Course ontology-based user’s knowledge requirement acquisition from behaviors within e-learning systems

Qingtian Zenga,*, Zhongying Zhaob,c,d, Yongquan Lianga

a College of Information Science and Engineering, Shandong University of Science and Technology, No.579 Qianwangang Road, Economic and Technical Development Zone, Qingdao 266510, PR China
b Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China
c Graduate School of Chinese Academy of Sciences, Beijing 100080, China
d Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518054, China

A R T I C L E   I N F O

Article history:
Received 12 October 2008
Received in revised form 20 April 2009
Accepted 29 April 2009

Keywords:
Learning communities
Simulations
Distance and telelearning

A B S T R A C T

User’s knowledge requirement acquisition and analysis are very important for a personalized or user-adaptive learning system. Two approaches to capture user’s knowledge requirement about course content within an e-learning system are proposed and implemented in this paper. The first approach is based on the historical data accumulated by an interactive question-answering process. The association space is proposed to record and formalize the historical interactive information which is used to compute user’s knowledge requirement. The second approach is based on user’s reading behavior logs in the process of reading e-documents. User’s reading actions including underline, highlight, circle, annotation and bookmark, are used to compute user’s knowledge requirement. Two experiments are conducted to implement the two proposed approaches and acquire the user’s knowledge requirement. The evaluation results show that the user models computed by two approaches are consistent and can reflect user’s real knowledge requirements accurately.

1. Introduction

Traditional educational approaches are usually teacher-centric, not student-centric, because they do not sufficiently take into account the difference in characteristics between different students (Berlanga & García, 2005; Bueno, Alonso, & Castillo, 2007). In order to enhance student-centric learning and instruction efficiency, instructors should understand the implicit knowledge requirement of students so as to prepare and design their teaching materials for the students. However, it is usually difficult for instructors to capture the students’ real requirements regarding course knowledge.

With the development of information and network technology, e-learning is changing the situation of the traditional education, and user-adaptive or personalized e-learning systems have recently become hot research topics (Dolog, Henze, Nejdl, & Sintek, 2004; Chen & Chen, 2005; Thyagarajan, 2007; Vasilyeva, 2007). To realize a user-adaptive or personalized e-learning system, user model and modeling are two of the key problems. Usually, the knowledge requirement of a user is an important component in the user model. In this paper, we propose two approaches to capture user’s knowledge requirement within an e-learning system. In our e-learning system, we take into account two aspects about the learning activities of students: the interactive question-answering (Q/A) process (discussion process) and the self reading process in their daily learning. A course ontology (or concept hierarchy), which presents the outline or skeleton about the course content, needs to be predefined by an instructor for e-learning systems.

Many Web-based user-interactive Q/A systems have been developed for users to interactively post and browse Q/A. A user-interactive Q/A system can be embedded into an e-learning system to promote collaborative learning (Hung, Wang, Yang, Chiu, & Yee, 2005; Wang & Huang, 2007). In the module of the Q/A process within an e-learning system, the course ontology can be used to generate the corresponding board structure to hold relevant questions. Users can exchange their knowledge by posting their questions on the related boards, or browsing to find the most interesting or favorite questions to answer. The e-learning system can record all the historical data for each user during the Q/A process, including the browsing records, questions and answers. All of these historical data contain a tremendous amount of information which can be used to capture the knowledge requirement of users.

* Corresponding author. Tel.: +86 532 86057126.
E-mail addresses: qtzeng@sdu.edu.cn, qingtianzeng@hotmail.com (Q. Zeng), zy.zhao@suh.siat.ac.cn (Z. Zhao), lyq@sdu.edu.cn (Y. Liang).
0360-1315/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
Besides considering the interaction activities in an e-learning system, we also focus on capturing the knowledge requirement of users by analyzing their reading behaviors in the reading process. The reading behaviors include actions such as underline, highlight, circle, annotation and bookmark, which imply abundant information about the knowledge requirement of users. The predefined course ontology can also be used as a reference to classify the e-documents that users have read. We propose a behavior table to record the reading behaviors for each user. A behavior matrix and weight matrix are introduced to obtain the relative quantity for each topic, which can be used to calculate the knowledge requirement of each user about the course ontology. Two experiments are conducted to implement the two proposed approaches and capture the knowledge requirement of users within e-learning systems. The evaluation results reveal that the model computed from our two approaches can reflect a user’s true knowledge requirement accurately.

The remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 presents the framework to acquire user's knowledge requirement. The formal definition for course ontology is presented in Section 4. Section 5 presents the method to acquire user's knowledge requirement from Q/A logs. Section 6 presents the method to acquire user's knowledge requirement from reading behaviors. Section 7 verifies the two proposed approaches with experiments and evaluation. Section 8 concludes the whole paper and presents our future work.

