

Photo Sharing Websites Using Personalized and Non-Personalized Image Search

V.Dhavamani¹, V.Shalini², Dr.K.P.Kaliyamurthie³

1,2 Department of Information Technology, Bharath University, Chennai=600073.

Dhavamani.vijayamani@gmail.com

3. Professor and Head, Dept. of I.T, Bharath University, Chennai-600 073.

kpkaliyamurthie@gmail.com

Abstract-Increasingly developed social sharing websites, like photobucket, pixable, picasa, piczo, instagram and Kodak gallery, allow users to create, share, annotate and comment medias. Photo sharing websites helps in publishing or transferring user's photos on-line, that enables the user to share their photos with others public/privately. Nowadays, there has been an increase in the photo sharing websites which allows users not only to create, share, annotate and comment multimedia contents, but also provide useful information to improve media retrieval and management. Experiments conducted on the COREL image dataset demonstrate the effectiveness and efficiency of the proposed integrated framework and the conversion strategies.

Index terms - Personalized image search, social annotation, tensor factorization, topic model.

I.INTRODUCTION

KEYWORD-BASED search has been the most popular search paradigm in today's search market. Despite of simplicity and efficiency, the performance of the keyword – based search is far from satisfying. Poor user experience has investigation on Google search, for 52% of 20 000 queries, the searchers did not find any relevant results. This is due to two reasons:

1) queries are in general short and nonspecific, e.g., the query of “IR” has the interpretation of both information retrieval and infra-red, and

2) users may have different intentions for the same query, e.g., searching for “Samsung laptop”

by a charger battery CD-Drive has a completely different meaning from searching. One thing is there to address these problems is, where the user specific information is considered to distinguish the exact intentions of the user queries and rerank the results lists.

Given the large and growing importance of search engines like google, yahoo etc., personalized search has the potential to significantly improve that searching experience. Compared with non-personalized search, the ranking type of a document like (web page, image, audio, message, video, etc.) in the result list is decided not only by the user query, but by the preference of user.



Figure.1: (top) non-personalized and (bottom) personalized search results for the query “Samsung laptop”.

Fig. 1 shows the example for non-personalized and personalized image search results from the search engines. The non-personalized search returned results only based on the user query relevance and displays Samsung laptop images as well as it can displays the Samsung charger battery on the above image in figure1. While personalized search results consider as both user query relevance and user preference, so the personalized results from an laptop lover rank the laptop images on the top.

This provides a natural two step solution scheme. Most of the existing work [2]–[5] follow this scheme and decompose personalized search into two steps: computing the non-personalized relevance score between the query, the document, and computing their personalized score by estimating the user’s preference over the document. After this, a merge operation is

conducted to generate a ranked list. It suffers from two problems.

- 1) The interpretation is less straight and its not that much convinced. The intuition of personalized search is to rank the returned documents by estimating the user’s preference over documents under the particular queries. In this we can directly analyzing the user query document correlation, the existing system scheme approximates it by separately computing a user query document relevance score and a user document relevance score.
- 2) How to determine the merge strategy is not trivial. In this research, we simultaneously considers the Typically a weighting parameter will be optimized to balance the two scores, or the learnt user preference is used to rerank the query relevance-based original list .

