PERSONALIZED TAG RECOMMENDATION FOR FLICKR USERS

Xiaoxiao Liu[1], Xueming Qian[1*], Dan Lu[1], Xingsong Hou[1], Liejun Wang[2]

[1]SMILES LAB, Xi'an Jiaotong University, [2]Xinjiang University.
qianxm@mail.xjtu.edu.cn; houxs@mail.xjtu.edu.cn

ABSTRACT

Social image share websites such as Flickr allow users to manually annotate their images with their own words, which can be used to facilitating image retrieval and other image applications. For the vast number of online images contributed by social users, existing methods on tag recommendation haven't taken users' characteristics and tagging habits into consideration. In this paper, we propose a personalized tag recommendation system for Flickr users. It can recommend users personalized tags for their newly uploaded photos based on the history information in their social communities. We carry out the personalized tag recommendation from three aspects. First, the tags we recommend to users are users' own vocabularies. Second, different recommendation methods are implemented to different users. Third, different users are recommended with different number of tags based on their tagging habits. The experimental results indicate that our personalized tag recommendation is effective.

***Index Terms*—**Social Community, Geo-Social Media, Tag Recommendation, User, Personalized, Flickr

1. INTRODUCTION

Online social image share websites such as Flickr has become more and more popular. It is not exaggerated to say that user's behavior and user-generated information are significantly influencing the way people manage and search multimedia resources.

User's history information recorded in Flickr is substantially rich. The geographical coordinates, taken time, and tags of the uploaded image are all recorded in its metadata. The metadata is of great value in our personalized tag recommendation system. Different from other tagging methods, the brilliant idea of our tag recommendation is that we take users' characteristics and tagging habit into consideration to realize the personalized services.

*Corresponding author. This work is supported in part by National Natural Science Foundation of China (NSFC) Project No. 61173109, No. 60903121, No. 61128007, and Microsoft Research.

At first, Flickr users may annotate images with their "unique" vocabularies. For example, a picture of a dog labeled with "apple" may be misleading to other users, but it is meaningful for the user himself. For "apple" is actually the dog's pet name. Furthermore, Flickr users' tagging behaviors may be influenced by different factors. For a Flickr who is fond of traveling, geographical location may affect his tagging results greatly. Last but not least, Flickr users may prefer different number of tags. Some users are used to giving many tags for their images. While others may prefer only a few tags. Those problems described above are all taken care of in our personalized tag recommendation system.

The contributions of this paper can be described as follows. (1) We propose a personalized tag recommendation system by recommending tags for Flickr users using their own vocabularies. (2) We analyze users' different characteristics to decide the specific recommendation methods for each user. (3) We study users' different tagging habits in order to make our recommendations more close to users' own tags.

The reminder of this paper is structured as follows. In Section 2, we review the related work on tag recommendation. Our recommendation system is stated in Section 3. The setup and performance of our experiments are shown in Section 4. Section 5 gives the conclusion and future work.

2. RELATED WORK

Various works are applied to recommend tags to Flickr users. The tag recommendation based on collective knowledge is proposed in [1]. The authors measured the similarity between tags by their co-occurrence information in the data collection, and used the top similar tags as recommendations. However, this kind of recommendation is based on single modality of tag co-occurrence on the whole dataset. In a broader context, various research projects have exploited the tag information in the Flickr community and other tagging systems to automatically extract useful semantics. The GPS and time stamps of photographs have been used along with the tags to extract events ("NY Marathon") and places ("Logan Airport") [2]. In our method, GPS and time stamps are also used in one of the steps of our system. In this step, we also use the visual features of images. In [3], Li et al. estimate the geo-locations of scenic spots by analyzing user-contributed social images.

