Courseware Recommendation in E-Learning System

Ge Liang, Kong Weining, and Luo Junzhou

School of Computer Science and Engineering, Southeast University, Nanjing, P.R. China
heavenbuyer@seu.edu.cn, kweining@21cn.com, jluo@seu.edu.cn

Abstract. E-learning systems , as an education pattern, are becoming more and more popular. In e-learning systems, courseware management is an indispensable part. As the number of various courseware increases, how to find the courseware or learning materials that are most suitable to users and users of e-learning systems are most interested in is a practical problem. In this paper, we apply the idea of knowledge discovery techniques to make personalized recommendation for courseware. We design the courseware recommendation algorithm which combines contents filtering and collaborative filtering techniques. Also we propose the architecture of courseware management system with courseware recommendation, which is seamlessly integrated in our E-learning system. The experiment shows that our algorithm is able to truly reflect users' interests with high efficiency.

Keywords: E-learning system, courseware management, recommendation, interests.

1 Introduction

As a novel education pattern, E-learning, which characterizes the huge information, the strong interactivity, the great coverage and no space-time restrictions [1][2], has been an important way to resolve the contradictions between the social needs and the relatively inadequate educational resources.

Among various elements in different E-learning systems, courseware management is being attached great importance. Courseware, according to the CELTS (China E-learning Technology Standards)-42[3], is defined as a program that implements a complete teaching process of one or several knowledge points. The conventional methodology of managing courseware is to store the courseware in databases and offer an interface for users to search and retrieve. However, as the amount of courseware is becoming larger and larger, how to find the courseware that users are most interested in and are most valuable to them is a problem faced in nearly every E-learning system. Different users may have different focus on courseware. As for this, inspired by knowledge discovery techniques in [6][7], we propose the idea of courseware recommendation, which recommends courseware that most reflects the true interests of different users. Courseware recommendation system collects and analyzes users' information and behavior in the e-learning system to figure out their interests and then makes an active recommendation on courseware.

About the various recommendation technologies that courseware recommendation system is based on, they are generally categorized into two kinds: recommendation

W. Liu, Q. Li, and R.W.H. Lau (Eds.): ICWL 2006, LNCS 4181, pp. 10 – 24, 2006.

based on rules and recommendation based on information filtering. The latter is furthermore divided into two types: filtering based on user contents and collaborative filtering.

Rule filtering recommendation systems use the predefined rules to filter information. The major problem is that the quality of rules cannot be guaranteed and rules cannot be updated dynamically. Moreover, when the number of rules increases over some extent, the system becomes overwhelmingly hard to manage. User contents filtering systems take considerations of the diversity of different users. They analyze the interests and features of users. As a result, the recommendation results are truly relevant to users. The precision and quality of recommendation is fairly good, but the user contents filtering only focus on a single user so that the results cannot be shared among other uses and reflect the comprehensive picture of user interests. Collaborative filtering works by building a database of preferences for items by users. A new user, Neo, for example, is matched against the database to discover neighbors, which are other users who have historically had similar taste to Neo. Items that the neighbors like are then recommended to Neo, as he will probably also like them. Although collaborative filtering is a promising technology, it has two fundamental drawbacks. One is sparsity problem, which is that ,at the very beginning of the system, recommendation system is unable to recommend anything because not enough evaluations are available in the system. The other problem is scalability of the collaborative filtering. As the users and resources increase, the system performance tends to decrease to some extent.

In the paper, we design algorithms to make the correct and effective recommendation of courseware to users. And we propose the courseware management architecture with courseware recommendation that combines the user contents filtering and collaborative filtering. The system has implemented the algorithms. The whole scheme has the following unique features:

- Recommendation algorithm: The algorithms combine the two kind of filtering technology. They not only maintain the precision and quality of recommendation, but also give comprehensive recommendations in view of the users' interests. Furthermore, the algorithms also consider the relations among courseware to improve the quality of recommendation.
- Integration: The whole courseware recommendation architecture and its implementation are smoothly integrated into our E-learning systems, which is the key project of China's Tenth Five-year plan: E-learning key technology and its demonstration. It goes well with other components of the E-learning system: the real-time teaching system, the non-real-time teaching system, the network teaching management system, the e-learning settlement system, the courseware making and intelligent question-answering system, the exercise and examination management system, and the educational resources management system.

