BOOST SEARCH RELEVANCE FOR TAG-BASED SOCIAL IMAGE RETRIEVAL

Dong Liu[†*], *Xian-Sheng Hua*[‡], *Meng Wang*[‡], *HongJiang Zhang*[§]

[†]School of Computer Sci.& Tec., Harbin Institute of Technology
[‡] Microsoft Research Asia
[§] Microsoft Advanced Technology Center

ABSTRACT

Social media sharing web sites like Flickr allow users to annotate images with free tags, which greatly facilitate social image search and browsing. However, currently tag-based image search on Flickr does not provide the option of relevance-based ranking, i.e., the search results cannot be ranked according to their relevance levels with respect to the query tag, and this has limited the effectiveness of tag-based search. In this paper, we propose a relevance-based ranking scheme for social image search, aiming to automatically rank images according to their relevance to the query tag. It integrates both the visual consistency between images and the semantic correlation between tags in a unified optimization framework. We propose an iterative method to solve the optimization problem, and the relevance-based ranking can thus be accomplished. Experimental results on real Flickr image collection demonstrate the effectiveness of the proposed approach.

Index Terms— Social Media, Relevance Ranking, Tag-based image search.

1. INTRODUCTION

There is an explosion of community-contributed multimedia content available online, such as Youtube, Flickr, Zooomr, etc. Rather than simply searching for and passively consuming information, such media repositories promote users to collaboratively create, evaluate and distribute media information, which underscores a transformation of the Web as fundamental as its birth [1]. Flickr [2], which is the earliest and most popular photo sharing website, claims to host over 200-million personal photos[1] annotated with descriptive keywords called tags. With the rich tags as metadata, users can freely organize, index and search the shared media content, which provides opportunities to make large-scale media retrieval work in practice.

Tag-based search, which returns all images annotated with a specific query tag, is an important way of searching or browsing images on Flickr. Currently, Flickr provides two options in the ranking for tag-based image search result[2]. One is "most recent", which ranks the most recently uploaded images on top and the other is "most interesting", which ranks the images by "interestingness", i.e., a measure that takes into account click-through, comments, etc. These two search options are both useful, but Flickr has not provided the option of ranking by relevance, which is an important requirement for image search [3].

The lack of relevance ranking for tag-based social image retrieval can yield search results that are inconsistent in terms of image relevancy. Fig. 1 shows the top 16 returned results of query tag bird ranked by "most interesting" option. As can be observed, almost half of the images in the ranking list are irrelevant or weakly relevant to user's query. This significantly limits tag-based information seeking application on Flickr, and this directly motivates our work.

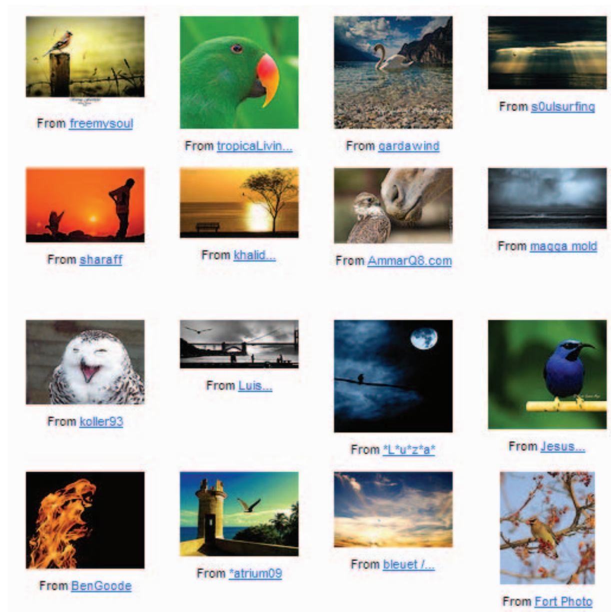

Fig. 1. Search results of query tag "bird" on Flickr, which are ranked by "most interesting". Almost half of the images are irrelevant or weakly relevant to "bird".