2. Related work

Numerous on-line Web-based tutoring systems and hypermedia in courseware have been developed (Brusilovsky & Vassileva, 2003; Chen, Liu, & Chang, 2006; Papanikolaou & Grigoriadou, 2002; Brusilovsky, 1999). To assist users to learn more efficiently, many powerful personalized/adaptive guidance mechanisms, such as adaptive presentation, adaptive navigation support, curriculum sequencing, and intelligent analysis of student's solutions, have been proposed. Boticario and Hernandez (2000) presented adaptive navigation support and adaptive collaboration support tasks in WebDL, and an interactive system for focusing teaching on student performance and resolving problems detected in the Internet is used for distance learning (Boticario & Hernandez, 2000). Alejandro Canales, Alejandro Pená, and Gutierrez (2007) carried out their adaptive and intelligent Web-based Education Systems (WBES) that take into account the individual student learning requirements, by means of a holistic architecture and framework for developing WBES (Alejandro Canales et al., 2007). Chen and Hsu (2007) proposed an approach to mine learner profiles utilizing association rule for Web-based learning diagnosis (Chen and Hsu, 2007). They applied the discovered association rules of the common learning misconceptions to tune courseware structure through modifying the difficulty parameters of courseware in the courseware database so that the learning pathway was appropriately tuned.

Nowadays, most adaptive/personalized tutoring systems have begun to consider learner’s preferences, interests and browsing behaviors for personalized services. Lee analyzed Web-based instruction learners’ adaptation styles and characteristics related to the styles by retrospectively assessing the perceptions of various aspects of Web-based instruction (Lee, 2001). Paper recommendation techniques for learners in an evolvable e-learning system were discussed in Tang and Mccalla (2003), in which the experiment was carried out using artificial learners for two techniques. The first one is based on the matching of the learner model to the papers (pure model-based recommendation), and the second technique is based on peer learner recommendation (hybrid collaborative filtering), which is relatively domain independent. The main objectives of Papanikolaou and Grigoriadou (2002) were to protect learners from cognitive overload and disorientation by supporting them to find the most relevant content and path in the hyperspace. In the approach presented in Papanikolaou and Grigoriadou (2002), a learner’s knowledge level and individual traits are used as valuable information to represent the learner’s current state and personalize the educational system accordingly, in order to facilitate the achievement of personal learning goals and objectives.

Although many Web-based learning techniques have been proposed to assist adaptive/personalized Web-based learning, few research results have attempted to acquire the knowledge requirement of users to develop Web-based tutoring systems (Papanikolaou & Grigoriadou, 2002; Lee, 2001; Tang & Mccalla, 2003). The knowledge requirement acquisition of users can help the tutoring system to determine user-adaptive/personalized instruction contents. Therefore, the knowledge requirement of learners can promote personalized learning performance especially in the recommendation process of the instruction course contents. Compared with the most famous work in past years (Papanikolaou & Grigoriadou, 2002; Lee, 2001; Tang & Mccalla, 2003), the knowledge requirement acquisition of users within an e-learning system addressed in this paper is also an important component of the user model.

The evaluation about the performance of a proposed user model is often difficult in traditional approaches (Papanikolaou & Grigoriadou, 2002; Lee, 2001; Tang & Mccalla, 2003). In this paper, we compared the two experiment results and then introduced the recommendation acceptance rate to evaluate the experimental results computed from the two proposed approaches.

3. Framework to acquire user's knowledge requirement

We first present the framework for capturing the knowledge requirement of users within e-learning systems, which is shown in Fig. 1. This framework can be embedded into any e-learning system as an assistant component to realize user-adaptive or personalized learning or instruction. The main components contained in the framework include the following:

- **Course ontology (or concept hierarchy):** The course ontology is predefined by the instructor with suitable granularity and scale. It presents the structure of the course content. Furthermore, it is the basis for generating the board structure for the Q/A process and classifying the e-documents read by the students.

- **User-interactive Q/A process:** Based on the predefined course ontology, the corresponding board structure of the user-interactive Q/A system can be generated. Within their favorite boards, users can post the most urgent questions on corresponding topics, browse their favorite answers, and select others’ questions to answer. All these Q/A logs can be recorded and accumulated as historical data to capture the knowledge requirement of users.

- **Self-reading process:** According to the predefined course ontology, the e-documents read by each user can be classified following some classification algorithm. The reading behavior logs of each user can also be recorded as historical data that includes the reading actions such as underline, highlight, circle, annotation and bookmark and so on. Both the Q/A logs and reading behavior logs will be stored in the historical data database used to capture the knowledge requirement of users.

