Image Search Reranking with Multi-latent Topical Graph

Junge Shen
Xidian University,
Xi'an 710071,
P. R. China
Email: shenjunge@gmail.com

Tao Mei
Microsoft Research Asia,
Beijing 100080,
P. R. China
Email:tmei@microsoft.com

Qi Tian
University of Texas
at San Antonio,
USA
Email:qitian@cs.utsa.edu

Xinbo Gao
Xidian University,
Xi'an 710071,
P. R. China
Email:xbgao@xidian.edu.cn

***Abstract*— Image search reranking has attracted extensive attention. However, existing image reranking approaches deal with different features independently while ignoring the latent topics among them. It is important to mine multi-latent topic from the features to solve the image search reranking problem. In this paper, we propose a new image reranking model, named reranking with multi-latent topical graph (RMTG), which not only exploits the explicit information of local and global features, but also mines multi-latent topic from these features. We evaluate RMTG over the MSRA-MM dataset and show that RMTG outperforms several existing reranking methods.**

I. Introduction

With the development of multimedia technologies and the tremendous success of social media, millions of images are uploaded and shared per day. Image search becomes more and more crucial to information retrieval. Currently, most of available Internet image searching engines are on the basis of "query by keyword". Due to the semantic gap between the textual and visual search, visual search reranking has attracted broad attentions in recent years to make up for the deficiencies of current text-based retrieval. Through a number of studies conducted in this field, we can summarize the following difficulties for reranking: 1) image document representation— it is an important foundation of the visual search system, as the representation of visual documents can affect the performance of the successive stages; and 2) reranking model—based on the initial search results, it is necessary to rerank the results according to some relevance model.

Various approaches have been proposed to tackle the above difficulties, where the reranking methods are mainly based on low-level features which are classified into global features and local features, [1], [2], [3]. However, there are several challenges for the above methods. If the similarities of the images are estimated only by global or local features, the returned images cannot be satisfied for all the queries. Figure 1 shows visual examples that each feature has its strengths and limitations. The first row is the reranked images with the query "apple" which gives good performance based on the global features, while the returned images with the query "butterfly" which gives good performance based on the local features. Hence, it is really difficult to determine which kind of feature is more suitable. For this reason, combining different visual

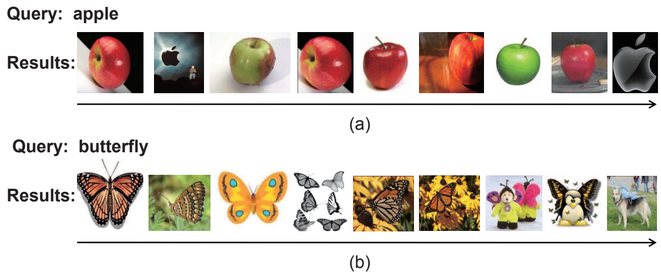

Fig. 1. Visual examples of reranking methods based on global feature and local feature. The reranking oder is the direction of the arrow. (a) is reranking of "apple" query results based on the global feature, and (b) is reranking of "butterfly" query results based on the local feature.

features will achieve significant and expected improvement over the visual search baseline with an individual feature.

Based on previous analysis, we proposed a new approach, called reranking with multi-latent topical graph (RMTG). We mine the multi-latent topical graph via different features with the inspiration of semi-supervised methods. The multi-latent topical link structure is represented by a connected graph. Figure 2 gives a visual example to show how to rerank with the graph when given the query "sports". Figure 2 (a) shows the explicit links between the images, and the solid lines represent the similarities of images which are weighed by the features. Unlike combining two kinds of feature matrices directly, we select matrix factorization to solve our reranking problem [4],[5], . The multi-latent topical feature vector should be learnt for the images by joining two features. Then, the multi-latent topical graph is constructed by the latent vector.

The novelties of the proposed image reranking approach can be listed as follows:

- Our approach can be used to rerank the top ranked images with semi-supervised machine learning.
- We incorporate two visual features into multi-latent topic analysis which can not only preserve the two kinds of visual features but also mine the information of latent feature .
- Our solution is efficient. Our method can be divided into two parts, online and off-line. Since the latent space graph is learnt off-line, given a query, we are able to achieve real-time image reranking.

