

Improving Adaptation in Web-Based Educational Hypermedia by means of Knowledge Discovery

Andrej Krištofič

Institute of Informatics and Software Engineering
Faculty of Informatics and Information
Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
kristofic@zoznam.sk

Mária Bieliková

Institute of Informatics and Software Engineering
Faculty of Informatics and Information
Technologies, Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
bielik@fiit.stuba.sk

ABSTRACT

Most adaptive web-based hypermedia systems adapt presentation of the content and/or navigation using predefined set of rules. Considering different behavior and preferences of each user it may be hard to generalize and construct all appropriate rules in advance. This problem is more noticeable in educational adaptive hypermedia systems, where adaptation to individual learning style of a student is important for the student to effectively assess particular domain. In this paper we present techniques for data mining, which can be used to discover knowledge about students' behavior during learning, as well as techniques, which take advantage of such knowledge to recommend students lessons they should study next. We also describe a process of recommendation based on knowledge discovery and present an architecture of a web-based system, which uses proposed approach to improve adaptation. Proposed architecture is independent of actual adaptive hypermedia system used.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Algorithms, Design, Experimentation

Keywords

Adaptive web-based educational hypermedia, knowledge discovery, adaptive navigation, concept recommendation, usage patterns

1. INTRODUCTION

Adaptive hypermedia systems have become popular in last few years in many application areas such as educational

hypermedia, on-line information systems or information retrieval hypermedia [4]. Their growth is supported also by widespread use of the web as a basis for making a 'new education' possible, because it is available *anyplace* and *anytime* [8]. Adaptive hypermedia application in the field of educational systems is natural, as each student generally has different characteristics related to learning (goals, learning styles, preferences, etc.). An adaptation to the student's characteristics can help the student to master topics more effectively.

However, it is quite difficult for an author of the course to develop appropriate adaptation model (usually represented using if-then rules). This requires respectable knowledge of students' behavior to create such adaptation rules, which will result into an effective adaptation of presented knowledge together with appropriate navigation in the course information space. We propose to use data collected by the educational adaptive hypermedia system during its usage to extract implicit, unknown and potentially useful information related to the adaptation from data. Methods and techniques of knowledge discovery provide us with discovering interesting knowledge, which can be employed to improve the adaptation model together with simplification of the adaptation model development.

The aim of this paper is to present a proposal to employ knowledge discovery for purpose of finding trends and patterns of a usage from data collected by the adaptive hypermedia educational system. These patterns reflect characteristic behavior of students (individuals or groups) during learning with the aid of the adaptive hypermedia system. Discovered patterns are used in the process of recommendation of relevant concepts to students, mainly for curriculum sequencing. We explore this proposal in the context of learning programming using programming exercises.

The rest of the paper is structured as follows. First, we give short overview of knowledge discovery in adaptive hypermedia in the context of related works. Next, in Section 3 we describe interesting data selected as a source for knowledge discovery in educational adaptive hypermedia. In Section 4 we present steps for preprocessing data from educational web-based system. Next, in Section 5 we discuss techniques appropriate for data mining with the aim of discovering interesting patterns in learners' behavior. In Section 6 we present the usage of discovered knowledge for recommendation in adaptive hypermedia. We describe design of a recommender system, which is based on discovering behavioral patterns and the usage of such discovered

knowledge. The paper concludes with a summary, and a description of future directions of our research.

2. RELATED WORKS

Knowledge discovery in educational adaptive hypermedia usage data is a complete process, rather than a particular algorithm. It is aimed to reveal interesting patterns hidden in data collected during student's education. The process consists of the basic steps identified for knowledge discovery (e.g., [9]) and later for web usage mining (e.g., [18]). The process begins with understanding application domain, stating the problem and collecting data. Next, acquired data is cleaned and preprocessed. This step may also include integration of data from various sources, selection of relevant attributes and data transformation for the next step – data mining. During the data mining step, various methods for pattern extraction or classification are applied. In the end of knowledge discovery process extracted knowledge is analyzed, interpreted and introduced into practical use (e.g., using for improving adaptation in educational hypermedia system).

Knowledge discovery for navigation recommendation is often based on usage data mining. Current approaches mainly fall on web usage mining area. Their primary goal is usually evaluating web site designs through an analysis of the browsing behavior of the web site users. Recently some authors focus on personalization issues rather than producing analytical knowledge (e.g., [17]).

There exists a variety of techniques for data mining, each with specific characteristics and possible usage. Commonly used techniques are cluster mining, association rules mining and sequential pattern mining. These techniques have drawn a lot of attention lately and a variety of algorithms have been developed (see survey in [14]). Other artificial intelligence approaches for discovering interesting patterns can be used. For example, in [19] the Education Prediction Rules tool for prediction rules discovery in adaptive hypermedia system using genetic algorithm is presented. Authors used adapted AHA! system as the source of usage data.

