An Optimization Technique for Image Search in Social Sharing Websites

Saurabh Trikande
Dept. of Computer Engineering, SCOE, University Of Pune India
saurabh.trikande@gmail.com

M.K. Kodmelwar
Dept. of Computer Engineering, SCOE, University Of Pune, India
mkkodmelwar.scoe@sinhgad.edu

P.R. Futane
H.O.D, Dept. of Computer Engineering, SCOE, University Of Pune, India

ABSTRACT
Social sharing websites allow users to personalize media. The user can generate, share, tag and comment on media. The large-scale user-generated meta-data not only facilitates users in sharing and organizing multimedia content, but provides useful information to improve media retrieval and management of the websites. Personalized search serves as one of such examples where the web search experience is improved by generating the returned list according to the modified user search intents. In this paper, we exploit the social annotations and propose a novel framework considering both the user and query relevance to learn to personalized image search. The basic premise is to embed the user preference and query-related search intent into user-specific topic spaces. Since the users' original annotation is too sparse for topic modelling, we need to enrich users' annotation pool before user-specific topic spaces construction. The proposed framework contains two components. Firstly a Ranking based Multi-correlation Tensor Factorization model that is proposed to perform annotation prediction, which is considered as users' potential annotations for the images. Secondly we introduce User-specific Topic Modelling to map the query relevance and user preference into the same user-specific topic space. For performance evaluation, two resources involved with users' social activities are employed. Experiments on a large-scale dataset are used to demonstrate the effectiveness of the proposed method.

General Terms
Image search, Optimization, Social sharing website, personalized search engine.

Keywords
Personalize search, Topic space, Ranking based multi-correlation tensor factorization (RMTF), User-specific topic modeling.

1. INTRODUCTION
It is needless to emphasis the rising trend of exploitation of the Social Web. Instead of simply searching for, and passively consuming, information, users prefer to use blogs and social media sites like Flickr and Youtube that are creating, evaluating, and distributing information. In the process of using these sites, users are generating not only content that could be of interest to other users, but also a large quantity of metadata in the form of tags and ratings, which can be used to improve Web search and personalization. Web personalization refers to the process of customizing Web experience to an individual user. Personalization is used by online stores to recommend relevant products to a particular user and to customize a user's shopping experience. It is used by advertising firms to target ads to a particular user. Search personalization has also been studied as a way to improve the quality of Web search by disambiguating query terms based on user's browsing history or by eliminating irrelevant documents from search results. Personalizing image search is an especially challenging problem, because, unlike documents, images generally contain little text that can be used for disambiguating terms. Consider, for example, a user searching for photos of "Jaguars", should the system return images of luxury cars or spotted animal to the user? In this context, personalization can help disambiguate query keywords used in image search or to weed out irrelevant images from search results. Therefore, if a user is interested in wildlife, the system will show her images of the predatory cat of South America and not of an automobile.

In this paper we explore a novel source of evidence – user-generated metadata – that can be used to personalize image search results. We perform a case study of the technique on the simulated social photo sharing site, which allows users to upload images and label them with freely-chosen keywords, known as tags. Tags are meant to help users organize content and make it searchable by themselves and others. In addition to describing and categorizing images, tags also capture user's photography interests. We use a machine learning method to find topics of a large corpus of tagged images returned by image search on the simulated site. We then use the learned topics to match images to an individual user's interests. This appears to be a promising method for improving the quality of image search results.

2. STATE OF THE ART ANALYSIS
The existing system is found to be less efficient as compared to the proposed system in following manner.

2.1 Following are the drawbacks included in existing system
In existing system, users may have different intentions for the same query, e.g., searching for "Jaguar" by a car fan has a completely different meaning from searching by an animal specialist. Only, user-specific information [1] is considered to

distinguish the exact intentions of the user queries and re-rank the list results.

Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience [2] [3]. The user specific data is only used to re rank the result list.

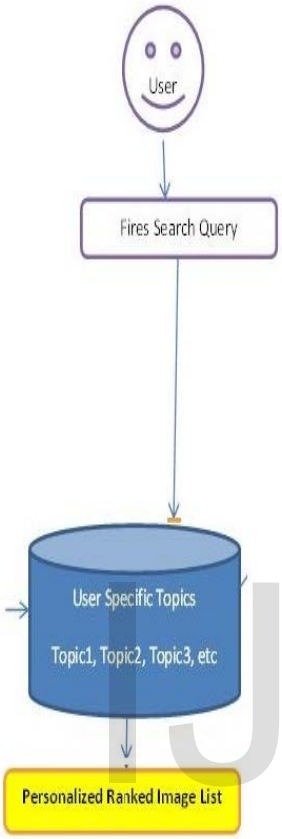

Fig. 1: Existing System

3. SCOPE OF THE PROPOSED SYSTEM

It is a web based system. It can be used to save response delay. It is also used by the companies for searching the data required from web which has a merit of being very fast thus reducing their infrastructure cost. It can also used on Airports for reservation purpose. It can be used to save user time of obtaining required data. It can be used to weed the irrelevant, query results. It can be used in recognize user intention of keyword used for searching.