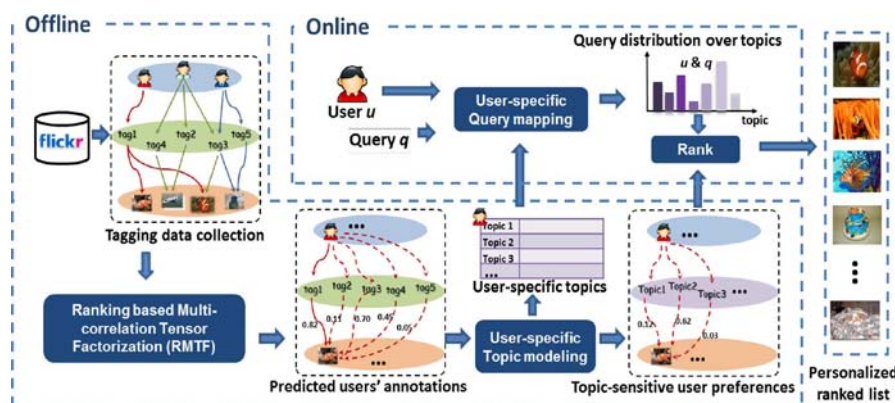


Fig 2-offline and online proposed framework

user and query dependence and present a novel framework to tackle the personalized image search problem. To investigate on user preference, perform user modeling, and the popular social activity of tagging also included. Combined tagging has become most popular for sharing and organizing resources to user preferences, its leading to a huge amount of user is generating the annotations. Online photo sharing websites using network, such as Flickr, Picasa, Photobucket, Pixable, Piczo, Zoomr, and Pinterest its allows a users as owners to tags the image as taggers, and comment the image as commenters for their contributed contents to interact and collaborate with each other in a social media dialogue. the original annotations available is not enough for user preference mining. Therefore, we can transfer the problem of personalized image search to users annotation prediction. Moreover, as user queries and tags do not follow the simple one to one communication, we build user specific topic spaces to exploit the relations between the queries and tags.

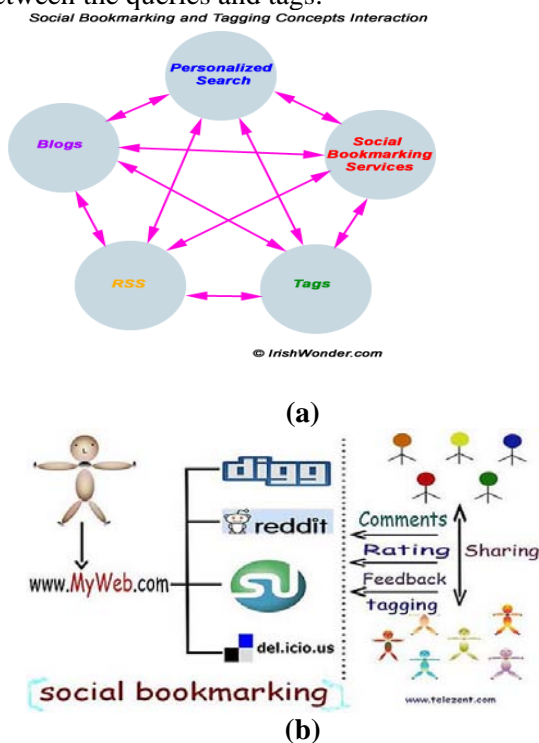


Fig (a) and(b) eg. For tag, blog, feedback, comment, share

A formal assumption is that, *the users is tagged actions is to reflect their personal relevance judgement.* For example, if a user tagged "Birthday" to an image, it is probable that the user will consider this image as relevant if he/she issues "Birthday" as a user query. Illustrated by this, the intuition of this research is that if the users annotations to the images are available, we can estimate directly to the users preference under certain queries.

A.Framework

The framework of this research is shown in Fig. 2. It contains only two - stages: first one is offline model training stage and the another one is online personalized search response stage.

For the offline stage, three types of data including users, two images and tags as well as their ternary interrelations and intra-relations are first collected. To alleviate the sparsity and noisy problem, we present a novel method Following the assumption, we can straightly utilize the predicted user annotations for personalized image search, if a user has a high probability to assign the tag to an image, the image should be ranked higher when the user issues query . However, this formulation has two problems:

1) it is unreasonable to assign the query to a single tag in the tag vocabulary, e.g., when a user searches "barbie dolls", he/she would like the images that he/she annotated with semantic related tag "barbiestills" are also ranked as higher, and 2) there are variations in individual user's tagging patterns and vocabularies, e.g., the tag "singam movie images" from an tamil movie specialist should be related to "Movies", while a Hindi singam movie images or Telugu singam imgaes will consider "singam movie images" more related to "films". To address the two problems, we perform *user-specific topic modeling* to build the semantic topics for every user. The user's annotation for an image is view like an *document*. The separate tag to the image is an *word*. User annotations for all the images constitute the *corpus*. Named (RMTF-*ranking-based multicorrelation tensor factorization*) to better leverage the observed tagging data for users' annotation prediction. The contributions of this research are summarized as three-folds. • We propose a novel personalized image search framework by simultaneously considering user and query information of the image. The user's preferences over images under certain query are estimated by how probable he/she assigns the query related tags to the images. • A tensor factorization model named RMTF is proposed to predict the users annotations using ranking based to the images. • We have build the user specific topics and map the queries as well as the users' preferences onto the learned topic spaces to represents the better query tag relationship.

II. RELATED WORK

In recent years, extensive efforts have been focusing on personalized search. Regarding the explicit user profile, relevance feedback , user history data (browsing log , click-through data , and social annotations etc.), context information is time and location, etc. and social network are also exploited For the implementation there are two basic stages

1) query refinement and 2) result processing.

III. RANKING BASED MULTICORRELATION TENSOR FACTORIZATION

In this section, we present the algorithm for annotation prediction. Table - 1 lists the key notations used in this research. There are three types of entities, we have to share our image to many websites like google, piczo, photobucket, flickr etc. The tagged data can be viewed as a set of triplets. Predicting the users' annotations to

the images are related to reconstructing the user-tag-image ternary interrelations. The low-rank approximation is performed to use Tucker decomposition in a general tensor factorization model. In this research, a model named RMTF is proposed to designed as objective function. We first introduce a novel ranking based optimization scheme for representation of the tagging data or tag the image to perform a better leveraged images. The users may among multiple inter relations images and tags are utilized as the smoothness constraints to take on it.

Table - 1

LIST OF KEY NOTATIONS

Symbol	Description
\mathcal{Y}, \mathcal{C}	user-image-tag tensor, and core tensor
U, I, T	represent user, image, tag factor matrices
$\mathbb{U}, \mathbb{I}, \mathbb{T}$	sets of users, images and tags, respectively
u, i, t	represent user, image, tag index
$\mathbf{u}, \mathbf{i}, \mathbf{t}$	represent user, image, tag feature vectors

Fig. 3(a) shows the tensor constructed from the running example in Fig. 2.

A. Ranking Based Optimization Scheme

However, under this type of situation of social image tagging data, the semantics of encoding all the unobserved data as 0 are incorrect, which is illustrated with the running example.

• Firstly, the fact that *user3* has not given any tag to *image2* and *image4* does not mean *user3* considering all the tags are bad for set forth the images. Maybe man/women does not want the image to tag or it has no chance to see the fifth image

• Secondly, *user1* annotates *image1* with only the third tag. It is also indefensible to assume that other tags should not be annotated to the image, as many concepts may be missing in the user generated tags and individual user may not be familiar to all the relevant tags in the large tag vocabulary.

According to the optimization function is

(3), *0/1 scheme* tries to predict 0 for both cases. the above two issues, in this research, we present a ranking optimization scheme which intuitively takes the user tagging behaviors into reflection. First of all we have to note that only the soft difference is important and fitting to the numerical values of 1 and 0 is unnecessary.

