) = \arg \max_{x \in \mathcal{D} - \mathcal{S}_i} \frac{r(x)}{i} Div(x) (C + Div(x)), \quad (7)$$

where

$$\mathcal{S}_i = \{\tau(1), \tau(2), \dots, \tau(i-1)\}, \quad (8)$$

$$C = \sum_{k=1}^{i-1} r(\tau(k))Div(\tau(k)). \quad (9)$$

Figure 4 illustrates implementation process of the DRR algorithm. Note that C can be viewed as constant in Eq. (7). So we can clearly see that the selection of the i th image will be determined by two factors: the relevance of the image and its difference with the previously selected images.

4 Relevance and Similarity of Social Images

In this section, we introduce the estimation of relevance scores and similarities of social images, which are the two necessary components of the DRR algorithm (see Fig. 4). The following notations will be used. Given a query tag t_q , denote by $\mathcal{D} = \{x_1, x_2, \dots, x_n\}$ the collection of images that are associated with the tag. For image x_i , denote by $\mathcal{T}_i = \{t_1^i, t_2^i, \dots, t_{|\mathcal{T}_i|}^i\}$ the set of its associated tags. The relevance scores of all images in \mathcal{D} are represented in a vector $\mathbf{r} = [r(x_1), r(x_2), \dots, r(x_n)]^T$, whose element $r(x_i) > 0$ denotes the relevance score of image x_i with respect to query tag t_q . Denote by \mathbf{W} a similarity matrix whose element W_{ij} indicates the visual similarity between images x_i and x_j .

4.1 Relevance Estimation

Our relevance estimation approach is accomplished by leveraging both the visual information of images and the semantic information of tags. Our first assumption is

Input:

$\mathcal{D} = \{x_1, x_2, \dots, x_n\}$; /*image collection*/
 $\mathcal{S} = \phi$; /*selected set*/
 $r(x_i), 1 \leq i \leq n$; /*relevance scores*/
 $s(x_i, x_j)$; /*the semantic similarity between two images*/
 $C = 0$

Output:

τ /*an order of the images*/

Begin:

$\tau(1) = \arg \max_{x \in \mathcal{D}} r(x)$;
 $\mathcal{S} = \mathcal{S} \cup \tau(1)$;
for $i = 2, 3, \dots, n$
 /*decide the i th sample*/
 $\tau(i) = \arg \max_{x \in \mathcal{D} - \mathcal{S}} \frac{r(x)}{i} \text{Div}(x)(C + \text{Div}(x))$
 $\mathcal{S} = \mathcal{S} \cup \tau(i)$;
 $C = C + r(\tau(i)) \text{Div}(\tau(i))$;
for each $x \in \mathcal{D} - \mathcal{S}$ /*update $\text{Div}(x)$ */
 $\text{Div}(x) = \min\{\text{Div}(x), 1 - s(x, \tau(i))\}$
end
end

Fig. 4 Pseudo-code of the proposed DRR algorithm

that the relevance of an image should depend on the “closeness” of its tags to the query tag. Thus we first have to define the similarity of tags. Different from images, which can be represented as sets of low-level features, tags are textual words and their similarity exists only in semantics. Recently, there are several works aim to address this issue [6, 32]. Here we adopt an approach that is analogous to Google distance [6], in which the similarity between tag t_i and t_j is defined as

$$\text{sim}(t_i, t_j) = \exp\left(-\frac{\max(\log c(t_i), \log c(t_j)) - \log c(t_i, t_j)}{\log M - \min(\log c(t_i), \log c(t_j))}\right), \quad (10)$$

where $c(t_i)$ and $c(t_j)$ are the numbers of images associated with t_i and t_j on Flickr, respectively, $c(t_i, t_j)$ is the number of images associated with both t_i and t_j simultaneously, and M is the total number of images on Flickr. Therefore, the similarity of the query tag t_q and the tag set of image x_i can be computed as

$$\text{sim}(t_q, \mathcal{T}_i) = \frac{1}{|\mathcal{T}_i|} \sum_{t \in \mathcal{T}_i} \text{sim}(t_q, t). \quad (11)$$

Figure 5 illustrates two examples to demonstrate the rationality of this approach. Figure 5(a) and (b) show two images associated with query tags “dolphin” and “eagle”, respectively. Intuitively, we can see that the images on the left are much more relevant than the images on the right. Then we can see that actually this fact can

Fig. 5 (a) Two images tagged with “dolphin”. (b) Two images tagged with “eagle”. We can see that the images on the left are more relevant with respect to the query tags, and their associated tags are also closer to the query tags



be inferred from the associated tag sets of the images. The tags of the left images are obviously closer to the query tags (for example, “dolphin” is strongly correlated with “ocean” and “water”, and this correlation can be reflected on their Google distance [6]).