Following discovered patterns various approaches have been proposed for navigation recommendation (global or local, direct or indirect guidance). In [16] authors have proposed to conduct the recommendation directly with discovered patterns. The active user session is used to search the patterns and find the matching ones. Suffixes of these patterns become candidates for the recommendation. Only the most recent part (called window – w) of the active user session is used for matching, and it is matched against the patterns with length $|w|+1$. When no recommendation can be made with current window, its size is decreased until it is possible to generate a recommendation.

Another approach [15] is aimed to group stored user sessions or discovered frequent item sets into session clusters using techniques of clustering. Representative usage profile is computed for each cluster. Active user session is then matched against computed usage profiles.

Mentioned approaches deal with one kind of patterns and provide support for this specific kind. We propose using several kinds of patterns (e.g., sequential and traversal patterns) distinguished during the process of recommendation and describe software framework for a recommender system,

¹where $|w|$ denotes the size of window

which extends architecture of educational adaptive web-based hypermedia applications.

3. DATA FOR KNOWLEDGE DISCOVERY

Basic goal of knowledge discovery in an educational adaptive hypermedia system is to discover characteristic patterns of behavior of students, or groups of students during learning. Source data are substantial for sound results of knowledge discovery. We have identified several types of data, which represent useful source for knowledge discovery in the educational adaptive hypermedia system:

- user activity logs,
- user knowledge level (performance value),
- domain model structure.

In the following paragraphs we explain importance of each type of data for the process of knowledge discovery. As we experimented with two adaptive educational systems: AHA! [7] and ALEA [13], we provide examples and discussion related to data that can be found in these two adaptive hypermedia systems. However, our observations are valid also for other systems because data related to the user behavior in adaptive hypermedia systems do not differ significantly (common ontologies such as [12] and standards such as [11] can be used).

3.1 User activity logs

Almost every adaptive hypermedia system maintains its usage logs, which contain data related to users' activities. These logs are the main source of data for mining. The data show us how the educational system was being used by students – which concepts (knowledge elements) students visited and which information they were presented with.

Log entries usually contain *timestamps*, which are helpful in estimating the time a student spent on certain concept. Time can be used also to evaluate similarity between different user learning activities [10]. Time usage this way may originate inaccuracies though, because timestamps denote only the time of concept accessing, but not the time of concept studying. We cannot say, whether the student really worked on displayed concept, or he was just idling. However, after preprocessing time access values (e.g., by filtering extreme values) it can be used as an auxiliary characteristic in the process of knowledge discovery. There exist approaches, which employ a client-side program to capture the time spent on information fragments more accurately [20].

Most of today's adaptive hypermedia systems are realized as web-based applications. Consequently, when there are no hypermedia system logs available, logs from the web server can be used instead. Naturally, additional data preprocessing and cleaning has to be carried out.

Our proposal considers following user action attributes usable in the process of knowledge discovery:

- *timestamp*
defines time when action occurred. It is used to order actions in time and to compute the time a user spent on certain concept;
- *user identification*
having each user uniquely identified, we can perform knowledge discovery in actions of individual user and thus allow personalization;

- *type*
allows reasoning on different levels. We defined following elementary action types (with possibility of an extension):
 - *Login* and *Logout* which are used to delimit user’s learning session;
 - *ConceptVisit* and *FragmentDisplay*, which carry information about visited concepts by particular user and information fragments he was presented with;
- *object*
specifies identification of an object on which the action is carried out. For example, for the *ConceptVisit* action it holds an identification of the concept visited by the user.

3.2 Domain model structure

Domain model structure is important source for discovering interesting patterns in usage data. Domain model of an educational hypermedia system is composed of a set of domain knowledge elements [5], which represent elementary fragments of knowledge for the given domain. They can be named differently in various educational hypermedia systems, e.g. concepts, knowledge items, topics, knowledge elements, learning outcomes. We denote elementary fragments of domain knowledge in this paper as concepts. Moreover, we distinguish concepts and fragments. Fragments represent actual information content, i.e. knowledge in case of educational hypermedia system. Differentiation of concepts and fragments enables more accurate data mining as various concept-fragment relations can exist. However, our approach is applicable also in situations where only concepts are considered.

One possible structuring of domain concepts in educational adaptive hypermedia systems is a hierarchy [5]. By mapping concepts to their parents in the concept hierarchy, we are able to work with concepts on different levels and thus discover more general usage patterns.

Several different types of concepts and fragments can exist in the educational adaptive hypermedia system. For example in the context of learning programming using programming exercises, the concept type can provide information whether the concept contains general knowledge (e.g., a programming scheme) or it contains specific knowledge (e.g., an exercise together with an example of source code – a specific application of the programming scheme). Providing availability of information about the concept type or fragment type we can look for more general behavioral patterns. Instead of working with concrete concepts, we can work with their types only on different levels of abstraction and generalization.

In the context of learning programming we use the following concept and fragment types (inspired by the ALEA system [13]):

- Concept types
 - Text
 - Programming scheme
 - Exercise
 - Test

- Fragment types
 - Text
 - Exercise definition
 - Exercise hint
 - Exercise solution
 - Source code

3.3 User knowledge level

For the task of knowledge discovery it is important to know knowledge level related to particular concept for each student. Usually, we are interested in mining data of “successful” students, i.e. those with high knowledge level of visited concepts (or good performance results), because our goal is to help students to successfully learn the domain. Estimating the user knowledge level is a complex task and it is usually performed by the educational adaptive hypermedia system itself. After the knowledge level for certain concept and user is determined it is stored into the user model.