The rest of this paper is organized as follows. Section 2 discusses the related work. Section 3 describes the architecture of the courseware recommendation. Section 4 gives the user aggregation algorithm. Section 5 talks about the courseware aggregation. Section 6 gives the detail flow of courseware recommendation algorithm. Section 7 presents some experiment results to show the benefit of the system and an instance of the system implementation. The last section makes a conclusion and talks about future work.

2 Related Work

Since the research on the courseware recommendation is relatively rare, here we briefly present some of the research literature related to collaborative filtering, recommender systems, data mining and personalization.

Tapestry [7] is one of the earliest implementations of collaborative filtering-based recommender systems. This system relied on the explicit opinions of people from a close-knit community, such as an office workgroup. However, recommender system for large communities cannot depend on each person knowing the others. Later, several ratings-based automated recommender systems were developed [8][9][10][11].

Other technologies have also been applied to recommender systems, including Bayesian networks, a Bayesian networks create a model based on a training set with a decision tree at each node and edges representing user information. The model can be built off-line over a matter of hours or days. The resulting model is very small, very fast, and essentially as accurate as nearest neighbor methods [12]. Bayesian networks may prove practical for environments in which knowledge of user preferences changes slowly with respect to the time needed to build the model but are not suitable for environments in which user preference models must be updated rapidly or frequently.

Clustering techniques work by identifying groups of users who appear to have similar preferences. Once the clusters are created, predictions for an individual can be made by averaging the opinions of the other users in that cluster. Some clustering techniques represent each user with partial participation in several clusters. The prediction is then an average across the clusters, weighted by degree of participation. Clustering techniques usually produce less-personal recommendations than other methods, and in some cases, the clusters have worse accuracy than nearest neighbor algorithms.

3 The Architecture of Courseware System with Courseware Recommendation

The architecture of Courseware management system with courseware recommendation is as the Figure 1 suggests. The whole architecture is composed of portal of e-learning system, e-learning login server, portal of courseware management system, courseware search engine, courseware database, courseware metadata database and the courseware recommendation module.

Normally, a typical process of using courseware management system is like this: A user logins on the portal of our e-learning system. After authenticated by the login server, the user enters the e-learning system. And then he visits the portal of courseware management system. On the portal, he inputs some key words to look for the courseware he wants learn. Then he will see some recommendation information about courseware. The information is generated in the Courseware Recommendation Module. There are two kind of recommendation information: ① The Top 5 most popular courseware in the user's major and degree.② The Top 5 courseware that other users in the same interest group of the user recommend. After getting these information, the user can choose to learn the courseware that the system recommends or search the courseware that he is really interested in. And the search engine will search the courseware database based on the key words the user just input. The search process is accelerated by the Courseware Metadata database which is built according to the CELTS-3,42.

