# Personalized Web Image Retrieval Based on User Interest Model

Zhaowen Qiu[1,*], Haiyan Chen[1], and Haiyi Zhang[2]

[1] Institute of Information and Computer Engineering, Northeast Forestry University, China
qiuzw@nefu.edu.cn
[2] Jodrey School of Computer Science, Acadia University, Canada
haiyi.zhang@acadiau.ca

**Abstract.** The traditional search engines don't consider that the users interest are different, and they don't provide personalized retrieval service, so the retrieval efficiency is not high. In order to solve the problem, a method for personalized web image retrieval based on user interest model is proposed. Firstly, the formalized definition of user interest model is provided. Then the user interest model combines the methods of explicit tracking and implicit tracking to improve user's interest information and provide personalized web image retrieval. Experimental results show that the user interest model can be successfully applied in web image retrieval.

**Keywords:** user interest model, personalized, interest learning, web image retrieval.

## 1 Introduction

Information retrieval services provided by traditional search engines do not take into account the user's differences lead to different users enter the same keyword, the same search results returned. But in reality, due to differences in background knowledge, interests and other aspects of the different needs of users tend to be different. With the rapid development of information technology at the same time, people in trouble among the massive data, which does not distinguish between the user's information retrieval will consume a lot of search time and reduce the efficiency of information retrieval.

In order to get a more accurate search results in line with the actual needs of different users, personalized Web image retrieval has become a research hotspot. The essence of personalized for different users with different service strategy to provide different services, and personalized Web image retrieval is interested in the initiative to learn and record the user based on the user feedback on the search results, suggesting that the user's interest demand [1]. So consider the differences in the user's search, you can greatly improve retrieval efficiency. To solve the above problems, this paper presents the user interest model, support the user's personalized Web image retrieval, in order to improve retrieval efficiency.

* Corresponding author.

H. Wang et al. (Eds.): ICYCSEE 2015, CCIS 503, pp. 188–195, 2015.

User interest model [2] is a personalized image retrieval core, which is used to store and manage user interest information. By collecting user feedback, establish initial user interest model, and then through a long learning, and constantly update the user interest model, and ultimately get the user's interest tends to provide personalized search services for different users. Currently, most studies of user interest model is still rare in the image field. Hsuetal[3] The user interest model used in medical image retrieval, but only for this type of image; Lietal[4] According to the user's interest to adjust the visual characteristics of the underlying weights, but ignores the Figure like the high-level semantic information. Through the establishment of user interest model, the use of explicit and implicit method of combining learning and to consider user interest by the time the impact of this factor, and constantly improve the user's interest in information provided by the user interest model personalized web image search service for users [5].

### 1.1 Image Semantic Feature Extraction

Low-level visual features of images by color, texture and shape features characteristic composition. Among them, the color characteristics of HSV color space based on 20 non-uniform quantization algorithm [6], a total of 20 dimensions. Four statistic texture feature extraction using GLCM, select the features that the problem: Contrast, consistency, relevance pixel gray scale and as entropy feature vector [7], at 00,450,900, 1350 extracting the four directions, four texture, composed of 16-dimensional feature vector [8]. Feature extraction using the seven-dimensional shape of Hu's invariant moments. In this paper, a method based on SVM semantic association to obtain semantic information of the image, as shown in Figure 1 specific method.

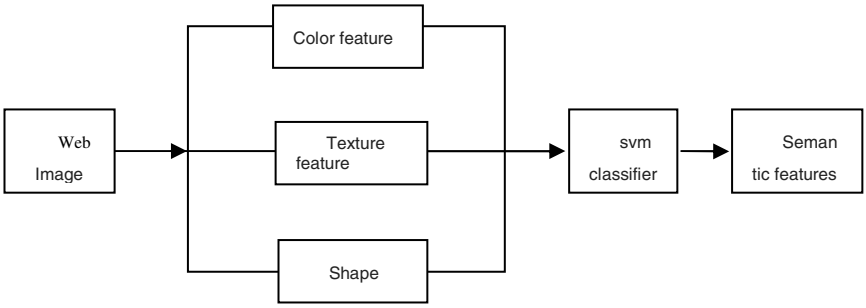


**Fig. 1.** Extraction of image high-level semantic feature

Based on SVM [9] semantic correlation method is to extract semantic information through the establishment of low-level image features to high-level features of the visual mapping. First calculate the underlying visual image feature vector, and use them as SVM input vectors, then the image class to learn to build low-level visual features of the image is mapped to the high-level semantics. Semantic correlation method, the semantics of a class associated with each one as a separate classification of the two, that is, for ei`ther a semantic category, the positive training examples are all the images included in the class, and is not a counter-example all other classes of image class.

