Computer Science and Information Systems 14(3):629–641 https://doi.org/10.2298/CSIS170212020J

Click-Boosted Graph Ranking for Image Retrieval

Jun Wu[1,2], Yu He[1], Xiaohong Qin[1], Na Zhao[2], and Yingpeng Sang[3]

[1] School of Computer and Information Technology, Beijing Jiaotong University
Beijing 10044, China
{wuj, 15120398, 14120420}@bjtu.edu.cn
[2] Logistics and E-commerce College, Zhejiang Wanli University
Ningbo, 315100, China
zhaona@zwu.edu.cn
[3] School of Information Science and Technology, Sun Yat-Sen University
Guangzhou 510275, China
sangyp@mail.sysu.edu.cn

Abstract. Graph ranking is one popular and successful technique for image retrieval, but its effectiveness is often limited by the well-known semantic gap. To bridge this gap, one of the current trends is to leverage the click-through data associated with images to facilitate the graph-based image ranking. However, the sparse and noisy properties of the image click-through data make the exploration of such resource challenging. Towards this end, this paper propose a novel click-boosted graph ranking framework for image retrieval, which consists of two coupled components. Concretely, the first one is a click predictor based on matrix factorization with visual regularization, in order to alleviate the sparseness of the click-through data. The second component is a soft-label graph ranker that conducts the image ranking by using the enriched click-through data noise-tolerantly. Extensive experiments for the tasks of click predicting and image ranking validate the effectiveness of the proposed methods in comparison to several existing approaches.

Keywords: Image Retrieval, Click-Through Data, Graph Ranking, Matrix Factorization.

1. Introduction

Graph ranking [34] has received increasing attention in recent years due to its superiority in various visual ranking tasks, such as natural image search [5], video search [7], shape retrieval [2], cross-media retrieval [30], 3D object retrieval [1], etc. Unlike traditional visual ranking that considers only the pairwise similarity between visual documents, graph-based visual ranking aims to explore the intrinsic manifold structure collectively hidden in visual documents, hoping to refine the similarity measure. Despite these successes, the performance of graph-based visual ranking is still limited by the well-known semantic gap existing between low-level image pixels captured by machines and high-level semantic concepts perceived by humans, especially when the visual targets are dispersed in the feature space.

In order to boost the performance of image retrieval and overcome the semantic gap, a relevance feedback mechanism [35] is incorporated into the graph-based ranking framework [4, 27], which encourages the user to label a few images returned as either positive or negative in terms of whether they are relevant to user's query or not. The labeled instances is then used to refine the ranking model towards the user's query intends. However, it is not easy to obtain sufficient and explicit user feedback as users are often reluctant to provide enough feedbacks to search engines. It is noted that search engines can record queries issued by users and the corresponding clicked images. Although the clicked images cannot reflect the explicit user preference on relevance of particular query-image pairs, they statistically indicate the implicit relationship between individual images in the ranked list and the given query [33]. Therefore, we can regard the click-through data associated with images as the 'implicit' feedbacks based on an assumption that, in a same query session, most clicked images are relevant to the given query, and the reliability of this assumption has been empirically validated by [18].

We consider a particular ranking scenario where the click-through data is sparse and inaccurate. Such a scenario is pervasively presented in the image retrieval problem for which some methods [8, 33, 11] employ the click-through data as implicit feedbacks for image ranking. Inevitably, the sparsity may lead to an underfitting ranker, and the inaccuracy may further mislead the ranker. Towards this end, this paper presents a Click-Boosted Graph Ranking (CBGR) approach for effective image retrieval based on a preliminary work [12], which incorporates click enriching with noise-tolerant graph ranking within an unified framework. The main contributions are two-fold. For one thing, a Visual Regularized Matrix Factorization (VRMF) method is proposed to enrich the click-through data. For another, a Soft-Label Graph Ranking (SLGR) technique is developed to leverage the enriched click-through data noise-tolerantly. An empirical study on the tasks of click predicting and image ranking shows encouraging results in comparison to several exiting approaches.

The rest of this paper is organized as follows. Section 2 reviews the related works. Section 3 presents the proposed CBGR method. Section 4 reports on the experiments. Finally, section 5 concludes this paper and raises the problem for future works.

2. Related Work

We briefly group related work into three dimensions: graph ranking, collaborative image retrieval and collaborative filtering, and introduce them separately in the following subsections.