There are two information clues that can be advantageous for boosting tag-based image search in terms of relevance. The first comes from tags associated with social images. Note that each social image has a set of tags, and their semantic correlation with the query tag measures the relevance of the image from semantic aspect. The second information clue is the distribution of visual similarities among images. Visual repetition in a large collection of social images is an important signal that can be used to infer a common "visual theme" for the query tag. Finding such consistent visual theme and their relative strength is the basis of the image ranking scheme proposed in this study. More specifically, the relevance levels of two visually similar images should be close.

The visual search ranking problem has been intensively studied in the literature. Some researchers have explored the Pseudo-

*This work was performed at Microsoft Research Asia.
[1]Flickr Blog, Jan. 2008. http://flickr.com/blog.
[2]http://www.flickr.com/search/?q=bird&m=tags

relevance feedback(PRF) [4] and reranking [5, 6] methods to boost search relevance for multimedia retrieval. However, most of them solely rely on visual signal, and thus ignore the potential semantic clues. In this paper, we propose a method to derive relevance-based ranking for tag-based social image search. A unified optimization framework that integrates both semantic correlation between tags and visual consistency between images is proposed to accomplish the ranking. The ranking process prioritizes images with consistent visual theme and is also biased with the preference towards images having higher semantic correlation with the query tag. Empirical results on real Flickr social image dataset show that the proposed method is able to rank search result according to their relevance levels with respect to the query tag.

The main contribution of this paper can be summarized as follows:

- We explore to utilize two complementary information clues associated with social images, i.e., semantic and visual clues, to infer a relevance-based ranking approach for tag-based social image search.
- We propose a unified optimization framework that simultaneously models the visual consistency between images and semantic correlation between tags. We also introduce an iterative method to solve it.

The rest of this paper is organized as follows. We formulate the relevance-based ranking problem in Section 2. In section 3, we provide experimental justification. In section 4, we conclude the paper.

2. RELEVANCE-BASED RANKING

In this section, we first give an overview of our relevance-based ranking approach and then introduce the estimation of semantic scores and the optimization framework.

The following notations will be used throughout the rest of this paper. Given a collection of social images $D = \{x_1, x_2, \ldots, x_n\}$ that are annotated with query tag t_q. Each social image x_i is annotated with a set of tags $T_{x_i} = \{t_1, t_2, ..., t_{|T_{x_i}|}\}$. The relevance scores of all images in D can be represented in a vector $\mathbf{F} = \{f_1, f_2, ..., f_n\}$, whose element $f_i > 0$ denotes the relevance score of image x_i with respect to query tag t_q. Denote by similarity matrix $\mathbf{W}$ whose element W_{ij} denotes the visual similarity between images x_i and x_j. We define another vector $\mathbf{Y} = \{y_1, y_2, ., y_n\}$ whose element $y_i > 0$ denotes the semantic relevance score of image x_i with respect to the query tag t_q(this score will be computed based on the semantic correlation of tags).

2.1. Overall approach

We accomplish the relevance-based ranking with an optimization framework that integrates both the visual and semantic information. We first compute semantic relevance scores of images that estimate their semantic correlation with the query tag by exploring the semantic similarity between any two tags mined from Flickr website. We then further integrate visual consistency of nearby images, and a regularization framework is derived. The solution of the optimization problem is finally adopted as the relevance scores for the given social image collection, based on which the relevance-based ranking can be achieved. In the next two sub-sections, we will detail the computation of semantic relevance scores and the optimization framework, respectively.

2.2. Estimating of semantic relevance scores

We estimate the relevance of each social image based on the semantic correlation of its tag set and the query tag. Firstly, we define the semantic similarity of two tags based on their co-occurrence on Flickr website. Analogous to Google distance [7], the semantic similarity between tag t_i and tag t_j is estimated as t_i and tag t_j is estimated as

$$G(t_i, t_j) = exp(-\frac{max(logf(t_i), logf(t_j)) - logf(t_i, t_j)}{logM - min(logf(t_i), logf(t_j))}) \quad (1)$$

where $f(t_i)$ and $f(t_j)$ are the numbers of images annotated with t_i and t_j respectively, $f(t_i, t_j)$ is the number of images annotated with both t_i and t_j simultaneously, and M is the total number of images on Flickr.