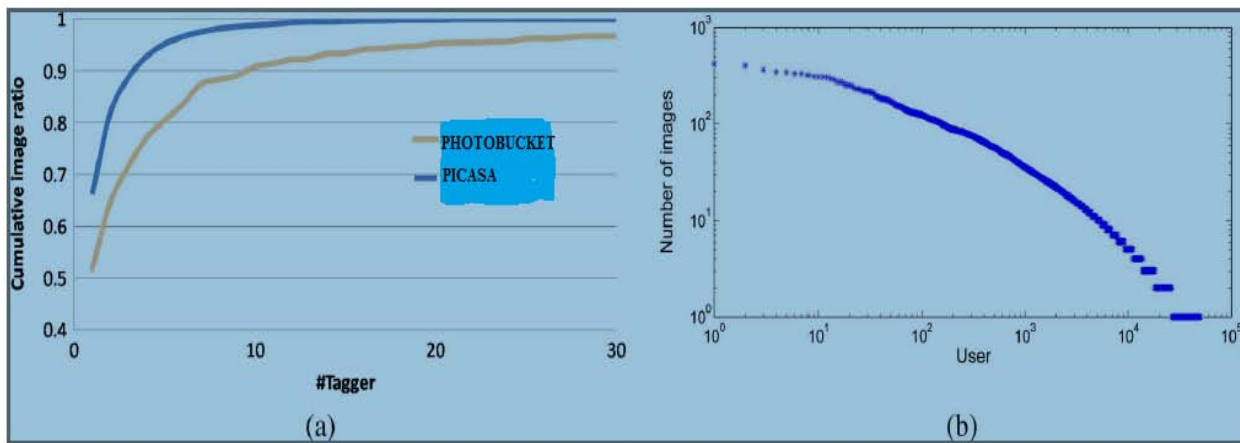


Fig. 4. (a) Cumulative distribution ratio-number of taggers for photobucket/picasa. (b) Number of tagged images per user for picasa.

B. Multicorrelation Smoothness Constraints

Photo sharing websites differentiate from other social tagging systems by its characteristic of self-tagging: most images are only tagged by their users or owners. Fig. 4(a) shows the #tagger statistics for picasa and the

webpage tagging system photobucket. We can see that in picasa, 90% images have no more than four taggers and the average number of tagger for each image in picasa is 1.9 . All the same , the average tagger in photobucket is 6.1. The severe default or lack of problem calls for external resources to enable information propagation.

**TABLE - 2
THE TWO EXAMPLE USERS FOR USER-SPECIFIC TOPICS**

User A	Topic 1	military, aircraft , battleship, navy, artillery, iraq, aircraftcarrier, barracks
	Topic 2	apple, computer, art, girl, cellphone, cool, vintage, digital
	Topic 3	athlete, basketball, baseball, actor, sports, art, film, black

User B	Topic 1	buddha, buddhist, temple, religion, buddhism, thailand, asia, ancient
	Topic 2	blossoms, blooms, nature, macro, flower, bravo, butterfly, spring
	Topic 3	airplane, boeing, aircraft , airport, aviation, jet, aeroplane, cockpit

IV. USER-SPECIFIC TOPIC MODELING

The reconstructed user-tag-image ternary conjunction or combination, so the personalized image search is to perform directly. When user submits a query, the rank of image is inversely proportional to the probability of

annotating with tag q. However in practice, the queries and tags do not follow one-to-one relationship query usually corresponds the tag vocabulary is to several related tag . In any case, the query-tag correspondence differs from user or owner to user or owner. Hence, we build topic spaces for each user to exploit this user-specific one-to-many relationship. We investigate on a

Flickr dataset of 270-K images that the average number of annotated images per user is only 30. From the user-specific topics, we can see:

- user’s interest profile, e.g., user is likely to be a art who also likes flowers, natural and butterfly, while user is keen at religion and interests in gardening and blossoms;
- the same tag may have different topic posterior distributions for different users, e.g., for user , “aircraft” occurs frequently in a military-related topic, while for user , “aircraft” returns to its literal sense of air vehicle.

V. EXPERIMENTS

In the research community of personalized search, is not an easy task since relevancy judgment can only be evaluated by the users or the searchers themselves. The most accepted approach is user study , where participants are asked to judge the search results. Obviously this approach is so costly. In addition, a main problem for user study is that the results are likely to be biased as the participants know that they are being tested. Another extensively used approach is by user query logs history. However, this needs a large scale real search logs, which is not available for most of the researchers.