2.1. Graph Ranking

Graph ranking has been extensively studied in the multimedia retrieval area. Its main idea is to describe the dataset as a graph and then decide the importance of each vertex based on local or global structure drawn from the graph. One canonical graph-based ranking technique is the Manifold Ranking (MR) algorithm [34], and He et al. [5] first applied MR to image retrieval. Its limitations are addressed by latter research efforts. For example, Wang et al. [24] improved the MR accuracy using a k-regular nearest neighbor graph that minimizes the sum of edge weights and balances the edges in the graph as well. Wu et al. [27] proposed a self-immunizing MR algorithm that uses an elastic kNN graph to exploit unlabeled images safely. Wang et al [25] proposed a multi-manifold ranking method, which jointly exploits multiple visual modalities to encode the image ranking results. Xu et al. [29] proposed an efficient MR solution based on scalable graph structure to handle large-scale image datasets. In addition, the users' feedbacks are easily exploited by MR method, and previous studies have shown that MR is one of the most successful ranking approaches for the image retrieval with relevance feedback [5, 27, 29].

Note that most existing methods of graph ranking receive the supervision signals provided by the users directly. Differently, in this work the supervision signal, derived from the click-through data, is noisy, and directly using that may degenerates the retrieval performance. Hence this work presents a soft-label graph ranking solution for the noise-tolerance purpose.

2.2. Collaborative Image Retrieval

Collaborative Image Retrieval (CIR) regards the click-through data associated with images as the long-term experience and leverages it to boost the short-term learning with relevance feedback. For example, Yin et al. [31] exploited the long-term experiences to select the optimal online ranker from a set of candidates based on reinforcement learning. Hoi et al. [6] regarded the query log as the 'side information', and then, taking that as constraints, learned a distance metric form a mixture of labeled and unlabeled images. Su et al. [19] suggested discovering the navigation patterns from query logs, and using the patterns to facilitate new searching tasks. Wu et al. [28] proposed a multi-view manifold ranking method, which simultaneously exploits the visual and click features to encode the image ranking results.

In contrast, our proposed approach requires no users' feedbacks once the query has been issued. Alternatively, it automatically derives implicit feedbacks from the click-through data. This is motivated by empirical evidence suggesting that few users are willing to perform any form of feedback to improve their search results.

2.3. Collaborative Filtering

Collaborative filtering (CF) [20] is a family of algorithms popularly-used in recommendation systems. Depending on how the data of user-item rating matrix are processed, two types of methods, neighbor based and latent factor based, can be differentiated.

Neighbor-based methods use similarity measures to select users (or items) that are similar to the active user (or the target item). Then, the prediction is calculated from the ratings of these neighbors. Most of neighbor-based approaches can be further categorized as user-based or item-based depending on whether the process of finding neighbors is focused on similar users [14] or items [16]. Latent factor based methods are an alternative approach that tries to explain the ratings by characterizing both users and items on a few factors inferred from the user-item rating matrix. Matrix Factorization (MF) [9] might be one of the most promising techniques due to its excellent performance, as witnessed by the Netflix contest. During the past years, plenty of research effort has been made to further improve its effectiveness and efficiency, including maximum margin MF [13], Bayesian MF [15], online MF [10] and parallel MF [3], etc. Besides user-item rating matrix, a current trend is to leverage the plentiful side-information around user and item dimensions to enhance MF performance [17].

In the scenario of this work, the user-image clicking matrix is very similar to the user-item rating matrix in recommender system. Inspired by recommender system, we consider

MF for the purpose of click prediction. Different from the traditional recommendation problems, the 'items' in our scenario are the database images, which have plentiful visual content. Considering this, we present a VRMF algorithm that can exploit the images' visual information to improve the prediction accuracy.

3. The Proposed CBGR Approach

Our CBGR approach is developed based on two intuitions. At first, a 'good' visual ranker should be able to exploit the implicit feedback (click-through data) rather than the explicit feedback for the purpose of alleviating the user's labeling burden. Furthermore, the ranker should be able to handle the sparse and noisy properties of the click-through data.

As mentioned, our CBGR approach consists of two key components, i.e., VRMF and SLGR. We start with the description of notations, then elaborate the details of VRMF and SLGR, and lastly present an algorithmic framework of our CBGR approach.

3.1. Preliminaries

Let $\mathcal{X} = \{\mathbf{x}_i, i = 1, \cdots, n\}$ denote the image dataset, where each $\mathbf{x}_i \in \mathbb{R}^d$ is a visual feature vector. To discover the geometrical subacute (manifold), we build a neighborhood graph on $\mathcal{X}$, and define $\mathbf{W} \in \mathbb{R}_+^{n\times n}$ as the corresponding the affinity matrix with element W_{ij} storing the weight of edge between $\mathbf{x_i}$ and $\mathbf{x_j}$. Normally the weight is calculated using a Gaussian kernel

$$W_{ij} = \exp\left(-\frac{d^2(\mathbf{x}_i, \mathbf{x}_j)}{\sigma^2}\right) \tag{1}$$

if $i \in \mathcal{N}(j)$ or $j \in \mathcal{N}(i)$, otherwise $W_{ij} = 0$, where $\mathcal{N}(i)$ denotes a k nearest neighbor set of image i. Typically, k is a small number (e.g. a small fraction of n), $d(\mathbf{a}, \mathbf{b})$ is a distance metric between two vectors $\mathbf{a}$ and $\mathbf{b}$ (suggested by [4], L1 distance is considered), and σ is the bandwidth parameter that can be tuned by local scaling technique, the effectiveness of which has been verified in the clustering [32] and ranking [27] tasks.