The semantic relevance of a social image x_i with respect to a query tag t_q is now defined as the average semantic similarity between t_q and the tags in T_{x_i}, i.e.,

$$y_i = \frac{1}{|T_{x_i}|} \sum_{t \in T_{x_i}} G(t_q, t) \quad (2)$$

For example, Fig.2(a) and Fig.2(b) are both returned images for query tag “bird”. As can be seen in Fig.2(a), the tags “nature”, “animal” and “wildlife” are closely correlated with ”bird“, and this indicates high relevancy of this image with respect to query “bird”. On the contrary, the image in Fig.2(b) is irrelevant with respect to query “bird” and this has been reflected in the weak semantic correlation between its associated tags and query tag “bird”. Note that the tag correlation can also be computed through other approaches, such as Wordnet [8] or Flickr distance [9]. On the other hand, the tag correlation can be more precisely estimated if we further differentiate each tag’s relevance with respect to the given image, such as in [10, 11].

Fig. 2. Two social images annotated with “bird”. The tags associated with the image in (a) are better correlated with query tag “bird” in comparison with (b). Thus the image in (a) is more relevant than that in (b) with respect to the query “bird”.

2.3. Formulation of relevance-based ranking

As stated previously, we accomplish the relevance-based ranking by integrating both visual consistency and semantic relevance. The visual similarity between images can be directly computed with Gaussian kernel function with a radius parameter σ, i.e.,

$$W_{ij} = exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) \quad (3)$$

Therefore, we compute the relevance scores of images based on a regularization framework [12] as follows

$$Q(\mathbf{F}) = \sum_{i,j=1}^{n} W_{ij}(\frac{f_i}{\sqrt{D_{ii}}} - \frac{f_j}{\sqrt{D_{jj}}})^2 + C\sum_{i=1}^{n}(f_i - y_i)^2 \quad (4)$$

and we aim to solve the optimization problem as

$$\mathbf{F}^\star = \arg\min_{\mathbf{F}} Q(\mathbf{F}) \quad (5)$$

where f_i is the relevance score of x_i and $D_{ii} = \sum_{j=1}^{n} W_{ij}$. The first term of the right-hand side in the cost function means that the relevance scores of visually similar images should be close. The second term is a fitting constraint, which means that the relevance scores are biased with preference to initial semantic relevance measurements. The trade-off between these two competing constraints is captured by a positive parameter C. We can write Eq. 4 in a matrix form as

$$Q(\mathbf{F}) = \mathbf{F}(\mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})\mathbf{F}^\top + C\|\mathbf{F} - \mathbf{Y}\|^2 \quad (6)$$

where $\mathbf{D} = Diag(D_{11}, D_{22}, ..., D_{nn})$

Take derivation of Eq. 6 with respect to **F**, we obtain

$$\frac{\partial Q}{\partial F}|_{\mathbf{F}=\mathbf{F}^*} = (\mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})\mathbf{F}^* + C(\mathbf{F}^* - \mathbf{Y}) = 0 \quad (7)$$

and we can derive that

$$\mathbf{F}^* = \frac{C}{1+C}(\mathbf{I} - \frac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})^{-1}\mathbf{Y} \quad (8)$$

This is the closed-form solution of our optimization framework. We can use the following iterative algorithm to solve **F** to avoid the intensive computation brought by the direct matrix inversion in Eq. 8:

1. Form the image affinity matrix **W** by Eq. 3 if $i \neq j$ and otherwise $W_{ii} = 0$.
2. Compute the semantic relevance scores y_i based on Eq. 2.
3. Iterate $\mathbf{F}(t+1) = \frac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}\mathbf{F}(t) + \frac{C}{1+C}\mathbf{Y}$ until convergence.

Now we prove the convergence of the proposed iteration optimization function defined above.

Theorem 1. *The iteration of our optimization method convergences to a fixed point* $\mathbf{F}^*$.