Our second assumption is that the relevance scores of visually similar images should be close. The visual similarity between two images can be directly computed based on Gaussian kernel function with a radius parameter σ , i.e.,

$$W_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right). \quad (12)$$

But it is worth mentioning that we can also adopt other similarity measures here, such as those proposed in [21, 30]. Note that this assumption may not hold for several images, but it is still reasonable in most cases. Based on the two assumptions, we formulate a regularization framework as follows

$$Q(\mathbf{r}) = \sum_{i,j=1}^n W_{ij} \left(\frac{r(x_i)}{\sqrt{D_{ii}}} - \frac{r(x_j)}{\sqrt{D_{jj}}} \right)^2 + \lambda \sum_{i=1}^n (r(x_i) - \text{sim}(t_q, \mathcal{T}_i))^2, \quad (13)$$

$$\mathbf{r}^* = \arg \min Q(\mathbf{r}),$$

where $r(x_i)$ is the relevance score of x_i , and $D_{ii} = \sum_{j=1}^n W_{ij}$. We can see that the above regularization scheme contains two terms. The first term is a smoothness term which means that the relevance scores between two visually similar images should

be close (i.e., $r(x_i)$ and $r(x_j)$ should be close if W_{ij} is large), and the second term is a consistency term which means that the relevance score should be consistent with the relevance of the tag set (i.e., $r(x_i)$ should be great if $\text{sim}(t_q, \mathcal{T}_i)$ is great. The above equation can be written in matrix form as

$$Q(\mathbf{r}) = \mathbf{r}^T (\mathbf{I} - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}) \mathbf{r} + \lambda \|\mathbf{r} - \mathbf{v}\|^2, \quad (14)$$

where $\mathbf{D} = \text{Diag}(D_{11}, D_{22}, \dots, D_{nn})$ and $\mathbf{v} = [\text{sim}(t_q, \mathcal{T}_1), \text{sim}(t_q, \mathcal{T}_2), \dots, \text{sim}(t_q, \mathcal{T}_n)]^T$.

Taking derivative of Eq. (14) with respect to \mathbf{r} , we obtain

$$\left. \frac{\partial Q}{\partial \mathbf{r}} \right|_{\mathbf{r}=\mathbf{r}^*} = 2(\mathbf{I} - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}) \mathbf{r}^* + 2\lambda(\mathbf{r}^* - \mathbf{v}) = 0 \quad (15)$$

and we can derive

$$\mathbf{r}^* = \frac{\lambda}{\lambda + 1} \left(\mathbf{I} - \frac{1}{1 + \lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^{-1} \mathbf{v}. \quad (16)$$

This is the closed-form solution of our optimization framework. However, we can see that the above solution involves the inversion of an $n \times n$ matrix, of which the computational cost scales as $O(n^3)$. Here we present a more efficient algorithm to solve \mathbf{r} in an iterative way:

- (1) Construct the image affinity matrix \mathbf{W} by Eq. (11) if $i \neq j$ and otherwise $W_{ii} = 0$.
- (2) Initialize $\mathbf{r}^{(0)}$. The initial values will not influence the final results.
- (3) Iterate $\mathbf{r}^{(t+1)} = \frac{1}{1+\lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \mathbf{r}^{(t)} + \frac{\lambda}{1+\lambda} \mathbf{v}$ until convergence.

The method can be viewed as a random walk process, and it will converge to a fixed point, i.e.,

Theorem 2 *The iterative process converges to the optimal \mathbf{r}^* in Eq. (16).*

Proof According to the iterative function

$$\mathbf{r}^{(t+1)} = \frac{1}{1 + \lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \mathbf{r}^{(t)} + \frac{\lambda}{1 + \lambda} \mathbf{v} \quad (17)$$

we have

$$\begin{aligned} \mathbf{r}^* &= \lim_{t \rightarrow \infty} \left(\frac{1}{1 + \lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^t \mathbf{v} \\ &\quad + \frac{\lambda}{1 + \lambda} \left(\sum_{m=1}^t \left(\frac{1}{1 + \lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^m \right) \mathbf{v}. \end{aligned} \quad (18)$$