The click-through data is represented by a user-image clicking matrix $\mathbf{R} \in \{0, 1\}^{m\times n}$ whose rows correspond to the users and columns correspond to the images. If an image has been clicked in a query session, the corresponding cell is assigned to value 1, i.e. $R_{ij} = 1$, otherwise $R_{ij} = 0$.

3.2. VRMF: Visual Regularized Matrix Factorization

We refer to the problem of click prediction as Matrix Factorization (MF), which is to learn low-rank representations (also referred as latent factors) of users and images from the information of the user-image clicking matrix, and then further employs the latent factors to predict new clicks between users and images. Also, to elevate the accuracy of click prediction, the visual information of images is taken as the regularization term and incorporated into the MF framework, thus called Visual Regularized MF (VRMF). Let $\mathbf{U} \in \mathbb{R}^{f\times m}$ and $\mathbf{V} \in \mathbb{R}^{f\times n}$ be two matrices of latent factors, and our VRMF method is to

Algorithm 1 Gradient Descent Process for Solving $\mathbf{U}_{*i}$ and $\mathbf{V}_{*j}$

1: Initialize $t = 0$, $\eta = 1$, $\mathbf{U}^{(0)}$ and $\mathbf{V}^{(0)}$;
2: **while** $t \leq T$ **do**
3: Update $\mathbf{U}_{*i}^{(t+1)} = \mathbf{U}_{*i}^{(t)} - \eta_t \frac{\partial F}{\partial \mathbf{U}_{*i}^{(t)}}$ and $\mathbf{V}_{*j}^{(t+1)} = \mathbf{V}_{*j}^{(t)} - \eta_t \frac{\partial F}{\partial \mathbf{V}_{*j}^{(t)}}$
4: If $F(\mathbf{U}_{*i}^{(t+1)}, \mathbf{V}_{*j}^{(t+1)}) < F(\mathbf{U}_{*i}^{(t)}, \mathbf{V}_{*j}^{(t)})$, $\eta_{t+1} = 2\eta_t$; otherwise, $\eta_{t+1} = \eta_t/2$;
5: $t = t + 1$;
6: **end while**

recovery the user-image clicking matrix $\hat{\mathbf{R}} = \mathbf{U}^T\mathbf{V}$ by solving the following optimization problem

$$\begin{aligned}\underset{\mathbf{U},\mathbf{V}}{\text{minimize}}\ F(\mathbf{U},\mathbf{V}) &= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij} - \mathbf{U}_{*i}^T\mathbf{V}_{*j})^2 \\ &+ \frac{\alpha}{2}\sum_{j=1}^{n}\sum_{k\in\mathcal{N}(j)} W_{jk}\|\mathbf{V}_{*j} - \mathbf{V}_{*k}\|^2 \\ &+ \frac{\beta}{2}(\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) \end{aligned} \tag{2}$$

where $\mathbf{U}_{*i}$ is a column vector of $\mathbf{U}$, representing the latent factors of user i, likewise, and $\mathbf{V}_{*j}$ represents the latent factors of image j. I_{ij} is an indicator function that is equal to 1 if $R_{ij} = 1$, otherwise 0. $\| \bullet \|_F$ denotes the Frobenius norm of a matrix, and α and β are two free parameters. The first term of above cost function is the fitting constraint that ensures the learned $\hat{\mathbf{R}}$ to be consistent with the observed user-image matrix, the second term is the smoothness constraint that makes the visually similar images having similar latent factors, and the last term is the regularizer that is to alleviate model overfitting.

We adopt a gradient descent process to solve Eq. (2). By differentiating F with respect to $\mathbf{U_{*i}}$ and $\mathbf{V_{*j}}$, we have

$$\frac{\partial F}{\partial \mathbf{U}_{*i}} = \sum_{j=1}^{n} I_{ij}(R_{ij} - \mathbf{U}_{*i}^T\mathbf{V}_{*j})\mathbf{V}_{*j} + \beta\mathbf{U}_{*i}, \tag{3}$$

and

$$\frac{\partial F}{\partial \mathbf{V}_{*j}} = \sum_{i=1}^{m} I_{ij}(R_{ij} - \mathbf{U}_{*i}^T\mathbf{V}_{*j})\mathbf{U}_{*i} + \alpha\sum_{k\in\mathcal{N}(j)} W_{jk}(\mathbf{V}_{*j} - \mathbf{V}_{*k}) + \beta\mathbf{V}_{*j}. \tag{4}$$

