International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.6 (2014), pp.31-48
http://dx.doi.org/10.14257/ijmue.2014.9.6.05

Research on User-personalized Image Retrieval Method

Yu Song, Jing-fei Ren, Mao-Zhu Jin* and Pei-Yu Ren

Business School of Sichuan University, Chengdu, China
jinmaozhu@scu.edu.cn

Abstract

With the rapid expansion of information resources, the amount of image data in the network shows an explosive growth trend. The traditional search engines have not considered users' different interests; therefore image retrieval efficiency is reduced. To solve the problem, this paper puts forward a research on user-based personalized image retrieval technologies. Firstly, this paper studies the user interest model, and provides its definitions and application strategies; secondly, it studies collaborative filtering algorithm based on K-means clustering, and solves the problem of sparse resources effectively; Finally, explicit tracking, implicit tracking and relevance feedback methods are adopted to learn and update user interest model constantly to meet the users' needs and improve retrieval accuracy and efficiency. Based on the above studies, this paper presents a kind of user-based personalized recommendation technology, and completes an image retrieval system based on user personalization, proving that this recommendation technology is able to provide users with better personalized recommendation service.

Keywords: *Personalized recommendation; Collaborative filtering; Image Retrieval; Relevance feedback*

1. Introduction

The information age has come as computer technologies develop rapidly. The information resources in this era are much more than ever before. As an important information resource, multimedia information such as videos, images and audios have become vital information media on the internet. Images play a major role in multimedia information due to its low producing cost, convenient storage and speedy transmission. As a result, the researches on image retrieval technologies become increasingly appealing to researchers and this also has practical significance to researches on personalized image retrieval technologies.

Personalization refers to the differentiated and specific service offered to users according to their varied demands. Personalized image retrieval means the initiative learning on users' interests based on information such as users' operation on image data and their retrieval histories, and according to the learning, users' demands and the image information they are about to search can be predicted [1]. Personalized recommendation technology play a vital role in helping users' acquire corresponding demanded information so that it is highlighted in the academic circle. At present, researches on personalized recommendation have deepened and personalized recommendation technology has been developed and applied greatly. By preprocessing user rating matrix, Fang Yuke [4] simplified the ranking question as a rating question that based on the nearest users, and applied integrating learning method boosting in recommending service. According to Wang Guoxia [5], the network was excavated

* Corresponding Author

ISSN: 1975-0080 IJMUE

intelligently by drawing diagrams and complex network theoretic technology, and under that circumstance the performance of the recommendation system was improved.

Among various recommendation technologies, collaborating filtering technology is the most classic personalized recommendation technology. Wang Qian [6] converted users' evaluation on a project into the calculations of users' preference on certain projects, therefore the nearest user group could be calculated. Li Feng [7] and Xia Xiufeng [8] put forward a personalized recommendation technology that based on product feature.

Nowadays, the major element that affects the recommendation accuracy of collaborating filtering technology is the so called sparse data, namely when considering about the nearest users' evaluation on resources, the evaluated resources are little compared to the total resources in the system, and this leads to a scarce and sparse evaluation data given by nearby users. The sparse data fails the system from accurately confirming the nearest user groups of the targeted users and therefore it cannot conduct a high-quality and high-efficiency personalized recommendation to targeted users [9].

Among personalized recommendation systems, user interest model [10] is the core of the system when offering personalized recommendation services to the users and it is established by recording all kinds of users' behavior information. As users search more, the system continues to amend user interest model. Hsu adopted user interest model in medical image retrieval and it worked well [11].

In personalized recommendations, relevance feedback technology is used to perfect user interest model so as to better reflect users' demands. The thought of relevance feedback technology is to adjust recommendation mechanism by using information of users' feedback on the results of the system. And the purpose is to provide more accurate and more reliable recommendation service. At present, in terms of image retrieval, the major relevance feedback technology adopted in personalized recommendation technologies is man-machine coordinated and interactive learning method [12]. Yin used users' retrieval logs as the feedback information to conduct image semantic clustering [13].

2. Research Foundation

2.1. Collaborative Filtering Technology

Collaborative filtering technology is a major kind of personalized recommendation technology. Its main idea is to predict targeted users' evaluation on resources that have not been visited by taking account of evaluations of the nearest users group of the target users. In this case, a personalized recommendation can be carried out. In other words, if user X and Y have similar evaluation or behavior towards some projects, then their opinions towards other projects are similar [14].

So far, collaborative filtering technology can be divided into two kinds: content-based collaborative filtering technology and project-based collaborative filtering technology. The major idea of former one is to calculate users' similarity through their evaluations on the projects. Then the nearest user group of the targeted users can be found, and through user groups' evaluation on projects, targeted users' evaluation on projects that have not been visited can be predicted. Finally the recommendation can be realized based on the prediction. Project-based collaborative filtering technology is to calculate project similarity by taking account of their attributes and then recommend according to the similarity level.

2.2. User Interest Model

User interest model is a kind of structured file layout that stores and manages users' interested information. It is mainly consisted of semantics that the users are interested in. Each semantic has its correspondent interest weight and the higher the weight, the more interest the user has towards it.

Nowadays, the user interest model that has been researched and applied more is based on semantic representation method. The semantic is obtained according to the key words that the users searched in the system as well as the system's analysis on users' operation behavior, therefore the semantic can authentically reflect users' interests. User interest model that based on semantic representation method is easy, stable and united, so this paper employs it.