Besides the interaction activities in an e-learning system, we also focus on capturing the knowledge requirement of users by analyzing their reading behaviors in the reading process. The reading behaviors include actions such as underline, highlight, circle, annotation and bookmark, which imply abundant information about the knowledge requirement of users. The predefined course ontology can also be used as a reference to classify the e-documents that users have read. We propose a behavior table to record the reading behaviors for each user. A behavior matrix and weight matrix are introduced to obtain the relative quantity for each topic, which can be used to calculate the knowledge requirement of each user about the course ontology. Two experiments are conducted to implement the two proposed approaches and capture the knowledge requirement of users within e-learning systems. The evaluation results reveal that the model computed from our two approaches can reflect a user’s true knowledge requirement accurately.

The remainder of the paper is organized as follows. Section 2 discusses the related work. Section 3 presents the framework to acquire user’s knowledge requirement. The formal definition for course ontology is presented in Section 4. Section 5 presents the method to acquire user’s knowledge requirement from Q/A logs. Section 6 presents the method to acquire user’s knowledge requirement from reading behaviors. Section 7 verifies the two proposed approaches with experiments and evaluation. Section 8 concludes the whole paper and presents our future work.

2. Related work

Numerous on-line Web-based tutoring systems and hypermedia in courseware have been developed (Brusilovsky & Vassileva, 2003; Chen, Liu, & Chang, 2006; Papanikolaou & Grigoriadou, 2002; Brusilovsky, 1999). To assist users to learn more efficiently, many powerful personalized/adaptive guidance mechanisms, such as adaptive presentation, adaptive navigation support, curriculum sequencing, and intelligent analysis of student’s solutions, have been proposed. Boticario and Hernandez (2000) presented adaptive navigation support and adaptive collaboration support tasks in WebDL, and an interactive system for focusing teaching on student performance and resolving problems detected in the Internet is used for distance learning (Boticario & Hernandez, 2000). Alejandro Canales, Alejandro Pená, and Gutierrez (2007) carried out their adaptive and intelligent Web-based Education Systems (WBES) that take into account the individual student learning requirements, by means of a holistic architecture and framework for developing WBES (Alejandro Canales et al., 2007). Chen and Hsu (2007) proposed an approach to mine learner profiles utilizing association rule for Web-based learning diagnosis (Chen and Hsu, 2007). They applied the discovered association rules of the common learning misconceptions to tune courseware structure through modifying the difficulty parameters of courseware in the courseware database so that the learning pathway was appropriately tuned.

Nowadays, most adaptive/personalized tutoring systems have begun to consider learner’s preferences, interests and browsing behaviors for personalized services. Lee analyzed Web-based instruction learners’ adaptation styles and characteristics related to the styles by retrospectively assessing the perceptions of various aspects of Web-based instruction (Lee, 2001). Paper recommendation techniques for learners in an evolvable e-learning system were discussed in Tang and Mccalla (2003), in which the experiment was carried out using artificial learners for two techniques. The first one is based on the matching of the learner model to the papers (pure model-based recommendation), and the second technique is based on peer learner recommendation (hybrid collaborative filtering), which is relatively domain independent. The main objectives of Papanikolaou and Grigoriadou (2002) were to protect learners from cognitive overload and disorientation by supporting them to find the most relevant content and path in the hyperspace. In the approach presented in Papanikolaou and Grigoriadou (2002), a learner’s knowledge level and individual traits are used as valuable information to represent the learner’s current state and personalize the educational system accordingly, in order to facilitate the achievement of personal learning goals and objectives.

Although many Web-based learning techniques have been proposed to assist adaptive/personalized Web-based learning, few research results have attempted to acquire the knowledge requirement of users to develop Web-based tutoring systems (Papanikolaou & Grigoriadou, 2002; Lee, 2001; Tang & Mccalla, 2003). The knowledge requirement acquisition of users can help the tutoring system to determine user-adaptive/personalized instruction contents. Therefore, the knowledge requirement of learners can promote personalized learning performance especially in the recommendation process of the instruction course contents. Compared with the most famous work in past years (Papanikolaou & Grigoriadou, 2002; Lee, 2001; Tang & Mccalla, 2003), the knowledge requirement acquisition of users within an e-learning system addressed in this paper is also an important component of the user model.

The evaluation about the performance of a proposed user model is often difficult in traditional approaches (Papanikolaou & Grigoriadou, 2002; Lee, 2001; Tang & Mccalla, 2003). In this paper, we compared the two experiment results and then introduced the recommendation acceptance rate to evaluate the experimental results computed from the two proposed approaches.

3. Framework to acquire user’s knowledge requirement

We first present the framework for capturing the knowledge requirement of users within e-learning systems, which is shown in Fig. 1. This framework can be embedded into any e-learning system as an assistant component to realize user-adaptive or personalized learning or instruction. The main components contained in the framework include the following:

- **Course ontology (or concept hierarchy):** The course ontology is predefined by the instructor with suitable granularity and scale. It presents the structure of the course content. Furthermore, it is the basis for generating the board structure for the Q/A process and classifying the e-documents read by the students.