Based on the fact that $0 < \frac{1}{1+\lambda} < 1$ and that the eigenvalues of matrix $\mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$ are in $(0, 1)$, we have

$$\lim_{t \rightarrow \infty} \left(\frac{1}{1 + \lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^t = 0 \quad (19)$$

and

$$\lim_{t \rightarrow \infty} \sum_{m=1}^t \left(\frac{1}{1+\lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^m = \left(\mathbf{I} - \frac{1}{1+\lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^{-1}. \quad (20)$$

Hence,

$$\mathbf{r}^* = \frac{\lambda}{1+\lambda} \left(\mathbf{I} - \frac{1}{1+\lambda} \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2} \right)^{-1} \mathbf{v}. \quad (21)$$

This is the same as the closed-form solution in Eq. (16). \square

4.2 Similarity Estimation

We define a semantic similarity for social images, which is estimated based on their associated tag sets. Note that we have obtained the similarity of tag pair in Eq. (10). Consequently, we estimate the semantic similarity of images x_i and x_j as

$$s(x_i, x_j) = \frac{1}{2|\mathcal{T}_i|} \sum_{k=1}^{|\mathcal{T}_i|} \max_{t \in \mathcal{T}_j} \text{sim}(t_k^i, t) + \frac{1}{2|\mathcal{T}_j|} \sum_{k=1}^{|\mathcal{T}_j|} \max_{t \in \mathcal{T}_i} \text{sim}(t_k^j, t). \quad (22)$$

We can see that the above definition satisfies the following properties:

- (1) $s(x_i, x_j) = s(x_j, x_i)$, i.e., the semantic similarity is symmetry.
- (2) $s(x_i, x_j) = 1$ if $\mathcal{T}_i = \mathcal{T}_j$, i.e., the semantic similarity of two images is 1 if their tag sets are identical.
- (3) $s(x_i, x_j) = 0$ if and only if $\text{sim}(t', t'') = 0$ for every $t' \in \mathcal{T}_i$ and $t'' \in \mathcal{T}_j$, i.e., the semantic similarity is 0 if and only if every pair formed by the two tag sets has zero similarity.

This method is close to Song et al.'s approach [25], which estimates the similarity of images based on their annotated semantic concepts. However, their method simply counts the overlapped concepts of two images and our approach further takes the relationship between different concepts into consideration. Now we explain why we do not use visual similarity which should be the most straightforward approach. First we emphasize that visual diversity and semantic diversity have both been investigated in many research works [25, 28, 31] and both have their own rationality. However, in our scheme search relevance will significantly degrade if we adopt visual similarity. This is because the relevant images are more aggregated in visual space in comparison with semantic space. To quantitatively demonstrate this fact, we first define the aggregation score of relevant images for a query as follows

$$A = \frac{\frac{1}{|\mathcal{R}|^2} \sum_{x_i, x_j \in \mathcal{R}} s(x_i, x_j)}{\frac{1}{|\mathcal{R}| |\bar{\mathcal{R}}|} \sum_{x_i \in \mathcal{R}} \sum_{x_j \in \bar{\mathcal{R}}} s(x_i, x_j)}, \quad (23)$$

where \mathcal{R} and $\bar{\mathcal{R}}$ are the sets of relevant and irrelevant images, respectively.

Table 1 The aggregation score comparison of using visual similarity and using semantic similarity for several queries. Higher aggregation score indicates that relevant samples are more aggregated in the space and the diversifying process is more likely to degrade the search relevance

	Using Visual Similarity	Using Semantic Similarity
Car	1.466	1.112
Forest	1.324	1.048
Hairstyle	1.960	1.013
Jaguar	1.333	1.074
Shark	1.588	1.296
Waterfall	1.097	1.052

Then we compare the aggregation scores using visual similarity and semantic similarity for several queries. The dataset and parameter settings will be described in the next section. Table 1 illustrates the results. From the table we can see that the obtained aggregation scores with visual similarity are much higher than those obtained with semantic similarity. This indicates that relevant images are more aggregated in visual space than semantic space (now we can revisit our second assumption in relevance estimation and see its rationality: the relevance scores of visually similar images should be close). Therefore, in our diverse relevance ranking approach it will be difficult to simultaneously maintain high relevance level and visual diversity. For example, in the process in Fig. 4 if previous images are relevant, then the next image will have high probability to be irrelevant as we enforce it to be visually different with the previous images. Empirical results in the next section will demonstrate this fact and user study will show the superiority of semantic similarity over visual similarity.