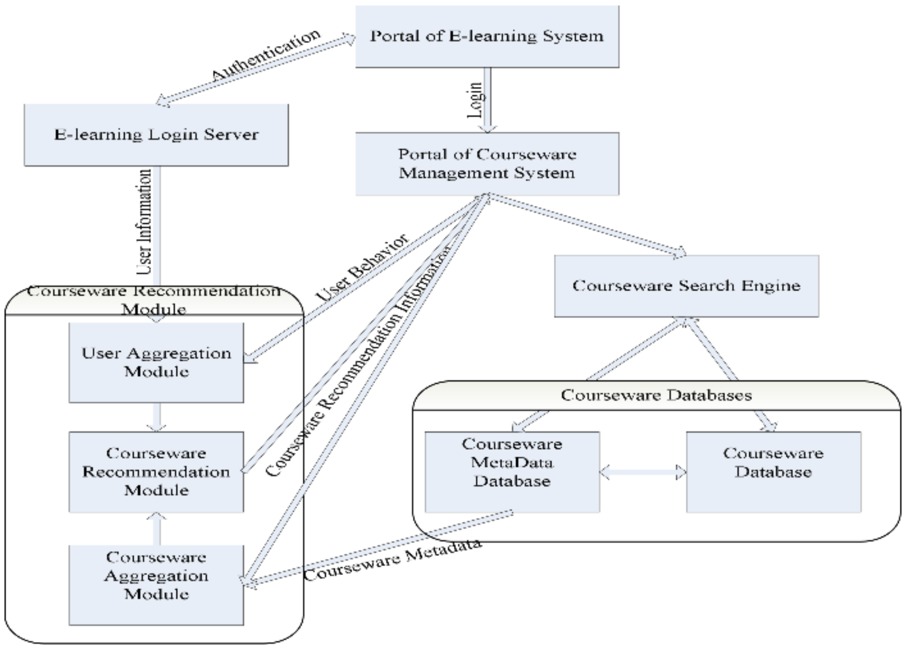

Fig. 1. The Architecture of Courseware System with Courseware Recommendation

The courseware recommendation module is composed of three parts: User Aggregation Module, Courseware Aggregation Module, and Courseware Recommendation Module. User Aggregation Module mainly collects user information from e-learning login server and user behavior from portal of Courseware Management System. After these data have been collected, the module uses User Aggregation Algorithm to figure out users' interests group, user information similarity and user credits. Courseware Aggregation Module takes data from Courseware Metadata Database to make a sort of courseware based on their evaluation by users. These sorting results are handed to Courseware Recommendation Module. Courseware Recommendation Module receives inputs from the two modules and implements Courseware Recommendation Algorithm to obtain the recommendation information about courseware. Then those information are transferred to portal of Courseware Management System.

4 User Aggregation Algorithm

The main function of user aggregation module is to implement user aggregation algorithm. The algorithm forms user aggregation which is the user group with similar interests on courseware. The module also calculates two values as the parameters of the courseware recommendation algorithm: Users Credit and User Information Similarity.

The module obtains user profile information from e-learning login server, courseware keywords and the evaluations of courseware by users from portal of courseware management system. The module also maintains a dataset of user login history and its relevant courseware keywords and evaluations. This dataset offers information for recommendation algorithm.

4.1 Some Definitions for User Aggregation Algorithm

Some definitions of the user aggregation algorithm:

- ***User_Info:*** According to Data Exchange Standard of our E-learning system[4], the user profile information is defined as follows: User_Info={user_id, user_name, name, type, school_id, major, class, station, e-mail, gender, national, language, telephone, address, postcode, birthplace, birthday, preschool, premajor, degree, description}
- ***User_Courseware Associate Matrix:*** Suppose there are m users in the list maintained by user aggregation module. And we pick up n different courseware key words.(Each user has at least one keywords). We create the User_Courseware Matrix $M_{m\times n}$. Each line vector M[i.] denotes the evaluations user u_i comments on all the n different courseware keywords. Each column vector M[.j] represents evaluations from all the users. And the element M_{ij} tells about what user u_i comments on courseware keywords.
- ***User Information Associate Matrix:*** In the same user group with similar interests about courseware, we maintain a matrix to associate users and its user_Info: $I_{m\times 9}$. Each line vector I[i.] denotes all the user profile information of user u_i . Each column vector I[.j] represents the j-th information of all users. And the element I_{ij} is the j-th information of user u_i .
- ***User Credit:*** user credit is defined as the degree of similarity between users about courseware. We define Credit(u_i ,u_j) as the user credit between user u_i and user u_j .

User Information Similarity: we define userinfo_sim (u_i , u_j) to denote the similarity between users about their profile information.

4.2 User Aggregation Algorithm

The flow of the algorithm is as follows:

(1) Calculate the similarity matrix of users: $M_{m\times n}^{Sim}$. We calculate $M_{m\times n}^{Sim}$ based on the User_Courseware Associate Matrix $M_{m\times n}$. The element of $M_{m\times n}^{Sim}$ is $Sim(i, j)$ .The formula (1) shows the calculation. In the formula,

$Sim(i, j)$ stands for the degree of similarity between user u_i and u_j about courseware. And $\overline{M_k}$ denotes the average evaluation that all users give to the k-th courseware keywords.

