Esta es la versión de autor de la comunicación de congreso publicada en:
This is an author produced version of a paper published in:


Copyright: © 2008 AACE

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription
Using Decision Trees for Discovering Problems on Adaptive Courses

Javier Bravo\(^1\), César Vialardi\(^2\) and Alvaro Ortigosa\(^1\)
\(^1\)Computer Science Department, Universidad Autónoma de Madrid, Spain
\(^2\)Computer Science Department, Universidad de Lima, Peru
{javier.bravo, alvaro.ortigosa}@uam.es
cvialar@correo.ulima.edu.pe

Abstract: Adaptive Hypermedia Systems personalize the learning experience of each user, by providing learning materials adapted to his/her needs, preferences, personal characteristics, etc. The goal is to make the learning process easier or more efficient. However, on the teacher side the improvement and evaluation of these systems are difficult tasks, especially when there are multiple student profiles or huge amount of interaction data of students. In this work, data mining methods, and specifically decision trees, are used for helping in both improvement and evaluation. Our work consists of analyzing two data sets by using decision trees. The first data set contains the interaction data of 24 real students, and the second data set is composed of synthetic data about 100 students. The results of these analyses demonstrated that 24 students is a small data set when decision trees are used. However, the tree showed information relating to the practical activities in which students had more problems for completing them providing useful feedback to the course designer.

Keywords: adaptive hypermedia, adaptive educational hypermedia, decision tree.

Introduction

Adaptive Educational Hypermedia (AEH) Systems (Brusilovsky, 2003) support customization to individual students by recommending teaching activities to each student according to his/her needs. These systems adapt the educational contents to different dimensions of each learner profile such as: level of knowledge, goals, educational context, or learning styles (Paredes & Rodríguez, 2004), among others. The final goal is to provide each student with a personalized learning experience oriented to improve the outcome of the process.

One of the main obstacles for a wider adoption of AEH systems is that the creation of adaptive courses is far from an easy task. In order to allow the adaptive system to know what to present at a given moment to each student, the author of the AEH system needs to structure the knowledge and to define the mapping to the educational material. This mapping should vary according to each student profile so as to facilitate their learning based upon their needs. Inevitably, doing so complicates the AEH design.

Evaluation of AEH systems is also a complex and time-consuming task. Ideally, teachers should analyze how adaptation is working for each of their students via their profiles. However, in most AEH systems, the teacher defines rather small knowledge modules and rules to relate these modules. Based on these rules, the system structures the material to be presented to every student on the fly, depending on his/her profile. As a result of this dynamic organization of educational resources, the teacher cannot visualize the complete course structure and how it varies for each student. Furthermore, the evaluation of the effectiveness of the course is hindered by lack of student feedback in on-line learning environments, where discussions that occur in a traditional classroom are missing. Even if the results of tests taken by distance learners are available, they only provide limited information regarding student knowledge. They may also provide little information about how student interacts with the system and also what information was presented to them.

As a consequence, there is a need for methods and tools specially designed to support development and evaluation of adaptive material. In this context, we have proposed the use of data mining techniques to assist on the authoring process (Vialardi et al., 2007; Bravo et al., 2007) with the teacher’s perspective in mind. The goal is to mining the data about distance education on a collective way, just as teachers would do in a classroom when they adapt a course for students (Zaïane, 2002).

In the previous works, we have shown how data mining techniques can be used to discover and present relevant pedagogic knowledge to the teachers (Bravo et al., 2007; Vialardi et al., 2007; Bravo et al., 2008). The
techniques were tested using synthetic data. **Simulog** (SIMulation of User LOGs), a tool able to simulate student behavior by generating log files according to specified profiles (Bravo & Ortigosa, 2006), generated the data.

Although synthetic data resulted in a good initial test for the method and tool developed, the next logical step was to test them with data from real students. For this reason, this paper aims to show how the proposed method works with real data. In this direction, we have analyzed the interaction data from another previous case study; the goal of the case study was to evaluate the effectiveness of an adaptive course, by contrasting the performance of students accessing the adaptive material against students accessing traditional, linear material (Muñoz & Ortigosa, 2007). In that original work, logs were collected as part of the normal functioning of the AEH System, but they were not analyzed.

