

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Mining Personalized User Profile Based on Interesting Points and Interesting Vectors

Zeze Wu, Qingtian Zeng and Xiaowen Hu

College of Information Science and Engineering, Shandong University of Science and Technology,
579 Qianwangang Street, Huangdao, Qingdao 266510, China

Abstract: To dig out the implicit meanings in user's multi-behavior sequences, a new approach of mining personalized user profiles is proposed. Firstly, the method is presented to mine user's interesting points and interesting vectors. A user's interesting profile is obtained by combining the interesting point group with interesting vector group together, which is denoted by a weighted directed graph. Then, an algorithm is proposed to calculate the similarity between such user profiles. To verify the effectiveness of the approach proposed in this study, personalized recommendation experiments are realized by using content-based filtering and collaborative filtering, respectively. The results show that the average not acceptance rates of these recommendation services are only 5.94% using content-based filtering recommendation and 3.7% using collaborative filtering. It indicates that the approach proposed in this study is quite available in mining personalized user profiles.

Key words: Personalized information system, user interesting profile, interesting point, interesting vector

INTRODUCTION

One of the most important technologies in the personalized information service system is to mine the personalized user profiles. User profiles mainly describe the characteristics and the relationships of the different users. Content-based filtering and collaborative filtering are the two kinds of most frequently-used approaches in the personalized recommendation system (Wu *et al.*, 2001). Content-based filtering identifies and provides the relevant data for the users based on the similarity between the data and the user profiles. Its advantage is simplicity and availability, but it can hardly find new interests for users (Balabanovic and Shoham, 1997). Collaborative filtering measures the similarity between the user profiles, then identifies the relevant users who own similar profiles and provides the data users interested in. It can find new interests for the users, but its main problem is cold-start-short of user and user's evaluations for the items at the beginning of the recommendation system is used. With the increase of the system users, the problem of the system expansibility becomes more and more serious (Sarwar *et al.*, 2001; Yan-Hong and Gui-Shi, 2008; Ying *et al.*, 2008). Since, both content-based filtering and collaborative filtering approaches are dominated by the user profile, building a high quality of user profile is quite necessary in the recommendation systems.

At present, there are no unified standards for building user profiles. Chun *et al.* (2002) summarized the

denotation and update approaches of different kinds of user profiles and compare several of prototypes systems in user profiles. In addition, they introduced the methods how to collect and update the user's personalized information in detail. Claypool *et al.* (2001) made a deep research on the different kinds of user behaviors to find out which behavior can represent user's interests and which can not. According to the research of Claypool *et al.* (2001), user behaviors such as reading, searching, downloading *et al.* can imply the user's interests, behaviors such as clicking, dragging *et al.* can not imply the interests.

Thompson (2004) designed a personalized conversational recommendation system (Adaptive Place Advisor) to help users choose an item from a large set all of the same basic type. In the Adaptive Place Advisor system, the user profile contains information about preferences for items and item characteristics. In the adaptive place advisor, user's personalized preferences is collected unobtrusively because users often cannot articulate their preferences clearly until they learn more about the domain. And this approach don't need users to participate in.

Guo-Lin *et al.* (2008) explored a 2-layers model to represent the user profiles. The first layer is a vector of categories composed of categories which visited frequently by the user and the second layer is a vector of keywords composed of keywords included in categories in the first layer. Chun *et al.* (2003) recorded the resource

categories visited frequently by the users and adopt the probability of these categories visited by users to represent interesting profiles. This approach can get better effect on describing the variety of user profiles compared with the traditional approaches. Wu *et al.* (2001) recorded the resource categories which are visited frequently and use these categories to represent user's interesting profile. In the work of Wu *et al.* (2001), the 2-category sets are used to represent the user's behavior profile. Finally, they form the user profile combining the interesting profile with the behavior profile. The experiment results show that this approach can get high accuracy in recommendation systems using content-based filtering or using collaborative filtering approach. Middleton *et al.* (2004) explored a novel ontological and describe the user profiles with ontology terms. The advantage of this kind of user profiles is convenient to infer the new interests of the users. Moreover, Middleton *et al.* (2004) show the profiles described by ontology terms to the users themselves. Thus, users can give a feedback according to the profiles shown to them. The approach proposed by Middleton *et al.* (2004) can enhance the accuracy and intelligibility of the user profiles and an external ontology can resolve the cold-start problem to an extent degree.