$$Sim(i, j) = \frac{\sum_{k=1}^{m}(M_{ik} - \overline{M_k})(M_{jk} - \overline{M_k})}{\sqrt{\sum_{k=1}^{m}(M_{ik} - \overline{M_k})^2}\sqrt{\sum_{k=1}^{m}(M_{jk} - \overline{M_k})^2}} \quad (1)$$

(2) User Aggregation: giving a threshold value ε. $\forall Sim(i, j)(n \le i, j \le m)$, if $Sim(i, j) \ge \varepsilon$, then we can make user u_i and all the other user u_j in one interest group. Apparently, one user can be grouped into more than one group. The threshold value ε is adjustable. If the user aggregation number is too large, we can increase the value of ε, on the contrary, we can decrease ε to increase number of user interest group.

(3) Calculate User Credit: After we obtain user aggregation group and degree of user interests similarity, we can compute User Credit using Formula(2):

$$Credit(u_i, u_j) = Sim(u_i, u_j) / \max Sim \quad (2)$$

5 Courseware Aggregation

Courseware aggregation module is mainly in charge of grouping courseware according to their metadata and ranking the courseware in each group. And then the module sorts the grouped courseware based on the rating that users give when using these courseware.

As the Figure 1 demonstrates, courseware aggregation module obtains information from Courseware Metadata Database. In our integrated E-learning system, every courseware must comply with CELTS-3 and CELTS-42. Also the CELTS-3 and CELTS-42 are compatible with Dublin Core Metadata Initiative [5]. XML is used to describe courseware. Figure 2 shows part of the XML file of a certain courseware. Primarily based on these metadata standards, we group those courseware.

According to CELTS-3 and CELTS-42, the courseware metadata have some mandatory elements: title, creator, subject, keywords, description, data, type, format, identifier, language, audience. Not every element can be the criteria of grouping courseware. Here we pick two elements that are mostly associated with users' interests. They are subject and audience. Subject is what courseware contents are about. The subject may be mathematics, computer science, etc. Audience is the degree of users. The audience may be bachelor, master, doctor, freshmen, sophomore, etc.

After grouping the metadata of courseware, the module will obtain information about the rating of the courseware and use it as the criteria to rate the various courseware in each groups. And the top 5 highest rating courseware metadata will be sent to the recommendation module to be recommended to the user.

```xml
<?xml version="1.0"?>
<ResourceEncode xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          <Title>物理简史 </Title>
          <Creator>
  <Namepersonal>葛亮 </Namepersonal>
  <namecoporate>东南大学 </namecoporate>
  <postal>210096</postal>
  <email>heavenbuyer@seu.edu.cn</email>
 </Creator>
 <Subject>history</Subject>
 <Keywords>history</Keywords>
  <Description>
  <tableOfContent>物理简史 </tableOfContent>
 </Description>
 <Contributor>
 <Namepersonal></Namepersonal>
 <namecorporate></namecorporate>
```

Fig. 2. Part of the XML Description of a Courseware

The rating of courseware in our E-learning system is based on the users' behavior when they are in courseware system. Different behaviors imply different attitudes of users toward courseware and therefore can be deemed as criteria of rating courseware. The user action and its meaning are shown in Table 1. The rating does not reflect the quality of the courseware but the interests of different users.

Table 1. Meaning of the User Action

User Action	Meaning	Rating value
Watch the courseware and add a mark	Very high positive	5
Watch the whole of the courseware	High positive	4
Watch part of the courseware	Moderate positive	2
Ignore the courseware	Low negative or set to zero	0

6 Courseware Recommendation Algorithm

As the core part of the Courseware Recommendation system, this module basically outputs two kind of information to the portal of courseware system: ① The Top 5 most popular courseware in the user's major and degree. ② The Top 5 courseware that other users in the same interest group of the user recommend.

The first kind of recommendation information is quite straightforward. Courseware recommendation module receives information from the courseware aggregation module about the top 5 most popular courseware in the current user's major and degree (bachelor, master, doctor). The second recommendation information is generated by the courseware recommendation algorithms.

The recommendation algorithm is to generate the second information, which recommends the courseware that is recommended by other users in the same interest group. The algorithm needs two parameters: one is the user credit, which is calculated in formula (2); the other is user information similarity degree, which is calculated in section 6.1.