- **User-interactive Q/A process:** Based on the predefined course ontology, the corresponding board structure of the user-interactive Q/A system can be generated. Within their favorite boards, users can post the most urgent questions on corresponding topics, browse their favorite answers, and select others’ questions to answer. All these Q/A logs can be recorded and accumulated as historical data to capture the knowledge requirement of users.

- **Self-reading process:** According to the predefined course ontology, the e-documents read by each user can be classified following some classification algorithm. The reading behavior logs of each user can also be recorded as historical data that includes the reading actions such as underline, highlight, circle, annotation and bookmark and so on. Both the Q/A logs and reading behavior logs will be stored in the historical data database used to capture the knowledge requirement of users.
User modeling: Driven by the accumulated Q/A logs, the system builds the association relations between the questions and answers and the course ontology. It then computes the knowledge requirement of each user about the course ontology. From the reading behavior logs, we can construct the behavior matrix and weight matrix that are used to compute the knowledge requirement of each user about the course ontology.

4. Course ontology

Before presenting the approaches for capturing users’ knowledge requirement, the formal definition for the related course ontology (or concept hierarchy) is introduced, which presents the outline or skeleton about the course content.

Definition 1. A course ontology is defined as a tree-like graph (or concept hierarchy), \( \text{TopicOnto} = (\text{TS}, \text{R}) \), where

1. The set of nodes \( \text{TS} = \{\text{term}_1, \text{term}_2, \ldots, \text{term}_n\} \) represents a set of terms.
2. The set of edges \( \text{R} \subseteq (\text{TS} \times \text{TS}) \) represents the binary relations on the set \( \text{TS} \). If two terms satisfy one of \( \text{IS\_A}, \text{Component\_Of} \), or \( \text{Part\_Of} \) relations, then there is a direct edge between the two terms.

In the course ontology, a term can be the title of a chapter and a section in the course, or the key concept of the course contents. The granularity of the course ontology is determined by the instructor. There can be many kinds of relations between the terms in the ontology (Guarino & Welty, 2000; Shadbolt, Middleton, & Roure, 2004). In this paper, only three kinds of well-known relations are taken into account, namely, \( \text{IS\_A}, \text{Part\_Of} \), and \( \text{Component\_Of} \), which are considered as the most important relations to capture the knowledge requirement of users about the course contents.

The course ontology can be represented by a directed graph. The node set in the graph represents the terms. If \( \text{term}_1 \) and \( \text{term}_2 \) satisfy \( \text{IS\_A}(\text{term}_1, \text{term}_2), \text{Part\_Of}(\text{term}_1, \text{term}_2), \) or \( \text{Component\_Of}(\text{term}_1, \text{term}_2) \), it means that \( \text{term}_2 \) is a (part of, or component of) \( \text{term}_1 \), thus there is a direct edge from \( \text{term}_1 \) to \( \text{term}_2 \) in the directed graph. A node is named as a leaf if its out-degree is zero. An example for part of the course ontology of Artificial Intelligence is shown in Fig. 2. It contains three chapters: Agent Theory (including two sections: Agent Architecture and Multi-agent System), Machine Learning (including three sections: Deductive Learning, Inductive Learning and Analogical Learning) and Knowledge Representation (including three sections: Predicate Logic, Semantic Network and Ontology). In Fig. 2, we can see that terms \( T_i (i = 1, \ldots, 8) \) are leaf nodes of the course ontology.

5. User’s knowledge requirement acquisition from Q/A logs

During the interactive Q/A process within an e-learning system, users can post their own questions, browse and answer questions asked by others. All these Q/A logs can be used as historical data to acquire the knowledge requirement of each user within the e-learning systems. First, a formal model for the Q/A historical logs is defined.

5.1. Model for Q/A historical logs

The Q/A space, describing the relations between questions and answers, can be constructed from the Q/A logs in historical data. The association space is defined to express the relations between a Q/A space and the course ontology.
Definition 2. A Q/A space is a five-tuple \( \text{QASpace} = (\text{QS}, \text{AS}, f_B, f_A, f_{CA}) \), where

1. \( \text{QS} \) is the set of questions.
2. \( \text{AS} \) is the set of answers.
3. \( f_B: \text{QS} \rightarrow \mathbb{Z} \) is a function from \( \text{QS} \) to \( \mathbb{Z} \), where \( \mathbb{Z} \) is the integer set. \( \forall Q \in \text{QS}, f_B(Q) \) is nonnegative that is the browsing frequency (or, the number of times being browsed) of the question \( Q \).
4. \( f_A: \text{QS} \rightarrow \mathcal{P}(\text{AS}) \) is a function from \( \text{QS} \) to the powerset of \( \text{AS} \). \( \forall Q \in \text{QS}, f_A(Q) \) is the set of the answers about \( Q \).
5. \( f_{CA}: \text{QS} \rightarrow \mathcal{P}(\text{AS}) \). \( \forall Q \in \text{QS}, f_{CA}(Q) \) is the set of the correct answers about \( Q \).