The results of on-line tests offered by many educational adaptive hypermedia systems can also be valuable (and possibly more precise) estimation of users’ knowledge. Along with students’ results provided externally (e.g., results of “paper” tests or overall score in the course) they can help to assess the overall knowledge level of students more precisely. Overall level of knowledge is used for determination of a subset of acquired data used in the next step of knowledge discovery process for data mining.

3.4 Examples of data for knowledge discovery

In this section we provide examples of data for knowledge discovery provided by two above mentioned adaptive hypermedia systems AHA! and ALEA.

AHA!

AHA! is an adaptive web-based hypermedia system developed at the Eindhoven University of Technology [7]. As of upcoming version 3.0, the system records information about users’ activities either into the database or into a separate XML file. Log entry is created every time a resource is being accessed. One entry contains a timestamp, identification of the user’s session, name of accessed resource, identification of the user and a flag, which marks an access to the fragment. Even though AHA! does not record the Login and Logout actions, learning session of the user can still be identified thanks to the user’s session field in the log entry. The AHA! system also logs changes in the user’s model. This allows considering the context in which the action occurred during the process of knowledge discovery.

AHA! allows building a complex hierarchy of concepts, since arbitrary concept relationship types can be defined. Moreover, it is possible to assign certain type to each concept. The system maintains a table of concepts, which is filled in from the XML file. This enables an author to define concept and fragment types as described in Section 3.2.

The system uses overlay user model, i.e. values of all concept attributes for particular user are stored and maintained. Concepts representing domain knowledge fragments usually define the *knowledge* attribute, which represents user’s level of knowledge about the concept. The value of the knowledge attribute (and the other attributes as well) is updated when the user reads related page. AHA! also offers possibility to

create an on-line test, which results are used to update the value of the knowledge attribute. This way we can assess the knowledge of the user more precisely than just considering visits of fragments.

ALEA

Educational adaptive web-based hypermedia system ALEA (Adaptive LEArning) is used in the course Functional and logic programming at the Slovak University of Technology [13] since academic year 2002/2003. We have used the data gathered by this system to verify an approach for knowledge discovery in educational adaptive hypermedia proposed in this paper. ALEA stores detailed information about users' actions during learning into the database. Each time the user selects a concept or fragment, visits the fragment or explicitly changes his preferences, a log entry is created. It contains a timestamp, identification of the user, type of the action and an optional parameter of the action. Unlike AHA!, changes in the user model are not logged in ALEA. ALEA maintains only the most recent version of the user model. It means that we are only aware of the outcome of actions – the resulting user model. We do not know in which context (i.e., state of the user model) particular action occurred. This can be considered as a disadvantage, though it does not influence knowledge discovery at the level proposed here. For performing more detailed analysis of data, it seems useful to have such information.

The domain model is represented using XML. The XML file contains definition of concepts and relations between concepts. ALEA employs several concept relation types and concept types, which can be easily extended. Information fragments are stored in separate XHTML files and their types along with their mappings to the concepts are stored in the relational database.

For each visited concept ALEA stores a level of user's knowledge related to domain knowledge represented by the concept and the time when the user last visited concept together with the overall number of visits. The ALEA system also maintains information whether the user marked the concept as "understood".

4. PREPROCESSING USAGE DATA

The goal of preprocessing is to identify learning sessions of individual students and to discover potential inconsistencies and errors found in usage data (e.g., when the user fails to log out correctly). Result of preprocessing is a sequence of concepts visited during learning, together with time spent on learning individual concepts and data about the user. We also provide some filtering of sequences, for example, following their length (short sequences are considered as a noise).

In [21] the author proposes to take following steps to preprocess data from educational web-based system:

1. remove irrelevant entries
2. identify access sessions
3. map access log entries to learning activities
4. complete traversal paths (missing parts)
5. group access sessions by learners to identify learning sessions

6. integrate with other (external) data about learners and groups of learners (e.g., results of external tests)

Mentioned approach is suitable when usage data are collected by a web server. As we deal with data that come directly from an adaptive hypermedia system the preprocessing is a little simpler since we can omit steps related to mapping log entries to learning activities (this mapping is already defined in the domain model of the adaptive hypermedia system).

5. MINING ADAPTIVE HYPERMEDIA USAGE DATA

Main source of data for knowledge discovery in an educational adaptive hypermedia system are logs of users' learning activities. These logs contain records about every single user action in the adaptive hypermedia system. As a result of preprocessing, learning sessions of all users will be identified (one learning session is delimited by the Login and Logout actions). The goal of data mining in educational adaptive hypermedia is to discover characteristic patterns of navigation in users' learning sessions.