Proof. We re-write the iterative function as follows

$$\mathbf{F}_{t+1} = \frac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}\mathbf{F}_t + \frac{C}{1+C}\mathbf{Y} \quad (9)$$

and thus we have

$$\begin{aligned}\mathbf{F}^* = &\lim_{t\to\infty}(\tfrac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})^t\mathbf{Y} + \\ &\tfrac{C}{1+C}(\textstyle\sum_{t=1}^{n}(\tfrac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})^t)\mathbf{Y}.\end{aligned} \quad (10)$$

Since $0 < \frac{1}{1+C} < 1$ and the eigenvalues of matrix $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ lies in $(0, 1)$, we have $\lim_{t\to\infty}(\frac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})^t = 0$ and $\lim_{t\to\infty}\sum_{t=1}^{n}(\frac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})^t = 0$. Hence, we have

$$\mathbf{F}^* = \frac{C}{1+C}(\mathbf{I} - \frac{1}{1+C}\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})^{-1}\mathbf{Y} \quad (11)$$

This is the same with the closed form solution in Eq. 8. The relevance-based ranking is accomplished exactly based on the order with f_i in descending order. □

Methods	Average NDCG@30
SRR	0.7563
VRR	0.7816
Proposed	0.8162

Table 1. Average NDCG@30 of different image ranking strategies.

3. EXPERIMENT

3.1. Experimental setup

We conduct experiments with the data collected from Flickr. We select ten most popular queries, including *cat*, *sky*, *mountain*, *automobile*, *water*, *flower*, *bird*, *tree*, *sunset* and *sea*, and use them as query keywords to perform tag-based image search on Flickr. The search results are displayed using the "ranking by interestingness" option, and then the top 5000 images for each query are collected together with their associated information including tags, user ID, etc. In this way, we obtain a social image dataset comprising of $50,000$ images and $106,565$ unique tags. But many of the raw tags are misspelling and meaningless. Hence, we adopt a pre-filtering process for 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, $13,330$ unique tags are kept for our experiment. For each image, we extract 353-dimensional features, including 225-dimensional block-wise color moment generated from 5-by-5 partition of the image and 128-dimensional wavelet texture feature. The radius parameter σ in Eq. 3 is set to the median value of all pair-wise Euclidean distances between images, and the parameter C in Eq. 4 is empirically set to 1.

We use NDCG(**N**ormalized **D**iscounted **C**umulative **G**ain) [13], which can handle multiple levels of relevance judgments, as the performance evaluation measure. With the ranking scores residing in **F**, we rank images annotated with the query tag by the scores in descending order. Each image is labeled as one of five levels with respect to the query: Most Relevant (score 5), Relevant (score 4), Partially Relevant (score 3), Weakly Relevant (score 2), and Irrelevant (score 1). Given an image ranking list, the NDCG is computed as

$$N_n = Z_n \sum_{i=1}^{n}(2^{r(i)} - 1)/log(1+i) \quad (12)$$

where $r(i)$ is the relevance level of the i-th image and Z_n is a normalization constant that is chosen so that the optimal ranking's NDCG score is 1. After computing the NDCG measures of the image ranking lists for each query, we can average them to obtain an overall performance evaluation.

3.2. Experimental results

We compare our image relevance-based ranking approach with the following two image ranking strategies:

- Visual information based Relevance Ranking (**VRR**). The relevance of each image is only estimated from the visual consistency between images, i.e., each element in vector **Y** is set to $1/n$.
- Semantic information based Relevance Ranking (**SRR**). In this method, we directly rank search results based on the semantic relevance scores computed in Eq. 2.

Our method can be viewed as the combination of the first two methods. The results are illustrated in Fig. 3. From the figure we can see that the proposed method outperforms the other two methods for most of queries (7 out of 10 queries). Table 1 illustrates the average NDCG of the three methods, from which we can clearly see the superiority of the proposed method. We also illustrate the NDCG performances of the three methods with different depths in Fig. 4, from which we can see that our method consistently outperforms the other two methods. Fig. 5 illustrates the top 16 results of query "bird" with our method. Comparing them with the results in Fig. 1 that are returned with interestingness-based ranking options, we can clearly see that our top results are more relevant with respect to the query.