User interest model that based on semantic representation method is manifested as a descriptive document about users' interests in personalized recommendation system [15]. Its main idea is to describe the relation between semantics and their weights. Weight can be shown by Boolean Value or Real Value to represent users' interestingness for a certain semantic. Weight can reflect various semantics' degrees of importance in a user interest model, so here comes the following definition:

Definition 1. User Interest Model

UIM is used to store and mange information which the users are interested in. It is manifested as a quintuple [15]:

$$M=(U, S, \delta, Q, F)$$

$U=(u_1, u_2, u_3 \ldots u_i \ldots, u_n)$, u_i represents user i, and U is the collection of the entire users. $S=(s_1, s_2, s_3 \ldots s_i \ldots, s_n)$, and $s_i=(s_{i1}, s_{i2}, s_{i3} \ldots s_{im} \ldots s_{in})$. S_{im} refers to semantic m that user i is interested in. $Q=(q_1, q_2, q_3 \ldots q_i \ldots, q_n)$, and $q_i=(q_{i1}, q_{i2}, q_{i3} \ldots q_{im} \ldots q_{in})$. Q_{im} refers to the initial interest weight of semantic m which user i is interested in. $F=(f_1, f_2, f_3 \ldots f_i \ldots, f_n)$, and $f_i=(f_{i1}, f_{i2}, f_{i3} \ldots f_{im} \ldots f_{in})$. F_{im} refers to the final interest weight of semantic m which user i is interested in. δ represents user behavior. During the retrieval process, the weight of users' interested semantic constantly changes, making the interest weight of the semantic converts from initial interest weight to final interest weight.

When a user retrieve image, the major user behavior includes[16]:
Inquiry: the user uses key words to retrieve correspondent images.
Scan: the user scans detailed information of his or her interested images.
Download: the user downloads his or her interested images.
Evaluation: the user gives feedback, comments or marks to the retrieved images.

Users' interestingness for images can be judged from their operation behavior. Considering user behavior's impact on semantic interest weight, Table 1 is provided as a standard to update the interest weight of the semantic.

Table 1. User Operation Behavior's Impact on Interest Weight of the Semantic

User's line δ	Semantic interest weight(q_{ij}, f_{ij})
Inquiry	$q_{ij} = f_{ij}$; $f_{ij} = q_{ij}+0.02$; $f_{ij}=0.3(q_{ij} \leq 0.3)$
Scan	$q_{ij} = f_{ij}$; $f_{ij} = q_{ij}+0.02$; $f_{ij}=0.2(q_{ij} \leq 0.2)$
Download	$q_{ij} = f_{ij}$; $f_{ij} = q_{ij}+0.04$; $f_{ij}=0.4(q_{ij} \leq 0.4)$
Evaluation	$q_{ij} = f_{ij}$; $f_{ij} = q_{ij}+\alpha*0.01$; $f_{ij}=0.3(q_{ij} \leq 0.3)$

q_{ij} refers to the initial interest weight of semantic m which user i is interested in, f_{ij} refers to the final interest weight of semantic m which user i is interested in and α refers to users' evaluation index to the semantic. The value of the index are 1, 2, 3, 4, 5. The larger the value is, the higher the evaluation index and the deeper the interestingness towards the semantic.

2.3. Learning User Interest Model

By collecting user information such as retrieval history and operation behavior and after a series of processing, the elements to set up user interest model can be converted and then being used to offer guidance for personalized service [17]. Only by learning a user's retrieval information and operation behavior can the user's interest model be built. User interest models will constantly changes as the user's retrieval information and operation behavior change. So it is necessary to keep learning and updating a user interest model. Firstly, a user interest model is set up according to the user's initial information; then the method of explicit tracking and implicit tracking are combined in learning a user's behavior so as to perfect the user interest model effectively.

2.3.1. Explicit Tracking Learning

Explicit tracking learning is to obtain and update image semantic information which a user is interested in by making use of the user's feedback and evaluation on the retrieval results. Users will have feedback on their retrieval results, for example to judge whether the images meet their demands, or to score for the images. These kinds of information can directly and easily get the information of whether the user is interested in the image or not. However, most users will not feedback in a serious and detailed way so that the feedback and evaluation information which the system obtains cannot truly reflect how much interest does a user has towards a certain image. In this case, the rating data in the system will become sparse, making the system unable to accurately confirm what interests a user. Therefore, it is difficult to carry out a high-quality and high-efficiency personalized recommendation to the targeted users.

2.3.2. Implicit Tracking Learning

Implicit tracking learning is to learn about a user's interest implicitly through his or her operation behavior towards images. A user's interest in resources can be largely reflected through his or her operation behavior towards the resources, enabling the system from obtaining accurate information on user interest. Implicit tracking learning can overcome shortcoming as fake information which will appear in explicit tracking learning so that it can better reflect a user's interest. When a user inquires about images by key words, then the user's interested semantic can be obtained through the retrieval key words; when a user scans or downloads an image, the interested semantic can be obtained through the image's label and based on this, the user's interestingness for the image can be predicted.

2.3.3. User Interestingness

By analyzing a user's scan behavior on the web, the user's interestingness for a certain webpage can be judged. So a user's interestingness for the resources of a webpage can be calculated by this analysis. According to the lines scanned by a user in the web, Cui *et al.* (2011) [18] offered the following definition:

Definition 2. A computational formula of interestingness based on the time a user spends on scanning the webpage, as shown in formula (2.1).

$$\mathrm{Int}_{ki} = \frac{\sum_{j=1}^{m} t_{kij}}{\sum_{i=1}^{n}\sum_{j=1}^{m} t_{kij}} \qquad \text{Formula (2.1)}$$

t_{kij} is the time user k spends on product j of type i when scanning the webpage, m represents resource numbers of a certain type and n represents the total types of the resources.