Learning sessions are similar to purchases (transactions) of electronic market customers. They both express which items (concepts or products, respectively) users prefer. Thus we focus on data mining techniques commonly used for analysis of customers' transactions. We propose using the following techniques to accomplish data mining task: association rules mining, sequential patterns mining and traversal patterns mining. Mentioned techniques have already been employed in web usage mining [15]. Association rules and sequential associations are two notable types of web patterns generally. Traversal patterns can be denoted as a special case of sequential patterns proposed for mining a navigation in both open or closed information spaces.

Association rules mining. This technique was originally proposed to analyze customer transactions in the market/basket domain. It is commonly used for example in analysis of contents of on-line shopping carts. In this case it helps to find groups of related products based on preferences of customers. In educational adaptive hypermedia systems we employ association rules mining to find relations between concepts. Even though the technique does not take into account the order of concepts, its results can still be used in the process of recommending relevant concepts. Nevertheless, discovered relations between concepts are more useful for the author of information space who can modify the course to better suit users' needs. There exist several algorithms for association rules mining. One of the first algorithms presented for this technique is *Apriori* algorithm [1].

Sequential patterns mining. Sequential patterns mining is technique similar in some sense to association rules mining. Unlike association rules this technique takes into account the order of visited concepts. However, the concepts do not necessarily have to be adjacent. That is why the interpretation of sequential patterns in context of navigation within the hypermedia information space could be quite difficult. We only know that users have frequently visited the concepts in certain order, but we cannot tell, which concepts they had visited in the meantime. These

patterns are still useful for recommending concepts, which user should visit in the future. There exist modifications of Apriori algorithm for sequential pattern discovery, e.g., *Apriori-All* and *Apriori-Some* algorithms [2]. The other algorithms (e.g., SPAM [3]) focus on improving performance for mining on large databases.

Traversal patterns mining. Traversal patterns, sometimes also referenced as contiguous sequential patterns, are used for discovering frequent subsequences of students' learning sessions. They are often used in web server logs analysis. Traversal patterns form the base for recommending relevant concepts in our approach. When properly visualized, they can be also helpful to the author of educational course as they show preferred "paths" that students tend to follow in the educational hypermedia system. The representative algorithms for traversal patterns mining are *Full Scan* and *Selective Scan* algorithms [6].

It is worth noting that patterns discovered using above mentioned techniques share some common traits. They all represent sequences or sets of items (concepts, in our case). The algorithms used for their discovery share the same idea of generating and pruning the set of candidate patterns. Thus, we can take advantage of this similarity and store discovered patterns in common knowledge base. We can also generalize the algorithms for pattern discovery in educational adaptive hypermedia and introduce common software framework shared between different algorithms.

6. RECOMMENDER SYSTEM FOR ADAPTIVE HYPERMEDIA

Patterns revealed during knowledge discovery process are used for recommendation of relevant concepts. The result of recommendation is a sequence of concepts, which student should probably visit next, together with numeric representation of suitability of these concepts.

Because we are dealing with several kinds of patterns, it is necessary to distinguish between recommendations based on each kind. We proposed ordering of pattern kinds by their suitability for recommendation in educational adaptive hypermedia systems (see Figure 1). The ordering is employed during construction of final recommendation, formed by merging partial recommendations based on three different kinds of patterns.

We consider traversal patterns the most reliable base for recommendation, because in educational adaptive hypermedia systems the proper order of concepts is crucial for the student to master a subject. Certainly, sequential patterns as well as association rules can be also used for recommendation but they provide rather broader guidance than traversal patterns. Therefore the suitability of recommended concepts has to be evaluated with respect to the kind of patterns on which the recommendation was based.

6.1 Recommender system architecture

Proposed recommender system as a back-end for an adaptive hypermedia system is based on modular software architecture. The software architecture (see Figure 2) follows the flow of two main processes carried out in the system:

- knowledge discovery and
- concept recommendation.

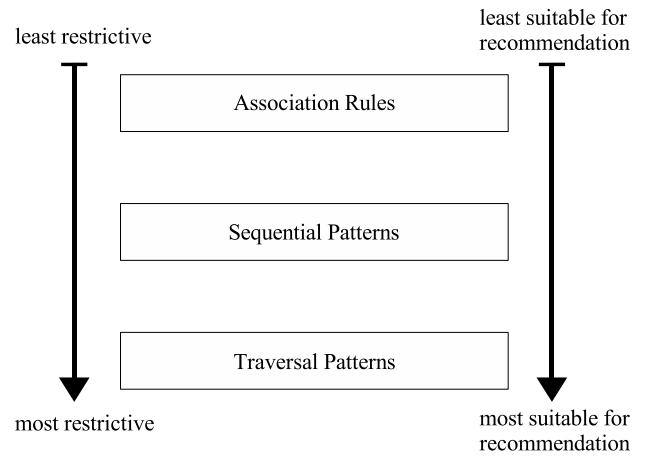


Figure 1: Suitability of various patterns for recommendation.