In the gradient descent process, we dynamically adapt the step-size η in order to accelerate the process while guaranteeing its convergence. Denote by $\mathbf{U}_{*i}^{(t)}$ and $\mathbf{V}_{*j}^{(t)}$ the values of $\mathbf{U}_{*i}$ and $\mathbf{V}_{*j}$ in the t-th turn of the iterative process. If $F(\mathbf{U}_{*i}^{(t+1)}, \mathbf{V}_{*j}^{(t+1)}) < F(\mathbf{U}_{*i}^{(t)}, \mathbf{V}_{*j}^{(t)})$, i.e. the cost function obtained after gradient descent is reduced, then we double the step-size; otherwise, we halve the step-size and do not recompute $\mathbf{U}_{*i}^{(t+1)}$ and $\mathbf{V}_{*j}^{(t+1)}$. The process is illustrated in Algorithm 1.

3.3. GRSL: Graph Ranking with Soft Labels

Our GRSL method is developed based on MR [34], and we slightly modify it for the noise-tolerance purpose. Let $\mathbf{r}: \mathcal{X} \rightarrow \mathbb{R}$ be a ranking function that assigns to each image instance $\mathbf{x}_i$ a ranking score r_i. We can view $\mathbf{r}$ as a vector $\mathbf{r} = [r_1, \cdots, r_n]^T$. We also define a label vector $\mathbf{y} = [y_1, \cdots, y_n]^T$, in which $y_i = 1$ if $\mathbf{x}_i$ is a query, and $y_i = 0$ otherwise. The cost function associated with $\mathbf{r}$ is defined to be

$$\underset{\mathbf{r}}{\text{minimize}}\ Q(\mathbf{r}) = \frac{1}{2}\sum_{i,j=1}^{n} W_{ij}\left(\frac{r_i}{\sqrt{D_{ii}}} - \frac{r_j}{\sqrt{D_{jj}}}\right)^2 + \frac{\lambda}{2}\|\mathbf{r} - \mathbf{y}\|^2 \tag{5}$$

where λ is a regularization parameter, and $\mathbf{D}$ is a diagonal matrix with $D_{ii} = \sum_{j=1}^{n} W_{ij}$. The first term in the cost function is a smoothness constraint, which make the nearby examples in visual space having close ranking scores. The second term is a fitting constraint, which makes the ranking result fitting to the label assignment.

By differentiating $Q(\mathbf{r})$ and set it to zero, we can easily get the following closed solution

$$\mathbf{r}^* = (\mathbf{I} - \mu\mathbf{S})^{-1}\mathbf{y} \tag{6}$$

where $\mu = 1/(1+\lambda)$, $\mathbf{I}$ is an identity matrix, and $\mathbf{S}$ is the symmetrical normalization of $\mathbf{D}$, i.e.

$$\mathbf{S} = \mathbf{D}^{1/2}\mathbf{W}\mathbf{D}^{1/2}. \tag{7}$$

We can directly use the above closed form solution to compute the ranking scores of examples. However, in large scale problems, we prefer to use the iteration solution

$$\mathbf{r}(t+1) = \mu\mathbf{S}\mathbf{r}(t) + (1-\mu)\mathbf{y}. \tag{8}$$

As illustrated by Eq. (6) and (8), it is noted that the label vector plays an important role in image ranking, and a dense $\mathbf{y}$ is desirable to derive $\mathbf{r}^*$. In the regular MR [34], only one labeled instance (i.e. the user's query) is concerned, which is hardly to achieve satisfactory ranking result. A few works, such as [4], [5], [27], and [28], etc., take the online relevance feedback mechanism into consideration for the label vector enrichment, but it is unpractical as mentioned before.

Different from previous studies, our idea is to enrich the label vector using the click-through data without any user intervention. The intuition behind our idea is that, when two images are clicked in a same query session, they often share a certain similar meaning and we expect different users to express similar judgments on them. Based on this assumption, the (hidden) correlation between any two images can be inferred by analyzing the judgements (clicks) made by different users on them. Given the user's query q, the correlation of it to each database image is defined by Jacquard coefficient based on the enriched query log matrix $\hat{\mathbf{R}}$

$$Sim(q,i) = \frac{|\mathcal{A}(\hat{\mathbf{R}}_{*q}) \cap \mathcal{A}(\hat{\mathbf{R}}_{*i})|}{|\mathcal{A}(\hat{\mathbf{R}}_{*q}) \cup \mathcal{A}(\hat{\mathbf{R}}_{*i})|} \tag{9}$$

where $\mathcal{A}(\mathbf{a})$ denotes a set composed of the nonzero elements of a binary vector, $\hat{\mathbf{R}}_{*i} \in \{0,1\}^m$ is a column vector of $\hat{\mathbf{R}}$, recording the clicks imposed by different users on image i, and $|\bullet|$ denotes the size of a set. Intuitively, we can directly predict $y_i = 1$ if $Sim(q,i)$ is highly positive. Although this idea can be straightly handled by MR, it may suffer from performance degradation as erroneous click-through data. In particular, in our scheme, more noises may be introduced by MF. To this end, we treat the labels in two different ways for the fault-tolerance purpose. In details, the user's query is treated as the 'hard-labeled' instance, while the images predicted by analyzing the click-through data are treated as 'soft-labeled' instances, i.e. $y_i = Sim(q,i) \in [0,1]$, where the magnitude of the label represents the confidence of the assigned label.