4. OBJECTIVE OF THE PROPOSED TECHNIQUE

Main objective is to Optimize Image Searching Environment so as to maximize efficiency of searching. Develop a System that ensures reliable and intelligent image searching.

5. THE SYSTEM ARCHITECTURE

We propose a novel personalized image search framework as shown in Fig. 2 by simultaneously considering user and query information. The user's preferences over images under certain query are estimated by how probable he/she assigns the query-related tags[4] to the images.

A ranking based tensor factorization model named RMTF is proposed to predict users' annotations to the images. To better represent the query-tag relationship, we have proposed building of user-specific topics and mapping the queries as well as the users' preferences onto the learned topic spaces.

In the former existing system, users may have different intentions for the same query, e.g., searching for "Jaguar" by a car fan has a completely different meaning from searching by an animal specialist. One solution to address these problems is personalized search, where user-specific information is considered to distinguish the exact intentions of the user queries and re-rank the list results. Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience.

5.1 Ranking – Multi Correlation Based

Photo sharing websites differentiate from other social tagging systems by its characteristic of self-tagging: most images are only tagged by their owners. For instance, the #tagger statistics for Flickr [5] [6]. We can see that in Flickr, 90% images have no more than 4 taggers and the average number of tagger for each image is about 1.9. However, the average tagger for each webpage in Del.icio.us is 6.1. The severe scarcity problem calls for external resources to enable information propagation. In addition to the ternary interrelations, we also collect multiple intra-relations among users, images and tags. We assume that two items with high affinities should be mapped close to each other in the learnt factor subspaces. In the following, we first introduce how to construct the tag affinity graph, and then incorporate them into the tensor factorization framework.

To serve the ranking based optimization scheme [7], we build the tag affinity graph based on the tag semantic relevance and context relevance as shown in Fig 2. The context relevance of tag is simply encoded by their weighted co-occurrence in the image collection.

5.2 User-Specific Topic module

Users may have different intentions for the same query, e.g., searching for "jaguar" by a car fan has a completely different meaning from searching by an animal specialist. One solution to address these problems is personalized search, where user-specific information is considered to distinguish the exact intentions of the user queries and re-rank the list results [8]. Given the large and growing importance of search engines, personalized search has the potential to significantly improve searching experience.

5.3 Personalized Image Search

In the research community of personalized search, evaluation is not an easy task since relevance judgment can only be evaluated by the searchers themselves. The most widely accepted approach is user study, where participants are asked to judge the search results. Obviously this approach is very costly. In addition, a common 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 or click through history. However, this needs a large-scale real search log, which is not available for most of the researchers.