Personalized recommendation is paid much attention recently [11-13]. Speaking of personalized tag recommendation, the authors are intended to use user's own vocabularies as the recommendations in [4] and [5]. However, compared with our work, they didn't take user's characteristic and tagging habit into consideration. They recommend the same number of tags to the Flickr using the same methods. Guan et al. propose a personalized tag recommendation method using graph-based ranking on multi-type interrelated objects in [6]. The work in [7] develops algorithms for automatic annotation of metadata. They transform the problem into a tag recommendation problem with a controlled tag library, and propose two variants of an algorithm for recommending tags.

3. OUR APPROACH

3.1. Problem Formulation

3.1.1. Image Representation

First, we introduce some notations. Let I,T,P,D,Z denote the image collections, the set of tags, GPS locations and image taken dates of a user u respectively. Let M denote the number of the total uploaded images by the user u. We have the user's history information $H=\{I,T,P,D,Z\}$ with

$I=\{I_i\}_{i=1}^M, T=\{T_i\}_{i=1}^M, P=\{P_i\}_{i=1}^M=\{(x_i,y_i)\}_{i=1}^M,$

$D=\{D_i\}_{i=1}^M,\ Z=\{Z_i\}_{i=1}^M$

where I_i means the i-th image; T_i is the tags of I_i; P_i is the GPS location of I_i; D_i *is* the taken date of I_i; Z_i is the visual features of I_i.

$T_i=\varnothing$ means no tags are provided by the user for the *i*-th image.

$P_i=\varnothing$ means no GPS locations are assigned to the *i*-th image.

The main information of the *i*-th image I_i can be a vector with four elements $s_i=\{\tau_i,p_i,d_i,z_i\}=\{\tau_i,(\mathrm{x}_i,\mathrm{y}_i),d_i,z_i\}$.

(1) τ_i is the tag set we recommend to the image I_i;

(2) p_i is the position the image I_i is taken;

(3) d_i is the taken date of the image I_i;

(4) z_i are the visual features of the image I_i.

We call the user u's image that we want to recommend tags to as input image. The input image is a newly uploaded image by user u, it has the GPS locations, taken time while has not been annotated by the user.

Among all these main information, user's name, taken position, taken time and visual features are the most useful ones to our methods. So the information of the input image can be written shortly as $s=\{\tau,p,d,z\}=\{\tau,(x,y),d,z\}$. Before utilizing our tagging method, the tag set τ of input image is empty.

The proposed approach recommends tags for the newly uploaded image according to the users' history information I,T,P,D,Z.

3.1.2. User Representation

For a user u, we have three parameters to measure his or her characteristic. We define the user as $u=(Eg\_u,Et\_u,Ev\_u)$.

(1) Eg_u is the experience point in the location domain of user u.

(2) Et_u is the experience point in the time domain of user u.

(3) Ev_u is the experience point in the visual domain of user u.

We view $u=(Eg\_u,Et\_u,Ev\_u)$ as a dot in a three-dimensional user space U showed in Fig.1. Every dot in the space U represents a user. Eg_u, Et_u, Ev_u are the coordinates in the three axis Eg, Et, and Ev respectively.

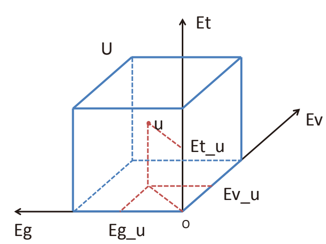


Fig. 1. The demonstration of user space U.

By comparing these three parameters, we decide which kind of personalized method we utilize to this specific user. Users in different regions of the space are recommended with different methods.

3.2. Overview of Our Approach

For tagging an input image $s=\{\tau,p,d,z\}=\{\tau,(x,y),d,z\}$ of user $u=(Eg\_u,Et\_u,Ev\_u)$, there are three steps in our personalized tag recommendation system. The detailed steps of our systemare demonstrated in Fig.2.

First, based on user's characteristic $u=(Eg\_u,Et\_u,Ev\_u)$, we select the best recommendation method (RM_u) for him or her. Second, we use the selected recommendation method to find corresponding neighbors of the input image. The neighbors might be GPS neighbors, time neighbors or visual neighbors.