The main result of this work is to show that data mining techniques can help on improving adaptability on AEH Systems when analyzing data from real students. It also shows limitations of the approach, as it cannot be applied with only small samples are analyzed.

The next section briefly describes the state of the art and related work. Section three shows the structure of interaction data of students. Section four explains the two experiments carried out in this work, and sections five and six describes conclusions and future work. Finally, section seven outlines the conclusions.

**State of the art**

Since the Internet has become a fundamental platform for information distribution, numerous attempts can be found on implementing different techniques with the goal of customizing information delivery. In particular, AEH systems provide to users more easy for learning such as platform independent and free locations. Therefore, the main challenge of these systems is to adapt the information showed to users regarding to personal features of them such as level of knowledge, type of culture, and motivation.

In the last years data mining has turned to the most used techniques not only in marketing decisions, but also in e-Learning arena. As a result, the last year a great amount of work has been dedicated to help teachers on finding student capabilities in an adaptive course. The most widespread techniques are **classification algorithms** (Bravo et al., 2007) and **association rules** (García et al., 2007). The first one supports the improvement of AEH courses, applying decision tree technique and finding the most relevant leaves of the tree afterwards. These leaves are presented to the teacher in order to improve the course design. Respect to association rules, García and Romero runs a comprehensive study of the advantages and disadvantages of using association rules for AEH courses.

Merceron and Yacef published another study where association rules are extracted from data of Logic-ITA (Merceron & Yacef, 2005). Logic-ITA is a Web-based tutoring tool used to offer courses, specifically to help students in their formal logic exercises. This system can also report the progress of students to their teacher. Similar to our study, Merceron and Yacef focused their work on detecting patterns of mistakes made by the students by using association rules. They found a relation between a given number of mistakes of students applying relevance methods such as Chi-Square, cosine and contrasting rules (Merceron & Yacef, 2007). In this study, the association rules are used to relate the categories of the student profile to the success or failure in practical exercises; nevertheless, Merceron uses the association rules to find errors as students solve exercises, displaying results in real time.

In a different research study, Superby (Superby et al., 2006) used data mining and statistical techniques to determine factors that influenced success in first year college students. The goal of this study was to classify students in three different groups: low, medium and high risk of failure in university level study. This study tries to explain the different academic performances of students by predicting the probability of success of any given student. Similar to our study, Superby uses ID3 algorithm (Quinlan, 1993), obtaining as in our case, a tree that has the advantage of being relatively simple to interpret.

**Data description**

In this work we used the logs that came from the research of the analysis of adaptation carried out by Muñoz and Ortigosa in 2006 (Muñoz & Ortigosa, 2006). The goal of that research was to evaluate the impact of using adaptive educational tools in the performance of secondary school students. The experience was carried out with 47 students, and the authors developed two courses with the goal of testing the effects of adaptation: one course had adaptation rules and the other provided the same content for all students. The contents of both courses were an introduction to
whole numbers for students of first grade of secondary mandatory education, according to the Spanish educational system. This course was divided into seven lessons: Ordenacion (Sorting), Valor Absoluto (Absolute Value), Suma (Addition), Resta (Subtraction), Parentesis (Operations with parentheses), Multiplicacion (Multiplication), and Operaciones Combinadas (Combining Operators). The students were assigned to the adaptive and non-adaptive version of the course randomly. As a result of this assignment, 24 students followed the adaptive version and 23 the non-adaptive. Since the present work is interested on supporting maintenance of adaptive courses, it focuses on analyzing the logs produced for these 24 students with the adaptive version of the course.