5 Empirical Study

5.1 Flickr Dataset

We evaluate our approach on a set of social images that are collected from Flickr. We first select a diverse set of popular tags from the tag list of [32], including *airshow*, *apple*, *beach*, *bird*, *car*, *cow*, *dolphin*, *eagle*, *flower*, *fruit*, *jaguar*, *jellyfish*, *lion*, *owl*, *panda*, *starfish*, *triumphal*, *turtle*, *watch*, *waterfall*, *wolf*, *chopper*, *fighter*, *flame*, *hairstyle*, *horse*, *motorcycle*, *rabbit*, *shark*, *snowman*, *sport*, *wildlife*, *aquarium*, *basin*, *bmw*, *chicken*, *decoration*, *forest*, *furniture*, *glacier*, *hockey*, *matrix*, *Olympics*, *palace*, *rainbow*, *rice*, *sailboat*, *seagull*, *spider*, *swimmer*, *telephone*, and *weapon*. We then perform tag-based image search with “ranking by most recent” option, and the top 2,000 returned images for each query are collected together with their associated information, including tags, uploading time, user identifier, etc. In this way, we obtain a social image collection consisting of 104,000 images and 83,999 unique tags. But many of the raw tags are misspelling and meaningless. Hence, we adopt a pre-filtering process on these tags. Specifically, we match each tag with the entries in a Wikipedia thesaurus and only the tags that have a coordinate

in Wikipedia are kept. In this way, 12,921 unique tags are kept for our experiment, and there are 7.74 tags associated with an image in average.

For each image, we extract 428-dimensional features, including 225-dimensional block-wise color moment features generated from 5-by-5 fixed partition of the image, 128-dimensional wavelet texture features, and 75-dimensional edge distribution histogram features. The ground truth of the relevance of each image is voted by three human labelers. The radius parameter σ in Eq. (12) is empirically set to the median value of all the pair-wise Euclidean distances between images, and the weighting parameter λ is empirically set to 0.1 for all queries.

5.2 Empirical Results

We first compare the following six ranking methods:

- (1) Time-based ranking, i.e., order the images according to their uploading time.
- (2) Relevance-based ranking, i.e., order the images according to their estimated relevance scores $r(x_i)$.
- (3) We cluster the images with Affinity Propagation [8], and the cluster exemplars are put forward in the ranking list and they are ordered according to their relevance scores. The non-exemplars are also ordered based on their relevance scores.
- (4) We first rank images according to their relevance scores and then perform the folding method proposed in [28]. The parameter ϵ is set to the mean value of pair-wise distances among all images.
- (5) Diverse Relevance Ranking (DRR) with visual similarity.
- (6) Diverse Relevance Ranking (DRR) with semantic similarity, i.e., the method proposed in this work.

For simplicity, these methods are denoted as “Time-Based Ranking”, “Relevance-Based Ranking”, “AP-Based Diversifying”, “Folding-Based Diversifying”, “DRR with Visual Similarity” and “DRR with Semantic Similarity”, respectively. The first two methods are baseline and the next two methods diversify search results with clustering approach on a relevance-based ranking list. Figure 6 illustrates the top results of exemplary query “waterfall” and “triumphal”, from which we can see that the results of our method are both relevant and diverse, whereas the results of the other methods are not satisfying in terms of either relevance or diversity. Figure 7(a) and Fig. 7(b) illustrate the AP and ADP measurements obtained by different methods, respectively. We also illustrate the mean AP (MAP) and mean ADP (MADP) measurements that are averaged over all queries. The MAP measurements of the six methods are 0.583, 0.684, 0.646, 0.621, 0.577 and 0.664, respectively, and their MADP measurements are 0.308, 0.361, 0.374, 0.334, 0.331 and 0.411, respectively. It can be found that relevance-based ranking achieves the highest AP measurement, but its ADP measurement is rather low. This indicates that it suffers from the lack-of-diversity problem. Although the AP-Based Diversifying, Folding-Based Diversifying and DRR with Visual Similarity methods can enhance the diversity, they