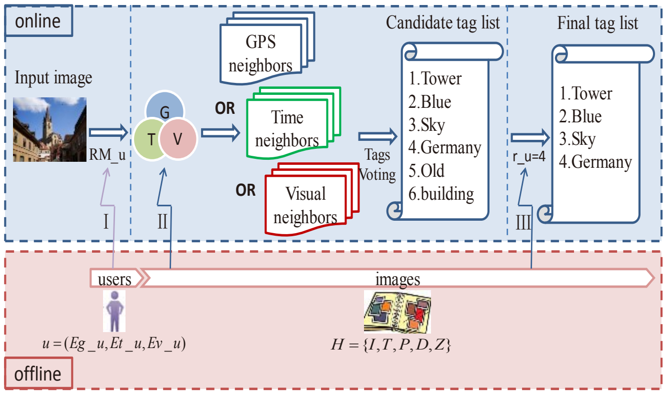


Fig. 2. The overview of our system. I is selecting recommendation method (RM_u)for the user. II is using RM_u to generate candidate tags for the input image. III is deciding the number of recommended tags (r_u) and generate the final results.

Then, we collect all the tags of these candidate neighbors and count their appearing times by tags voting. Third, by analyzing user's history information $H=\{I,T,P,D,Z\}$, we decide the number of recommended tags r_u to this user. Then the first r_u tags with higher appearing times are our final recommendations.

3.3. Recommendation Method (RM_u) Selection

Our recommendation methods for the input image s highly rely on its geographical coordinates p, taken date d and visual features z. p, d and z are applied to search GPS neighbors (G), time neighbors (T) and visual neighbors (V) respectively. The three elements G (geographical coordinates), T (taken time), and V (visual features) are used independently (UG, UT, UV) or fused into a unified tagging framework (UGT, UGV, UTV) in our recommendation methods.

For tagging an input image $s=\{\tau,p,d,z\}=\{\tau,(x,y),d,z\}$ of user $u=(Eg\_u,Et\_u,Ev\_u)$, UG only use the geographical information (x,y) to find its GPS neighbors. UT only utilizes the taken time d to find time neighbors. UV only processes the image features z to find visual neighbors. UGV uses the geographical information (x,y) and image features z to find GPS&Visual neighbors. UGT uses geographical information (x,y) and the taken time d to find GPS&Time neighbors. UTV makes use of taken time d and visual features z to search for Time&Visual neighbors.

We select the specific recommendation method for the user (RM_u) by analyzing user's own characteristics.

$$Eg\_u=\frac{Ng\_u}{Ng} \tag{1}$$

$$Et\_u=\frac{Nt\_u}{Nt} \tag{2}$$

$$Ev\_u=\frac{Nv\_u}{Nv} \tag{3}$$

where Ng_u, Nt_u, Nv_u are the number of locations, taken dates, visual clusters in user u 's history image collection respectively. The visual clusters are acquired by implementing self-adaptive K-means to the image collection. Ng, Nt, Nv are the average number of locations, taken dates, visual clusters among all users. So, for user u, we can find which factor plays a more important role in the process of tagging by comparing Eg_u, Et_u, Ev_u.

Then the recommendation method for user u is selected based on the following rules.

$$RM\_u=UP1=\begin{cases}UG,\text{if } Eg\_u=\max(Eg\_u,Et\_u,Ev\_u)\\ UT,\text{if } Et\_u=\max(Eg\_u,Et\_u,Ev\_u)\\ UV,\text{if } Ev\_u=\max(Eg\_u,Et\_u,Ev\_u)\end{cases} \tag{4}$$

$$RM\_u=UP2=\begin{cases}UTV,\text{if } Eg\_u=\min(Eg\_u,Et\_u,Ev\_u)\\ UGV,\text{if } Et\_u=\min(Eg\_u,Et\_u,Ev\_u)\\ UGT,\text{if } Ev\_u=\min(Eg\_u,Et\_u,Ev\_u)\end{cases} \tag{5}$$

UP1 means we only use one element in G, T, and V. UP2 means we use two elements in G, T, and V.