The course was delivered through the TANGOW system (Carro et al., 1999). TANGOW provides a flexible support for the creation of courses with different adaptive features. This system stores the interaction of students with the system in a log file for each student. Each entry of the log file is composed by the following attributes: activity id, type of activity, complete, grade, number of visits, timestamp and action of user. The table 1 shows more details of the parameters.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Activity</strong></td>
<td>Activity id</td>
</tr>
<tr>
<td><strong>Complete</strong></td>
<td>It represents the level of completeness of the activity. If the activity is composed, it takes into consideration the completeness of all sub-activities. It is a numeric parameter that ranges from 0 to 1. Value 0 indicates the activity was not completed and value 1 indicates the activity was fully completed</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td>The grade given to each activity. In practical (P) activities it is usually calculated from a formula provided by the teacher. In composed activities it is the arithmetic mean of sub-activity grades. Value 1 indicates the activity was finished with success and value 0 indicates the activity was finished with failures.</td>
</tr>
<tr>
<td><strong>NumVisit</strong></td>
<td>Number of times the student has visited the activity.</td>
</tr>
<tr>
<td><strong>Action</strong></td>
<td>The action executed by the student; these are defined by the TANGOW system: &quot;START-SESSION&quot;: beginning of the learning session. &quot;FIRSTVISIT&quot;: first time an activity is visited. &quot;REVISIT&quot;: any visit to the activity following the first one. &quot;LEAVE-COMPOSITE&quot;: the student leaves the (composed) activity. &quot;LEAVE-ATOMIC&quot;: the student leaves the (atomic) activity.</td>
</tr>
<tr>
<td><strong>TimeStamp</strong></td>
<td>The time when the student visits the activity.</td>
</tr>
<tr>
<td><strong>ActivityType</strong></td>
<td>The type of activity: theoretical (T), exercises (P) and examples (E).</td>
</tr>
</tbody>
</table>

**Table 1: Description of parameters of an instance of log file**

An example of the interactions of the student a1 with the system is showed in table 2. This table shows the student starts the course named curso-enteros (main composed activity of the course – course of whole numbers). After two minutes he/she leaves (action LEAVE-COMPOSITE) this activity and he/she starts (action FIRSTVISIT) the activity Ordenacion (Sorting) which is a theoretical activity (activityType=T). Parameters complete and grade have values 0.0 because it is the first time the student visits this activity.

| <entry activity="curso_enteros" activityType="T" complete="0.0" grade="0.0" numvisits="1" timestamp="2005-12-14T11:14:53.938+01:00" type="FIRSTVISIT" /> |
| <entry activity="curso_enteros" activityType="T" complete="0.0" grade="0.0" numvisits="1" timestamp="2005-12-14T11:16:27.828+01:00" type="LEAVE-COMPOSITE" /> |
| <entry activity="ordenacion" activityType="T" complete="0.0" grade="0.0" numvisits="1" timestamp="2005-12-14T11:16:27.828+01:00" type="FIRSTVISIT" /> |
Table 2: Interactions of a given student

Afterwards, he/she visited the activity named o1 and started the first practical activity called eo1_n1. The student needed for finishing this activity 44 seconds, and he completed it with success (complete=1 and grade=1). It is worth mentioning that the practical activity’s name is composed by the following notation:

\[ e[Name\_of\_Lesson][Number\_of\_Chapter]_[Experience][Number\_of\_Chapter] \]

In addition, the abbreviations of the parameters Name_of_Lesson, and Experience are described in the following table (Tab. 3).

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Name_of_Lesson</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>Sorting</td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>Absolute Value</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>Addition</td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>Subtraction</td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>Operations with parentheses</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>Multiplication</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>Combining Operators</td>
<td>Novice</td>
</tr>
<tr>
<td>b</td>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td>Advanced</td>
</tr>
<tr>
<td>a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Abbreviation of the parameters

For example, the exercise eo1_n1 can be read as “exercises of the first chapter of Sorting for normal students”, whereas the activity o1 (a theoretical activity) means “the first chapter of Sorting”. Furthermore, the table 3 demonstrates three different adaptations in the activities: novice, normal, and advanced. Though our work does not include advanced students, it is possible that novice and normal students can carry out activities for advanced students. For example, if a novice student received a grade of the previous lesson more than 0.7; the system presents this student the advanced version of the current activity.