6.1 User Information Similarity Degree

When users login in e-learning system, the users' profile information are transferred to user aggregation module to calculate user information similarity degree: userinfo_sim(u_i , u_j). Although there are 21 information elements in the user profile, we only pick 9 elements to calculate the similarity degree. The detail element similarity calculation is like the Table 2 suggest.

Table 2. The Element Similarity Degree Calculation

Element	Condition	Similarity Degree
School_ID	same school	1
	different school	0
Major	same major	1
	different major but same field(Arts Science)	0.5
	different major, different field	0
class	same class	1
	different class but same major same grade	0.7
	different class, different major	0
station	same station	1
	different station	0
gender	same gender	1
	different gender	0
birthday	same year	1
	older or younger x year	1-0.1*x
	older or younger 10 years or more	0
working place	Same place	1
	different place	0
Premajor	same major	1
	different major	0
Type	Same type	1
	higher or lower x in type series(*)	1-0.4*x
	higher or lower 3 in type series	0

*Type series: elementary school, junior high school, senior high school, bachelor, master, doctor.

Using Table 2, userinfo_sim can be calculated in formula (3):

$$\text{userinfo\_sim}(u_i, u_j).=\sum_{k=1}^{9}\{\omega_k \times sim(u_i\_Elem_k, u_j\_Elem_k)\} \quad (3)$$

$u_i\_Elem_k$ stands for user u_i information element k of Table 2. $sim(u_i\_Elem_k, u_j\_Elem_k)$ is the similarity degree of user u_i's element k and user u_j ω_k represents the weight of k-th element. ($\sum_{k=1}^{9}\omega_k = 1$, userinfo_sim(u_i , u_j)$\in (0,1]$)

6.2 Courseware Recommendation Algorithm

(1) The user u_i enters some keywords on the portal of courseware management system.

(2) Courseware Recommendation Module finds within the same user interest group of u_i the k courseware with the same or similar keywords that others choose. This is done by searching in the history record of portal.

(3) For each courseware p, calculate its recommendation degree R_p like this:

$$R_p = \sum_{m=1}^{L}\{[\omega_m \times userinfo\_sim(u_i, u_m) + Credit(u_i, u_m)]\} \times V(p, u_m) \quad (4)$$

R_p stands for recommendation value of courseware p. $Credit(u_i, u_m)$ is defined in section 3.1 as the trust degree between u_i and u_m . $V(p, u_m)$ is the evaluation of courseware p addressed by u_m . L is the number of users that system find in the same interest group of u_i . ω_m is weight of each userinfo_sim. After calculation, sort the k courseware according to its R_p .

(4) Output the top 5 recommended courseware in the sorted sequence to the portal of courseware management system.

7 Experiment and Analysis

The experiment is carried out in our E-learning system, which is the China national key project of the Tenth Five-year plan. The courseware management system with courseware recommendation is part of the integrated system. Here are some snapshots of the system environment: Figure 3 shows the portal of E-learning system. Figure 4 shows the courseware management system and Figure 5 gives a snapshot of a running courseware.