Although, the existing technologies of mining user profile own their respective advantages, there are two common weaknesses in these approaches:

- Failing in finding the implying meanings in the sequences of the categories or the keywords which appear frequently in the user's log file. For example, when a user enters the recommendation system he/she reads the papers about the Database and the Artificial Intelligence first and then reads the papers of other categories, it implies the user is interested in the knowledge about Database and the Artificial Intelligence
- Failing in considering the user's multi-behaviors, while to mine the user's interests. Most of the traditional mining approaches only consider one kind of behavior, while to mine the user's interests, such as reading behavior (Wu *et al.*, 2001; Middleton *et al.*, 2004; Guang-Qiu and Yong-Mei, 2008). If considering the multi user behaviors together (such as reading papers, answering questions, downloading resources), we can get more complete profiles

In this study, we mainly resolve the above two common weaknesses existed in most of the technologies to mine user profiles. Wu *et al.* (2001) recorded the

resource categories and 2-category sets visited frequently by the users to represent user profiles, which can describe the user interests well. However, the facts are ignored in the work of Wu *et al.* (2001), that the different categories and 2-category sets should have different weights to affect the user interests. Moreover, the approach proposed by Wu *et al.* (2001) failed in finding the implying meanings in the sequences of resource categories visited by users. Based on the idea about user interesting point and interesting vector, this study first extends the model for the user profile proposed by Wu *et al.* (2001) and then proposes a new mining approach for the personalized user profile.

We firstly introduce the approach to mine user interesting points and user interesting vectors from the user's multi-behavior log files. The approaches to compute the weights of each user interesting point and interesting vector are introduced later. Then, we introduce the algorithm to calculate the similarity between different profiles in detail. The mining technology proposed in this study emphasizes the implying meaning in the category sequences. To evaluate the model and the mining approach, QARSS (Question Answer and Resource Sharing System), an experimental platform has been designed and developed to collect the user's multi-behavior log files. The user's multi-behaviors in the QARSS include questioning, answering, browsing, downloading, uploading and ordering. The QARSS collects the information about the resource category sequences visited by the users from their behaviors and these category sequences are used to mine the user's interest points and interest vectors.

AN EXTENDED USER INTERESTING PROFILE

In our new algorithm, when users enter the recommendation information system, all kinds of their behaviors are recorded, such as questioning, answering, browsing, downloading, uploading and ordering. Further more, the category sequences visited by the users can be recorded and collected. These category sequences are named as User's Visit Sequence of Resources (UVSR). For example, $A \rightarrow B \rightarrow C \rightarrow A \rightarrow D \rightarrow C$ is a UVSR, which means that a user visited the content of the resource category A, B, C, A, D, C in order.

In order to mine the user's interests from the UVSRs, the definition of Transaction is introduced first.

Definition 1: A Transaction is a UVSR without any repeated categories.

The transitions can be obtained from the User Resource Visit Sequence (UVSR) using Algorithm 1.

Algorithm 1

Input: a UVSR

Output: Transactions:

- If there are no repeated categories in a UVSR, the UVSR can be regarded as a transaction
- If there are repeated categories in a UVSR, the transactions can be obtained by the following steps
- **Step 1:** Find the first repeated category *c* and split the UVSR into two parts using *c*. The category *c* is split into the second part and the first part can be regarded as a transaction
- **Step 2:** Take out the category *c* and its previous category from the UVSR (i.e., the first part) and regard this 2-category sequence as a transaction
- **Step 3:** The second part split from Step 1 is regarded as a new UVSR. If there are no repeated categories in the second part, the algorithm is over and the second part is also a transaction. Otherwise, go to Step 1

Obviously, a Transaction is a category sequence and there are no repeated categories in a transaction. If the UVSR includes repeated categories, it can be split into several transactions.

For example, one user enters the recommender system twice in a period of time. In the first time, the UVSR is A→C→E and in the second time the UVSR is A→B→C→A→B→A→D→E→F→G. We can see that there are no repeated categories in the first UVSR, so T₁(A→C→E) is a transaction. However, in the second UVSR, the category A and category B are both repeated. According to the definition of transaction, the transactions can be obtained using Algorithm 1: T₂(A→B→C), T₃(C→A), T₄(A→B), T₅(B→A), T₆(A→D→E→F→G).

With the running of the recommendation information system, we can get an aggregate of each user's Transactions from their logs. For example, four Transactions of User 1 are shown in Table 1.