Data analysis

The research carried out by Muñoz and Ortigosa obtained relevant conclusions. The students that followed the adaptive version of the course enhanced their performance in the Mathematics course. However, a preliminary analysis of logs indicates that 61% of the activities were unsuccessful (Tab. 4 shows the details of this analysis).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Number of instances</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>220</td>
<td>74</td>
</tr>
<tr>
<td>13,14</td>
<td>80</td>
<td>26</td>
</tr>
<tr>
<td>Experience</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>286</td>
<td>95</td>
</tr>
<tr>
<td>Novice</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 0.5</td>
<td>182</td>
<td>61</td>
</tr>
<tr>
<td>&gt; 0.5</td>
<td>118</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 4: Preliminary analysis of data

Therefore, this analysis indicates potential problems in the adaptation of the course. In other words, the adaptation of certain activities was possibly not appropriated for different profiles of students. Therefore this symptom might indicate possible problems in the course definition or in the contents of activities. Thus, the following experiments aim to discover the activities in which certain profiles of students showed more problems in practical activities.

For the next experiments Weka tools and Simulog were used. Weka tools (Witten & Frank, 2005) are data mining tools which provide several classification methods such as a modified C4.5 algorithm of decision trees developed by Quinlan (Quinlan, 1993).

**First Experiment**

The first experiment consisted of analyzing interaction data of 24 students. The goal was to find potential problems in the adaptation. With this purpose data mining tools and the method presented in (Bravo et al., 2007) were used. In particular, decision trees were used in this research study.

Classification algorithms learn a model based on the instances of the dataset, where each instance is described as a collection of attributes. Using the model learnt, it is possible to classify new instances, whose class is unknown a priori.

Nodes in decision trees involve testing a particular attribute of the instance to be classified. Depending on the value of attribute, the corresponding descendant branch is followed. This procedure is recursively applied until a leaf is reached. Usually, each leaf has a label with the class to be assigned to the instances that reach that leaf. Therefore decision trees have the advantage of explicit representation of the knowledge acquired.

In order to build a decision tree, it is necessary to select several attributes and the class attribute. In our case, selected attributes are related to student profiles (age and experience attributes) and activities (name of the activity). In our scientific research, classification attribute is the success variable, which indicates success or failure in a practical activity.

The results of this analysis show that the decision tree resulting from analyzing the logs of the 24 students has only one level (Fig. 1). For this reason, there is no information about profiles of students who had problems in practical activities. However, the tree shows the activities completed by the students. Therefore the figure 1 presents information about whether students had success or failure, in general, in a given activity. For example, the tree shows that the students had several problems in activities ep1_b1 (exercises of the first chapter of Operations with Parentheses for novice students), em2_b1 (exercises of the second chapter of Multiplication for novice students) and ec3_a1 (exercises of the third chapter of Combining Operators for advanced students). It is worth noting that activity is not the same than instance, because one activity of each student can generate several instances in the interaction data. The reason of this difference is because, when a student fails a given activity, it is presented again in the future by TANGOW system. Due to the student repeats this activity until he/she passes it. For example, the branch of the activity ec3_a1 (exercises of the third chapter of Combining Operators for advanced students) shows one misclassified instance out of 22. That is to say, 21 instances (they are not from 21 students) of previous activity
contain the “no” value in the success variable. Thus, the percentage of failures in this activity (i.e. success=no) is more than 95%. In addition, the activities ep1_b1 (exercises of the first chapter of Operations with Parentheses for novice students) and em2_b1 (exercises of the second chapter of Multiplication for novice students) presented more than 85% of failures.