978-1-4673-5762-3/13/$31.00 ©2013 IEEE

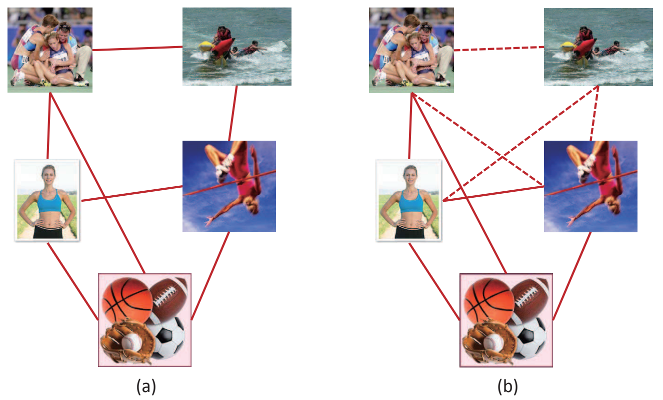

Fig. 2. An example of our reranking rules is shown by the top ranked results of "sports" query. (a) is a connected graph formed by using the similarities of images based on the explicit features. (b) is a latent space graph connected by multi-latent topical feature and the latent links are shown by dotted lines.

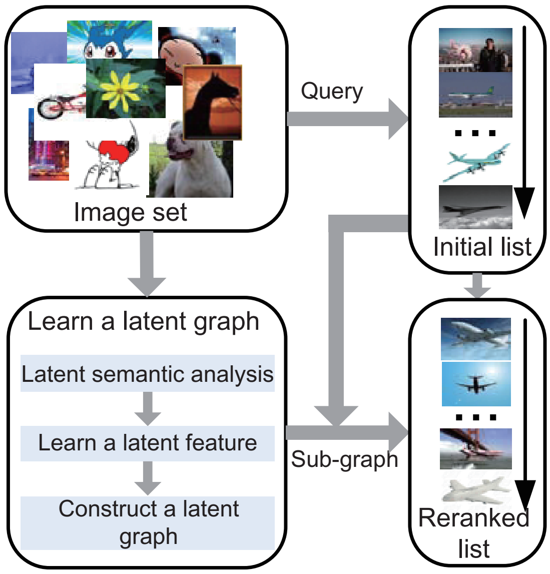

Fig. 3. The framework of reranking with multi-latent topical graph (RMTG).

The rest of paper is organized as follows. Section 2 introduces the framework and our approach in detail. Section 3 describes dataset and evaluations of the experiments. Finally, we conclude the paper in Section 4.

II. MULTI-LATENT TOPICAL GRAPH FOR IMAGE SEARCH RERANKING

A. Approach Overview

The purpose of our model is to mine multi-latent topical features between the global and local visual features, and then the multi-latent topic can be used to rerank the search results. Figure 3 illustrates the framework of our model. We first extract the global and local features separately. Secondly, the multi-latent topical feature can be mined by the latent semantic analysis [6] which is formulated as an optimization problem. Then the multi-latent graph is constructed. Given a textual query, an initial reranking list is obtained by current search engine and a sub-graph can be extracted from the latent graph by indexing the original images. Finally, the optimal reranked list can be obtained.

B. Problem Definition

Suppose we have an image set $M = \{m_1, m_2, \dots m_i, \dots\}$ to be reranked when given a query q. Let r'_j and r_j denote the initial ranking score and the reranking score for image m_j. Each image can be represented by a feature vector $\Sigma \in R^m$. And let $G = (V, E)$ be a directed graph, where the node-set V represents the images and the edges E represents the latent links between images. Assume that $W = \{w_{ij}\}$ is the $n \times n$ adjacency matrix, in which w_{ij} denotes the weight between m_i and m_j. And D is a diagonal matrix where $D_{i,i} = \sum_j w_{i,j}$. In terms of the reranking rules, we can formulate the reranking problem by minimizing the following loss function:

$$Q(R, q, G) = \frac{1}{2}\sum_{i,j=1}^{n} \omega_{ij} \left\| \frac{r(m_i, q)}{\sqrt{D_{ii}}} - \frac{r(m_j, q)}{\sqrt{D_{jj}}} \right\|^2 + \mu \sum_{i=1}^{n} \left\| r(m_i, q) - r'(m_j, q) \right\|^2 \quad (1)$$

where $r'(m_j, q)$ is the initial ranking score, and $r(m_j, q)$ is the reranking score. The initial ranking and reranking score vector can be denoted as R' and R respectively.