Social sharing websites provide rich resources that can be exploited for personalized search is evaluate. User’s social activities, such as rating as rate the image , tagging as tag the image or document to our friends and commenting as comment the image, indicate the user’s interest and user preference in a specific document. Nowadays, two types of such user feedback

are utilized for personalized search evaluation. The social annotations is used to be a first approach. The documents has tagged by user with tag will be considered relevant for the personalized query its behind the main assumptions. The other evaluation approach is proposed for personalized image search on Flickr, where the images marked *Favorite* – based by the user are treated as relevant when the queries issues. The two valuation approaches have their supplement for each other. We have to use both approaches in our experiments and list the results in the following. At the baseline we select two state of the art model.

- Topic-based: topic-based personalized search using folksonomy.
- Preference-based: personalized image search by predicting user interests-based preference. Note that both methods follow the two-step scheme: the overall ranking is decided by separately computing query relevance and the user preference. In addition, we also compared the performances of the proposed model with variety of settings.

- TF-Tensor Factorization 0/1 LDA- Latent Dirichlet Allocation : Tensor Factorization without smoothness constraints, optimization under the *0/1 scheme*, using user-specific topic modeling.
- MTF – Multicorrelation Tensor Factorization 0/1 LDA Latent Dirichlet Allocation: Tensor factorization with multicorrelation smoothness constraints, optimization under the *0/1 scheme*, using user-specific topic modeling.
- RMTF – Ranking Based Multicorrelation Tensor Factorization LDA - Latent Dirichlet Allocation, the proposed model: annotations predictions by Ranking based tensor factorization, using user-specific topic modeling.

- RMTF: Directly using the RMTF-based predicted annotations for personalized rank .

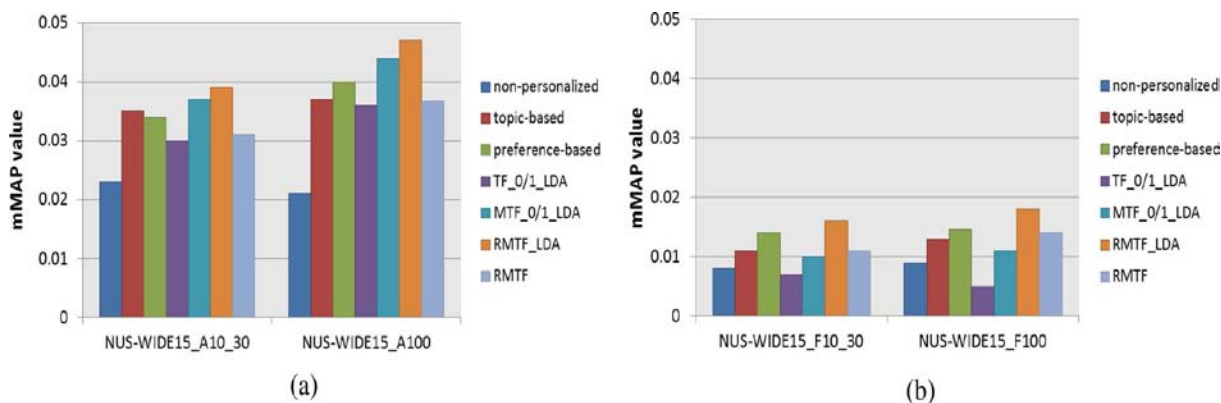


Fig. 7. mMAP value of personalized search for different methods (a) Annotation - based. (b) Favorite - based.