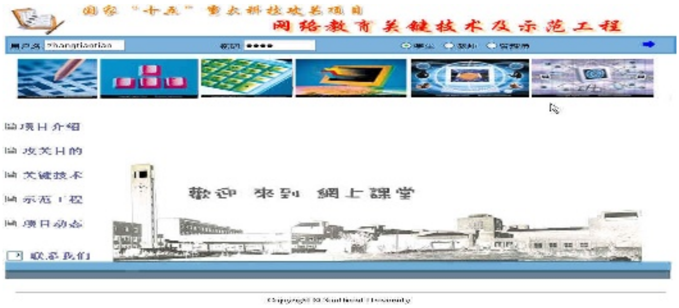

Fig. 3. Snapshot of the Portal of our E-learning System

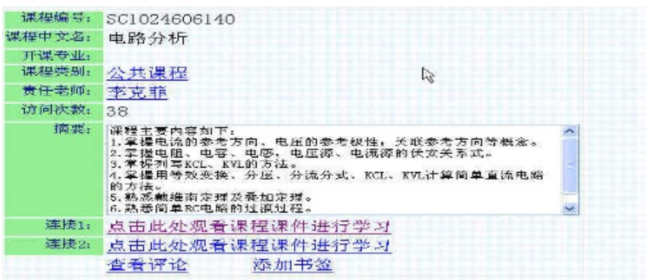

Fig. 4. Snapshot of Courseware Management System

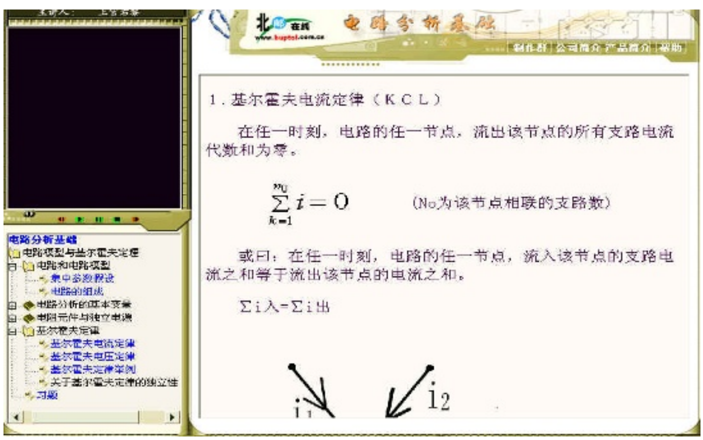

Fig. 5. Snapshot of a Running Courseware

7.1 Experiment Data Source

- **User Data:** From our e-learning system, we random pick up three classes from two distinct stations (station is the place where users are registered to have their distant learning using our system). We name them class A, class B, class C. The total user number is 100, class A has 20 users, class B has 30 users and class C has 50 users. Their user profile information is clearly and fully filled. And users are required to mark the courseware they choose to show whether the courseware reflect their real interests. The mark is ranged from -2 to 2. 2 stands for very interested, -2 stands for not interested at all.

- **Courseware Data:** In our courseware management system, we pick up 50 courseware ranged from computer science, mathematics, Chinese history to Chinese literature, marketing, wireless communication, etc. The users we choose will only pick the courseware that attracts them most from those courseware.

7.2 Evaluation Metrics

As [6] suggests recommender systems research has used several types of measures for evaluating the quality of a recommender system. One of the most widely used metrics is statistical accuracy metrics. They evaluate the accuracy of a system by comparing the numerical recommendation scores against the actual user ratings for the user-item pairs in the test dataset. *Mean Absolute Error* (MAE) between ratings and predictions is a widely used metric. MAE is a measure of the deviation of recommendations from their true user-specified values. For each ratings-prediction pair $< p_i, q_i >$ this metric treats the absolute error between them, equally. The MAE is computed by first summing these absolute errors of the N corresponding ratings-prediction pairs and then computing the average. Formally,

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \tag{5}$$

The lower the MAE, the more accurately the recommendation system predicts user ratings. We used MAE as our choice of evaluation metric to report experiments because it is most commonly used and easiest to interpret directly.

7.3 Experiment Procedures

(1) Firstly we use user aggregation algorithm to get user groups with similar interests. The results is shown in Table 3

Table 3. Results of User Aggregation

User group with similar interests ID	Number of group
1	20
2	30
3	24
4	26

(2) Courseware recommendation and feedback: Every user group login to our e-learning system and enter the courseware management system. Initially, there should have some training data set for our courseware recommendation system,

since our system has been used by other users with similar interests, we skip the training phase. When one individual user enters the portal of courseware management system, he will see two kinds of recommended courseware: the top 5 courseware in the user major and top 5 courseware other users recommend in the same interests group. The individual user will mark these 10 courseware. And the marks are collected.

7.4 Experiment Result and Analysis

The results of four interest groups are shown in Figure 6 to Figure 9.