According to the definition of a Q/A space, we can derive that \( f_{CA}(Q) \subseteq f_A(Q) \). A directed graph can be used to represent the relations between \( \text{QS} \) and \( \text{AS} \). \( \text{QS} \) and \( \text{AS} \) are represented by different nodes, and there are directed edges from the answers to the question \( Q \) if \( |f_A(Q)| > 1 \). \( \forall Q \in \text{QS} \), \( Q \) is named as a closed question if \( |f_{CA}(Q)| > 1 \) or its answer deadline is reached.

Given the course ontology and the Q/A space, we can build the association space \( \text{AssSpace} \) to express the relations between the historical data and the related topics. If a question \( Q \) is posted on the board corresponding to one term \( T \), then \( (Q, T) \in \text{AssSpace} \). The formal definition of the association space is not presented here. Fig. 3 shows a sketch map of an association space.

5.2. Acquiring user’s knowledge requirement from Q/A logs

The association space between the Q/A space and the course ontology indicates the relations between the historical data (browsing activities, questions, answers, and correct answers) and the course ontology. Within an e-learning system, the association space can be
constructed for each user to obtain the user’s knowledge requirement. In this subsection, the approach to acquire the user’s knowledge requirement from Q/A logs is proposed based on the association space.

First, we compute each user’s knowledge requirement about every question according to the association space. The user’s knowledge requirement about the whole course ontology is then computed, i.e., the knowledge requirement about each section of the course content. By common sense, the answers given by each user can only reflect whether the user has grasped the corresponding knowledge, and cannot indicate the knowledge requirement of a user. Therefore, we are not concerned with whether the user has given answers or whether the answers are correct. The knowledge requirement of a user about a question $Q_j$ is mainly determined by the browsing frequency $f_b(Q_j)$ and the average browsing frequency $Mean_B$. In our approach, for any question $Q_j \in QS$, the knowledge requirement of the $i$th user about the question $Q_j$ is calculated by:

$$KR_i(Q_j) = \arctan(f_b(Q_j) - Mean_B) / \pi + \arctan(Mean_B) / \pi.$$

where $f_b(Q_j)$ is the browsing frequency of the $i$th user about the question $Q_j$, $Mean_B$ represents the average browsing frequency of the $i$th user.

According to the knowledge requirement $KR_i(Q_j)$ about each question and the association space $AssSpace$, the knowledge requirement of the $i$th user about $T_k$ can be calculated by:

$$KR_i(T_k) = \sum_{j=1}^{n} KR_i(Q_j) / n,$$

where $(Q_j, T_k) \in AS$ and $n = |\{(Q_j, T_k) \in AS\}|$.

Based on the course ontology and the proposed approach above, we can see that $KR_i(T_k)$ represents the knowledge requirement of the $i$th user about the section $T_k$. Algorithm 1 presents the approach to capture each user’s knowledge requirement about the course contents from the Q/A historical logs.

**Algorithm 1.** Capture the knowledge requirement of each user about the course contents from the Q/A logs.

**INPUT:** Course ontology $TopicOnto = (TS, R)$

**OUTPUT:** The knowledge requirement of each user about the course ontology

**Step 1:** Generate the board structure according to the course ontology $TopicOnto = (TS, R)$ to accumulate the historical interactive Q/A data. Usually, each term in $TopicOnto$ corresponds a Q/A board.

**Step 2:** Construct Q/A space $QSspace = (QS, AS, f_b, f_s, f_A)$ for each user when they have posted and answered questions on the corresponding boards.

**Step 3:** Construct the association space $AssSpace$ between the Q/A space and the course ontology.

**Step 4:** Compute the knowledge requirement of each user about each question in the Q/A space $QSspace = (QS, AS, f_b, f_s, f_A)$.

**Step 5:** Compute the knowledge requirement of each user about each term in the course ontology based on $AssSpace$.

**Step 6:** Output the knowledge requirement of each user about the term contents $(T_k, KR_i(T_k))$.

### 6. User’s knowledge requirement acquisition from reading behavior logs

Within an e-learning system, there are many e-documents provided for students to read and refer to. During their reading process, users often markup the e-document with their own reading actions such as underline, highlight, circle, annotation, bookmark and so on. Such reading behaviors also reflect abundant information about the user’s requirement, so they should also be taken into account as resource data to acquire the knowledge requirement of users. First, one table and two matrixes used to store the reading behavior data of each user are introduced.