The results are shown in Fig. 7(a). *Non-personalized* denotes the non-personalized rank result by only considering the query relevancy. We can see that all the personalized methods

Outperform to the non-personalized scheme. Comparison between the two test scenarios of NUS-WIDE15 A10 30 and NUS-WIDE15 A100, the performances of personalized methods improve as the test users original annotations increase. This is reasonable as these methods utilize the social annotation resources and the more user feedback is available, the more accurate user preferences can be estimated. What is interesting is that the preference-based model and the proposed model are more sensitive to the amount of original annotations. The reason may be that and our methods extract topic spaces by explicitly exploiting the tagging data, while in the topic-based model, the topic space is predefined and the original annotation is just used to generate the topic vector. Focusing on either test scenario, the performance of the proposed RMTF/ LDA, even MTF 0/1 LDA, is superior than the baseline methods, which demonstrate the advantage of simultaneously considering query relevance and user preference over the separate schemes. Depending on one-to-one query-tag assumption, the performance of RMTF deteriorates dramatically without the user-specific topic modeling. Moreover, RMTF LDA outperforms MTF 0/1 LDA, showing the advantage of the proposed ranking scheme over the conventional 0/1 scheme. Without smoothness priors, TF 0/1 fails to preserve the affinity structures and achieves inferior results. The metric of mMAP is utilized to evaluate the performance and the

results are demonstrated in Fig. 7(b). We have the following observations.

- The mMAP is relatively low compared with Annotation - based evaluation. This phenomenon reflects the problem of Favorite-based evaluation scheme: the *Favorite-based* images are considered relevant for all the test queries. As no query information is involved, for those queries non-relevant with the topic of the *Favorite based* images, the AP tends to be low;
- Comparing between the two test scenarios, the average performance of NUS-WIDE15 F100 also improves over NUS-WIDE15 F10 30, but not as significant as in Annotation-based evaluation. One possible reason for the improvement is that those users having more *Favorite* marks are active users who are likely to also attend more interest groups and tag more images. While, the improvement is not so significant demonstrates that the Favorite based evaluation scheme is less Sensitive to the amount of original annotations.
- Another obvious difference from the results of annotation based evaluation is that the performance of TF 0/1 and MTF 0/1 LDA degrade dramatically. The mMAP of TF 0/1 is even lower than the non-personalized method.