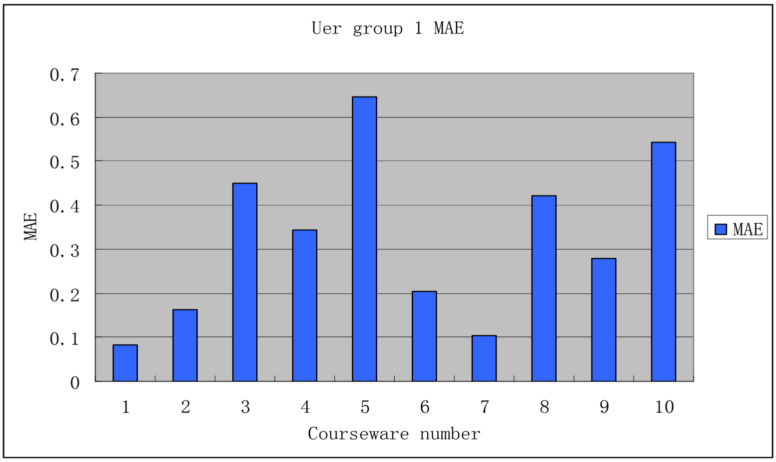

Fig. 6. The Result MAE of User Group 1

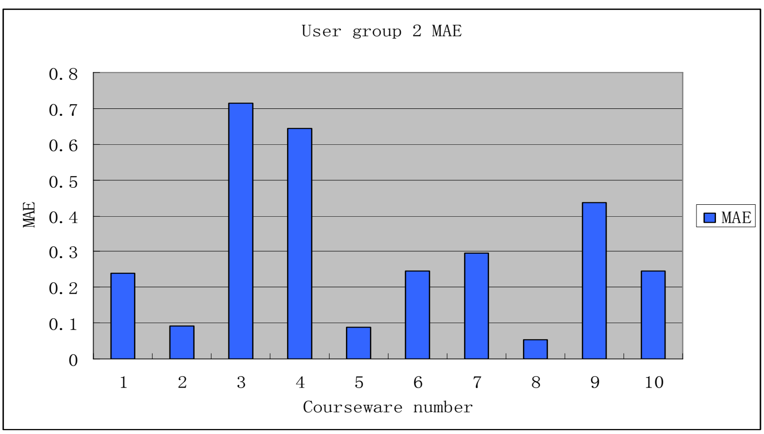

Fig. 7. The Result MAE of User Group 2

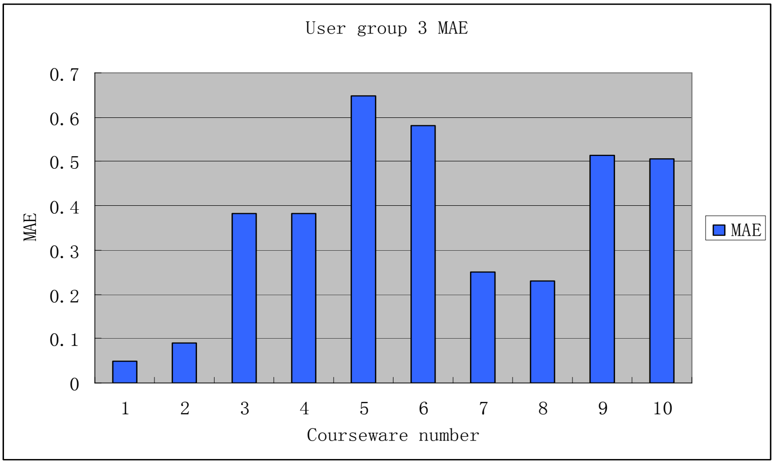

Fig. 8. The Result MAE of User Group 3

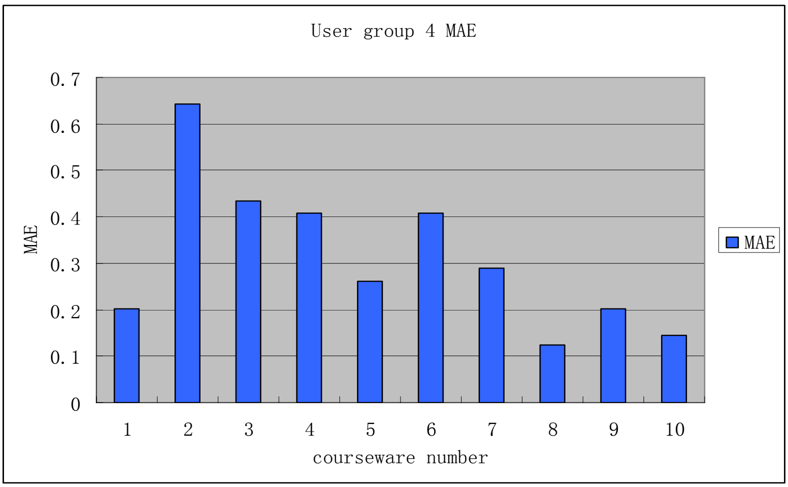

Fig. 9. The Result MAE of User Group 4

The four figures show that each group responds to the courseware in our system to express their interests about these recommended courseware. From the figures, we know that although the MAE towards individual courseware varies, the MAE is well under 1.0, which means that the recommended courseware satisfy users' interests. That is what the system designed for.

8 Conclusions

In this paper, we describe the architecture of the courseware management system with courseware recommendation. We especially present the algorithms that are used in the courseware recommendation module. The algorithms combine contents filtering, which recommends courseware solely from single user information, and collaborative filtering, which recommends courseware from other users perspectives. The experiment and system implementation shows that the system in use is able to reflect user's full interests in courseware selection. And the module is seamlessly integrated in our e-learning system.

Recommendation techniques are based on personalization technology and data mining. With the application of our e-learning system, users of our system will increase enormously and their demand and interests may differ greatly. How to find those diversified interests and reflect in the system is our next-step focus. And also the system performance under the circumstances that user amount is rather large is worthy of future research.

Acknowledgement. This work is supported by National Natural Science Foundation of China under Grants No. 90412014 and 90604004, the Key Scientific and Technological Program for the Tenth Five-Year Plan of China under Grants No. 2001BA101A12, China Specialized Research Fund for the Doctoral Program of Higher Education under Grants No. 20030286014, Jiangsu Provincial Key Laboratory of Network and Information Security under Grants No. BM2003201, Jiangsu Provincial High-Tech Research Program under Grants No. BG2004036 and ChinaGrid Program of Ministry of Education of China under Grants No. CG2003-CG007 and CG2003-GA004-01.