From Table 1, the first appearance of each category (denoted by First(*c*)) and the last appearance (denoted by Last(*c*)) can be recorded and the count of each category appeared in the UVSR can also be recorded (denoted by Count (*c*)). For example, the first time of the category F appears in T₂, so First (F) = 2 and we can see that Count (F) = 2.

The recorded results are shown in Table 2. According to the information in Table 2, Wu *et al.* (2001) proposed an approach to compute the user's interesting degree (Support₁) to each resource category .

Table 1: Four Transactions of user 1

No.	Transactions
1	T ₁ (A→C→E)
2	T ₂ (B→C→E→F)
3	T ₃ (D→E→F)
4	T ₄ (B→C→D→E)

Table 2: Interesting degree of user 1

Category	First	Last	Count.	Support ₁	Support ₂
A	T ₁	T ₁	1	Ignore	Ignore
B	T ₂	T ₄	2	66.7%	61.7%
C	T ₁	T ₄	3	75.0%	72.5%
D	T ₃	T ₄	2	100.0%	95.0%
E	T ₁	T ₄	4	100.0%	100.0%
F	T ₂	T ₃	2	66.7%	61.7%

Definition 2: The user's interesting degree on a resource category can be computed by the Eq. 1 (Wu *et al.*, 2001):

$$\text{Support}_1(c) = \frac{\text{count}(c)}{T - \text{First}(c) + 1} \quad (1)$$

where, *c* is a resource category and T means the number of the current Transactions.

Using Eq. 1, we can set higher interesting degree to all the categories which are frequently visited by the user lately. For example, from Table 2 both category D and category B appear two times in the UVSR, but category D appears two times successively in the latest two Transactions, which means the frequency to visit the category D is more than that of the category B recently. Therefore, we can see that Support₁(D) > Support₁(B) from Table 2.

There exists one problem in Eq. 1. The category D appears two times successively and the category E appears 4 times, however we can see that Support₁(D) = Support₁(E) from Table 2. It indicates that the approach proposed by Wu *et al.* (2001) can not distinguish the interesting degree of such categories appeared successively but their frequencies are different.

To resolve this problem, we propose an advanced approach to compute the user's interesting degree (denoted by Support₂) to a category.

Definition 3: The user's interesting degree of a category can be computed by Eq. 2:

$$\text{Support}_2(c) = \frac{\text{count}(c)}{T - \text{First}(c) + 1} - \lambda \frac{T - \text{count}(c)}{T} \quad (2)$$

where, λ→[0, 1] and the value of λ is in verse radio with T. The interesting degree of User1 is computed using Eq. 1 and 2 are shown in Table 2, respectively, where, λ = 0.1.

From Table 2, it is obvious that Eq. 2 can describe the user's different interesting degree to the different categories in detail.

Table 3: Five transactions

No.	Transactions
1	T ₁ (A→C→E)
2	T ₂ (B→C→E→F)
3	T ₃ (D→E→F)
4	T ₄ (B→C→D→E)
5	T ₅ (F→G)

To ensure we can get the accurate result using Eq. 2, there are another adjustment parameters to be defined before we compute the Support₂ for each resource categories.

Firstly, the following example is presented. From Table 3, T₅ (F→G) is a new transaction and the category G appears for the first time. Using Eq. 2, we can get Support₂(G) = 92%. It means that the category such as G is with a high interesting degree (near to 100%) when such category appears for the first time in the aggregate of transactions. This fact is not reasonable. Thus, a threshold called the minimal count (denoted by r) is introduced to filter out categories with the small counts (Wu *et al.*, 2001). From Table 2, we can see that the category A appears only once, if we set r = 2, Support₁(A) and Support₂(A) can be ignored.

On the other hand, we observe the Support₂ of a category is very low if its first appearance transaction is far from the last Transaction. For example, a category c first appears in T₁ and disappears in T₂ and T₃, but appears continually in T₄ and T₅. Using Eq. 2, its Support₂ is 56% (λ = 0.1). However, the fact that this category appears successively in the recent two Transactions implies that the user is very interested in the resource category recently. Therefore, another threshold called the expired time (denoted by m) is defined to restrict the interval between the First(c) and the Last(c) (Wu *et al.*, 2001). As a category c, if its first appearance and last appearance are T_i and T_j such that j-i ≥ m, then T_i is replaced by T_j. For example, as the example above, if m = 3, 4-1 = 3, therefore T₁ is replaced by T₄. In other words, T₄ is regarded as the first appearance of the category c.