The system is designed to be independent of underlying adaptive hypermedia application as much as possible. Therefore the architecture incorporates a wrapper, which acts as a facade between an adaptive hypermedia system and the recommender system itself. This approach allows using various adaptive hypermedia systems as sources of data for knowledge discovery without modification of the recommender system itself. The architecture is not restricted by application area of the adaptive hypermedia system (i.e., its usage is broader than educational adaptive hypermedia).

Wrapper

All the adaptive hypermedia system dependent tasks are performed in the wrapper module. This includes usage data retrieval as well as tasks related to the processing and deploying generated adaptation rules, and the user model update. The wrapper module hence defines common interface for the access to the adaptive hypermedia system resources and data. Using the recommender system with several adaptive hypermedia systems requires a wrapper module for every adaptive hypermedia system. This module forms also the interface to an authoring tool, where adaptation knowledge is defined by the author.

Data preprocessing

The data preprocessing module takes the stream of user actions as its input. These actions are preprocessed and searched for potential errors and inconsistencies. The module also takes care of identification of users' sessions. Sessions are filtered according to their length (short sessions are dropped) and/or overall knowledge level of respective user. Finally, the module stores sessions into database of the recommender system. Information about the domain structure and types of concepts and fragments is also stored here.

Data mining

In the data mining module various algorithms for pattern mining are applied on data from the database. Traversal patterns, sequential patterns and association rules are discovered using different views on data (e.g., different levels of concepts in the domain hierarchy, concept types). Discovered patterns are stored into the knowledge base.

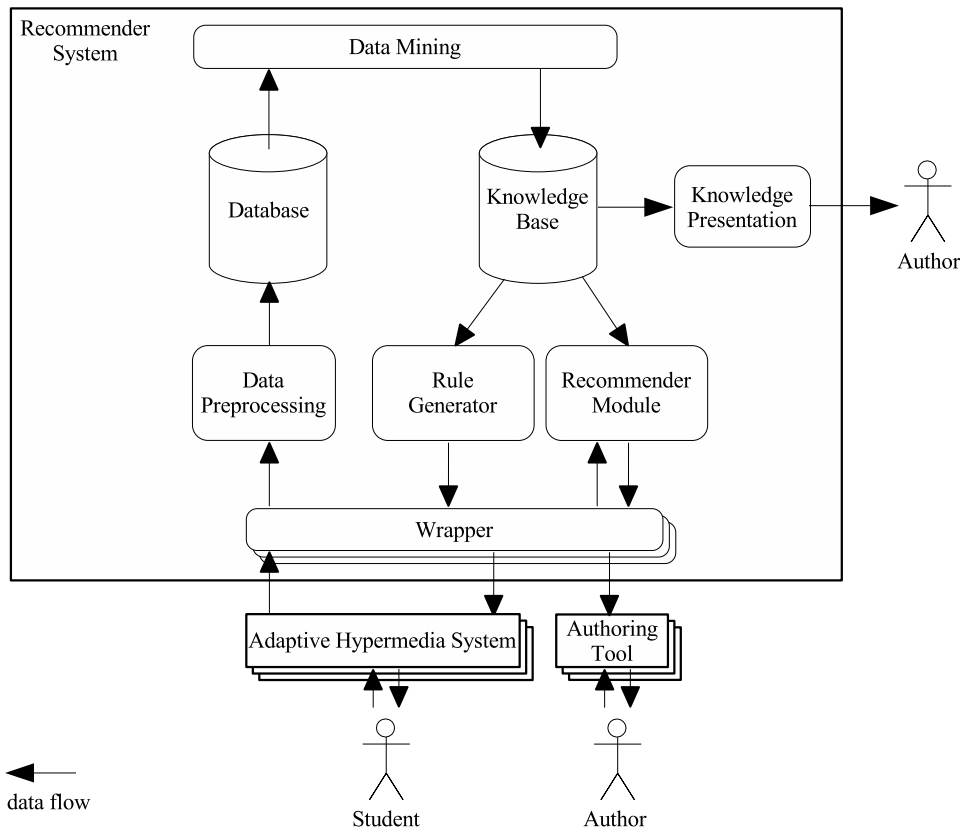


Figure 2: Architecture of recommender system.

Knowledge presentation

The main purpose of the knowledge presentation module is to present selected discovered knowledge (e.g., association rules) to the author of a course. The author can then make modifications to the domain structure and the contents of the information fragments.

Rule generator

The task of the rule generator module is to construct static adaptation rules (i.e., rules where current learning session of a student is not considered) based on patterns in the knowledge base. These rules are deployed to the adaptive hypermedia system through the wrapper.

Recommender module

The recommender module creates a sequence of recommended concepts based on the current session of a user and the knowledge stored in the knowledge base. The concepts in the current user session are matched against the ones in patterns of the knowledge base. When appropriate (similar) patterns are found, the concept recommendation is carried out. The result is a sequence of recommended concepts and the evaluation of their suitability and relevance for the user. This sequence is generated on demand of the adaptive hypermedia system (each time the user moves to another concept or fragment, or after some time period or at the beginning of each session).