References

1. Anane R., Crowther S., Beadle J. et al.: eLearning Content Provision, Proceedings of the 15th International Workshop on Database and Expert Systems Applications (DEXA'04), Aug.-Sept. (2004) 420-425
2. Mallinson, B., Sewry D.: Elearning at Rhodes University – a case study, Proceedings of the IEEE International Conference on Advanced Learning Technologies (ICALT'04), Aug.-Sept. (2004) 708-711
3. http://www.celtsc.edu.cn/ CELTS-3,42: China E-learning Technology Standard
4. Data Exchange Standard and Interfaces in E-learning integration platform
5. http://dublincore.org/ Dublin Core Metadata Initiative
6. Badrul S., Georage K., Joseph K., John R.: Item-Based Collaborative Filtering Recommendation Algorithms[J], Proceedings of the tenth international conference on World Wide Web, ACM Press, Hong Kong, (2001)285-295
7. Zeng C., Xing C., Zhou L.: A Personalized Search Algorithm by Using Content-Based Filtering, Journal of Software Vol.14, No.5, (2003)
8. Zhao Z., Yuan W.: Research on the User Collaborative Recommendation Algorithm in Personalized Search Engine, Microelectronics and computer, Vol. 22, No.6, (2005).
9. Goldberg, D., Nichols, D., Oki, B. M., and Terry, D: Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, December, (1992).
10. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of CSCW '94, Chapel Hill, NC, (1994).

11. Konstan, J., Miller, B., Maltz, D., Herlocker, J.,Gordon, L., and Riedl, J.: GroupLens: Applying Collaborative Filtering to Usenet News. Communications of the ACM, 40(3), (1997)77-87.
12. Shardanand, U., and Maes, P.: Social Information Filtering: Algorithms for Automating' Word of Mouth'. In Proceedings of CHI '95. Denver,Co, (1995).
13. Hill, W., Stead, L., Rosenstein, M., and Furnas, G.: Recommending and Evaluating Choices in a Virtual Community of Use. In Proceedings of CHI '95, (1995).
14. Breese, J. S., Heckerman, D., and Kadie, C.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, (1998) 43-52.