If we set the thresholds λ = 0.1, r = 2 and m = 4, we can compute the Support₂ for category A, B, C, D, E, F and G. The results are shown in Table 4.

Now we introduce the definition of user interesting point, interesting point group, interesting vector and interesting vector group, respectively.

Definition 4: Let support₂(c) be the interesting degree of one user about the category c. c is called an interesting point (denoted by IP) of the user if support₂(c) ≥ α, where α is a given threshold.

Definition 5: Let support₂(c) be the interesting degree of one user about the category c. The interesting point group (denoted by IPG) of the user is:

Table 4: Interesting degree for User1 from the Transactions in Table 3

Category	First	Last	Count	Support ₂
A	T ₁	T ₁	1	Ignore
B	T ₂	T ₄	2	44.0%
C	T ₁	T ₄	3	56.0%
D	T ₃	T ₄	2	60.7%
E	T ₁	T ₄	4	78.0%
F	T ₂	T ₃	3	71.0%
G	T ₅	T ₅	1	Ignore

Table 5: 2-category set supports of five Transactions for User 1

2-category set	First	Last	Count	Support ₂
A→C	T ₁	T ₁	1	Ignore
C→E	T ₁	T ₂	2	34.0
B→C	T ₂	T ₄	2	44.0
E→F	T ₂	T ₃	2	44.0
D→E	T ₃	T ₄	2	60.7
C→D	T ₄	T ₄	1	Ignore
F→G	T ₅	T ₄	1	Ignore

$$IPG = \{c | Support_2(c) \geq \alpha, c \in Category\}$$

where, α is a given threshold and Category is the set of all the categories in the system.

Obviously, IPG ⊆ Category. For example, in Table 4, the interesting point group of User 1 is IPG_{User1} = {C, D, E, F} if α = 50%.

Before giving the definition of the interesting vector and interesting vector group, we introduce the 2-category set first. For example, from Table 3, T₂ (B→C→E→F) can be split to three 2-category sets B→C, C→E and E→F. Obviously, T₁, T₃, T₄ can be split to several 2-category sets in the same way. Therefore, we can compute the interesting degree to 2-category sets of a user by Eq. 2, denoted by Support₂(x→y), where, x, y ∈ Category. The results are shown in Table 5.

Definition 6: Let Support₂(x→y) denote a user's interesting degree to the 2-category set (x→y). If Support₂(x→y) ≥ β where, β is a given threshold, the 2-category set (x→y) is called an interesting vector (denoted by IV) of the user.

Definition 7: Let Support₂(x→y) denote the support of the 2-category set (x→y). Given a threshold β, the interesting vector group of a user (denoted by IVG) is defined as follows:

$$IVG = \{x \rightarrow y | Support(x \rightarrow y) \geq \beta, x \rightarrow y \in Vector\}$$

where, Vector represents all the 2-category sets of the user.

Obviously, IVG ⊆ Vector. For example, if we set B = 40%, from Table 5, the interesting vector group of User 1 is shown as follows:

$$IVG_{User1} = \{B \rightarrow C, E \rightarrow F, D \rightarrow E\}$$

Finally, we give the definition of the user profile based on the user interesting point group and user interesting vector group.

Definition 8: A user interesting profile based on IPG and IVG can be defined as a weighted directed graph $G = (V, E, W)$, where (1) $V = V_1 \cup V_2$, $V_1 \subseteq \text{Category}$ and $V_2 \subseteq \text{Category}$ where, Category represents all the resource set. V_1 represents the IPG of the user and V_2 represents the categories in the IVG; (2) $E \subseteq V_2 \times V_2$ represents the IVG of the user; (3) $W = F \cup g$, represents the function of the interesting degree about the interesting points and the interesting vectors of a user, in which, (3.1) $f: V_1 \rightarrow [0, 1]$ represents the function of the interesting degree to each interesting point. For each $v \in V_1$, $f(v) \geq \alpha$ where, α is the threshold to filter the interesting points. (3.2) $g: E \rightarrow [0, 1]$ represents the function of the interesting degree about each interesting vector. For each $e \in E$, $g(e) \geq \beta$ where, β is the threshold to filter the interesting vectors. For example, the weighted directed graph for the user's interesting profile of User1 is shown in Fig. 1. In the graph, to distinguish the nodes in IPG and IVG, the rectangle nodes are used to represent the interesting vectors and the circular nodes are used to represent the interesting points.