---

\[
\begin{align*}
\text{activity} &= \text{ep1}_\text{b1}: \text{yes} \quad (24.0/3.0) \\
\text{activity} &= \text{ev1}_\text{ni}: \text{no} \quad (24.0/6.0) \\
\text{activity} &= \text{col}_\text{ni}: \text{no} \quad (24.0/11.0) \\
\text{activity} &= \text{es2}_\text{ni} \\
& \quad | \text{age} \leq 12: \text{no} \quad (18.0/8.0) \\
& \quad | \text{age} > 12: \text{yes} \quad (6.0/2.0) \\
\text{activity} &= \text{cr1}_\text{bl}: \text{yes} \quad (23.0/6.0) \\
\text{activity} &= \text{er2}_\text{bl}: \text{no} \quad (25.0/10.0) \\
\text{activity} &= \text{ep1}_\text{bl}: \text{no} \quad (28.0/4.0) \\
\text{activity} &= \text{em2}_\text{bl}: \text{no} \quad (24.0/3.0) \\
\text{activity} &= \text{ec3}_\text{bl}: \text{no} \quad (22.0/10.0) \\
\text{activity} &= \text{ec3}_\text{al}: \text{no} \quad (2.0) \\
\text{activity} &= \text{cr1}_\text{al}: \text{yes} \quad (12.0) \\
\text{activity} &= \text{er2}_\text{al}: \text{no} \quad (12.0/1.0) \\
\text{activity} &= \text{cr1}_\text{pr}: \text{no} \quad (12.0) \\
\text{activity} &= \text{er1}_\text{nl}: \text{yes} \quad (2.0) \\
\text{activity} &= \text{cr2}_\text{nl}: \text{no} \quad (2.0) \\
\text{activity} &= \text{ep1}_\text{al}: \text{no} \quad (9.0/1.0) \\
\text{activity} &= \text{ev1}_\text{bl}: \text{yes} \quad (2.0/1.0) \\
\text{activity} &= \text{col}_\text{bl}: \text{yes} \quad (2.0) \\
\text{activity} &= \text{col}_\text{bl}: \text{yes} \quad (2.0) \\
\text{activity} &= \text{es2}_\text{bl}: \text{yes} \quad (2.0) \\
\text{activity} &= \text{es2}_\text{bl}: \text{no} \quad (1.0) \\
\end{align*}
\]

Number of Leaves : 23
Size of the tree : 25

---

Figure 1: Decision tree of interaction data in experiment 1

Second Experiment

The first experiment demonstrated that a group of students had serious problems with three exercises in the adaptive course (Fig. 1). Unfortunately the resulting decision tree does not contain information of student profiles. The incomplete information might be due to the size of interaction data, since 300 instances (24 students) was not sufficient for applying data mining methods in this experiment. For this reason, this experiment is split in two parts. The first phase consisted of enlarging the previous data set by using simulating methods. The analysis of simulated data by using data mining methods is carried out in the second phase.

The first step in the first phase was to establish the proportion and profiles of students who had failed activities ep1_b1, em2_b1 and ec3_al by analyzing the interaction data. Table 5 presents the results of analysis of the data. As illustrated in the following table 75% of students with “age=12” and “experience=normal” had failures in the activity “exercises of the first chapter of Operations with Parentheses for novice students” (ep1_b1). This proportion represents the percentage of students with a given profile compare to all students (i.e. 24 students in this case). Therefore 95% of students had problems successfully completing this activity. The second step was to generate interaction data by using a simulator tool (Simulog). The proportions and profiles presented in table 5 are utilized as inputs in the simulator. Thus, at the end of this phase 100 students were generated. Figure 3 illustrates the interface of Simulog when 100 students are generated through this tool. On one hand, the left side of Simulog presents the proportion and profiles of students. On the other hand, the right side contains the problems in the adaptive course (anomalies or symptoms of potential problems in the adaptation). In addition, nine anomalies were set in the simulator, however only the first one is showed in figure 3.
<table>
<thead>
<tr>
<th>Activity</th>
<th>Age</th>
<th>Experience</th>
<th>Students with failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>ep1_b1</td>
<td>12</td>
<td>normal</td>
<td>75%</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>novice</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>normal</td>
<td>16%</td>
</tr>
<tr>
<td>em2_b1</td>
<td>12</td>
<td>normal</td>
<td>54%</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>novice</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>normal</td>
<td>21%</td>
</tr>
<tr>
<td>ec3_a1</td>
<td>12</td>
<td>normal</td>
<td>58%</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>novice</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>normal</td>
<td>16%</td>
</tr>
</tbody>
</table>

Table 5: Analysis of activities with failures of real interaction data

Second phase consisted of finding patterns of student profiles that had problems in the activities (see Tab. 5). The same methods were used in this experiment as the first one. Also selected variables and the classification variable for decision tree were the same. In other words, selected attributes were: age, experience and activity; and classification variable was the success attribute.