2.4. Similarity

Sim(D_1,D_2) refers to the similarity between document D_1 and D_2 in a vector space model. If the document is shown by vector quantity, then similarity between documents can be calculated by a distance formula of the vector quantities. As a result, the distance between document vector quantities can be used to calculate the similarity between documents. The nearer distance between the correspondent vector quantities of the documents, the higher similarities between the documents and vice versa [19]. This method can be adopted to calculate similarities among users. At present, there are three major methods to calculate the similarity, namely Cosine Similarity, Pearson Correlation Similarity and Amended Cosine Similarity [20].

Definition 3. Computational formula of cosine distance, as shown in formula(2.2)

$$\mathrm{sim}(\mathrm{User}_1, \mathrm{User}_2) = \frac{\sum_{k=1}^{n}(W_{1k} \times W_{2k})}{\sqrt{\sum_{k=1}^{n} W_{1k}^2} \times \sqrt{\sum_{k=1}^{n} W_{2k}^2}} \qquad \text{Formula (2.2)}$$

W_{1i} represents $User_1$'s score for project k and W_{2k} represents$User_2$'s score for project k.

3. User-personalization-based Recommendation Technologies

3.1. Vocabulary-based User Interest Model

To set up a user interest model, this paper adopted a vocabulary-based user interest model and it is stored as a data base. The vocabularies in this paper's model are from the retrieval key words that the users type, key words when scanning images and users' feedback information towards retrieval results and recommendation results. The interest weight of these vocabularies is updated according to Table one.

If user A's user interest model is M, and according to definition 1, this paper defines user interest model which it adopts as follows.

Definition 4. User interest model:

$$M=(U, S, \delta, H)$$

U represents user A's Id. $S=(s_1, s_2, s_3 \ldots s_i \ldots, s_n)$, and s_i represents vocabulary i which user A is interested in. $H=(h_1, h_2, h_3 \ldots h_i \ldots, h_n)$, and h_i represents the interest weight of vocabulary i which user A is interested in. δ represents user behavior, and the conversion of interest weight of the vocabularies is based on Table one.

With the help of this model, the relation between users' interested vocabularies and their weights can be directly described. In addition, a user's interestingness for a certain vocabulary can be described in a relatively united way. According to user behavior that δ represents and in combination with Table one, the user interest model can be timely and conveniently updated so that the model can be optimized to meet the real demands of users.

3.2. Learning Vocabulary-based User Interest Model

Since users' interests are changeable, the system needs to learn and update user interest model constantly based on users' retrieval histories and operation behavior. In terms of this problem, this paper will use explicit tracking learning as a supplement and implicit tracking learning as a major method to learn about users' retrieval histories and operation behavior so as to perfect user interest model.

3.2.1. Image-retrieval-based Explicit Learning: This paper employs explicit tracking learning to gradually perfect user interest model by using relevant information of the users obtained through relevance feedback and key-word-based image retrieval.

With regard to image retrieval, the major step is to analyze the semantic information that the image carries to reflect a user's interested field. So in explicit tracking learning, firstly an initial user interest model is set up based on relevant information a user filled when he or her registers, for example hobbies and interests; then regard vocabulary on the hobbies and interests as the initial vocabulary information in a user interest model and set its interest weight at 0.5.

When a user conducts key-word-based image retrieval, his or her interested vocabulary can be directly obtained by tracking retrieval key words. Meanwhile, this vocabulary information will be added into this user's interest model. And if the vocabulary information is already in the model, then its interest weight should be updated according to Table 1.

When the system offers a user the retrieval results and recommendation results, the user's interestingness for this image can be acquired by his or her feedback on those results. The feedback can be mainly divided into dislike, somewhat dislike, okay, somewhat like and like, and their vocabulary interest weight are 1,2,3,4,5 respectively. Then the vocabulary information the image carries and its correspondent interest weight is added into the user's interest model. If the vocabulary information is already existed in the model, then its interest weight should be updated according to Table 1.

According to a user's interested vocabulary, a user interested vocabulary matrix can be obtained. This paper employs A (M,N) matrix to reflect user interested vocabulary matrix which is got through explicit tracking learning. In another word, it represents the interest weights of N words from M users, as shown in Table 2. $A_{i,j}$ represents user i's vocabulary interest weight of word j.

Table 2. Vocabulary Interest Weight Matrix

$A_{1,1}$	$A_{1,2}$	$A_{1,3}$	……	$A_{1,j}$	……	$A_{1,n-1}$	$A_{1,n}$
$A_{2,1}$	$A_{2,2}$	$A_{2,3}$	……	$A_{2,j}$	……	$A_{2,n-1}$	$A_{2,n}$
……	……	……	……	……	……	……	……
$A_{i,1}$	$A_{i,2}$	$A_{i,3}$	……	$A_{i,j}$	……	$A_{i,n-1}$	$A_{i,n}$
……	……	……	……	……	……	……	……
$A_{m,1}$	$A_{m,2}$	$A_{m,3}$	……	$A_{m,j}$	……	$A_{m,n-1}$	$A_{m,n}$

However, in real life, only a few users will give the feedback in a serious and detailed manner so that this paper employs implicit tracking learning to get relevant information of the user.

3.2.2. Image-retrieval-based Implicit Learning

With regard to image retrieval, the major user behaviors of the implicit tracking learning in this paper are scanning and downloading.

A user's operation behavior reflects his or her interested image and the semantic the image carries can represent the user's interest to some degree. The system records a user's operation behavior implicitly to get the semantic of the image which the user scans right then, and the user interest model is updated and perfected according to Table 1.