## 2 User Interest Model

User interest model describes the user's interests and hobbies, plays a central role in personalized retrieval system, is the foundation of personalized service. User interest model is mainly visited by statistical and learning to get the user information, the need to constantly update [9].

User interest model (User interest model, UIM) is defined as six group:

$M = (U, S, Q, F, T, \delta,)$ Among them, $U = (u_1, u_2, ..., u_n)$ is a collection of user;

$S = (S_1, S_2, ..., S_i ...., S_n)$ is a collection of interest to the user semantics; $S_i = (s_{i1}, s_{i2}, ..., s_{im})$ ,it show that M i represents the user interest semantic;

$Q = (Q_1, Q_2, ..., Q_i, ..., Q_n)$ is interested in the semantics of the initial weights corresponding set of users, $Q_i = (q_{i1}, q_{i2}, ..., q_{im})$ ,it represents the initial weights for the user i of interest semantic of m;

$F = (F_1, F_2, ..., F_i, ..., F_n)$ is that who interested in the right to terminate the corresponding user set value semantics. $F_i = (f_{i1}, f_{i2}, ..., f_{im})$ ,it represents that user i value the right to terminate the m interested in semantics;

$T = (T_1, T_2, ..., T_i, ..., T_n)$ record that users interested in the semantics of the time it first appears, Important parameters that affect the long-term interests of the users, $T_i = (t_{i1}, t_{i2}, ..., t_{ij}, ..., t_{im})$ ,it represents that Users interested in the semantics of m i for the first occurrence of the time.

$\delta$ is a collection of user behavior, after a series of operations in the user's right to the semantic values of interest to the user corresponding to the set value set by the initial weight value is converted to the right to terminate collection.

The main behavior of users as follows:

Query: keyword search query the user submitted images;

Click on browse: the user clicks on a thumbnail, browse image-related information;

Favorites: Users will add the current page to your favorites;

Download: Users download and save the image; Rating: user feedback on the search results and evaluation.

Users interested in the extent of the image is mainly reflected by the access behavior of users, depending on the access behavior of users, for users interested in semantic weights updated accordingly, as shown in Table 1.

Different users have different semantics for the same degree of interest, and therefore the evaluation made by different feedback $\delta$ , weights based on the user interest semantic semantics will be the degree of interest of users is divided into A, B, C, D shown, E5 a rating scale, specific criteria for the classification are shown in Table 2,the lager $\delta$ weight, show more interest to users of the semantic category, corresponds to the higher degree of interest.

A user interest corresponding to the highest degree, E corresponds to the lowest degree of user interest.

**Table 1.** Semantics weight value influence from user visiting action

| user access behavior $\delta$ | Image semantic weights $(q_{ij}, f_{ij})$ |
|---|---|
| Query | $q_{ij} = f_{ij} + 0.1 \quad f_{ij} = q_{ij}$ |
| Click on Browse | $q_{ij} = f_{ij} + 0.2 \quad f_{ij} = q_{ij}$ |
| Favorites | $q_{ij} = f_{ij} + 0.3 \quad f_{ij} = q_{ij}$ |
| Download | $q_{ij} = f_{ij} + 0.4 \quad f_{ij} = q_{ij}$ |
| Evaluation | $q_{ij} = f_{ij} + \alpha \times 0.1 \; f_{ij} = q_{ij}$ |

**Table 2.** User interest degree to semantics

| The degree of user interest | A | B | C | D | E |
|---|---|---|---|---|---|
| Weights $\alpha$ | 0 | 1 | 2 | 3 | 4 |

## 3 Users Interested in Learning

Due to the volatility of the user's interest has, therefore need to access the user's behavior, the use of interest corresponding learning algorithm to adapt to constantly update the user's corresponding changes [10]. To address this problem, we first establish the initial user interest model based on the information and the user's first visit to the behavior of user-submitted, and user access behavior using explicit and implicit learning method combining constantly improve the user interest model.