3.4. Algorithmic Framework

So far, we can assemble the proposed VRMF and GRSL methods into the CBGR framework for image retrieval, which ranks the database images with respect to the user's query based on visual features and click-through data. We outline this algorithmic framework in below.

1. Build a neighborhood graph on $\mathcal{X}$, and compute the corresponding affinity matrix $\mathbf{W}$ by Eq. (1) and the normalized one $\mathbf{S}$ by Eq. (7);
2. Compute the enriched user-image matrix $\hat{\mathbf{R}}$ based on $\mathbf{R}$ and $\mathbf{S}$ using Algorithm 1;
3. Compute the soft-label vector $\mathbf{y}$ based on q and $\hat{\mathbf{R}}$ by Eq. (9);
4. Compute the ranking-score vector $\mathbf{r}$ based on $\mathbf{S}$ and $\mathbf{y}$ by Eq. (6) or (8);

Note that the step 1 and step 2 can be implemented offline, and therefore our CBGR approach can be quite efficient in processing online queries. Note that our CBGR approach mainly focuses on processing the in-sample queries, but it can be easily extended to handle the out-of-sample queries. For example, when a completely new query arrives, we can apply a strategy named one-click query expansion [21] to transform the out-of-sample query into a in-sample query with very few user efforts.

4. Experiments

In this section, we first introduce our experimental settings, and then present the experimental results that validate the effectiveness of our approach. The experiments actually contain two parts. In the first part, we will compare our VRMF method with those CF methods that can be used for the task of click prediction. In the second part, we compare our CBGR approach with several existing graph ranking methods for the task of image retrieval.

4.1. Experimental Setup

We employ the '10K Images' dataset[1] which is publicly available on the web to make our experiments reproducible. The images are from 100 semantic categories, with 100 images per category. Three kinds of visual features are extracted to represent the images,

[1] http://www.datatang.com/data/44353. The dataset was firstly used in [26].

Table 1. The P@N measures of our VRMF method compared with several exiting CF approaches.

	N=10	20	30	40	50
user-based	0.659	0.582	0.508	0.462	0.423
item-based	0.681	0.612	0.527	0.448	0.38
regular MF	0.722	0.661	0.596	0.532	0.473
ItemVisual	0.734	0.682	0.61	0.558	0.52
VRMF	**0.75**	**0.702**	**0.652**	**0.593**	**0.534**

Table 2. The F1@N measures of our VRMF method compared with several exiting CF approaches.

	N=10	20	30	40	50
user-based	0.158	0.248	0.294	0.324	0.343
item-based	0.163	0.26	0.305	0.312	0.307
regular MF	0.18	0.294	0.38	0.387	0.384
ItemVisual	0.178	0.298	0.385	0.398	0.395
VRMF	**0.185**	**0.31**	**0.392**	**0.435**	**0.43**

including a 64-dimensional color histogram, an 18-dimensional wavelet-based texture and a 5-dimensional edge direction histogram [26].

A click-through dataset consisting of 1000 query sessions is used in experiments, which is simulated based on the ground truth of image dataset. The average number of clicks in each query session is 20. Also, we randomly add 12% noise into the click-through dataset to approach the real-world search scenario [1].

Essentially, our click prediction solution acts for the image recommendation task, while the image ranking problem is equivalent to the image retrieval task. Many measures are commonly used to evaluate both recommendation and retrieval tasks, such as precision and recall. In the top N recommendation scenario, precision and recall are often summarized as the F1 measure. Similarly, in the retrieval scenario, PR (Precision-Recall) graph is widely used to depict the relationship between precision and recall, and it could be further summarized as the MAP (Mean Average Precision) measure. In addition, for many web applications, only the top returned images can attract users' interests, so the precision at top N (P@N) metric is significant to evaluate the image recommendation and retrieval performance.

To evaluate the average performance of image retrieval methods, a query set with 200 images is equally sampled from all semantic categories, i.e. two images are randomly picked from each category.

4.2. Experimental Results for The Task of Click Prediction

In this part, we compare our VRMF method with several existing CF approaches that can be used for the task of click prediction, including user-based CF [14], item-based CF [16],

[1]Empirical study reported by [18] showed that the satisfaction rate of the image click-through data is approximately equal to 88%.