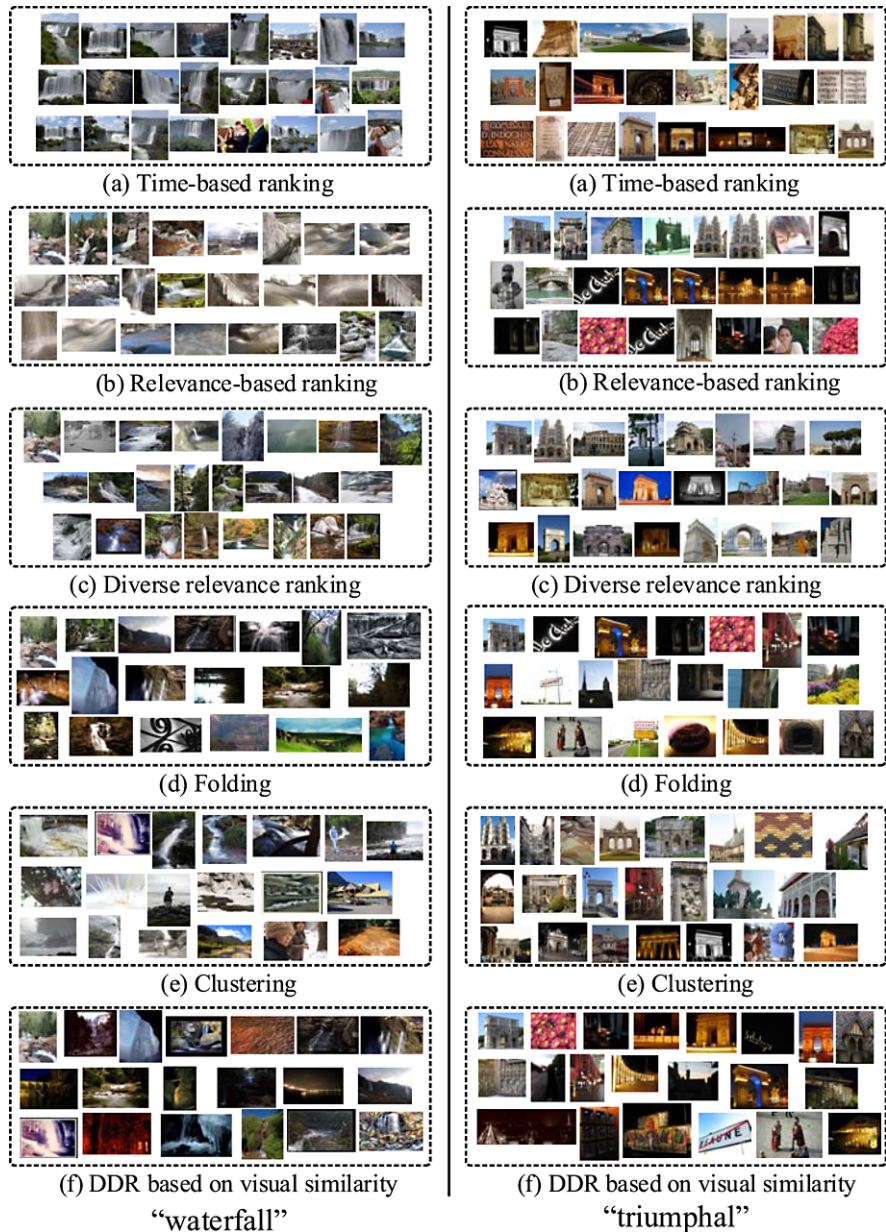


Fig. 6 The top results of different ranking methods of query “waterfall” and “triumphal”

degrade search relevance much in comparison with relevance-based ranking (we have analyzed why DRR with Visual Similarity will degrade search relevance in Sect. 4.2) and thus we can see that their ADP measurements are not high. The DRR