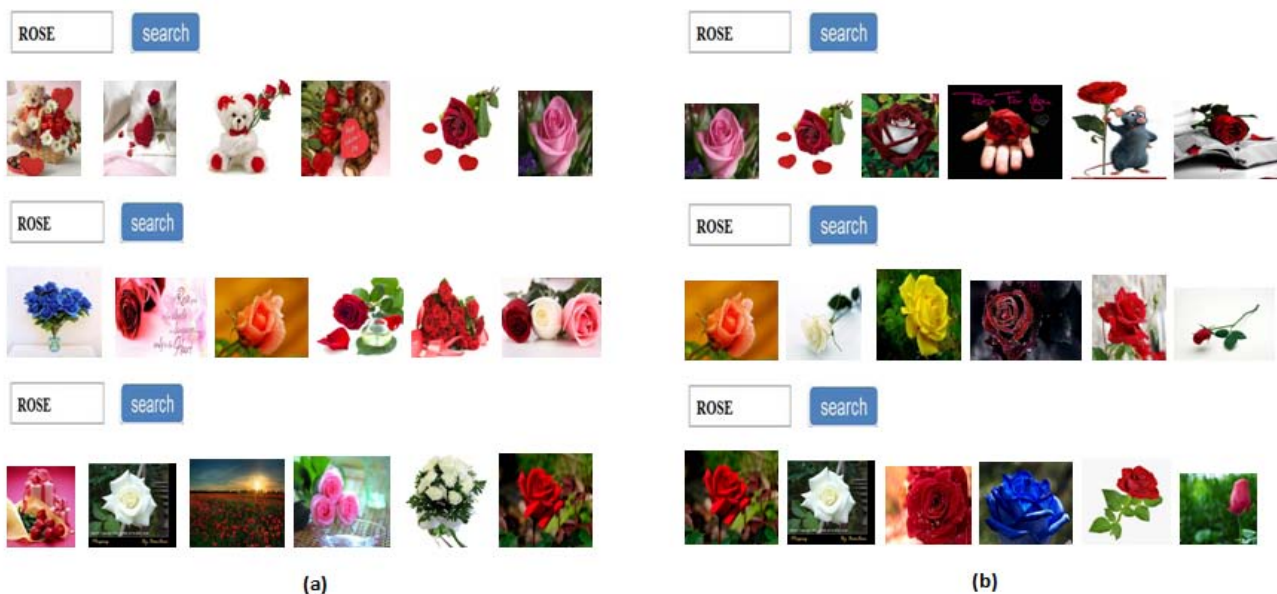


Fig. 8. Example of non-personalized (top) and personalized (middle for User 1 and bottom for User 2) search results for query “rose” (a) Topic-based method;(b) RMTF LDA.



Fig. 9. Example of non-personalized (*top*) and personalized (*middle* for User 1 and *bottom* for User 2) search results for query “computer.” (a) Topic-based method.(b) RMTF /LDA.

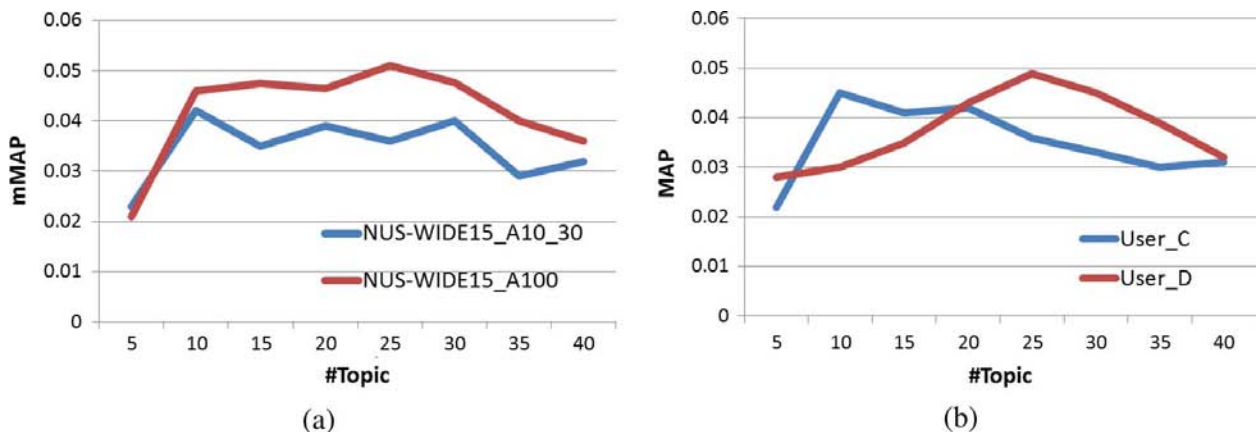


Fig. 10. Influence of topic number K (a) mMAP for all test users; (b) MAP for two users from NUS-WIDE15 A100.

Parable results due to the implicit prior knowledge provided by the original annotations. By utilizing the *Favorite based* marks, a heterogenous resource to eliminated the implicit prior for evaluation.

Fig. 8 displays exemplary search results for the query “rose”. The top six non-personalized results and the personalized results of User A and User B . We can also considering the query relevance and user information, to the proposed system (RMTF – Ranking based multicorrelation tensor factorization) and (LDA-Latent Dirichlet Allocation) captures the user’s preference under certain topics. As a result of mapping “art , flower etc” to Topic 2 of Table II, the top search results for user A mainly focus on blossom, butterfly

etc. While, for user B, the above search results are basically military related, which coincides with user B preference. For the baseline method which is having the separate query relevance and user preference, sometimes its very hard to interpret the search results. For example, the 2nd image and 3rd images for user B in Fig. 8(a) are ranked as higher because user B has a major interest in navy and butterfly. From these images is having the little relation with natural. We note that for some general queries which search intents have clear, its tends to fail the personalized search. In Fig. 9 one of such examples, With “computer” having common understanding to the varient users, incorporating user information will generate confusing search results. There are literatures are discussing the

issue about when to perform personalization. Benefit of personalization is highly dependent on the ambiguity of the user query. Since there is no conclusion to this problem, in this research we focus on the problem of how to perform personalization and discussion of when to perform personalization is beyond the scope of this research.