User interested vocabulary matrix will be got based on a user's interested vocabulary. This paper employs B(M,N) matrix to reflect user interested vocabulary matrix which is got through implicit tracking learning. In another word, it represents the interest weights of N words from M users, as shown in Table 3. $B_{i,j}$ represents user i's vocabulary interest weight of word j.

Table 3. Vocabulary Interest Weight Matrix

$B_{1,1}$	$B_{1,2}$	$B_{1,3}$	……	$B_{1,j}$	……	$B_{1,n-1}$	$B_{1,n}$
$B_{2,1}$	$B_{2,2}$	$B_{2,3}$	……	$B_{2,j}$	……	$B_{2,n-1}$	$B_{2,n}$
……	……	……	……	……	……	……	……
$B_{i,1}$	$B_{i,2}$	$B_{i,3}$	……	$B_{i,j}$	……	$B_{i,n-1}$	$B_{i,n}$
……	……	……	……	……	……	……	……
$B_{m,1}$	$B_{m,2}$	$B_{m,3}$	……	$B_{m,j}$	……	$B_{m,n-1}$	$B_{m,n}$

By combining Table 2 and Table 3, a user's final vocabulary interest weight matrix P(M,N) can be obtained, as shown in Table 4.

Table 4. Final Vocabulary Interest Weight Matrix

$P_{1,1}+A_{1,1}+B_{1,1}$	$P_{1,2}+A_{1,2}+B_{1,2}$	…	$P_{1,j}+A_{1,j}+B_{1,j}$	…	$P_{1,n}+A_{1,n}+B_{1,n}$
$P_{2,1}+A_{2,1}+B_{2,1}$	$P_{2,2}+A_{2,2}+B_{2,2}$	…	$P_{2,j}+A_{2,j}+B_{2,j}$	…	$P_{2,n}+A_{2,n}+B_{2,n}$
……	……	…	……	…	……
$P_{i,1}+A_{i,1}+B_{i,1}$	$P_{i,2}+A_{i,2}+B_{i,2}$	…	$P_{i,j}+A_{i,j}+B_{i,j}$	…	$P_{i,n}+A_{i,n}+B_{i,n}$
……	……	…	……	…	……
$P_{m,1}+A_{m,1}+B_{m,1}$	$P_{m,2}+A_{m,2}+B_{m,2}$	…	$P_{m,j}+A_{m,j}+B_{m,j}$	…	$P_{m,n}+A_{m,n}+B_{m,n}$

3.3. K-means-based Collaborative Filtering Algorithm

K-means-based collaborative filtering algorithm firstly will confirm the nearest user group by user similarity. As for calculating user similarity, it is related with users' interestingness for a same word. Then vocabulary interestingness will be calculated and it is through the analysis and processing of the vocabulary and its weight in user interest model. Finally, if the nearest user group of the targeted users is confirmed, then the resources, which the nearest user group speaks highly of, can be recommended.

The structure of K-means-based collaborative filtering algorithm is shown as Figure 1:

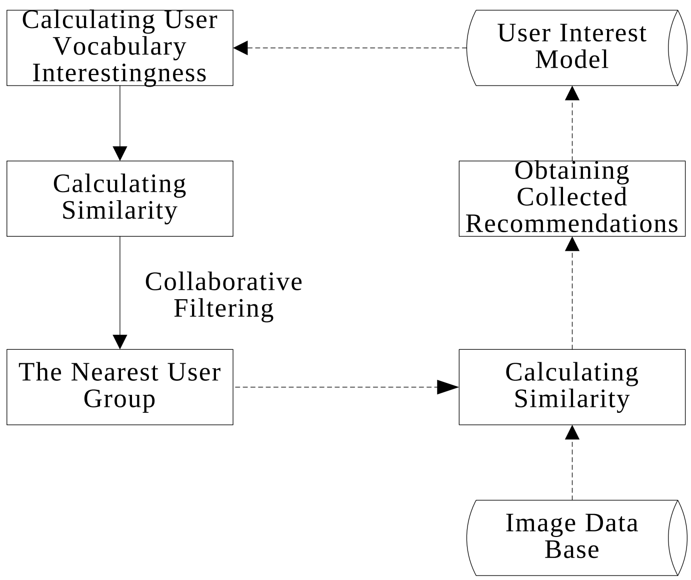

Figure 1. Algorithm structure

3.3.1. User Interest Vocabulary: A user interest model which is based on interestingness regards users' vocabulary interest weight as the foundation in searching users' interested semantics of the images. As a result, to obtain users' final interest vocabulary of a certain kind of resources in an image retrieval system, the users' interestingness for the semantics which the images carry should be calculated by formula (2.1).

In a user interest model, provided that altogether there are K interested words, then all the words in the model constitute a user interest vocabulary set Tsouce, and the interest weights of all the vocabularies can be seen as a series, so here comes the following definition.

Definition 5. Computational formula to calculate a user's interestingness for vocabulary, as shown in formula (3.1).

$$\text{Interest}(Word_i) = \frac{\text{Weight}\,(Word_i)}{\sum_{i}^{Tnum} \text{Weight}\,(\text{Word}_i)} \qquad \text{Formula (3.1)}$$

Tnum represents the total number of words in a user interest model, and Weight($Word_i$) represents user i's interest weight on $Word_i$.

On the basis of user interest model, by calculating users' interestingness for vocabularies, vocabularies that users truly interested in will be confirmed.

3.3.2. The Nearest User Group: This paper regards vocabulary-interestingness-based Cosine Similarity sim_Interest(User,$User_i$) as the way to measure user similarity in K-means algorithm, and only takes targeted users as the center. Based on formula(2.2), a method to measure vocabulary-interestingness-based similarity is put forward and here comes the definition.