3.4. Candidate Tag Generation

3.4.1 Searching GPS neighbors

In this paper, we recommend user related tags for the image according to users' history information of $H=\{I,T,P,D,Z\}$. We choose the user's own images as GPS neighbors. Only in this way can we tag user's image with his or her own vocabularies. We determine whether the image I_i is GPS neighbor of input image or not by comparing its GPS location (x_i,y_i) with (x,y).

$$G(i)=\begin{cases}1,\text{if } sg(x_i-x,\alpha)=0\ \&\ sg(y_i-y,\alpha)=0\\ \quad 0,\text{otherwise}\end{cases} \tag{6}$$

where $sg(x,\alpha)=0$, if the integer portion and the first α decimal places of x are all 0; otherwise $sg(x,\alpha)=1$. G(i)=0 means that the image is not a GPS neighbor, while G(i)=1means it is a GPS neighbor. In this paper, the total number of GPS neighbors in these M images is defined as $N_G=\sum_{i=1}^{M}G(i)$. NG=0means that there is no GPS neighbors for the newly uploaded image taken at (x,y). This is the case for the new users or the users capture photos at some new places. NG≠0, which means that the user has already uploaded images taken in the same place with the newly uploaded image. α is a parameter that indicates the accuracy of geographical coordinates. When α is 5, the position of (x,y) and (x_i,y_i) are less than 1.1 meters apart.

3.4.2Searching Time neighbors

We determine whether the image is time neighbor of input image or not by comparing the taken time with the taken time of the input image.

$$T(i)=\begin{cases}1, \text{ if } t(d_i,d)\le\beta \\ 0, \text{otherwise}\end{cases} \tag{7}$$

where $t(d_i,d)$ means the time intervals between d_i and d. $T(i)=0$ means that the i-th image I_i is not a time neighbor, while $T(i)=1$ means it is a time neighbor of input image. In this paper, the total number of time neighbors in these M images is defined as $N_T=\sum_{i=1}^{M}T(i)$. $N_T=0$ means that there is no other images taken in the same date d with the newly uploaded input image.

3.4.3 Searching Visual neighbors

For each image in the user's image collection I, we compare its low level features with the input image. Features z in $s=\{\tau,p,d,z\}=\{\tau,(x,y),d,z\}$ of our approach are described as by the grid based color moment and the hierarchical wavelet packet descriptor.

Color feature has been proved to be the most GPS-informed feature [8]. It is used as global feature representation for the image in our method. An image is divided into four equal sized blocks and a centralized image with equal-size. For each block, a 9-D color moment is computed, and thus the dimension of color comment for each image is 45. The 9-D color moment of an image segment is utilized, which contains values of mean, standard deviation and skewness of each channel in HSV color space.

Texture feature has been shown to work well for texture description of image and for scene categorization and image recognition. The texture feature in our method is described by hierarchical wavelet packet descriptor (HWVP) [9]. A 170-D HWVP descriptor is utilized by setting the decomposition level to be 3 and the wavelet packet basis to be DB2.

The visual similarity between images is measured by the Euclidean distance of two images as follows:

$$D(i)=\|z_i-z\| \qquad i=1,2,M \tag{8}$$

where z_i and z are the low-level feature of the image I_i and the input image. In this paper, we rank the distances in ascending order and select the top ranked 10 images as its visual neighbors under the constraints that the visual similarity of two images are sufficient large.

3.4.4 Tags Voting

We use the tags appeared in the image neighbors to annotate the newly uploaded image by ranking repetition times of tags of the GPS, time and visual neighbors.

3.5. Recommended Tag Number (r_u) Prediction

Different users have different tagging habits. Some users are used to giving many tags for their images. While others may prefer only a few tags.

In this step, we try to predict the number of recommended tags to approach the real number of the input image. That is to say, the best result of our prediction is that r_u=o.

For user u, we can can use the following three ways to predict the number of recommended tags r_u.