Figure 2: Part of decision tree of logs of second experiment

The resulting decision tree is composed of three levels, i.e. two levels for student profile (experience and age) and one for activity. Only relevant branches related to three previous activities are presented in the figure 2. As shown in this figure the tree provides a lot of information on the student profile for the activity ep1_b1. Since, the tree shows that students with normal experience and age 12 years old had several difficulties in completing successfully the activity ep1_b1. However, there is less information about the students profile for “exercises of the chapter 2 of Multiplication for novice students” (em2_b1) activity. Due to the tree does not provide information about the experience of students, but it shows the failures of the 12 year old students in this activity. Finally, there is
not information of student profile for last activity, \textit{ec3\_a1}. However, the tree illustrates that a huge group of students had problems in this activity.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Simulog.png}
\caption{Interface of Simulog}
\end{figure}

\section*{Conclusions}

This work demonstrates the utility of data mining methods, especially decision trees, when they are applied for assessing the effectiveness of AEH systems. A method based on decision trees has already been proposed (Bravo et al., 2007), but the novelty of this work is to show its performance with real student data.

With this purpose, two experiments were carried out. The first experiment showed that the resulting decision trees contained information related to practical activities in which students had more problems. However, there was no information in the tree regarding which student profiles had more failures in activities. This lack of information could be related to the small size of data set (the sample consisted of data from 24 students).

The second experiment confirmed that the lack of information in the tree is really due to the size of sample. In this second case synthetic data from 100 students were analyzed; the synthetic data set contained around four times more interactions than the data set of real students. Synthetic data was generated by simulating students with the same distribution of attribute values and behaviors (interacting with the system) that the original sample of real students. This is to say, interaction data of simulated students were slightly different but with similar behaviors than interaction data of real students. With this new data set it was possible to identify through classification trees the problematic student profiles or, in other words, which was the profile (or profiles) of students that had more problems in the adaptive course. These patterns are symptoms of potential problems in the adaptation of the adaptive course.

Therefore, these experiments demonstrated how decision trees, usually a classification tool, can also be used for providing teachers with information about unusual patterns in student behavior, as proposed in (Bravo et al., 2007). Related to adaptive courses, these behavioral patterns can possibly indicate failures in adaptation of contents or problems in the structure of a given course.
Nonetheless, processing the decision trees can be a difficult task. Detecting problematic patterns in the trees is a complex task and it is increased with the growth of size of the tree. For example, the adaptive course used for this work is composed of 22 practical activities, which required a lot of work to extract the relevant information from the tree. Therefore it is extremely complicated to find adaptive problems by just taking a look at the resulting decision tree. Currently, we are working in finding a possible solution for solving this weakness of the proposed method and it is also part of future work.

Future work

The accuracy of the results provided by decision trees depends on the quality of data (interaction data in our case). In other words, proportion of student profiles, number of practical activities of each student and size of interaction data might change significantly the result of the tree. Therefore part of our future work is centered on supporting the results of decision trees with other data mining methods such as association rules and clustering. The idea is to overcome the problems derived from the lack of quality data by combining information obtained from different sources or through different methods.

In this research, we have found problems in processing the decision tree due to its size (it was composed by more than 22 leaves). In addition, the teachers do not have enough knowledge for understanding the data mining results. Therefore, other step in our future work will consist of developing a tool for assisting instructors in improving and maintaining adaptive courses. An initial version of this tool, called ASquare (Vialardi et al., 2007), has already been implemented and a second improved version was presented at the 2008 Hypertext conference (Bravo et al., 2008).

The sample of real data analyzed in this work is small: only 24 students took the adaptive course. In addition, each hypermedia adaptive system has different types of interaction data and in this work only one adaptive system was used (the TANGOW system). Thus, future work includes collaborations with other research groups, in order to be able to analyze logs generated from different AEH systems.

References


Acknowledgements

This work has been funded by Spanish Ministry of Science and Education through the HADA project TIN2007-64718. Cesar Vialardi is also funded by Fundación Carolina. We would like to thank Félix Muñoz for allowing us to use the interaction data of his students. We would also like to thank to Elisabeth Rodríguez who contributed in this work with her helpful comments.