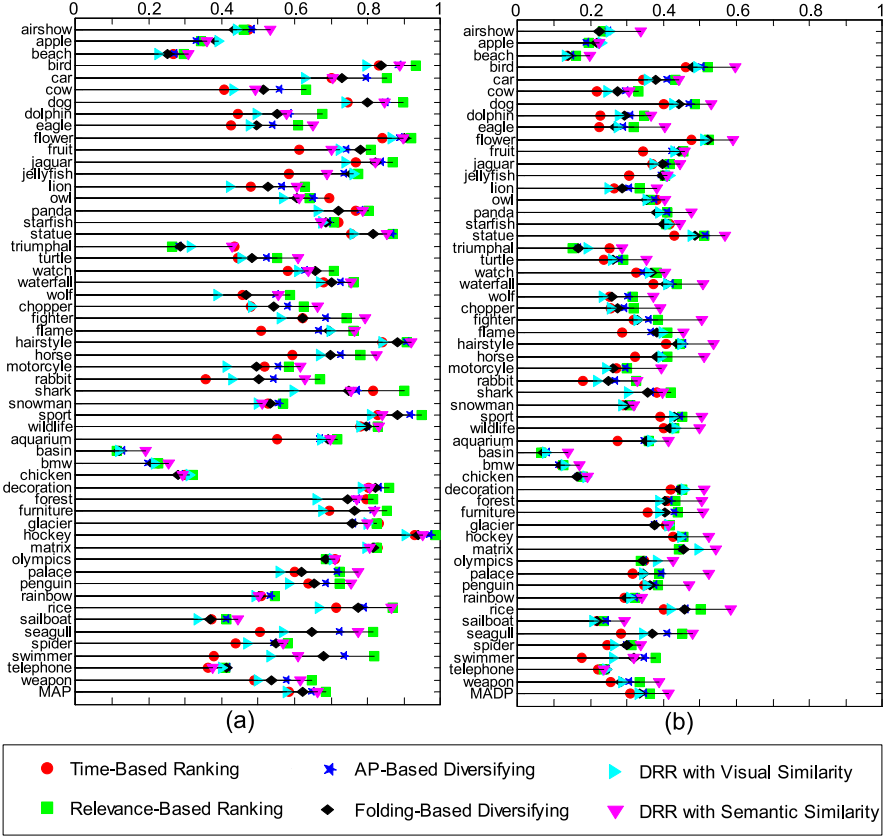


Fig. 7 (a) The comparison of AP measurements of different approaches. (b) The comparison of ADP measurements of different approaches

with Semantic Similarity achieves the best ADP measurements and it only performs slightly worse than Relevance-Based Ranking in terms of AP. This shows that it is able to achieve a good trade-off between relevance and diversity.

We then conduct a user study to compare the six ranking schemes. To avoid bias, a third-party data management company is involved. The company invited 30 anonymous participants, who declare they are regular users of Internet and familiar with image search and media sharing websites. They were asked to freely choose queries and observe image ranking lists. They compared DRR with Semantic Similarity, i.e., our proposed approach, with each of the other five ranking approaches in terms of search relevance and diversity.¹ The users are asked to give the com-

¹It is worth noting that diversity is not directly related to a user's search requirements. Therefore, actually the users are asked to take search relevance and comprehensiveness into account. For search comprehensiveness, we asked them to imagine different search intentions when they posed these queries for themselves, and then it is better if the top results in a list cover more possibilities.

Table 2 The left part illustrates the average rating scores and variances converted from the users study on the comparison of “DRR with Semantic Similarity” and “Time-Based Ranking”. The right part illustrates the ANOVA test results. The p -values show that the difference of the two ranking schemes is significant, and the difference of users is insignificant

DRR with Semantic Similarity vs. Time-Based ranking		The factor of ranking schemes		The factor of users	
DRR with Semantic Similarity	Time-Based ranking	F -statistic	p -value	F -statistic	p -value
2.40 ± 0.386	1.03 ± 0.033	108.57	2.58×10^{-11}	0.656	0.894

Table 3 The left part illustrates the average rating scores and variances converted from the user study on the comparison of “DRR with Semantic Similarity” and “Relevance-Based Ranking”. The right part illustrates the ANOVA test results

DRR with Semantic Similarity vs. Relevance-Based ranking		The factor of ranking schemes		The factor of users	
DRR with Semantic Similarity	Relevance-Based Ranking	F -statistic	p -value	F -statistic	p -value
2.40 ± 0.455	1.133 ± 0.189	46.74	2.5×10^{-11}	0.25	0.999

Table 4 The left part illustrates the average rating scores and variances converted from the user study on the comparison of “DRR with Semantic Similarity” and “AP-Based Diversifying”. The right part illustrates the ANOVA test results

DRR with Semantic Similarity vs. AP-Based Diversifying		The factor of ranking schemes		The factor of users	
DRR with Semantic Similarity	AP-Based Diversifying	F -statistic	p -value	F -statistic	p -value
2.43 ± 0.392	1.10 ± 0.162	62.7	9.85×10^{-9}	0.3	0.999

parison results using “>”, “ \gg ” and “=”, which mean “better”, “much better” and “comparable”. To quantify the results, we convert the results into ratings. We assign score 1 to the worse scheme and the other scheme is assigned a score 2, 3 and 1 if it is better, much better and comparable than this one, respectively. Since there will be disagreements among the evaluators, we perform an ANOVA test [16] to statistically analyze the comparison. The five comparison results are illustrated in Table 2, 3, 4, 5 and 6, respectively. The results demonstrate the superiority of our approach over the other methods. ANOVA test shows that the superiority is statistically significant and the difference of the evaluators is not significant. This further confirms the effectiveness of our approach.