1) $r_u=N$
2) $r_u=N_u$
3) $r_u=N_{up}$

where N is the average number of tags among all images in our dataset. According to our statistic, the average number is approximately 6.76. So we set $N=6$. N_u is the average number of tags among all user u's images. N_{up} is the average number of tags among user u's images taken in different period of time.

We categorize user's image collection into eight periods based on the taken date: spring day, spring night, summer day, summer night, autumn day, autumn night, winter day, and winter night. User's tagging habit may change over time, so we introduce the third way. In the third way, first we assign the input image into one of the eight periods. Then we use the average number of tags in this class as r_u.

4. EXPERIMENTS

4.1. Dataset

In order to evaluate the performance of our methods, we randomly crawled more than 6 million images together with their tags from the image sharing site Flickr.com through its public API. The initial data includes 6,715,251 images uploaded by 7,387 users and their related files recording the information of tags and geographical ordinates. We remove the information of images that have no tags and no geographical ordinates. We have made a statistic about all the images and users. The result is shown in Table 1.

As we can see in Table 1, the remaining data contains 1,903,089 images uploaded by 6,582 users. That is to say, most users have the habit to give their images tags or geographical ordinates. For every user, we choose the image uploaded most recently as input image for testing, and we view other images as the training set of this user. Among these testing images, about 14.8% of them are in batch tagging mode. We remove these batch tagging images to avoid their influence on our tagging methods. So, it turns out that there are 5,607 images for testing the performances of the proposed tagging approaches.

Table 1. Flickr users and their uploaded images

		ALL	With tag	With GPS	With GPS+tag
User	Num.	7387	7276	7276	6582
	%	100%	98.50%	98.50%	89.09%
Image	Num.	6715251	5317909	2144661	1903089
	%	100%	79.19%	31.94%	28.34%

4.2. Criteria of Performance Evaluation

For the input image, the user has annotated o tags $t=\{t_1,t_2,...,t_o\}$. When the input image has been uploaded by this user, o is an invariant value. For each input image we recommend r tags $\tau=\{\tau_1,\tau_2,...,\tau_r\}$ to the user. By comparing tags in these two sets $t=\{t_1,t_2,...,t_o\}$ and $\tau=\{\tau_1,\tau_2,...,\tau_r\}$, we find that some are the same with the original tags but the others are not. In this paper we use Recall, Precision and F1 to measure tagging performance of a test image, which are defined as follows.

$$Recall=\frac{c}{c+m}\times 100\%=\frac{c}{o}\times 100\% \tag{9}$$

$$Precision=\frac{c}{c+f}\times 100\%=\frac{c}{r}\times 100\% \tag{10}$$

$$F1=\frac{2\times Recall\times Precision}{Recall+Precision}\times 100\% \tag{11}$$

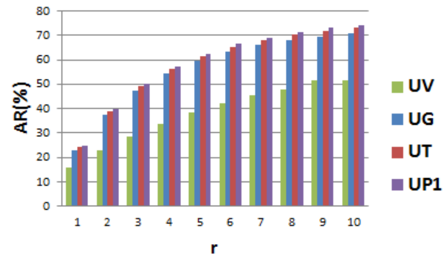
where c, f, and m are the number of correct, false and missed tags. We use the average recall (AR), average precision (AP) and average F1 (AF) of 5,607 users under different r for evaluating tagging performance.

4.2. Result of Our Experiments

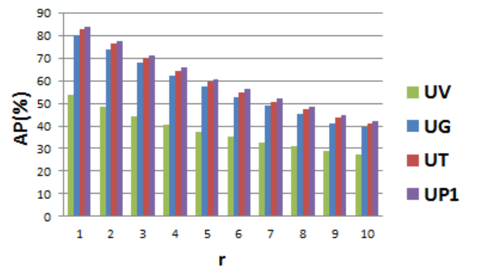
4.2.1 The Performance of recommendation method selection

In this part, our experiments are carried out in order to demonstrate whether we have selected better method for each user or not. So we set the number of recommended tags r from one to ten. Then we compare the results of different methods under the same r. In UP1 and UP2, users are recommended by different methods based on their different characteristics. In UP1, all users have selected one of the three methods UG, UT, and UV. In UP2, all users have selected one of the three methods UGT, UTV, and UGV. In UG, we set the parameter $\alpha=5$. In UT, we set the parameter $\beta=1$ day.