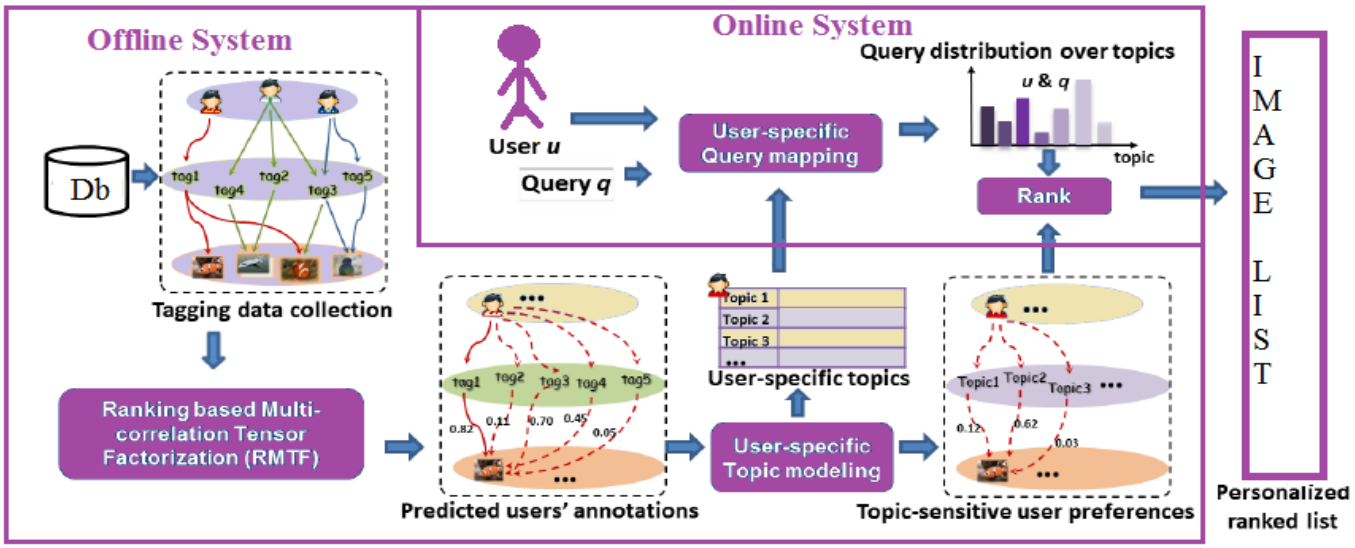

Fig. 2: Optimize Image Search Architecture of proposed system diag.

6. THE SYSTEM DESIGN

Three types of data including users images, tags as well as their ternary interrelations and intra-relations are first collected. Thereafter we performed users annotation prediction using Ranking based Multi-correlation Tensor Factorization mode. There are variations in individual user's tagging pattern and vocabularies. To overcome this problem the Use Specific Topic Modelling [9] is built for creating the semantic topics for each user.

As the original annotation is too parsed for topic modeling, we used the reconstructed ternary relations as the document collections. When a user *u* submits a query *q*, we first map the query *q* to user *u*-specific topics. The query distribution is then sent to the rank module and employed as the weight on topics to calculate the user *u*'s topic sensitive preferences over the images. Finally, the images are ranked according to the calculated user's preferences, which simultaneously consider the query and user information.

7. WORKING OF THE SYSTEM

In this paper, an application for image searching process optimization is generated which allow user to avoid irrelevant search results during keyword based search in social photo sharing website. This application uses rich metadata generated by user during his regular visits to the sites, i.e. application uses large amount of metadata created by user. The application allows weeding out the irrelevant images from the image search result.

8. CONCLUSION

Users of Web 2.0 sites face irrelevant data results while searching for data on search engines which causes wastage of resources. This problem is tackled in this proposed system utilizing the user created metadata, in the form of tags and social networks which describes their interests, tastes and preferences in searching the images on the social photo sharing websites.

This system can help user for searching the query based result in image searching. This system can predict user intention related to image, which are demanded by the user. In this system, only the simple case of one word-based query is considered. Actually, the construction of topic space provides a possible solution to handle the complex multiple words-based queries. During the user-specific topic modelling process, the obtained user-specific topics represent the user's distribution on the topic space and can be considered as that of the user. For batch of new data (new users or new images), the RMTF and user-specific topic modelling process is directly restarted.

9. ACKNOWLEDGEMENT

The authors would like to thank Principal S.C.O.E. and University of Pune for their valuable support and guidance.

10. REFERENCES

[1] B.Smyth, "A community-based approach to personalizing web search," Computer, vol. 40, no. 8, pp. 42–50, 2007.

[2] S. Xu, S. Bao, B. Fei, Z. Su, and Y. Yu, "Exploring folksonomy for personalized search," in SIGIR, 2008, pp. 155–162.

[3] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'El, I. Ronen, E. Uziel, S. Yogev, and S. Chernov, "Personalized social search based on the user's social network," in CIKM, 2009, pp. 1227–1236.

[4] Y. Cai and Q. Li, "Personalized search by tag-based user profile and resource profile in collaborative tagging systems," in CIKM, 2010, pp. 969–978.

[5] D. Lu and Q. Li, "Personalized search on flickr based on searcher's preference prediction," in WWW (Companion Volume), 2011, pp. 81– 82.

[6] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search?" in WSDM, 2008, pp. 195–206.

[7] Jitao Sang, Changsheng Xu, Senior Member, IEEE, Dongyuan Lu, "Learn to Personalized Image Search from the Photo Sharing Websites" IEEE Transactions on Multimedia, Volume:PP,Issue:99,2012.

[8] D. Zhou, J. Bian, S. Zheng, H. Zha, and C. L. Giles, "Exploring social annotations for information retrieval," in WWW, 2008, pp. 715–724.

[9] J. Tang, S. Yan, R. Hong, G. Qi and T. Chua, "Inferring Semantic Concepts from Community-Contributed Images and Noisy Tags," in ACM Multimedia, 2009, pp. 223–232.