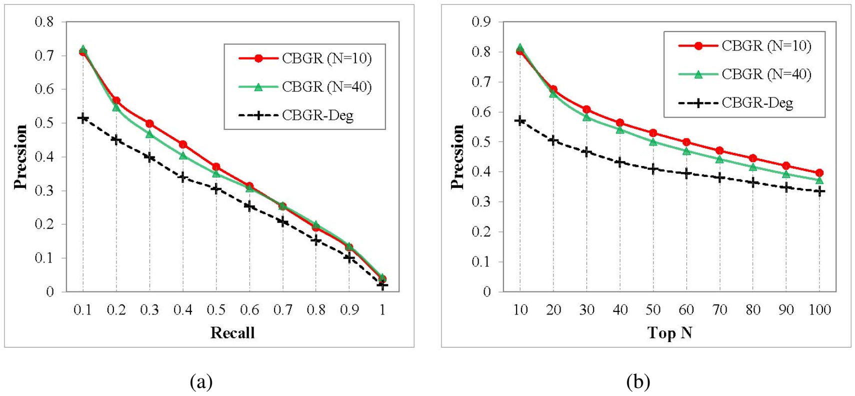


Fig. 1. The performance of our CBGR approach on different enriching scales in terms of (a) PR graphs and (b) P@N curves.

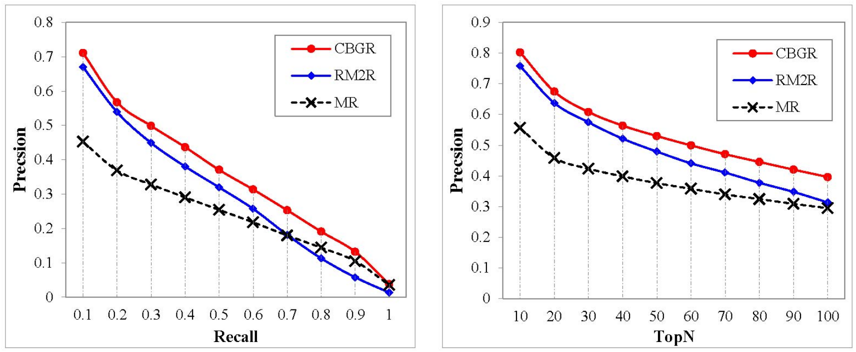
and regular MF method [9]. In addition, the method appeared in our previous conference version [12] is also included in the comparisons. Essentially our previous method is a item-based CF solution with visual side-information, so we term it as ItemVisual.

For the proposed VRMF method, there are two parameters, i.e., α and β (see Eq. (2)). We tune the two parameters through 5-fold cross-validation, and the best settings are $\alpha = 0.06$ and $\beta = 0.04$. For the iteration runs T (see the Algorithm1), we set it to 1000. In our experiments, we found that this value can lead to a well convergence of the optimization process.

Table 1 and Table 2 respectively print the P@N and F1@N measures of our VRMF method compared with several other approaches, where the best performance has been boldfaced. From the experimental results, the following interesting observations are revealed. First, by examining all methods, the two methods using the visual side-information (VRMF and ItemVisual) outperform the three baseline methods (user-based, item-based and regular MF), which verifies the usefulness of the visual side-information. Second, by comparing the three baseline methods, the regular MF performs better than the user-based and item-based CF, which demonstrates the superiority of the latent factor models. At last, the proposed VRMF method achieves the best performance among all comparing approaches. It well demonstrates that combining the latent factor model with the visual side-information is effective and beneficial to the task of click prediction.

4.3. Experimental Results for The Task of Image Retrieval

Furthermore, we evaluate the performance of our CBGR approach on two enriched click-through datasets which are respectively attained when $N = 10$ (with highest precision) and $N = 40$ (with highest F1 measure). To verify whether the click prediction solution is beneficial to image retrieval or not, we compare our CBGR approach with a its degenerated variant termed CBGR-Deg. The CBGR-Deg method is almost same with the CBGR approach, except the former directly use the click-through data without enrichment.



(a) (b)

Fig. 2. The performance of our CBGR approach compared with several existing methods in terms of (a) PR graphs and (b) P@N curves.

For the our CBGR approach, there is only one parameter, i.e., the μ used in Eq. (8). For convenience, we set $\mu = 0.01$, consistent with the previous experiences [34, 4, 29, 27].

Figure 1 print the PR graphs and P@N curves of all comparing methods. By examining the experimental results, we observe that both CBGR ($N = 10$) and CBGR ($N = 40$) approaches outperform the CBGR-Deg method, which verifies the effectiveness of the click prediction solution used by our CBGR approach. Further, by comparing the CBGR ($N = 10$) and CBGR ($N = 40$) methods, we found that their performance curves are very similar to each other. Based on this observation, we prefer to set $N = 10$ for the computational efficiency purpose.