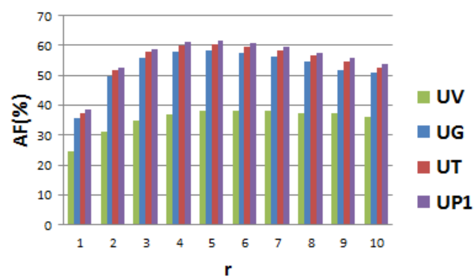
As it is shown in Fig.3, UT outperforms UG and UV on average of all the 5607 users. However, it doesn't mean UT is the best for every user. So the performance of UP1 is even better than UT. At the same time, UP2 outperforms UGT, UTV, and UGV. This result proves that the process of method selection is effective. That is to say, by analyzing user's characteristic, we manage to choose a better recommendation method for our user.



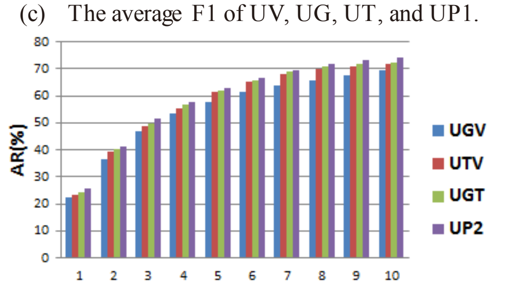
(a) The average recall of UV, UG, UT, and UP1.



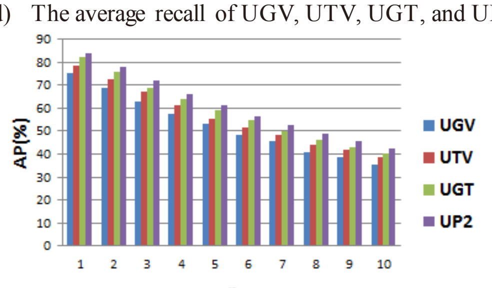
(b) The average precision of UV, UG, UT, and UP1.



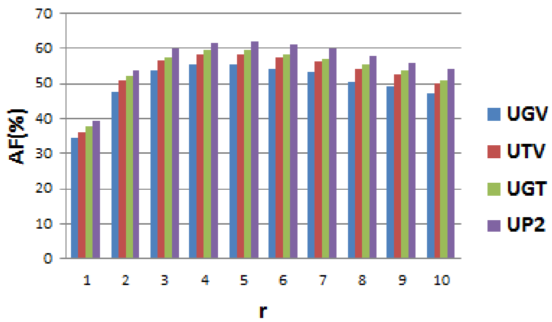
(c) The average F1 of UV, UG, UT, and UP1.



(d) The average recall of UGV, UTV, UGT, and UP2.



(e) The average precision of UGV, UTV, UGT, and UP2. (to be continued)



(f) The average F1 of UGV, UTV, UGT, and UP2.

Fig. 3. AR, AP, and AF of our methods.

Different users are recommended with different methods. We also make a statistic about users' distribution in method selection. The result is shown in Table 2.

Table 2. User's distribution in method selection.

Method	UT	UG	UV
Percentage	71.6%	18.1%	10.3%

In UP1, 71.6% of users have chosen the method UT. Meantime, 18.1% of users are recommended by method UG. 10.3% of users are recommended by method UV. Since UT outperforms the other two methods, it is reasonable that most of the users have chosen the method UT.

4.4.2The Performance of tag number prediction

We make a statistic about the average number of tags for 6,582 users. The result is demonstrated in Table 3. According to this table, we can see that user's tagging habit in tag number differs. About 52.64% users annotate their images with no more than 5 (denoted by [1,5]) tags on average. About 17.05% users give their images more than 10 (denoted by >10) tags on average. For the rest of the users, their average number of tags is between 6 and 10 (denoted by [6,10]).