6.2 Use case scenario

Sequential diagram in Figure 3 describes typical usage scenario of proposed recommender system. While users (students in case of educational adaptive hypermedia system) work with the adaptive hypermedia system, it logs and stores all their actions (e.g., logins, logouts, concept visits). When scheduled, the recommender system retrieves usage data, preprocesses it and stores it into database. Data mining algorithms are executed on stored usage data and usage patterns are discovered and stored in the knowledge base.

With knowledge base filled up, the recommender system can recommend a sequence of relevant concepts for particular user, when asked to do so. An adaptive hypermedia system provides the recommender system with the user identification and his current session and the recommender system returns recommended sequence of concepts, which the user should visit next. The adaptive hypermedia system may annotate concepts in the sequence according to suitability evaluation provided by the recommender system. It should be noted that it is up to adaptive hypermedia system how the recommendation will be handled. It can be just annotated and displayed to the user (like it is shown in Figure 4) or it can become a part of recommendation which takes multiple aspects (e.g., learning goals or skills of student) into account.

When author of the content decides to improve it, visualization of discovered knowledge can help to decide which modifications should be made. The author can also com-

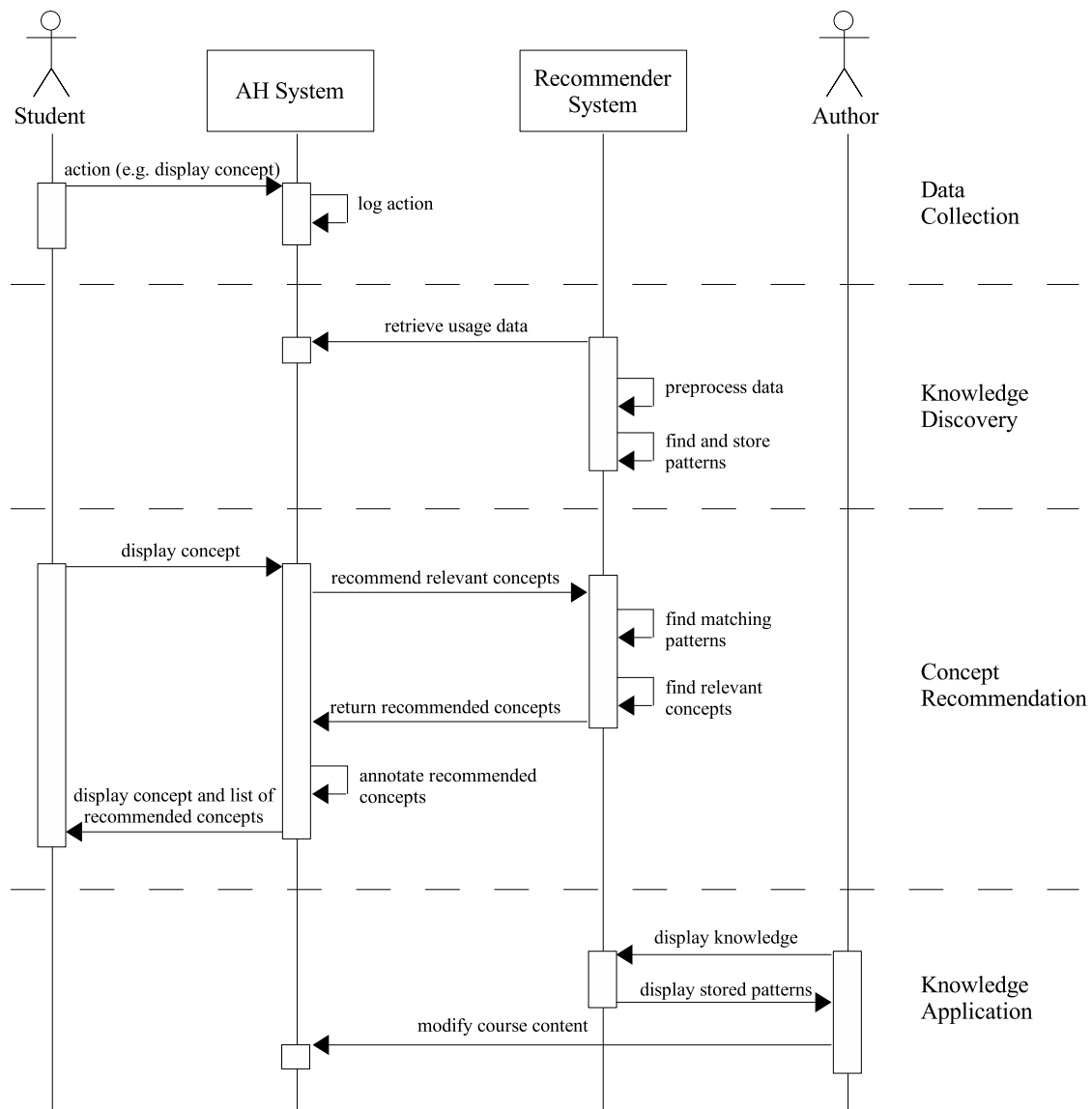


Figure 3: Typical scenario of the recommender system usage.

pare historical records and evaluate the effect of his previous changes to the adaptive hypermedia system content.