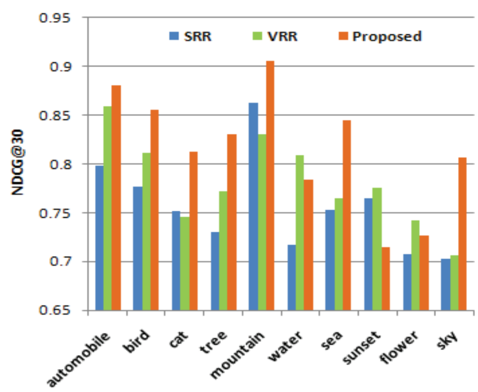

Fig. 3. NDCG@30 values of different image ranking methods across ten query.

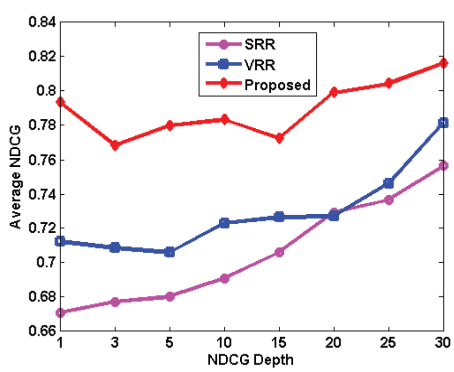

Fig. 4. Average NDCG comparison with varied depths using different ranking methods.

4. CONCLUSION

In this paper, we have proposed an approach to accomplish relevance-based ranking for social image search. It integrates both the visual consistency between images and the semantic correlation between tags in a unified optimization framework. Then an iterative solution method is adopted to compute the relevance scores. Our experimental results on real-world social image data demonstrate that the proposed method is able to order social images according to their relevance levels.

It is worth noting that although we have only used Flickr data in this work, the proposed relevance ranking method for tag-based search result is a general approach and can be applied for other social media sources (e.g., Youtube) as well.

Fig. 5. The top results of query "bird" with the proposed relevance-based ranking strategy.

5. REFERENCES

[1] P. Anderson. What is Web 2.0? Ideas, Technologies and Implications for Education. In *JISC Technology and Standards Watch*, 2007.

[2] Flickr. *http://www.flickr.com.*

[3] Y. S. Jing and S. Baluja. VisualRank: Applying PageRank to Large-Scale Image Search. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

[4] R. Yan, A. Hauptmann and R. Jin. Multimedia Search with Pseudo-Relevance Feedback. In *Proceedings of International Conference on Content-based Image and Video Retrieval*, 2003.

[5] W. Hsu, L. Kennedy and S. F. Chang. Novel Reranking Methods for Visual Search. In *IEEE Multimedia*, 2007.

[6] J. J. Liu, W. Lai, X. S. Hua, Y. L. Huang and S. P. Li. Novel Video Search Reranking via Multi-graph Propagation. In *Proceeding of 12th ACM International Conference on Multimedia*, 2007.

[7] R. Cilibrasi, P. M. B. Vitanyi. The Google Similarity Distance. In *IEEE Transactions on Knowledge and Data Engineering*, 2007.

[8] Y. Jin, L. Khan, L. Wang, and M. Awad. Image Annotation by Combining Multiple Evidence & WordNet. In *Proceeding of 12th ACM International Conference on Multimedia*, 2005.

[9] L. Wu, X. S. Hua, N. H. Yu, W. Y. Ma and S. P. Li. Flickr Distance. In *Proceeding of 15th ACM International Conference on Multimedia*, 2008.

[10] D. Liu, X. S. Hua, L. J. Yang, M. Wang and H. J. Zhang. Tag Ranking. In *Proceeding of ACM International World Wide Web Conference*, 2009.

[11] X. R. Li, C. G. M. Snoek, and M. Worring. Learning Tag Relevance by Neighbor Voting for Social Image Retrieval. In *Proceedings of the ACM International Conference on Multimedia Information Retrieval*, 2008.

[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston and B. Schölkopf. Learning with Local and Global Consistency. In *Advances in Neural Information Processing Systems*, 2004.

[13] K. Jarvelin and J. Kekalainen. Cumulated Gain-Based Evaluation of IR Techniques. In *ACM Transcations on Information System*, 2002.