Finally, given a query q, the initial score and the latent space graph, R can be evaluated. The reranking score vector is given by

$$R = (I - \xi S)^{-1} R' \quad (2)$$

where ξ ranges from 0 to 1 [7], [8], [9].

C. Learning a latent space graph

1) Latent semantic analysis: Latent Semantic Analysis (L-SA) [6] is to create vector-based representations of documents with reduced dimensionality and has been applied in many multimedia applications. Citing our local-feature matrix as an instance, an $n \times m$ matrix of the BoW, denoted as B, whose rows represent visual words and columns correspond to images, can be disposed by singular vector decomposition (SVD):

$$B = TSD^T \quad (3)$$

where both T and D have orthonormality, i.e. $T^T T = D^T D = I$, and S is the diagonal matrix which has the singular value of B. A new diagonal matrix S_0 is obtained by introducing zeros into S, and the representation can be simplified with the k singular values. The matrix $\Sigma = T_0 S_0$ ($\Sigma \in R_{n\times k}$) is a new representation, and each row of the matrix is the latent feature vector of each image. Finally, the matrix can be represented as

$$B \approx B' = T_0 S_0 D_0^T = \Sigma D_0^T. \quad (4)$$

The corresponding optimization problem can be represented as

$$\min_{\Sigma, D_0} \left\| B - \Sigma D_0^T \right\|_F^2 + \gamma \left\| D_0 \right\|_F^2 \quad (5)$$

where the matrix $D_0 \in R^{m\times k}$, γ is a small positive value, and $\|.\|_F$ is the Frobenius norm. And k needs large enough to fit all the data and small enough to work well for reducing dimensionality.

2) Joint optimization: Suppose we have two feature matrixes. The local-feature matrix is an $n \times p$ matrix of images, denoted as B. The global feature matrix, image-CM [4],[10] , is described as matrix C of size $n \times l$, whose columns represents color information of each image. Intuitively, the shared representation Σ [8] should preserve both the structures of the matrix B and C. Hence, they share the same Σ, and following is the optimization problem to be solved [8],

$$\min_{\Sigma, D_B, D_C} \left\| B - \Sigma D_B^T \right\|_F^2 + \gamma \left\| D_B \right\|_F^2 + \lambda \left(\left\| C - \Sigma D_C^T \right\|_F^2 + \gamma \left\| D_C \right\|_F^2 \right) \tag{6}$$

3) Construct latent space graph: The goal of the optimized matrix Σ, is to construct the multi-latent topical graph [9]. Given the optimized matrix Σ, the latent graph can be constructed with K nearest neighbors (KNN). Let the edge weight w_{ij} denote the similarity of two images, m_i and m_j. Hence, we can formulate the weight-matrix by

$$w_{ij} = exp^{-\|\varepsilon_i - \varepsilon_j\|^2 / 2\sigma^2}, \tag{7}$$

where σ is a parameter for the heat kernel, and ε_i is the vector of matrix Σ. Then W is normalized with the formula $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$. The latent space graph is denoted as S.

D. Reranking

Since we are interested in the top ranked results instead of all the retrieval results, the top reranking score will be more sensible than the whole one. Hence, we extract the latent space sub-graph S' from S for the high efficiency. The sub-graph S' is formed by indexing the top-ranked images. Such a reranking task is formalized as:

$$\widehat{R} = (I - \xi S')^{-1} \widehat{R'}, \tag{8}$$

where $\widehat{R}$ is a $t \times 1$ vector, and t is the number of the top-ranked images.

III. EXPERIMENTS

A. Dataset

We use the images collected from the MSRA-MM (Microsoft Research Asia Multimedia) dataset [10].To demonstrate the effectiveness and efficiency of our approach, we randomly selected 73 queries from these categories which contain $54,474$ images in total to learn the latent space graph. And 20 queries are selected to evaluate our reranking model.

B. Evaluations

We evaluate the performance of our proposed model compared with the following methods: 1) CrowdReranking [11]. This representative method mines relevant visual patterns from search results crowded from some search engines with query examples. 2) IB Reranking [12]. This is a representative model to discover relevance-consistent patterns from the initial ranked list without any query examples or pre-trained concept detectors.