#### 6.1. Behavior table

A two-dimension table is used to store the reading behavior data of each user. It contains eight attributes: User ID ($UID$), document ID ($DID$), corresponding topics ($T$), the number of underlines ($UDL$), the number of highlights ($HLT$), the number of circles ($CIR$), the number of annotations ($ANT$) and the number of bookmarks ($BMK$), where $T$ is one of the leaf terms in the course ontology. An example of a behavior table is shown in Table 1. Each line in the table records the number of the reading actions while user$_i$ read the related documents. For example, user user$_1$ left 1 underline, 8 highlights, 3 circles, 2 annotations and 9 bookmarks while user$_1$ read document $D_1$ that is about term $T_1$.

Let $RT$ represent the average browsing time of one document page of each user, which is calculated by $RT = DTime / PD$, where $DTime$ is the browsing time on the whole document and $PD$ is the page numbers of the document. To ensure that users have really browsed the documents, not merely opened them and then immediately closed them, or left the document open without reading it, we define a timing interval $[min, max]$ for each document $D_j (j = 1, 2, \ldots)$. That is, if $D_j$ belongs to behavior table, it must satisfy $min \leq RT(D_j) \leq max$.

The e-documents read by each user can be classified by a document clarification algorithm, for example, the IBk classifier (Aha, Kibler, & Albert, 1991), which is boosted by the AdaBoostM1 algorithm (Freimd and Schapiew, 1996). The IBk classifier is a k-Nearest Neighbor type

<table>
<thead>
<tr>
<th>$UID$</th>
<th>$DID$</th>
<th>$T$</th>
<th>$UDL$</th>
<th>$HLT$</th>
<th>$CIR$</th>
<th>$ANT$</th>
<th>$BMK$</th>
</tr>
</thead>
<tbody>
<tr>
<td>user$_1$</td>
<td>$D_1$</td>
<td>$T_1$</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>user$_1$</td>
<td>$D_2$</td>
<td>$T_1$</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
classifier that uses example documents to add into a term-vector space. Example documents in the training set can be manually labeled using the class names within the course ontology. AdaBoostM1 has been shown to improve the performance of the weak learner algorithms, particularly for the stronger learning algorithms like $k$-Nearest Neighbor. It has also been shown that the boosted IBk classifier has a higher classification performance (Middleton et al., 2004). It is thus a sensible choice to use the AdaBoostM1 classifier in our system.

6.2. Behavior matrix and weight matrix

In order to calculate the knowledge requirement of each user, we employ a set of behavior matrixes $BMS = \{BM(T_1)_{n_1,5}, BM(T_2)_{n_2,5}, \ldots\}$ to store the data in behavior tables. Each element $BM(T_k)_{n_k,5}$ represents the reading behavior data about the documents related to topic $T_k$ and it is shown as:

$$BM(T_k)_{n_k,5} = \begin{pmatrix}
UDL_1 & HLT_1 & CIR_1 & ANT_1 & BMK_1 \\
UDL_2 & HLT_2 & CIR_2 & ANT_2 & BMK_2 \\
UDL_3 & HLT_3 & CIR_3 & ANT_3 & BMK_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
UDL_k & HLT_k & CIR_k & ANT_k & BMK_k
\end{pmatrix},$$

where $n_k$ represents the document number about $T_k$, and 5 represents the five reading behaviors including $UDL$, $HLT$, $CIR$, $ANT$ and $BMK$.

It is a fact that different people have different reading habits, and the reading habits of each user are usually stable. For example, annotation may be the most interested behavior for some learners but less important for others. Therefore, the behavior weights corresponding to $UDL$, $HLT$, $CIR$, $ANT$ and $BMK$ here are specified by each user respectively, which permits flexibility for various users and can help the system capture the knowledge requirement of each user precisely. An online e-survey system has been designed to obtain the behavior weight of each user. The interface for reading behavior survey is shown in Fig. 4.

Another matrix denoted as $WM_{5,1}$ is adopted to represent the behavior weight such that $WM_{5,1} = \left(\begin{array}{c}
w_1 \\
w_2 \\
w_3 \\
w_4 \\
w_5
\end{array}\right)$, where $w_i (i=1,2,3,4,5)$ represents the weight of $UDL$, $HLT$, $CIR$, $ANT$ and $BMK$, respectively.

Let $RM(T_k)_{n_k,1} = BM(T_k)_{n_k,5} \times WM_{5,1}$, then we can get

$$RM(T_k)_{n_k,1} = \begin{pmatrix}
RD_1 \\
RD_2 \\
\vdots \\
RD_4
\end{pmatrix}$$

where $RD_j$ represents the relative number on document $D_j$, $RD_j = w_1 \times |UDL| + w_2 \times |HLT| + w_3 \times |CIR| + w_4 \times |ANT| + w_5 \times |BMK|$.