At last, we compare our CBGR approach with a CIT scheme named RM2R [28]. RM2R is a two-view graph ranking model which exploits both visual and click features to encode the ranking results. Moreover, the conventional MR method [5] as a baseline is also included in comparisons to verify whether exploiting the click-through data is beneficial to the task of image retrieval or not. To be fair, all above three graph ranking methods take the local scaling trick [32] to tune the bandwidth parameter used by the Gaussian kernel.

Figure 2 plots the PR graphs and P@N curves of our approach compared with other two methods. We found that the methods using both visual feature and the click-through data perform better than the baseline MR method, which verifies the benefit of exploiting the click-through data for the task of image retrieval. It is impressive that the performance of our CBGR approach is always the best among all comparing methods. It is worth noting that the performance of the RM2R method evaluated in our experiments is not as good as the results reported by [28]. The main reason is that we evaluate it without using relevance feedback as done in [28], and thus its performance drops. This observation reveals that leveraging the click-through data as supervision signal is more effective than as feature set, when only a user's query is available.

5. Conclusions

Existing image search engines usually suffer from imperfect results caused by the well-known semantic-gap. Although many studies on learning with the click-through data have been conducted to address this problem, the improvement in performance is limited as little effort on investigating both sparseness and noisiness of the click-through data. This paper presents a novel CBGR method that aims at noise-resistantly leveraging the click-through data to boost graph-based visual ranking. Concretely, we first propose a VRMF method to enrich the click-through data, and then develop a GRSL solution for fault-tolerant image ranking. Experimental study validates the superiority of the proposed techniques in comparison to some representative approaches.

In the future, we will take more side-information (e.g., the social relationships between users and the surrounding text information of images) into consideration in order to further enhance the effectiveness of the click predicting and visual ranking. In addition, inspired by the data stream mining techniques [22, 23], another extension of this work is to study the incremental solutions to the tasks of click predicting and graph ranking.

Acknowledgment. The authors would like to thank the anonymous reviewers for their constructive suggestions. This work was supported in part by the 'Natural Science Foundation of China' (61370070 and 61671048), the 'Fundamental Research Funds for the Central Universities' (2015JB-M029), and the 'Research Foundation for Talents of Beijing Jiaotong University' (2015RC008). The corresponding author of this paper is Na Zhao (email: zhaona@zwu.edu.cn).

References

1. Bai, S., Bai, X.: Sparse contextual activation for efficient visual re-ranking. IEEE Trans. Image Process. 25(3), 1056–1069 (2016)
2. Bai, X., Wang, B., Yao, C., Liu, W., Tu, Z.: Co-transduction for shape retrieval. IEEE Trans. Image Process. 21(5), 2747–2757 (2012)
3. Chin, W., Zhuang, Y., Juan, Y., Lin, C.: A fast parallel stochastic gradient method for matrix factorization in shared memory systems. ACM Trans. Intel. Syst. Tec. 6(1), 2:1–2:24 (2015)
4. He, J., Li, M., Zhang, H., Tong, H., Zhang, C.S.: Manifold-ranking based image retrieval. In: ACM Multimedia. pp. 9–16 (2004)
5. He, J., Li, M., Zhang, H., Tong, H., Zhang, C.: Generalized manifold-ranking-based image retrieval. IEEE Trans. Image Process. 15(10), 3170–3177 (2006)
6. Hoi, S., Liu, W., Chang, S.F.: Semi-supervised distance metric learning for collaborative image retrieval. In: CVPR. pp. 1–7 (2008)
7. Hsu, W.H.: Video search reranking through random walk over document-level context graph. In: ACM Multimedia. pp. 971–980 (2007)
8. Jain, V., Varma, M.: Learning to re-rank: Query-dependent image re-ranking using click data. In: WWW. pp. 277–286 (2011)
9. Koren, Y., Bell, R., Volinsky, C., et al.: Matrix factorization techniques for recommender systems. Computer 42(8), 30–37 (2009)
10. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. J Mach. Learn. Res. 11, 19–60 (2010)
11. Pan, Y., Yao, T., Mei, T., Li, H., Ngo, C.W., Rui, Y.: Click-through-based cross-view learning for image search. In: ACM SIGIR. pp. 717–726 (2014)
12. Qin, X., He, Y., Wu, J., Sang, Y.: Leveraging click completion for graph-based image ranking. In: PDCAT. pp. 155–160 (2016)