Definition 6. A computational formula to measure the Cosine Similarity between the targeted User and user i, sim_Interest(User,$User_i$), is shown as formula(3.2).

$$\mathrm{sim\_}Interest(\mathrm{User},\mathrm{User_i}) = \frac{\sum_{k=1}^{n}(Interest(\mathrm{W}_k)\times Interest(\mathrm{W}_{i,k}))}{\sqrt{\sum_{k=1}^{n} Interest(\mathrm{W}_k)^2}\times\sqrt{\sum_{k=1}^{n} Interest(\mathrm{W}_{i,k})^2}} \quad \text{Formula(3.2)}$$

Interest(W_k) represents targeted User's interestingness for word k and Interest($W_{i,k}$) represents user i's interestingness for work k.

Considering about the number of people who have the same vocabulary interestingness, different methods are employed to confirm user-based nearest user group under two kinds of circumstances. If there are less than five people who have the same vocabulary interestingness, then this group is directly seen as one nearest user group. If not, formula (3.2) is used to calculate polymerization degree among the users to form a nearest user group.

3.4. User-based Personalized Image Recommendation Algorithm

Provided that there are some users in the system and the system establishes user interest model for each user, on the basis of the above analysis on each part of user-based personalized image retrieval technologies, the user-based personalized image recommendation process and algorithm defined by this paper is as follow:

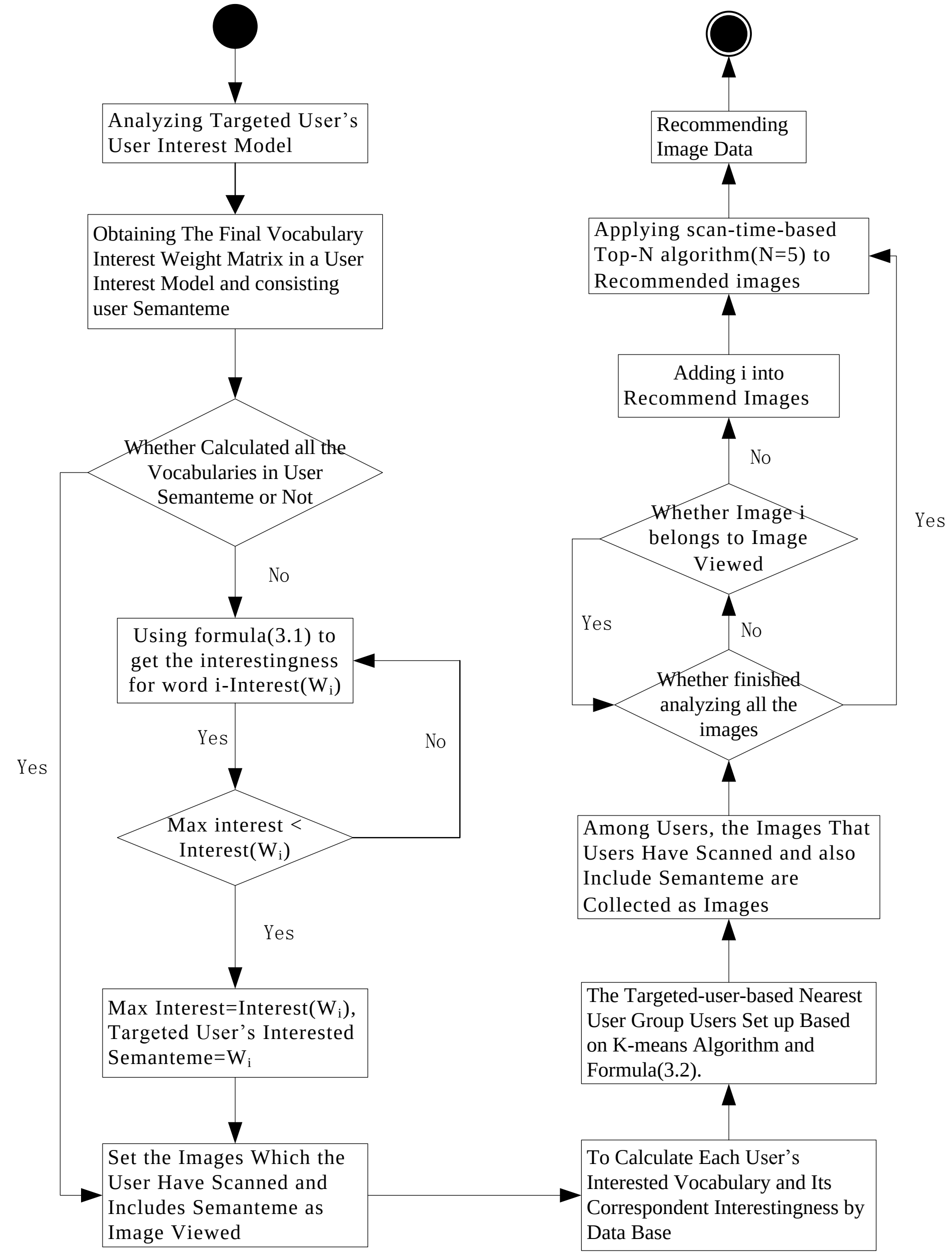

Figure 2. Algorithm Process

Algorithm name: user-based personalize image recommendation algorithm
Algorithm function: to confirm recommended images for targeted users
Input:user behavior sets
Output: the recommended M resources
Algorithm process:
①obtaining User Semanteme
②calculating interestingness for user interest vocabulary:
To set the images which targeted users have scanned and include Semanteme as Image Viewed.

To calculate other users' interestingness for user interest vocabulary.

③calculating the nearest user group-Users

The images that targeted-user-based Users have scanned and include Semanteme are collected as Images.