USER PROFILE SIMILARITY COMPUTING

Here, we propose a user profile based on user interesting points and user interesting vectors. In this section, the approach is proposed to calculate the similarity between the different user interesting profiles. To calculate the similarity, there are three main steps: (1) to calculate the similarity between user interesting point groups; (2) to calculate the similarity between user interesting vector groups and (3) to calculate the similarity between user interesting profiles.

The similarity between user interesting point groups: In most of the information recommendation systems, all the information resources have been already classified strictly and the number of the information categories is fixed in a period of time. Thus, the cosine similarity is used to calculate the similarity between different user interesting point groups. First, the vector of the interesting point group is defined.

Definition 9: Let $\text{Category} = \{c_1, c_2, c_3, \dots, c_n\}$ be the information resource categories in the recommendation system and $\text{IPG}_x \subseteq \text{Category}$ be the interesting point group of one user User_x . The vector of IPG_x can be defined as:

$$\text{VIPG}_x = \langle w(c_1), w(c_2), w(c_3), \dots, w(c_n) \rangle$$

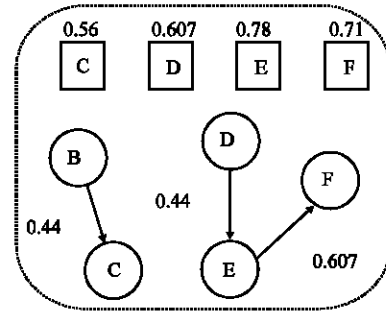


Fig. 1: Interesting profile of User1

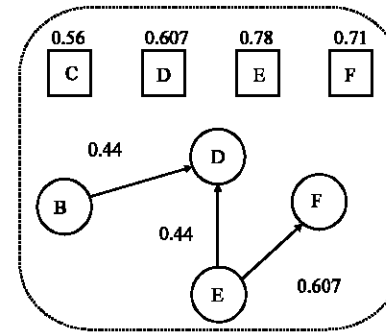


Fig. 2: Interesting profile of User2

Where:

$$w(c_n) = \begin{cases} f(v_n) & v_n \in \text{IPG}_x \\ 0 & \text{otherwise} \end{cases}$$

and $f: \text{IPG}_x \rightarrow [0, 1]$ is the function of the interesting degree about the interesting points.

For example, assuming that the information resources are $\text{Category} = \{A, B, C, D, E, F, G\}$ and let $\text{IPG}_{\text{User1}}$ represent the interesting point group of User1. According to Fig. 1, the vector of $\text{VIPG}_{\text{User1}}$ is $\text{VIPG}_{\text{User1}} = \langle 0, 0, 0.56, 0.607, 0.78, 0.71, 0 \rangle$.

Definition 10: Let IPG_x and IPG_y be the interesting point group of user x and user y , respectively. The similarity between IPG_x and IPG_y can be calculated by the following equation:

$$\text{Sim}(\text{IPG}_x, \text{IPG}_y) = \frac{\text{VIPG}_x \bullet \text{VIPG}_y}{|\text{VIPG}_x| |\text{VIPG}_y|} \tag{3}$$

where, VIPG_x and VIPG_y represents the vectors of IPG_x and IPG_y , respectively.

For example, the interesting profile of User2 is shown in Fig. 2. Assuming the information resources are $\text{Category} = \{A, B, C, D, E, F, G\}$. According to Definition 9, the vector of the interesting point group of User2 is $\text{VIPG}_{\text{User2}} = \langle 0, 0.56, 0, 0.607, 0.78, 0.71, 0 \rangle$.

$$\begin{aligned} \text{Sim}(IvA, IvB^T) &= \frac{IvA \bullet IvB^T}{|IvA||IvB^T|} \\ &= \frac{0.44 \times 0.44}{\sqrt{0.44^2 + 0.44^2 + 0.607^2} \bullet \sqrt{0.44^2 + 0.44^2 + 0.607^2}} \\ &\approx 0.256 \end{aligned}$$

Thus, if $\chi = 0.5$, the similarity of the interesting vector group between User 1 and User 2 is:

$$\begin{aligned} \text{Sim}(IVGx, IVGy) &= \\ \text{Sim}(IvA, IvB) + 0.5\text{Sim}(IvA, IvB^T) &= \\ = 0.616 \end{aligned}$$

The similarity between user interesting profiles: Based on the similarity between user’s interesting point group and the similarity between the interesting vector group, the approach is proposed to calculate the similarity between user interesting profiles.