Table 3. The average number of user's tags

Tag Number	[1,5]	[6,10]	>10
User Number	3465	1995	1122
Percentage	52.64%	30.31%	17.05%

To measure the effectiveness of our tag number prediction, we use the result of UP1 when $r_u = o$ as the benchmark (i.e. the number of initial tags labeled by user). It is the best result that UP1 can get with our predicted number of tags. The experimental comparison is shown in Table 4.

Table 4. The results under different number of r_u

UP1	$r_u = o$	(1) $r_u = N$	(2) $r_u = N_u$	(3) $r_u = N_{up}$
AR	72.81%	66.37%	67.24%	68.26%
AP	72.81%	56.17%	66.56%	68.97%
AF	72.81%	60.85%	66.90%	68.61%

According to the experimental comparison illustrated in Table 4, the third way has the best performance. (3) is better than (2). This explains user's tagging habit indeed changes over time. (1) has the lowest performance. This shows recommending different users with different number of tags is effective in some extent.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a personalized tag recommendation system for Flickr users. We can recommend user's own vocabularies to him or her when an image is newly uploaded. To realize the personalized service, we analyze user's characteristic and tagging habit. Different users are recommended with different methods. What's more, different users are recommended with different number of tags. Experimental results show our system is effective. However, there is still much work to be done. First, we will dig deep on how to recommend tags in a more personalized way. Second, we will work on studying user's behavior with more aspects. At last, we will try to apply some learning-based methods on personalized tag recommendation.

REFERENCES

[1] B. Sigurbjörnsson, R. V. Zwol, "Flickr tag recommendation based on collective knowledge", In WWW, 2008.

[2] T. Rattenbury, N. Good, M. Naaman, "Towards automatic extraction of event and place semantics from Flickr tags", SIGIR, 2007, pp.103-110, 2007.

[3] J. Li, X. Qian, Y. Y. Tang, L. Yang, T. Mei, "GPS estimation for places of interest from social users' uploaded photos", IEEE Trans. Multimedia 2013, vol.15, no.8, pp.2058-2071.

[4] X. Qian, X. Liu, C. Zheng, Y. Du, X. Hou, "Tagging photos using users' vocabularies", Neurocomputing, vol.111, 2013, pp.144-153.

[5] N. Garg, I. Weber, "Personalized, interactive tag recommendation for flickr", ACM ResSys, 2008.

[6] Z. Guan , J. Bu , Q. Mei , C. Chen , C. Wang, "Personalized tag recommendation using graph-based ranking on multi-type interrelated objects", SIGIR, 2009.

[7] S. Tuarob, L. C. Pouchard, C. L. Giles, "Automatic tag recommendation for metadata annotation using probabilistic topic modeling", ACM/IEEE-CS joint conference on Digital Libraries, 2013.

[8] J. Li, X. Qian, Yuan Yan Tang, L. Yang, C.Liu, "GPS estimation from users' photos", In Proc. MMM2013, pp.118-129.

[9] X. Qian, G. Liu, D. Guo, Z. Li, Z. Wang, H. Wang, "Object Categorization using Hierarchical Wavelet Packet Texture Descriptors", in Proc. ISM 2009, pp.44-51.

[10] J. Shen, D. Tao, and X. Li, "QUC-Tree: Integrating Query Context Information for Efficient Music Retrieval", IEEE Trans. Multimedia, vol.11, no.2, 2009, pp.313-323.

[11] Z. Cheng and J. Shen, "Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation," ACM ICMR, 2014.

[12] J. Shen, and Z. Cheng, "Personalized video similarity measurement", Multimedia Syst., vol.17, no.5, vol.2011.

[13] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized Recommendation Combining User Interest and Social Circle", IEEE Trans. Knowledge and Data Engineering, 2014.