To set collection Common and recommendation collection as Recommend Images.

④generating of the recommended images

⑤updating user interest model

End

The complexity of this algorithm is O(N*M), in which N represents the number of users in the nearest user group and M represents the number of the analyzed interested words of targeted users.

This paper will give a detailed description for some sub-algorithm in this algorithm in the following lines.

For sub-algorithm ①:

To obtain the user's user interest model through inquiring data base;

To inquire five semantics with have the highest vocabulary interest weight in the user interest model, and gather these words and their weights to form a User Semanteme;

For sub-algorithm ②:

```
To obtain the user's user interest model through inquiring data base;
To inquire the weight for all the vocabularies-Wset in the user interest model;
While(each word i∈User Semanteme)  do{
      Set the interest weight of word i as W and interestingness as X;
While(each weightWi∈Wset) {
               X+=W/Wi;
}
If(Max Interest<X){
Max Interest = X;
Interest vocabulary of targeted users, Semanteme= i;
}
}
```

For sub-algorithm ③:

```
To obtain the user's user interest model through inquiring data base;
To inquire the weight for all the vocabularies-Wset in the user interest model;
While(word n∈UserSemanteme)do{
      The interest weight of word n is N, and its interestingness is X;
While(each weight Wi∈Wset) do{
               X+=N/Wi;
               If(Max Interest Weight<X){
                        Max Interest Weight=X;
                        Max Interest Words = n;
                }
}
}
If(Max Interest Words==Semanteme) {
User Count++;
```

```
    Add the user into userSet;
  }
  If(user Count <5){
  The nearest user group based on targeted users, Users=userSet;
  }
  Else{
      If(targeted user's interested word-SourceW==any user's interested word-W){
      calculate the Distance between targeted user's interestingness of the interest word and W;
      put Distance and other users' Id into UserDistance;
  }
  Choose five smallest Distance in UserDistance to form the nearest user group based on targeted users.
  }

  For sub-algorithm④:
  While (each image i∈Images)  do{
    If(i∉ImageViewed){
  put i into Common;
  }
  }
  If(Common.length ≤ 5){
  RecommendImages=Common;
  }
  Else{
  applying scan-time-based Top-N algorithm(N=5) to Common so as to get a new set Common'.
  RecommendImages= Common';
  }
  Recommending image data in RecommendImages to users;

  For sub-algorithm⑤:
  obtain user's retrieval key word W;
  If(W∈UserSemanteme)
  Updating vocabulary interest weight matrix according to Table 1 and updating user interest model as well;
  Else add W to UserSemanteme,set h=0.5;
  While (each image Image_i∈RecommendImages) do{
  Users offer feedback on Image_i;
  Updating vocabulary interest weight matrix and user interest model according to relevant user feedback after scanning Image_i;
  }
```

4. Experiment and Analysis

To verify whether the user-based personalized image recommendation algorithm which is put forward by this paper can provide users qualified personalized service or not, this paper will design a paper prototyping system, namely user-based personalized image retrieval system. This system is based on user interest model, K-means-based collaborative filtering technology and user-based personalized recommendation algorithm.

This system is a user-based personalized image retrieval system and can provide users with personalized image retrieval and recommendation service. Therefore, in terms of developing model, B/S model is adopted to facilitate the separation of user information and system data. This system employs a Tomcat 7.0-based and JDK1.7-based integrated developing environment; the foreground developing adopts My Eclipse 10.0 developing application program; the system's design language is Jsp and JavaScript, and the background data base is MySql.

4.1. System Design

The user-based personalized image retrieval system in this paper includes foreground and background. The foreground mainly includes user registration and log in module, personal information management module, user interest model module, image recommendation module and image information feedback module. Whereas the background includes image management module. The structure of this system is shown as Figure 3.

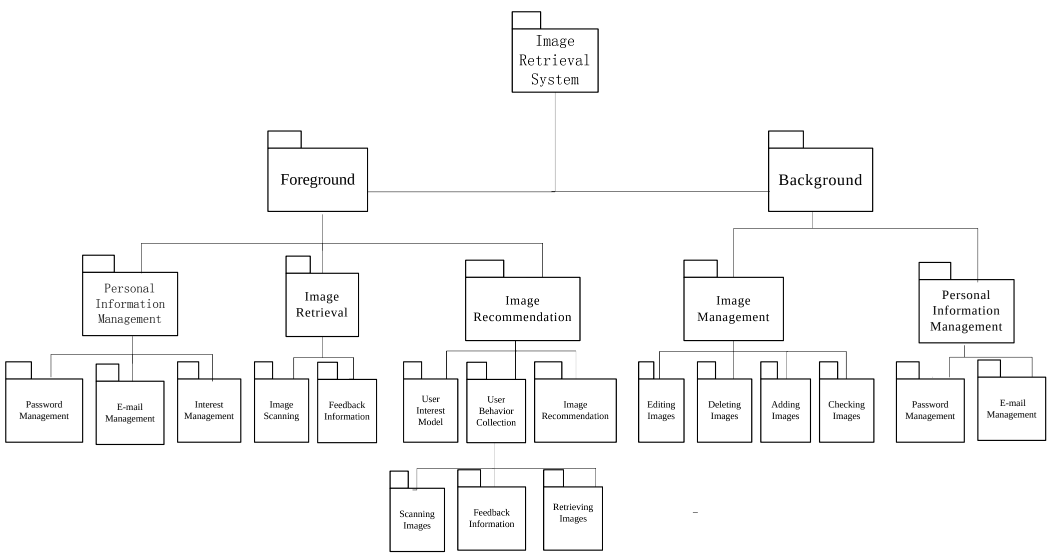

Figure 3. System Structure

When a user logs in the system by inputting the username and password, he or she can editing the password, e-mail, hobbies and interests; the user can acquire images by key words and at the same time the user can click the acquire results to see the details. When scanning the details, the user can give feedback of whether the images are interested or not; if the user interest model reaches a certain scale, namely the user interest vocabulary matrix of the model reaches P (5,5), and when the user-based personalized image recommendation technology in this paper is employed, then the users can obtain their interested image information through image recommendation set when the system recommends images to them. By analysis on users' different operation on the retrieval results and recommendation results, the system will adopts a learning method which takes explicit tracking as a supplement and implicit as a dominant to track users' retrieval histories and operation behavior so as to update and perfect the user interest model. In this case, the system can timely follows users' interests and offers them convenient and accurate personalized recommendation service.