### 3.1 Explicit Learning

Explicit learning refers users to choose their own interest or the semantics of the retrieved images for feedback and evaluation, so as to achieve the purpose of learning [11]. According to the semantics of the user is interested, you can get the user rating matrix, using an M × N matrix to represent, as shown in equation (1). M wherein M represents the row number of users, N of the N columns represent the semantics of a user element represents a semantic score

$$R(M,N) = \begin{pmatrix} R_{1,1} & \cdots & R_{1,j} & \cdots & R_{1,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{i,1} & \cdots & R_{i,j} & \cdots & R_{i,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{M,1} & \cdots & R_{M,j} & \cdots & R_{M,N} \end{pmatrix} \tag{1}$$

The value of $R_{i,j}$ is 0,1,2,3,4, the larger the value of $R_{i,j}$, the more that the user interest semantic. Users in image retrieval process, you can explicit feedback

information, that after a user browse images, images of scores presented to illustrate the extent of the image of the user interested. We define the rating scale is divided into five, as shown in Table 3-2. Through the study of user feedback information, the system to dynamically update the user interest information.

In practice, the user's interest is constantly changing, and each user generally only a small part of all the information on the entries were evaluated, all due to the $R(M,N)$ become an extremely sparse matrix. User rating for sparse matrix problems, Based on the user's implicit interest in learning to solve.

### 3.2 Implicit Learning

Since the display user data collected to learn there may be some false information or non-real feedback and evaluation, we can analyze the behavior of the user's access to the user's interest in implicit learning, in order to constantly improve the user interest model, to provide users personalized service [12].

Users interested in implicit learning is automatic, so the system can automatically update the user interest model based on user access behavior. Users to perform queries, click Browse, when collections, download and other actions, the user interested in semantics corresponding weights will increase. When utilizing user interest model retrieval of image information, priority will weights greater semantic class returned to the user. User access behavior reflects the degree of interest to users of the resource, the impact of user access to semantic degree of interest shown in Table 1. According to the behavior of users can get access to the following scoring matrix:

$$R(M,N)=\begin{pmatrix} f_{1,1} & \cdots & f_{1,j} & \cdots & f_{1,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{i,1} & \cdots & f_{i,j} & \cdots & f_{i,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{M,1} & \cdots & f_{M,j} & \cdots & f_{M,N} \end{pmatrix} \tag{2}$$

When a user interested in for a resource, the user when browsing this page will consume more time and will often repeat visits to the page, so we can use the access time and access frequency to calculate the user a certain degree of interest in semantics, and users will be affected interest of time, with the loss of time and gradually weakened, and therefore the degree of user interest can be defined as:

$$\mathrm{F}=(\frac{n}{N}+\frac{t}{T})\times\frac{l}{L}\times e^{\frac{\log_2(T_l-T)}{hl}} \tag{3}$$

Where n is the number of visits to the semantic-based nodes, the total number of N-based visits, t is access to the semantic node consumes time, T is the total time of the site visit used, l-oriented semantics visit nodes , L is the total number of points

semantic festival website. $e^{\frac{\log_2(T_l - T)}{hl}}$ is amnesia factor,h1 is Life Cycle parameters, generally, One week after exposure to new knowledge will begin to forget, so h1 equals seven, T for users interested in the semantics of the first occurrence time, Tl is the last time the user behavior. Degree of interest in using the above model of equation (1) to be updated, user rating matrix obtained, see equation (4).

$$R(M,N) = \begin{pmatrix} R_{1,1} + F_{1,1} & \cdots & R_{1,j} + F_{1,j} & \cdots & R_{1,N} + F_{1,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{i,1} + F_{i,1} & \cdots & R_{i,j} + F_{i,j} & \cdots & R_{i,N} + F_{i,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{M,1} + F_{M,1} & \cdots & R_{M,j} + F_{M,j} & \cdots & R_{M,N} + F_{M,N} \end{pmatrix} \tag{4}$$

Combining the formula (2) and (4) update the user to obtain the final score matrix, see equation (5).

$$R(M,N) = \begin{pmatrix} R_{1,1} + F_{1,1} + f_{1,1} & \cdots & R_{1,j} + F_{1,j} + f_{1,j} & \cdots & R_{1,N} + F_{1,N} + f_{1,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{i,1} + F_{i,1} + f_{i,1} & \cdots & R_{i,j} + F_{i,j} + f_{i,j} & \cdots & R_{i,N} + F_{i,N} + f_{i,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ R_{M,1} + F_{M,1} + f_{M,1} & \cdots & R_{M,j} + F_{M,j} + f_{M,j} & \cdots & R_{M,N} + F_{M,N} + f_{M,N} \end{pmatrix} \tag{5}$$

## 4 Results

Experiment, Corel Photo gallery and join the network video material in the library portion of the image, a total of 6000 images, these images into 80 semantic categories, such as mammals, birds, reptiles, plants, vehicles, buildings and so on. Select a picture in each of the semantic class as a representative, and labeling. Taking into account the large amount of computation problems, the experiment only 80 randomly selected semantic class 100 image as a query test set, the image retrieval model based on user interest and user interest model does not use image retrieval done comparative experiments, the test results are shown in Figure 2.