13. Rennie, J.D.M., Srebro, N.: Fast maximum margin matrix factorization for collaborative prediction. In: ICML. pp. 713–719 (2005)
14. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: CSCW. pp. 175–186 (1994)
15. Salakhutdinov, R., Mnih, A.: Bayesian probabilistic matrix factorization using markov chain monte carlo. In: ICML. pp. 880–887 (2008)
16. Sarwar, B.M., Karypis, G., Konstan, J.A., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: WWW. pp. 285–295 (2001)
17. Shi, Y., Larson, M., Hanjalic, A.: Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. ACM Comput. Surv. 47(1), 3:1–3:45 (2014)
18. Smith, G., Brien, C., Ashman, H.: Evaluating implicit judgments from image search clickthrough data. J. Am. Soc. Inf. Sci. Tec. 63(12), 2451–2462 (2012)
19. Su, J., Huang, W., Yu, P., Tseng, V.: Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. IEEE Trans. Knowl. Data Eng. 23(3), 360–372 (2011)
20. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Adv. Artificial Intellegence 2009, 421425:1–421425:19 (2009)
21. Tang, X., Liu, K., Cui, J., Wen, F., Wang, X.: Intentsearch: Capturing user intention for one-click internet image search. IEEE Trans Pattern Anal.Mach. Intell. 34(7), 1342–1353 (2012)
22. Wang, H., Wu, J., Pan, S., Zhang, P., Chen, L.: Towards large-scale social networks with online diffusion provenance detection. Comput. Netw. 114, 154 – 166 (2017)
23. Wang, H., Zhang, P., Zhu, X., Tsang, I., Chen, L., Zhang, C., Wu, X.: Incremental subgraph feature selection for graph classification. IEEE Trans. Knowl. Data Eng. 29(1), 128–142 (2017)
24. 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)
25. Wang, Y., Cheema, M.A., Lin, X., Zhang, Q.: Multi-manifold ranking: Using multiple features for better image retrieval. In: PAKDD. LNAI, vol. 7819, pp. 449–460 (2013)
26. Wu, J., Shen, H., Li, Y.D., Xiao, Z.B., Lu, M.Y., Wang, C.L.: Learning a hybrid similarity measure for image retrieval. Pattern Recogn. 46(11), 2927–2939 (2013)
27. Wu, J., Li, Y., Feng, S., Shen, H.: Self-immunizing manifold ranking for image retrieval. In: PAKDD. LNAI, vol. 7819, pp. 92–100 (2013)
28. Wu, J., Yuan, J., Luo, J.: Robust multi-view manifold ranking for image retrieval. In: PAKDD. LNAI, vol. 9652, pp. 92–103 (2016)
29. Xu, B., Bu, J., Chen, C., Wang, C., Cai, D., He, X.: Emr: A efficient manifold ranking model for content-based image retrieval. IEEE Trans. Knowl. Data Eng. 27(1), 102–114 (2014)
30. Yang, Y., Xu, D., Nie, F., Luo, J., Zhuang, Y.: Ranking with local regression and global alignment for cross media retrieval. In: ACM Multimedia. pp. 175–184 (2009)
31. Yin, P.Y., Bhanu, B., K.-C., C., Dong, A.: Integrating relevance feedback techniques for image retrieval using reinforcement learning. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1536–1551 (2005)
32. Zelnik-Manor, L., Perona, P.: Self-tuning spectral clustering. In: NIPS. pp. 1601–1608 (2004)
33. Zhang, Y., Yang, X., Mei, T.: Image search reranking with query-dependent click-based relevance feedback. IEEE Trans. Image Process. 23(10), 4448–4459 (2014)
34. Zhou, D., Weston, J., Gretton, A., Bousquet, O., Scholkopf, B.: Ranking on data manifolds. In: NIPS. pp. 169–176 (2003)
35. Zhou, X.S., Huang, T.S.: Relevance feedback in image retrieval: A comprehensive review. Multimedia Syst. 8(6), 536–544 (2003)

Jun Wu received the Ph.D. degree in computer science from the Dalian Maritime University, China, in 2010. He was a visiting scholar with the Department of Computer Science,

University of Rochester, from 2015-2016. He is currently an Associate Professor with the School of Computer and Information Technology, Beijing Jiaotong University, China. His current research interests include multimedia retrieval, recommendation and large-scale computing.

Yu He received the B.E. degree in software engineering from the Dalian Jiaotong University, China, in 2015. He is currently pursuing the Masters degree with the School of Computer and Information Technology, Beijing Jiaotong University, China. His research mainly focuses on the machine learning algorithms for recommender systems.

Xiaohong Qin received the Master degree in computer science from the Beijing Jiaotong University, China, in 2017. She is currently a software engineer with the China Unicom Corporation, China. Her research mainly focuses on the image retrieval.

Na Zhao received the Ph.D. degree in transportation planning and management from the Dalian Maritime University, China, in 2008. She is currently an Associate Professor with the School of Logistics and E-Commerce, Zhejiang Wanli University, China. Her current research mainly focuses on the optimization problems in the transportation planning.

Yingpeng Sang received her Ph.D. degree in computer science from the Japan Advance Institute of Science and Technology, in 2007. He was also a Postdoctoral Research Fellow with the University with the School of Computer Science, University of Adelaide, Australia, from 2007 to 2010. He is currently an Associate Professor the School of Data and Computer Science, Sun Yat-sen University, China. His research interests include privacy-preserving problems in databases, data mining, and other networking scenarios.

Received: February 12, 2017; Accepted: July 23, 2017.