When a manager logs in the system by inputting the username and password, he or she can check and alter users' personal information, for example to alter the passwords and e-mails; the manager can also add, delete, alter and check the images in the system to ensure the reliability of the system's image data.

4.2. Experiment Results and Evaluation

The experiment in this paper is conducted in the user-based personalized image retrieval system which has been already designed. During the experiment, 10 users are selected to be included in a contrast experiment, in other words, they will experience both the user-based personalized image retrieval system and non-user-based system. In this contrast experiment, users' interest theme A will be recorded explicitly. And by contrasting users' retrieval on A in seven days, the performance of the personalized recommendation technology can be tested.

The system will record the 10 users' retrieval, download and feedback information and automatically analyzes their interests. The total number of the system's recommended images, user's retrieved images and users' accepted images will be recorded separately. And during the seven days, users' retrieval, scanning and download information will be tracked.

This paper mainly adopts two parameters to evaluate the accuracy of the system's recommendation-recommendation accuracy Precision and retrieval times Count, and their computational formulas are as definition 7 and 8.

Definition 7. Recommendation accuracy. Precision is the proportion between the resource amounts that are accepted by the users and the total resources that the system recommends. It is used to reflect the performance of this system's recommendation technology, and its computational formula is shown as (4.1).

$$\mathrm{Pr}\,ecision = \frac{\mathrm{accept}Num}{\mathrm{recommend}Num} \times 100\% \qquad \text{Formula(4.1)}$$

acceptNum represents the visited times of the recommended images, and recommendNum represents the total number of the recommended images.

Definition 8. Retrieval times. Count is the average times of each user to click and search the key words which are being typed in the system's search bar.

Each user can choose their interested theme A according to their own interests. During the seven-day experiment, they can retrieve A in both the user-based personalized image retrieval system and the traditional image retrieval system. The experiment information of each day will be recorded.

The results of the experiment is shown in Figure 4 and 5.

Figure 4. Recommendation Accuracy

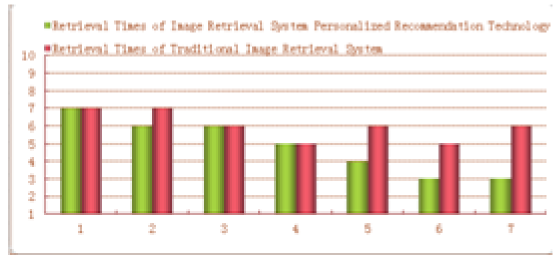

Figure 5. Contrast of Retrieval Times

By analyzing Precision, in the early stage of image retrieval, the recommendation accuracy of the image retrieval system that this paper designs is relatively low. But as time goes by, when the system gets a certain amount of users' retrieval histories and operation behavior three days later and builds a relatively comprehensive user interest model, the system is able to analyze users' interests accurately and obtains users' interested vocabularies to gradually improve the recommendation accuracy.

By analyzing Count and the experiment results, it is found that in the early stage, the retrieval times in image retrieval system which uses user-based personalized recommendation technology is almost the same with that in the traditional image retrieval system. However, as more users retrieve, it turns out that the retrieval times of the former system is far less than the latter one which demonstrates that user retrieval history and operation behavior can perfect user interest model and make the recommendation more accurate. By optimizing user interest model, the users can directly get their interested images through the system's recommended images and reduce their retrieval times. Under this circumstance, users' retrieval cost can be reduced and the retrieval efficiency can be improved as well.

5. Conclusion and Prospect

With the rapid development of computer technologies and the internet, the traditional information retrieval technologies can no longer meet the real demands of the users and user-based personalized recommendation technology is gaining more attention from researchers. This paper firstly researches on the background and development situation of image retrieval technology and personalized recommendation technology, and then takes image retrieval as an example to conduct a deeper analysis and research on user-based personalized recommendation technology. This paper has researched the related key technologies and offered some suggestions to improve them.

On the basis of this paper, in-depth researches on the following work still need to be done. Firstly, to research on various information's functions on Web image retrieval. For image information, the image's characteristics should be analyzed and reflected. Moreover, semantic information and image characteristics should be combined to further improve recommendation efficiency. Secondly, user behaviors that this paper covers are just key word retrieval, scanning, downloading and evaluation. However, there are far more behaviors than these, for example, collecting, marking and printing. All these behaviors can somewhat reflect users' interestingness for images, therefore it is necessary to research on information that can reflect users' interests in the Web, for example other operation behaviors. In the end, because the system structure of this paper is relatively simple and the data is not adequate enough, its demand for the system's efficiency is not high. However when encountered a complex image retrieval system which carries a large amount of data, the clustering algorithm this paper employs may not be applicable.