As can be seen from the experimental results, the model-based image retrieval of user interest, and do not use the results of user interest model approach compared to average precision has been significantly improved. After a user submits a query image keywords, correct image appears before the first one is the probability of 10 up to 96%, compared with not using user interest model higher than 6.9%. Visible, user interest model presented in this paper can be more accurate characterization of the user's interest, applied to Web image retrieval, effectively improve the efficiency of the image retrieval.

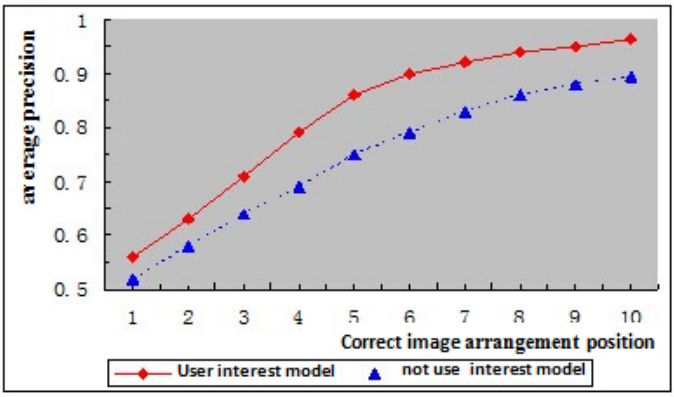


**Fig. 2.** Image semantics annotation accuracy by using user interest model

## 5 Conclusions

User interest model is the key technology of personalized image retrieval through user feedback speculated that the user needs to effectively compensate for the semantic gap problem underlying visual features and high-level semantics between. This paper presents an improved user interest model, and gives the formal definition, then the use of explicit and implicit tracking combined method for tracking user interest for learning, and constantly improve the user interest information [13]. Can provide personalized Web image retrieval service for users through the user interest model. The experimental results show that the model of interest to the user to effectively improve the efficiency of the image retrieval. Will further study among users with similar interests hobby migration issues, has studied the use of user interest information to improve the new registered users interested in information retrieval quickly provide personalized service for new users.

**Acknowledgment.** (Foundation: Supported by the Fundamental Research Funds for the Central Universities(DL10CB01), Heilongjiang Province Natural Science Fund(LC2012C06), Special Funds for Harbin Innovative Talent in Science and Technology(2012RFLXG022)).

## References

1. He, L., Jing, Z., Lan, S.: Sun Image retrieval construction method personalized user interest model. Computer Engineering and Applications 45(31), 168–171 (2009)
2. Martin-Bautista, M.J., Kraft, D.H., Vila, M.A., et al.: User profiles and fuzzy logic for Web retrieval issues. Soft Computing 6, 365–372 (2002)
3. Hsu, C.C., Chu, W.W., Taira, R.K.: A Knowledge-Based Approach for Retrieving Images by Content. IEEE Transactions on Knowledge and Data Engineering 8(4), 522–532 (1996)

4. Li, J.H., Liu, M.S., Cheng, Y.: User Interest Model Based Image Retrieval Technique. In: Proceedings of the IEEE International Conference on Automation and Logistics, China, pp. 2265–2269 (2007)
5. Adda, M.: A Formal Model of Information Retrieval Based on User Sensitivities. Procedia Computer Science, 428–436 (2013)
6. Zhaowen, Q., Tian, Z.: A new image color feature extraction method. Harbin Institute of Technology Journal 36(12), 1769–1701 (2004)
7. Zhaowen, Q.: User-oriented Web image retrieval key technology research. Doctoral dissertation. Harbin Institute of Technology (2009)
8. JBC. Based on Web Mining user personalized recommendations Research and realization. Beijing University of Posts and Telecommunications (2013)
9. Xiaohua, L., Lansun, S., Huanqin, L.: Estimation of crowd density based on wavelet and support vector machine. Transactions of the Institute of Measurement and Control 28(3), 299–308 (2006)
10. Park, Y.J., Chang, K.N.: Individual and group behavior-based customer profile model for personalized product recommendation. Expert Systems with Applications 36(2), 1932–1939 (2009)
11. Huang Huadong based personalized search user model study. East China University of Technology (2013)
12. Alberto, D., Pablo, G.: User-model Based Personalized Summarization. Information Processing & Management 43(6), 1715–1734 (2007)
13. Jing, Y.: User interest model and real-time personalized recommendation algorithm. Nanjing University of Posts and Telecommunications (2013)