Definition 14: The similarity of the interesting profiles between user x and user y is calculated by the following Eq. 5:

$$\begin{aligned} \text{Sim}(\text{User}_x, \text{User}_y) &= \\ = \kappa_1 \text{Sim}(\text{IPG}_x, \text{IPG}_y) + \kappa_2 \text{Sim}(IVG_x, IVG_y) \end{aligned} \quad (5)$$

where, κ_1 and κ_2 are regulation parameters such that $\kappa_1 + \kappa_2 = 1$.

For example, if $\kappa_1 = 0.5$ and $\kappa_2 = 0.5$, the similarity of the interesting profiles between User1 and 2 is:

$$\begin{aligned} \text{Sim}(\text{User}_1, \text{User}_2) &= \\ = \kappa_1 \text{Sim}(\text{IPG}_{\text{User}_1}, \text{IPG}_{\text{User}_2}) + \kappa_2 \text{Sim}(IVG_{\text{User}_1}, IVG_{\text{User}_2}) &= \\ = 0.5 \times 0.825 + 0.5 \times 0.616 = 0.7205 \end{aligned}$$

EVALUATION OF THE USER PROFILE

QARSS-the prototype system for personalized information recommendation: In order to verify the user profile proposed in this study, QARSS (Question Answer and Resource Sharing System) is designed and developed, which is convenient for users to learn knowledge and share resource online. The framework of the QARSS is shown in Fig. 4, which is composed by two kernel systems: one is a multi-user interactive QA system and the other is a multi-user resource sharing system:

- Multi-user interactive QA System. In the QA system, users can ask questions, answer questions and gather or order questions which they are interested in. If one user has questions, he/she can login the QA system to post the questions. Other users can browse or answer the questions they are

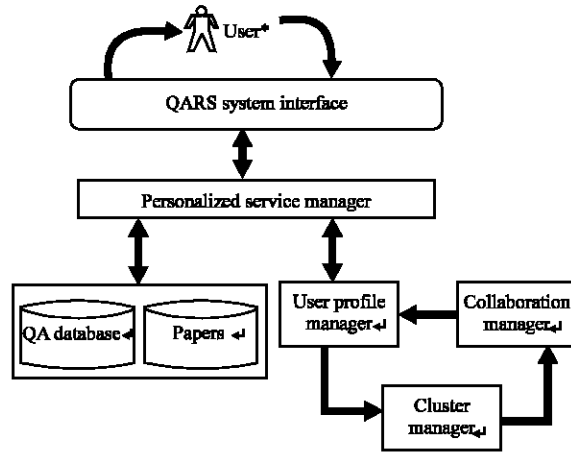


Fig. 4: The Structure of QARSS

interested in. The users can also manage their personalized information and to find friends with similar interests. In addition, users can also see the operation results and search questions in the QA system

- Multi-User Resource Sharing System. In resource sharing system, users can browse, download and order their interested resources (including studies and questions), or upload the resources to the relative resource classification of the system. Same as the structure of the QA system, the resource sharing system also has a personalized interface, in which users can manage the resources they have ordered, resources recommended by their friends, resources recommended by the recommendation system automatically. In addition, users is permitted to order resources, download resources, upload resources and browse resources which they are interested in conventionally through clicking the hyperlink

Most of the traditional recommendation systems merely consider one kind of user behaviors to mine the user’s interests, for example, OTS merely considers the behavior about reading (Wu *et al.*, 2001). However, the user profiles mining though a single behavior is not enough to describe user’s interests well. Different behaviors may imply different meanings in real life. For example, users browsing the papers maybe mean they need get some information from the papers. If they answer a question in the QA system maybe mean that they have ability to answer this kind of questions. In other words, they master something well in this domain. Therefore, multi-behaviors have stronger ability to describe user’s interests.

The purpose of integrating a QA system and a resource sharing platform is to collect browsing, reading, answering questions, uploading and downloading resources and other different behaviors and the sequences of categories visited from such behaviors are recorded to mine the user profiles. The QARSS applies both content-based filtering and collaborative filtering to recommend questions and resources to users according to their profiles.