6.3 Experiments

To verify our approach to knowledge discovery in adaptive web-based hypermedia we developed a prototype of the recommender system. We implemented three data mining algorithms for pattern discovery (traversal patterns, sequential patterns and association rules) suitable for educational hypermedia applications.

In some experiments with traversal patterns mining we created also maximal forward references [6] where returns to previously visited fragments are ignored. We have found that in this way loss of information is not acceptable. This is a consequence of the fact that the navigation in educational hypermedia has different characteristics as a navigation (browsing) in the web information space. The “return

back” action in learning session is important part of behavioral patterns in educational hypermedia systems and should not be ignored.

Results of the knowledge discovery are markedly dependent on source data. We chose the ALEA educational adaptive hypermedia system [13] for our experiments since we have had available usage logs gathered during 3 years of its usage in the Functional and logic programming course. We successfully preprocessed usage logs and were able to identify learning sessions of individual users. At time of experiments the database contained 1 170 sessions.

To evaluate recommendations carried out by the system, we split available data into two sets. Larger one (with size of 90% of original one) was used for pattern discovery (i.e., training). User sessions contained in smaller set were used to evaluate recommendations. Prefix of each session in the test set was passed to the recommendation algorithm. Result of

The screenshot shows the ALEA system interface. At the top, there is a navigation bar with 'Logout', 'Index', 'History', 'Options', 'Change password', and 'Bookmarks'. Below this is a 'Parents' section with 'Schema Display'. The 'Neighbors' section lists several examples, with 'Example product' selected and circled in black. The 'Recommendation' section is also circled in black and lists various options like 'Schema Display', 'Schemas', 'Schema Filter', 'Schema Reduction', 'Schema Test', 'Example sum', and 'Example max'. The main content area shows the 'Example product' details, including a 'problem statement', 'hint', and 'solution' tab. The 'problem statement' tab is active, displaying the predicate $product(+Vector, +Number, ?Result)$ and two examples with their solutions. There is an 'Add to bookmarks' button and an 'I have understood' button. At the bottom, there is a 'Comments' section with a table showing a comment by Tomáš Búci on 28.11.2002 at 17:34:54. A 'Show all' button and an 'Add' button are also present.

Figure 4: Recommendations incorporated into ALEA system.

the recommendation was compared with a concept following right after the prefix in the session (i.e., the concept that student actually had visited).

The results of experiments stated in Table 1 show *hit ratio* and *average ranking* of recommendations based on three different kinds of patterns as well as final recommendation (combination of first three mentioned recommendations), where

- *hit ratio* represents number of recommendations which recommended the concept that the student actually had visited next, and
- *average ranking* represents average position of such concept in the sequence of recommended concepts (lower ranking is better).

It is worth noting that high hit ratio used does not necessarily indicate a successful recommendation. Successful recommendations help the student to master whole domain. To evaluate whether recommendations were successful or not, the feedback from the student is needed. Nevertheless, the hit ratio is still a good measure for comparing recommendations based on different kinds of patterns.

Amongst recommendations based on single kind of patterns, those based on sequential patterns had the highest hit ratio (though their ranking was poor). The final recommendation performed best, suggesting that recommendations based on different kinds of patterns partly complement each other. Strong contribution of sequential patterns to the final recommendation can be seen – average ranking of the final recommendation was quite high.

Pattern kind	Hit ratio	Average ranking
Association Rules	22.67%	2.22
Sequential Patterns	50.64%	7.09
Traversal Patterns	34.32%	1.27
Final Recommendation	57.63%	5.97

Table 1: Evaluation of recommendations based on different kinds of patterns.

7. CONCLUSIONS AND FUTURE WORK

The adaptation in most of current educational adaptive hypermedia systems is driven by a fixed set of rules. The construction of such rules is a complex task and the author may not be able to completely assess all intricacies found in different learning styles of students. Techniques for data mining can provide knowledge needed either to recommend students concepts according to their characteristics or to assist the author of the course to improve the structure of the domain.

Main contribution of this paper is a proposal of adaptation model development support using data mining techniques. We analyzed data provided by educational adaptive hypermedia systems, which can be useful in the process of knowledge discovery. Usage data (i.e., logs of users' actions) are valuable source for knowledge discovery. We selected three well-known techniques for mining patterns in such data. In our approach, association rules, sequential patterns and traversal patterns represent automatically generated knowledge about students' behavior. Based on a student's current learning session and discovered patterns,

the recommender system is able to recommend a sequence of concepts, which the student should study next.

Important outcome presented in this paper is architecture of the recommender system, which connects the process of knowledge discovery with on-line recommendation of concepts based on discovered knowledge. The architecture of system allows using different underlying adaptive hypermedia systems, enabling us to share common knowledge (not specific to certain adaptive hypermedia system and its content). The architecture also incorporates the module for visualization of discovered knowledge helping the author of the course to prepare modification to the content.