6.3. Acquiring user’s knowledge requirement from reading behavior logs

Taking user’s reading behavior weight into account, if one user browsed and read a document $D_j$ times, we consider that the user has browsed this document $RD_j$ times. Similarly, we can calculate the relative number on each topic $T_k$ by $R(T_k) = \sum_{j=1}^{n_k} RD_j$.

Fig. 4. The interface for the reading behavior survey.
Thus, the total relative number of documents that one user browsed can be calculated by \( R_{User} = \sum_i R(T_i) \) where \( T_i \) is the topic that user browsed.

In our approach, according to \( R(T_i) \) and \( R_{User} \), the knowledge requirement of the \( i \)th user about each \( T_i \) is calculated by \( \text{KR}_i(T_i) = R(T_i) / R_{User} \) where \( T_i \) is one of the leaf terms in the course ontology. Algorithm 2 presents the method to compute the knowledge requirement of each user from the reading behavior logs.

**Algorithm 2.** Compute the knowledge requirement of each user about the course ontology from reading behavior logs

**INPUT:** Course ontology \( \text{TopicOnto} = (T_S, R) \)

**OUTPUT:** The knowledge requirement of each user about the course Ontology

**Step 1:** Based on the course ontology defined, classify the e-documents read by the users.

**Step 2:** Record the reading behavior data for each user including actions UDL, HLT, CIR, ANT and BMK, and store them into the behavior table.

**Step 4:** Construct the weight matrix \( WM_{n \times 1} \) by surveying the behavior weights of each user.

**Step 5:** Calculate \( RM(T_i)_{n \times 1}, R(T_i), \) and \( R_{User} \).

**Step 6:** Compute the knowledge requirements of each user about each leaf term in the course ontology and output \( (T_i, \text{KR}_i(T_i)) \).

### 7. Experiment and evaluation

In order to verify the proposed approaches and capture the knowledge requirement of the users, two experiments are designed and implemented. The predefined course ontology shown in Fig. 2 is used for these two experiments. The behavior data for these two experiments are collected from 10 students majoring in AI (Artificial Intelligence) course in the spring of 2008. Although there are other users attending the experiments, they are not selected as samples for the experiments. It is a fact that most of the other users not majoring in AI course have no or little knowledge requirements about AI, so the collected data about their reading behaviors and the Q/A information are not enough for the experiments. In order to ensure our approach can capture the real knowledge requirements of the users, we select all those 10 students majoring in AI course for our experiments. It takes one term to collect the data about the reading behaviors and Q/A information of each student.

In order to verify the knowledge requirement of each user, we collect the feedback through the recommendation acceptance rate during our evaluation process. In this section, we present the experiments and the related evaluation for the experimental results.

#### 7.1. Experiment 1: knowledge requirement acquisition from user’s Q/A logs

The first experiment is conducted in a Q/A platform on our research communication Web. The 10 students registering for the Artificial Intelligence course in the spring of 2008 are invited to attend the first experiment. All these 10 students are encouraged to post their urgent questions on their favorite boards. Then, other helpful students who have mastered the related knowledge are invited to answer those posted questions. During this process, all of these 10 students are also encouraged to browse and answer questions of others. After one term, we obtain the accumulated Q/A historical data used for the first experiment. The first proposed approach is applied to compute the knowledge requirement of each student from her/his own Q/A logs. The knowledge requirement of each student about the Artificial Intelligence course ontology is shown in Table 2 and in Fig. 5.

#### 7.2. Experiment 2: knowledge requirement acquisition from user’s reading behavior logs

To verify the second proposed approach, another experiment is implemented with the same 10 students that have attended the Q/A experiment. This experiment is also based on the course ontology shown in Fig. 2. In this experiment, more than 100 e-documents are provided and indexed by the document title. The students are asked to browse and select their favorite documents to read. Their reading behaviors are observed and recorded by a behavior table. After one term, we obtain the accumulated reading behavior data used for the second experiments. The knowledge requirement of each student about the course ontology is computed and shown in Table 3 and Fig. 6.

#### 7.3. Evaluation on the experimental results

To examine the reliability of the experimental results, we first make a comparison between the knowledge requirement of each student acquired from the Q/A logs and that from the reading behavior logs. The recommendation acceptance rate is then used to evaluate the two proposed approaches.