Acknowledgements

This research was supported by the Major International Joint Research Program of the National Natural Science Foundation of China (No.71020107027), the National Natural Science Foundation of China (No 71371130), Project 985 and 211 of Sichuan University, Doctoral Fund of Ministry of Education of China (No 20110181110034), and Social sciences program of Chengdu (No.SKNHR13-07).

References

[1] L. He, J. Zhang and L. S. Shen, "Constructing user preference profile in personalized image retrieval", Computer Engineering and Applications, vol. 45, no. 31, (**2009**), pp. 168-171.
[2] A. Tösher, M. Jahrer and R. Bell, "The BigChaos solution to the Netflix grand prize", Citat a la pàgina, 25, (**2009**).
[3] J. Sill, G. Takács and L. Mackey, *et al*., "Feature-weighted linear stacking", arXiv preprint arXiv:0911.0460, (**2009**).
[4] Y. K. Fang, Y. Fu and J. L. Zhou, "Boosting algorithm for personalized recommendation", Computer Engineering and Applications, Computer Engineering and Applications, , vol. 47, no. 1, (**2011**), pp. 1-4.
[5] G. X. Wang and H. P. Liu, "Survey of personalized recommendation system", Computer Engineering and Applications, vol. 48, no. 7, (**2012**), pp. 66-76.
[6] Q. Wang, L. Y. Yang and D. L. Yang, "Collaborative filtering algorithm based on rating distribution of attributes faced user preference", Journal of Systems Engineering, vol. 25, no. 4, (**2010**), pp. 561-568.
[7] F. Li, J. H. Li and R. L. Wang, *et al.* "Personalized recommendation algorithm based on product features", Computer Engineering and Applications, vol. 43, no. 17, (**2007**), pp. 194- 197.
[8] X. F. Xia, Q. Dai and L. H. Cong, "Personalized recommendation algorithm based on multi-attribute model of attitude of user salient belief", Computer Engineering and Applications, vol. 47, no. 16, (**2011**), pp. 144-146.
[9] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques", Advances in artificial intelligence, (**2009**), 2009: 4.
[10] M. J. Martín-Bautista, D. H. Kraft, M. A. Vila, *et al.*, "User profiles and fuzzy logic for web retrieval issues", Soft computing, vol. 6, no. 5, (**2002**), pp. 365-372.
[11] C. C. Hsu, W. W. Chu and R. K. Taira, "A knowledge-based approach for retrieving images by content", Knowledge and Data Engineering, IEEE Transactions on, vol. 8, no. 4, (**1996**), pp. 522-532.
[12] T. Qin, X. D. Zhang and T. Y. Liu, *et al.*, "An active feedback framework for image retrieval", Pattern Recognition Letters, vol. 29, no. 5, (**2008**), pp. 637-646.
[13] X. Yin, M. Li and L. Zhang, *et al.* "Semantic image clustering using relevance feedback", Proceedings of the 2003 International Symposium on Circuits and Systems, 2003, ISCAS'03, IEEE, (**2003**), 2: II-904-II-907 vol. 2.
[14] L. Candillier, F. Meyer and M. Boullé, "Comparing state-of-the-art collaborative filtering systems", Machine Learning and Data Mining in Pattern Recognition", Springer Berlin Heidelberg, (**2007**), pp. 548-562.
[15] Y. J. Park and K. N. Chang, "Individual and group behavior-based customer profile model for personalized product recommendation", Expert Systems with Applications, vol. 36, no. 2, (**2009**), pp. 1932-1939.
[16] Y. H. Li, "Study of user interest model based on browser log and user behavior", Shanghai, Shanghai Jiao Tong University, (**2008**).
[17] A. Díaz and P. Gervás, "User-model based personalized summarization", Information Processing and Management, vol. 43, no. 6, (**2007**), pp. 1715-1734.
[18] C. S. Cui, Q. Z. Wu and Y. Wang, "Study on user interest level for clustering analysis in recommender systems", Computer Engineering and Applications, (**2011**), vol. 47, no. 7, pp. 226-228.
[19] F. Yu, "Research on personalized recommendation based on search engine", Haerbin, Harbin University of Science and Technology, (**2010**).

[20] L. Yu L and L. Liu, "Research on personalized recommendations in E-business", Computer Integrated Manufacturing Systems, vol. 10, no. 10, (**2004**),pp. 1306-1312.

Authors

Yu Song, He is a postgraduate student in Sichuan University for majoring in management, and a member of Information and Business Management Institute of Sichuan University. His researches mainly relate to information retrieval, electronic commerce, online travel etc. He has taken part in five projects, such as the Major International Joint Research Program of the National Natural Science Foundation of China (No.71020107027), the National Natural Science Foundation of China (No 71371130), Project 985 and 211 of Sichuan University, Doctoral Fund of Ministry of Education of China (No 20110181110034), and Social sciences program of Chengdu (No.SKNHR13-07).

Jingfei Ren, she received her Ph.D. in the Business School of Sichuan University, Chengdu,China, in 2010 and 2013. Her researches mainly relate to management information system, strategy management, service science and supply chain management. In recent years, she has published about ten academic papers at home and abroad, some of which are retrieved by SCI, EI or CSSCI.

Maozhu Jin, he received his M.S. degree in the Department of Electronics Science and Technology and Ph.D. in the Department of Business from University of Huazhong Science and Technology, Wuhan, China, in 2005 and 2008, respectively. In April 2009, he joined Business School of Sichuan University, Chengdu, China. His research interests include multi-objective optimization, game theory, service science and supply chain management.

Peiyu Ren, he is a professor of Sichuan University. In recent years, he has presided over and completed five programs of NSFC. He has published 6 monographs and more than one hundred peer reviewed academic papers. His research interests include multi-objective optimization, game theory, service science and supply chain management.