Design and analysis of experiment: To estimate the accuracy of the user profiles based on interesting points and interesting vectors, 97 users who login the QARSS frequently are invited to make comments on the quality of the recommendation. The 97 users are divided into 10 groups according to their original information (including user's major, age *et al.*) submitted by themselves. The three grades of comments are good, moderate and bad (Wu *et al.*, 2001). Firstly, the related parameters are estimated rationally before the experiment:

- Parameter λ is an inverse ratio with the numbers of the transactions. The number of transactions gathered in the experiment is about 100, therefore we set $\lambda = 0.1$
- According to the number of the transactions gathered in the experiment, we set $m = 20$ and $r = 10$
- The thresholds are set as $\alpha = 30\%$ and $\beta = 20\%$, respectively

In this study, we design two parts of different experiments to verify the user profile using content-based filtering and collaborative filtering recommendation, respectively. In the experiments, we collect and classify all the questions and resources (mainly papers) into 16 categories according to different subjects and each category has its own detailed description and keywords collection.

Design and analysis of content-based filtering recommendation experiment: The process of the recommendation experiment based on content-based filtering is simple. To judge whether a question or a paper should be recommended to a user or not, it only needs to check whether the category of the question or the paper belongs to the user's interesting point group, because all the resources in the QARSS are classified strictly. The visited resources are not permitted to be recommended to the same user again. If there are too many resources waiting to be recommended in a user's recommendation table, the top-n resources visited more frequently are first recommended to the user.

The results of the content-based filtering recommendation experiment are shown in Table 6. From

Table 6: Results of the content-based filtering recommendation experiment

10 group users	Comments (%)		
	Good	Moderate	Bad
1	39.30	53.70	7.00
2	44.10	51.00	4.90
3	32.50	61.30	6.20
4	24.60	66.20	9.20
5	41.60	52.10	6.30
6	46.50	46.80	6.70
7	53.10	43.70	3.20
8	42.40	49.40	8.20
9	37.80	59.10	3.10
10	38.30	57.10	4.60
Average	40.02	54.04	5.94

Table 6, we can see the comments from the 10 group users. The average of bad-rate is only 5.94%. In the study of Wu *et al.* (2001), the similar experiments are realized and the average of the bad-rates ranges from 5.4 to 28.5%. It means that the proposed approach to mine user profiles is feasible in the content-based recommendation system.

Design and analysis of collaborative filtering recommendation experiment: In the collaborative filtering recommendation experiment, the results of the recommendations might be different if we set different values to the three parameters (χ , κ_1 and κ_2) and the values of κ_1 and κ_2 should play more important roles in finding the similar users. Therefore, there are 18 experiments implemented in two parts. For the first part, we set different values to parameters κ_1 and κ_2 , while keep χ invariable in the first nine experiments. When the first nine experiments are completed, we pick out the best combination of the value of κ_1 and κ_2 that can get the highest recommendation accuracy-rate. For the second part (the second nine experiments), we set different values to the parameter χ while keep the values of κ_1 and κ_2 in the best combination picked out from the first nine experiments. From the second part of the experiments, we can also pick out the best combination of the value of κ_1 , κ_2 and χ . The results of the experiment are shown in the Table 7.

For the first part of experiments, the highest recommendation accuracy-rate is obtained when the values of the κ_1 and κ_2 are set 0.7 and 0.3, respectively. The result indicates it is hard to form the high similarity interesting vectors in the QARSS if the number of users is not large. Therefore, the low recommendation accuracy-rate is obtained if a high weight is set to the similarity of interesting vectors when we calculate the similarity between user profiles.

In the second part of the experiment, we set the values of κ_1 and κ_2 are 0.7 and 0.3, respectively. The highest recommendation accuracy-rate is obtained when the value of the χ is set as 0.7. This result means two users with the reverse sequence of the interesting vectors are not absolutely the same. In other words, the

Table 7: Results of the collaborative filtering recommendation experiment

Experiment Sequence	Parameters			Comments		
	κ_1	κ_2	κ_3	Good	Moderate	Bad

The first nine experiments						
1	0.1	0.9	0.6	14.80	53.10	32.10
2	0.2	0.8	0.6	14.40	58.70	26.90
3	0.3	0.7	0.6	22.30	58.10	19.60
4	0.4	0.6	0.6	26.50	55.30	18.20
5	0.5	0.5	0.6	31.40	58.70	9.90
6	0.6	0.4	0.6	37.20	58.30	4.50
7	0.7	0.3	0.6	40.50	55.40	4.10
8	0.8	0.2	0.6	39.20	54.90	5.90
9	0.9	0.1	0.6	36.90	56.10	7.00
Average				29.24	56.51	14.24
The second nine experiments						
10	0.7	0.3	0.1	29.10	60.10	10.80
11	0.7	0.3	0.2	30.70	61.20	8.10
12	0.7	0.3	0.3	30.80	62.10	7.10
13	0.7	0.3	0.4	35.40	57.80	6.80
14	0.7	0.3	0.5	39.20	56.10	4.70
15	0.7	0.3	0.6	41.10	55.10	3.80
16	0.7	0.3	0.7	43.10	53.20	3.70
17	0.7	0.3	0.8	38.90	55.90	5.20
18	0.7	0.3	0.9	36.80	57.10	6.10
Average				36.12	57.62	6.26