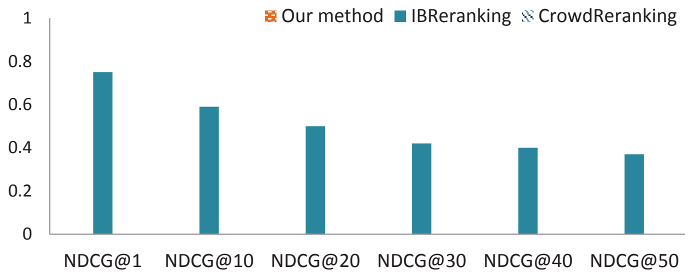

Fig. 5. The performance measured by NDCG@p of varying the parameter p.

We use the manually labeled relevance as ground truth to evaluate our reranking model with the metric NDCG [13]. It is defined as:

$$NDCG@p = C_p \sum_{p}^{i=1} \left(2^{rel_i}\right) / log(1+i)) \tag{9}$$

where rel_i is the scaled relevance level of the $i-th$ reranked image. C_p is a constant to normalize the value of $NDCG@p$. We choose BoW feature as the local feature and color moment as the global feature. The experimental results are shown in Figure 4. The performance of some queries has significant improvement. The latent visual feature contains not only semantic analysis and color information but also latent information between them. Since we mine the latent feature which incorporate local and global features, in CrowdRanking and IB reranking, we should calculate NDCG with local feature and global feature respectively. And the NDCG value of each method is the average value. All the evulated experiments set $k = 200$ in terms of $NDCG@10$. As a result, our model is of great importance in image reranking with good performance.

Furthermore, a few numbers of parameters in our model affect the performance. Some experiments are designed to see the variation of performance when we vary them respectively.

- **Parameter** p. Figure 5 shows the comparison of reranking methods in terms of NDCG. Each result is the average of 20 queries.
- **Parameter** k. In our method, the local feature is a matrix in size $n \times 2000$, and the global feature is an $n \times 255$ matrix, where n corresponds to the number of images. When k ranges from 100 to 200, we evaluate the performance of our model with different k in terms of $NDCG@10$. Finally, we set $k = 150$.
- **Parameter** μ **and** λ. As we known, both μ and λ range from 0 to 1. And the values of them reflect the weight of the feature in a certain way. And also, the experiments set $k = 200$ in terms of $NDCG@10$. To set the parameter, we not only use empirical values and also consider the efficiency of the experiments. Finally, μ ranges from 0.35 to 0.6, and λ ranges from 0.1 to 0.2. In our experiments, we set $\mu = 0.5$, and $\lambda = 0.15$.

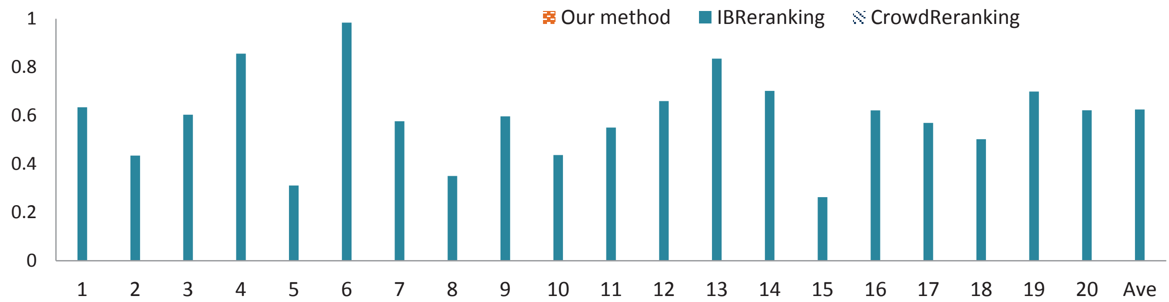

Fig. 4. Performance of each query in the dataset measured by NDCG@10 and the averaged performance of different methods.