**Table 2** The knowledge requirement of each student acquired from the Q/A logs.

<table>
<thead>
<tr>
<th>UserID</th>
<th>KR(T1)</th>
<th>KR(T2)</th>
<th>KR(T3)</th>
<th>KR(T4)</th>
<th>KR(T5)</th>
<th>KR(T6)</th>
<th>KR(T7)</th>
<th>KR(T8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>0.75</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.21</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#2</td>
<td>0.00</td>
<td>0.41</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.53</td>
</tr>
<tr>
<td>#3</td>
<td>0.00</td>
<td>0.39</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.67</td>
</tr>
<tr>
<td>#4</td>
<td>0.22</td>
<td>0.68</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.24</td>
<td>0.00</td>
<td>0.00</td>
<td>0.56</td>
</tr>
<tr>
<td>#7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.63</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#8</td>
<td>0.00</td>
<td>0.35</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.53</td>
</tr>
<tr>
<td>#9</td>
<td>0.65</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>#10</td>
<td>0.00</td>
<td>0.68</td>
<td>0.00</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
The comparison of the two experimental results is shown in Fig. 7. From Fig. 7, we can see that the knowledge requirement of each student acquired from the Q/A logs is generally consistent with that acquired from the reading behavior logs. There are only a few differences between the experimental results. In the instruction or learning process, however, we often pay more attention to the rank of the knowledge requirement rather than the exact value of it. This means that what a user-adaptive or personalized system is most concerned with are which elements of course content are required by a student and which one is the most urgent. From this viewpoint,
we can say that the knowledge requirement acquired from the reading behaviors has the same distribution with that acquired from the Q/A logs. Thus, it can be deduced that the experimental results generated by the two different approaches are consistent.

It still cannot ensure the efficiency of the two proposed approaches, although their experimental results generated are consistent. In order to prove the proposed approaches can capture the knowledge requirement of each user precisely, the models are evaluated by the content recommendation. According to the knowledge requirements obtained by the experiments, we recommend new available e-documents, new questions and new answers in the Q/A system to the 10 students involved in the experiments. If one user requires the recommended content, he/she clicks the recommended link to browse it. In this case, we consider that the user accepts the system’s recommendation. Otherwise, the recommendation is refused. The recommendation acceptance rate of the 10 students is calculated every day. If the two proposed approaches are adaptive and can express the knowledge requirement of each user precisely, the recommendation acceptance rate should be stable and retain a high percentage. After 60 days for the content recommendation, the recommendation acceptance rate about the 10 students is obtained and shown in Fig. 8. From Fig. 8, we can see that the recommendation acceptance rate is stable and the changing spectrum is always higher than 0.7. Thus, the experimental results show that the two proposed approaches can capture the knowledge requirement of the users precisely in the e-learning systems.

8. Conclusion and future work

Users’ knowledge requirement acquisition is very important for personalized course content design and recommendation within an e-learning system. In this paper, two approaches for capturing the knowledge requirement of the users about the course content within e-learning systems are proposed and implemented. The first approach is to acquire the knowledge requirement of the users from their
historical Q/A logs. In this approach, a course ontology can be used to generate the corresponding board structure to hold relevant questions. The user-interactive log data for each user are accumulated and used to build the association space between the user’s historical data and the course ontology. Based on the association space, the knowledge requirement of each user about the course ontology can be computed. The second approach is to acquire the knowledge requirement of the users from their reading behavior data collected from their daily learning and reading. The predefined course ontology is used to classify the e-documents read by each user. A behavior table is introduced to record the user’s reading behaviors including actions such as the underline, highlight, circle, annotation and bookmark. A behavior matrix and a weight matrix are introduced to compute the user’s knowledge requirement about the course ontology.

Two experiments are conducted to implement our proposed approaches and obtain the knowledge requirement of the users. The evaluation combined with the experiment results indicate that the user models computed by our methods can reflect their real knowledge requirements.

Once the user’s knowledge requirements are acquired, there are many potential applications within adaptive e-learning systems. The potential applications include but are not limited to the design of the course instruction contents and the course content recommendation. In our future work, we will study design-based research (Barab & Squire, 2004; Sandoval & Bell, 2004; Wang and Hannafin, 2005) and other current evaluative and adaptive paradigms (Hunter, 1997) to make our system more flexible.

Acknowledgements

This research is supported by the National Science Foundation of China under Grant Nos. 60603090, 90718011 and 50875158, the Excellent Young Scientist Foundation of Shandong Province of China under Grant No. 2006BS01019, the Research Foundation of Shandong Educational Committee under Grant No. J08LJ77, and the Taishan Scholar Program of Shandong Province. The authors would like to thank all the students for their efforts in our experiments. We are also appreciate the reviewers and editor for their valuable suggestions and comments. The authors would like to thank all the students for their efforts in our experiments. We are also appreciate the reviewers and editor for their valuable suggestions and comments. The authors would like to thank all the students for their efforts in our experiments. We are also appreciate the reviewers and editor for their valuable suggestions and comments.