Future development of the recommender system will focus on enhancing the set of data mining algorithms, e.g., to use clustering techniques in order to discover user clusters according their learning style. We plan to evaluate quality of recommendations carried out by our recommender system using data produced by other educational hypermedia systems. The evaluation will be based on a feedback from students as well as on results of the recommendation performed on a testing set of data. We also intend to conduct more experiments with the other educational adaptive hypermedia systems and with their different content with primary attention to AHA! considering also data related to user model changes.

8. ACKNOWLEDGMENTS

This work has been partially supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic grant No. VG1/ 0162/03, and by Science and Technology Assistance Agency under the contract No. APVT-20-007104.

9. REFERENCES

- [1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In J. B. Bocca, M. Jarke, and C. Zaniolo, editors, *Proc. of 20th Int. Conf. Very Large Data Bases, VLDB*, pages 487–499. Morgan Kaufmann, 12–15 1994.
- [2] R. Agrawal and R. Srikant. Mining sequential patterns. In P. S. Yu and A. S. P. Chen, editors, *11th Int. Conf. on Data Engineering*, pages 3–14, Taipei, Taiwan, 1995. IEEE Computer Society Press.
- [3] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu. Sequential pattern mining using a bitmap representation. In *Proc. of the 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pages 429–435, 2002.
- [4] P. Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, Ten Year Anniversary Issue, 11(1/2):87–110, 2001.
- [5] P. Brusilovsky. *Authoring Tools for Advanced Technology Learning Environments*, chapter Developing Adaptive Educational Hypermedia Systems: from Design Models to Authoring Tools, pages 377–409. Kluwer Academic Publishers, 2003.
- [6] M. S. Chen, J. S. Park, and P. S. Yu. Data mining for path traversal patterns in a web environment. In *Proc. of 16th Int. Conf. on Distributed Computing Systems*, pages 385–392, 1996.
- [7] P. De Bra, A. Aerts, B. Berden, B. de Lange, B. Rousseau, T. Santic, D. Smits, and N. Stash. AHA! the adaptive hypermedia architecture. In *HT'03: Proc. of the 14th ACM Conf. on Hypertext and Hypermedia*, pages 81–84. ACM Press, 2003.
- [8] P. De Bra, L. Aroyo, and A. Cristea. *Web Dynamics: Adapting to Change in Content, Size, Topology and Use*, chapter Adaptive Web-Based Educational Hypermedia, pages 387–410. Springer Berlin, 2004.
- [9] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. E. Uthurusamy. *Advances in Knowledge Discovery in Data Mining*. AAAI Press/MIT Press, 1996.
- [10] S. Gündüz and M. Tamer Özsu. A web page prediction model based on click-stream tree representation of user behavior. In *KDD'03: Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pages 535–540. ACM Press, 2003.
- [11] IEEE. IEEE P1484.2/D7, 2000-11-28. Draft Standard for Learning Technology. Public and Private Information (PAPI) for Learners. Available at <http://ltsc.ieee.org/>, 2000.
- [12] G. Klyne, F. Reynolds, C. Woodrow, H. Ohto, J. Hjelm, M. Butler, and L. Tran. Composite Capability/Preference Profiles (CC/PP): Structure and vocabularies 1.0. W3C Recommendation, 2004.
- [13] R. Kostelnik and M. Bieliková. Web-based environment using adapted sequences of programming exercises. In M. Beneš, editor, *Proc. of Int. Conf. on Information Systems Implementation and Modelling*, pages 33–40, 2003.
- [14] M. Koutri, N. Avouris, and S. Daskalaki. *Adaptable and Adaptive Hypermedia Systems*, chapter A survey on web usage mining techniques for web-based adaptive hypermedia systems. Idea Publ. Inc., 2004.
- [15] B. Mobasher, R. Cooley, and J. Srivastava. Automatic personalization based on web usage mining. *Communications of the ACM*, 43(8):142–151, 2000.
- [16] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa. Using sequential and non-sequential patterns in predictive web usage mining tasks. In *Proc. of the IEEE Int. Conf. on Data Mining*, pages 669–672, 2002.
- [17] B. Mobasher and M. Nakagawa. Impact of site characteristics on recommendation models based on association rules and sequential patterns. In *Proc. of the IJCAI'03 Workshop on Intelligent Techniques for Web Personalization*, 2003.
- [18] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. Spyropoulos. Web usage mining as a tool for personalization: A survey. *User Modeling and User-Adapted Interaction*, 13(4):311–372, 2003.
- [19] C. Romero, S. Ventura, P. De Bra, and C. de Castro. Discovering prediction rules in AHA! courses. In *Proc. of the User Modeling Conf.*, pages 25–34, 2003.
- [20] C. Shahabi, A. M. Zarkesh, J. Adibi, and V. Shah. Knowledge discovery from users web-page navigation. In *RIDE '97: Proc. of the 7th Int. Workshop on Research Issues in Data Engineering (RIDE '97) High Performance Database Management for Large-Scale Applications*, page 20. IEEE Computer Society, 1997.
- [21] O. R. Zaïane. Web usage mining for a better web-based learning environment. In *Proc. of Conf. on Advanced Technology for Education (CATE'01)*, pages 60–64, 2001.