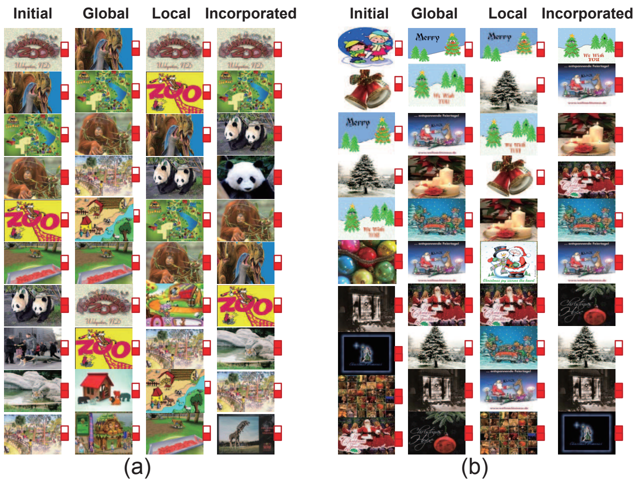

Fig. 6. Visual examples of different features.

C. Examples

In order to show the good performance of the multi-latent topical graph learnt from local and global features. Figure 6 shows some visual examples of the top 10 images of our model with local features, global features and latent features. Latent Semantic Analysis is explored to construct the graph using local and global features respectively. The latent feature gives better results, so we can easily see our proposed method gets the most satisfying.

D. Discussion

The effectiveness and efficiency of our approach come from two parts. Firstly, the off-line part uses the image collection to learn a latent space graph. In our experiment, we use joint link-content matrix factorization, the benefit of which can leverage the global and local features. In the process of learning multi-latent topical graph, it spends plenty of time on optimization. However,the online part can perform well with the high efficiency, for the subgraph extraction speed up the performance. Thus, the off-line and online parts can seek a balance between effectiveness and efficiency.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we present a novel visual reranking model based on multi-latent topical graph. Reranking with multi-latent topical graph (RMTG) not only exploits the explicit information of local and global features, but also mines multi-latent topical information. Moreover, the reranking scores are refined by the multi-latent topic space graph. Experiments on an image collection show that our model not only makes innovations but achieves good performance. It still remains some problem for further study. We will speed up the process of learning a latent space graph when the training set is quite large. Furthermore, the framework can be further extended to add more information like Wikipedia and Google for visual reranking.

REFERENCES

[1] L. Kennedy and M. Naaman, “Generating diverse and representative image search results for landmarks,” in *Proceedings of WWW*, 2008.

[2] Y. Liu and T. Mei, “Optimizing visual search reranking via pairwise learning,” *IEEE Transactionsons on Multimedia*, vol. 13, pp. 280–291, April 2011.

[3] Y. Liu, T. Mei, X. Wu, and X.-S. Hua, “Multi-graph-based query-independent learning for video search,” *IEEE Transactionsons on Circuits and Systems for Video Technology*, vol. 19, no. 12, pp. 1841–1850, December 2009.

[4] H. Deng, M. R. Lyu, and I. King, “Effective latent space graph-based re-ranking model with global consistency,” in *WSDM*, 2009, pp. 212–221.

[5] Y. Liu, T. Mei, G.-J. Qi, X. Wu, and X.-S. Hua, “Query-independent learning for video search,” in *In Proceedings of ICME*, June 2008, pp. 1249–1252.

[6] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman, “Indexing by latent semantic analysis,” *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.

[7] Q. Mei, D. Cai, D. Zhang, and C. Zhai, “Topic modeling with network regularization,” in *Proceedings of the 17th International Conference on World Wide Web*, 2008, pp. 101–110.

[8] S. Zhu, K. Yu, Y. Chi, and Y. Gong, “Combining content and link for classification using matrix factorization,” in *Proceedings of the 30th ACM SIGIR Conference on Research and Development in Information Retrieval*, 2007, pp. 487–494.

[9] X. Zhu, “Semi-supervised learning with graphs,” in *Ph.D. thesis*, Carnegie Mellon Universit, 2005.

[10] H. Li, M. Wang, and X.-S. Hua, “Msra-mm 2.0: A large-scale web multimedia dataset,” in *International Workshop on Internet Multimedia Mining, in Association With ICDM*, 2009.

[11] Y. Liu, T. Mei, and X.-S. Hua, “Crowdreranking: Exploring multiple search engines for visual search reranking,” in *Proceedings of ACM SIGIR Internation Conference on Information Retrieval*, Boston, USA, 2009.

[12] W. Hsu, L. Kennedy, and S. Chang, “Video search reranking via information bottleneck principle,” in *ACM Multimedia*, Santa Babara, CA, USA, 2006.

[13] W. Croft, D. Metzler, and T. Strohman, *Search engines information retrieval in practice*. PEARSON, 2010.