In the user study we also found several failure cases of our approach, such as the top images are irrelevant or not diverse enough. One major reason is the noises of

Table 5 The left part illustrates the average rating scores and variances converted from the user study on the comparison of “DRR with Semantic Similarity” and “Folding-Based Diversifying”. The right part illustrates the ANOVA test results

DRR with Semantic Similarity vs. Folding-Based Diversifying		The factor of ranking schemes		The factor of users	
DRR with Semantic Similarity	Folding-Based Diversifying	F -statistic	p -value	F -statistic	p -value
1.97 ± 0.240	1.13 ± 0.195	33.26	3.02×10^{-6}	0.15	1.0

Table 6 The left part illustrates the average rating scores and variances converted from the user study on the comparison of “DRR with Semantic Similarity” and with “DRR with Visual Similarity”. The right part illustrates the ANOVA test results

DRR with Semantic Similarity vs. DRR with Visual Similarity		The factor of ranking schemes		The factor of users	
DRR with Semantic Similarity	DRR with Visual Similarity	F -statistic	p -value	F -statistic	p -value
2.20 ± 0.441	1.10 ± 0.093	46.37	1.77×10^{-7}	0.37	0.996

tags which result in inaccurate relevance and semantic similarity estimation. Performing a tag refinement step [15, 18, 19] to reduce the noisy tags should further benefit our approach.

5.3 Complexity Analysis

According to the introduction in Sect. 3 and Sect. 4, we can see that the computational costs of relevance estimation, semantic similarity computation and the DRR algorithm all scale as $O(n^2)$. However, the relevance and similarity estimation can be accomplished off-line (an image-tag relevance matrix and a sparse image similarity matrix can be stored). We also do not need to generate the full ranking list using DRR. Actually, we can only generate the list of the first k images with the proposed algorithm, and then the rest images are simply ranked by relevance scores. In our experiments, it needs about 1.2 seconds to accomplish the ranking with 2000 images (Pentium4 3.0G CPU and 2 G memory), and a study on web users shows that the tolerable waiting time for web information retrieval is about 2 seconds [20]. The process can still be speeded up by adopting several strategies. For example, we can rank the images in a piecewise manner, such as first ranking the most relevant 500 images with DRR and put them on the top, and then rank the next most relevant 500 images with DRR, and so forth.

6 Conclusion

This chapter presents a diverse relevance ranking scheme for social image search, which is able to simultaneously take relevance and diversity into account. It leverages both visual information of images and semantic information of the associated tags. The ranking algorithm optimizes an Average Diverse Precision (ADP) measure, which is generalized from the conventional AP measure by integrating with diversity. Experimental results have demonstrated the effectiveness of the approach.

References

1. Agrawal, R., Gollapudi, S., Halverson, A., Leong, S.: Diversifying search results. In: *Proceedings of ACM International Conference on Web Search and Data Mining* (2009)
2. Buckley, C., Voorhees, E.M.: Retrieval evaluation with incomplete information. In: *SIGIR* (2004)
3. Cai, D., He, X., Li, Z., Ma, W.-Y., Wen, J.-R.: Hierarchical clustering of WWW image search results using visual, textual and link information. In: *Proceedings of ACM Multimedia*, pp. 952–959 (2004)
4. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: *Proceedings of SIGIR*, pp. 335–336 (1998)
5. Chen, H., Karger, D.R.: Less is more: probabilistic models for retrieving fewer relevant documents. In: *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, p. 436. ACM, New York (2006)
6. Cilibrasi, R., Vitanyi, P.M.B.: The google similarity distance. *IEEE Trans. Knowl. Data Eng.* **19**, 370–383 (2007)
7. Clarke, C.L.A., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 659–666. ACM, New York (2008)
8. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. *Science* **315**, 972–976 (2007)
9. Goffman, W.: A searching procedure for information retrieval. In: *Information Storage and Retrieval*, vol. 2, pp. 73–78 (1964)
10. 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)
11. Jaimes, A., Chang, S.-F., Loui, A.C.: Detection of non-identical duplicate consumer photographs. In: *Proceedings of ACM Multimedia*, pp. 16–20 (2003)
12. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **20**(4), 446 (2002)
13. Jing, F., Wang, C., Yao, Y., Deng, K., Zhang, L., Ma, W.-Y.: IGroup: web image search results clustering. In: *Proceedings of ACM Multimedia*, pp. 587–596 (2006)
14. Kennedy, L.S., Chang, S.F., Kozintsev, I.V.: To search or to label? predicting the performance of search-based automatic image classifiers. In: *Proceedings of MIR*, pp. 249–258 (2006)
15. Kennedy, L., Slaney, M., Weinberger, K.: Reliable tags using image similarity: mining specificity and expertise from large-scale multimedia databases. In: *WSMC '09: Proceedings of the 1st Workshop on Web-scale Multimedia Corpus*, pp. 17–24. ACM, New York (2009)
16. King, B.M., Minium, E.W.: *Statistical reasoning in psychology and education*. Wiley, New York (2003)
17. Li, J., Wang, J.: Real-time computerized annotation of pictures. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(6), 985–1002 (2008)