interesting vector ($x \rightarrow y$) is not completely same with the vector ($y \rightarrow x$), but the two vectors have a high similarity to each other.

For example, the interests of the users always read the papers about database firstly and then they read papers about AI (artificial intelligence) are not completely same to that of the users who always read papers about AI firstly and then read papers about database, but these two kinds of users have extremely similar interests.

From the experiments, we can see that the bad rate is only 3.7% in the collaborative filtering recommendation if we set the right values to κ_1 , κ_2 and χ . Therefore, the proposed approach in this study is also effective in the collaborative filtering recommendation system.

CONCLUSION

In this study, we discuss the advantages and disadvantages of the traditional technologies in mining the user interesting profiles. An improved approach to mine user interesting profiles is proposed to overcome the disadvantages of the existing approaches. Moreover, an effectual method is presented to calculate the similarity between two different user profiles. QARSS, a prototype system for personalized information recommendation based on user's multi-behaviors, is built to collect the interesting sequences of the users and to verify the accuracy of the user interesting profiles based on IPG and IVG. The experiment results show that the user interesting profile mining approach proposed in this study is adapt to recommendation systems using content-based filtering or collaborative filtering.

ACKNOWLEDGMENT

This research is supported by the National Science Foundation of China under Grant (No. 60603090, 90718011 and 50875158), the Research Foundation of Shandong Educational Committee under Grant (No. J08LJ77) and the Taishan Scholar Program of Shandong Province of China.

REFERENCES

Balabanovic, M. and Y. Shoham, 1997. Content-based collaborative filtering recommendation. *Commun. ACM*, 40: 66-72.

Chun, Z., X. Chun-Xiao and Z. Li-Zhu, 2002. A survey of personalization technology. *J. Software*, 10: 1952-1961.

Chun, Z., X. Chun-Xiao and Z. Li-Zhu, 2003. A Personalized search algorithm by using content-based filtering. *J Software*, 14: 999-1004.

Claypool, M., P.L.M. Waseda and D. Brown, 2001. Implicit interest indicators. *Proceedings of the ACM Intelligent User Interfaces Conference (IUI)*, Jan. 14-17, Santa Fe, New Mexico, USA., pp: 14-17.

Guang-Qiu, H. and Z. Yong-Mei, 2008. Approach to collaborative flittering recommendation based on HMM. *Comput. Appl.*, 28: 1601-1604.

Guo-Lin, P.U., Y. Qing-Ping and W.G.Q. Yu-Hui, 2008. Personalized model of user interests based on semantics. *Comput. Sci.*, 35: 181-184.

Middleton, S.E., N.R. Shadbolt and D.C. Deroure, 2004. Ontological user profiling in recommender systems. *ACM Trans. Inform. Syst.*, 22: 54-88.

Sarwar, B., G. Karypis, J. Konstan and J. Reidl, 2001. Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th International World Wide Web Conference*, May 1-5, Hong Kong, pp: 285-295.

Thompson, C.A., M.H. Göker and P. Langley, 2004. A personalized system for conversational recommendations. *J. Artificial Intell. Res.*, 21: 393-428.

Wu, Y.H., Y.C. Chen and A.L.P. Chen, 2001. Enabling personalized recommendation on the web based on user interests and behaviors. *Proceedings of the 11th International Workshop on Research Issues in Data Engineering*, Apr. 1-2, IEEE Xplore, London, pp: 17-17.

Yan-Hong, G. and D. Gui-Shi, 2008. Improved personalized recommendation algorithm in collaborative filtering. *Appl. Res. Comput.*, 25: 39-41.

Ying, G., Q. Hong, L. Jie and L. Dayou, 2008. A collaborative filtering recommendation algorithm combining probabilistic relational models and user grade. *J. Comput. Res. Dev.*, 45: 1463-1469.