VI. CONCLUSION AND FUTURE WORK

How to effectively utilize the rich user metadata in the social sharing websites for personalized search is challenging as well as significant. In this research we propose a novel framework to exploit the users social activities for personalized image search, such as annotations and the participation of interest groups. The query relevancy and user preference are simultaneously integrated into the final rank list. Experiments on a large-scale Flickr data set show that the proposed framework greatly out performs well. In future, we will develop our current work along with the four directions.

1) In this research, we only consider the simple case of one letter based queries. The actual construction of topic space provides a possible solution to handle the complex multiple letter-based queries. We will leave it for our future work.

2) During the user-specific topic modeling process, the obtained user specific topics represent the user's distribution on the topic space and can be considered as user's interest profile. Hence, this framework can be extended to any applications based on user's interest profiles.

3) For batch of new data (new users or new images), we can directly restart the RMTF- ranking multicorrelation tensor factorization and user-specific topic modeling process. While, for a small bulk of new data, designing the appropriate update rule is another future direction.

4) Utilizing large tensors brings challenges to the computation cost. We plan to turn to parallelization (e.g., parallel MATLAB) to speedup the RMTF converging the process. Moreover, the distributed storing mechanism of parallelization will provide a convenient way to store very large matrices and further reduce the storage cost.

REFERENCES

[1] J. Bian, S. Zheng, H. Zha, and C. L. Giles, "Exploring social annotations."
[2] Jia Hu and Philip K. Chan personalized web search by using learned user profiles in re-ranking.
[3] L. B. Marinho, A. Hotho, "Tag recommendations in social bookmarking systems,"

[4] Search engines for visual search re-ranking," in *Proc. SIGIR*, 2009, pp. 500–507.

[5] A. T. Campbell, "Hapori: Context-based local search for mobile phones using community behavioural modelling.

[6] E. Adar, and T. M. Breuel, "Personalized search," *Commun.*

[7] H. A. L. Kiers, "A fast method for choosing the number of components in Tucker3 analysis,".

[8] Chandrama G. Thorat personalized image search using hybrid re ranking based on ontology.

[9] Maryam Shokri1, Hamid R. Tizhoosh1, Mohamed Kamel2 Pattern Analysis and Machine Intelligence Laboratory reinforcement learning for personalized image search.

[10] S. Rendle, "Factorization machines," in *Proc. ICDM*, 2010, pp. 995–1000.

[11] D. M. Blei, "Sharing clusters among related groups:

[12] Yi Zhuang, Dickson K. W. Chiu, Nan Jiang, Guochang Jiang, Zhiang Wu personalized clustering for social image search results.

[14] Wei Liu[†], Yu-Gang Jiang[†], Jiebo Luo[‡], Shih-Fu Chang[†] noise resistant graph ranking for improved web image search.

[15] Changyun Zh, Kun Liy, Qin Lvz, Li Shangy, Robert P. Dickx personalized multi-modality image search for mobile devices.

[16] Yuli Gao¹, Chunlei Yang², Yi Shen², Jianping Fan² human computer interaction for large scale image retrieval.

[17] Neeraj Kumar, Student Member, IEEE, Alexander C. Berg, Member, IEEE, Peter N. Belhumeur, and Shree K. Nayar, Member, IEEE Describable Visual Attributes for Face Verification and Image Search