18. Li, X.R., Snoek, C.G.M., Worring, M.: Learning tag relevance by neighbor voting for social image retrieval. In: *Proceedings of MIR*, pp. 180–187 (2008)
19. Liu, D., Wang, M., Yang, L., Hua, X.-S., Zhang, H.-J.: Tag quality improvement for social images. In: *Proceedings of ICME*, pp. 350–353 (2009)
20. Nah, F.F.-H.: A study on tolerable waiting time: how long are web users willing to wait. *Behav. Inf. Technol.* **23**(3), 153–163 (2004)
21. Qi, G.J., Hua, X.S., Rui, Y., Tang, J.H., Zha, Z.J., Zhang, H.J.: A joint appearance-spatial distance for kernel-based image categorization. In: *Proceedings of CVPR*, pp. 1–8 (2008)
22. Robertson, S.: The probability ranking principle in IR. *J. Doc.* **33**(294), 294–304 (1977)
23. Rui, Y., Huang, T.S.: Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans. Circuits Syst. Video Technol.* **8**(5), 644–655 (1999)
24. Santos, R.L.T., Macdonald, C., Ounis, I.: Exploiting query reformulations for Web search result diversification. In: *Proceedings of the 19th International Conference on World Wide Web*, pp. 881–890. ACM, New York (2010)
25. Song, K., Tian, Y., Huang, T., Gao, W.: Diversifying the image retrieval results. In: *Proceedings of ACM Multimedia*, pp. 707–710 (2006)
26. Srinivasan, S.H., Sawant, N.: Finding near-duplicate images on the web using fingerprints. In: *Proceedings of ACM Multimedia*, pp. 881–884 (2008)
27. Sun, A., Bhowmick, S.S.: Image tag clarity: in search of visual-representative tags for social images. In: *WSM '09: Proceedings of the First SIGMM Workshop on Social Media*, pp. 19–26. ACM, New York (2009)
28. Van Leuken, R.H., Garcia, L., Olivares, X., Zwol, R.: Visual diversification of image search results. In: *Proceedings of WWW*, pp. 341–350 (2009)
29. Wang, B., Li, Z., Li, M., Ma, W.-Y.: Large-scale duplicate detection for web image search. In: *Proceedings of ICME*, pp. 353–356 (2006)
30. Wang, M., Hua, X.-S., Tang, J., Hong, R.: Beyond distance measurement: constructing neighborhood similarity for video annotation. *IEEE Trans. Multimed.* **11**(3), 465–476 (2009)
31. Weinberger, K.Q., Slaney, M., Van Zwol, R.: Resolving tag ambiguity. In: *MM '08: Proceeding of the 16th ACM International Conference on Multimedia*, pp. 111–120. ACM, New York (2008)
32. Wu, L., Hua, X.-S., Ma, W.-Y., Yu, N., Li, S.: Flickr distance. In: *Proceedings of ACM Multimedia*, pp. 31–40 (2008)
33. Yahoo key scientific challenges program. <http://research.yahoo.com/ksc/multimedia>
34. Yang, K., Wang, M., Hua, X.-S., Zhang, H.-J.: Social image search with diverse relevance ranking. In: *International MultiMedia Modeling Conference (MMM)* (2010)
35. Zhai, C., Cohen, W.W., Lafferty, J.: Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In: *Information Processing and Management*, pp. 10–17 (2006)
36. Zhai, C., Lafferty, J.: A risk minimization framework for information retrieval. *Inf. Process. Manag.* 31–55 (2006)
37. Zhao, W.L., Ngo, C.W.: Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection. *IEEE Trans. Image Process.* **18**(2), 412–423 (2009)
38. Zhu, J., Hoi, S.C.H., Lyu, M.R., Yan, S.: Near-duplicate keyframe retrieval by nonrigid image matching. In: *Proceedings of ACM Multimedia*, pp. 41–50 (2008)



<http://www.springer.com/978-0-85729-435-7>

Social Media Modeling and Computing

Hoi, S.C.H.; Luo, J.; Boll, S.; Xu, D.; Jin, R.; King, I. (Eds.)

2011, XIII, 286 p., Hardcover

ISBN: 978-0-85729-435-7