A User Modeling Approach to Web Based Adaptive Educational Hypermedia Systems

Ilhami Colak*, Member IEEE, Seref Sagiroglu**, Member IEEE, and H. Tolga Kahraman*,

*Gazi University, Faculty of Technical Education, Electrical Education Department
**Gazi University, Faculty of Engineering and Architecture, Computer Engineering Department
{icolak, ss, hamditolga}@gazi.edu.tr

Abstract

In this paper, a new user modeling approach has been developed to determine the knowledge status of user in Adaptive Educational Hypermedia System (AEHS). The approach is based on forming the domain model and determining relations among elements of that model to decide the knowledge status of user from domain model independently. The proposed method provides flexibility to individual learning or studying steps. In comparison to the literature, the proposed approach enables quick and powerful adaptation for matching instructional needs of users. The difficulties faced in developing such a system are purifying the gathered data, obtaining and evaluating useful data and providing a powerful adaptation effect in a short time.

Key words: User Modeling Approach, Adaptive Educational Hypermedia Systems, AEHS

1 Introduction

Use of web-based applications in education has been increased recently and has caused some problems [1]. For example, the required information can not be found by users, or the content and the structure of web-site can not be understood because of the materials in web site adaptively not submitting according to users’ knowledge levels and preferences [2]. In order to overcome these problems Adaptive Educational Hypermedia System (AEHS) has been developed by researchers mainly in the areas of education and computer science [3-4].

Kosba et. al [1] developed an adaptive feedback generation framework for students and teachers. The framework provides an understanding about students by knowing problems of students and persons who can help them. Hollink et. al [2] proposed a method that divides the pages of a web site to stages under basis of user’s logged navigation. The method finds optimal stages and each page refers only one stage. It aims to give better recommendations for users or to adapt the materials for user’s goals. The model is useful to provide navigational support and guide for web applications having unstructured material layer.

AEHS has three models essentially. These are domain model, user model and adaptation (teaching) model. The domain model represents the hypermedia material space and relations among the elements in space. The user model represents state of knowledge about the domain model and the preferences of user. It is the most important component of the AEHS because it has important effect in the adaptation process [5-7]. The adaptation model submits materials to the users using both of domain and user models.

To provide fast and powerful adaptation effect, the developers of AEHS must adapt machine learning algorithms and pedagogical approaches in User Modeling Component (UMC) and must establish a manageable and compatible domain model network. In addition, developers need data about users to create user models. Moreover they must obtain useful information to construct a powerful user model. For this purpose, the actions of users are logged in the user’s database. To obtain useful information from logged data about the user in UMC is a significant problem. In order to overcome this problem, an UMC has been developed that can be easily applied in AEHS. UMC firstly applies a rule-based inference to convert logged data about the user into a useful info and a Naïve Bayes (NB) classification algorithm was then applied to classify users according to knowledge status employing the obtained info. Users are classified according to status of knowledge levels such as "beginner", "intermediate" and "advanced" using an NB classifier. There are many reasons for selecting NB classification algorithm in UMC.
The NB classifier can easily be applicable to user modeling tasks, represent the results independently according to the factors in the classification process. It is also suitable for real-time adaptations or applications. Basic capabilities in AEHS are presented below as:

- include suitable materials for users, who have different knowledge levels [2],
- submit the materials with suitable goals, the knowledge level and the learning style of user [5, 6-8],
- include materials in the different formats (watching, listening, and reading) based on user's preferences [7, 9],
- give support to the facilitates of user's navigation in the web platform [2, 7-10],
- offer explanations and guidance for users [1],
- determine the knowledge levels of users for each element in the domain model [6,7],
- gather information about the users and convert them to the meaningful information [1-2, 11-12].

In this paper, as briefly mentioned above user modeling process has been essentially introduced. Additionally, designing the main components of application has been presented because of close relations of the components with UMC. Also, user domain model has been established to match the general needs of AEHS specified above. Moreover, a user model has been created. An adaptation model was also designed for adaptive material submitting to users as indicated in [6, 7, 9].

2 Establishing User Domain Model

The user domain model consists of two areas in AEHS. The former as called static area is created by the developed UMC in the process of user's membership. Users fill a form in this stage. Firstly users declare the preferences, the state, the graduates, and the identical info about themselves. The UMC, then automatically creates user static model and user domain model according to the declared values. Stored procedures and related classes are used in the process. The latter is user's knowledge status model. It is also called as dynamic model that continuously changes along with user interaction [6]. The power and efficiency of user domain model are depending on two conditions. The first is the degree the representation of user domain model for the domain model of hypermedia material space also called as domain model network in the AEHS. The second is the adopted approach to determine the knowledge state of user about the elements in the domain model network. "Overlay model" approach was adopted in the creation of user domain model as reported in [1, 13]. The reason for selecting the overlay model was to represent the user knowledge for each element in the domain model network independently, and to provide a flexible and powerful structure [5, 14].

The developed domain model network for an application is given in Fig 1. The user domain model is a copy of the subject layer of domain model. This model mainly contains elements of concepts, and topics defined by the goals. The elements are related to each other in the adaptive hypermedia. Relations can be pre-requested, propagated and hierarchical. The domain model network takes place after the relations defined in the references [1, 7, 10]. Link styles are changed [8], and additional explanations are added for each link [7] to give the navigational link support for students in the AEHS. The structure of the domain model is formatted by the expert of domain. Elements of the domain model are connected with set of learning goals. Additional explanations, texts, videos, audio, animations, simulations and examples are prepared for each element and placed to the material layer in the AEHS. Layer of domain model is arranged three layers in the AEHS [15–17].

![Fig. 1. The developed domain model network](attachment:fig1.jpg)

The goals are suitably defined according to users' instructional needs in the first layer. Related elements and degree of these relations are defined in the "Goals Specification XML Document". It enables a flexible
description of the relations. The topics and concepts are placed into the second layer. The difficulties faced in this work are to determine the topic levels, the relations of pre-requisite and propagation among the topics, and the process time [1] by expert knowledge. Related concepts and kinds of these relations are described in the "Topics Specification XML Document". "Concepts Specification XML Document" is similarly arranged as topic. The materials about topics and concepts are stored in the third layer. These materials are suitably prepared in different formats as audio, video and text to submit the learning styles and the preferences of users. The "Material Specification XML Document" is prepared to extend the impact of material layer by the application. The details of "Topics Specification XML Document" are illustrated in Figure 2.

<?xml version="1.0" encoding="utf-8" ?>
<TopicList>
  <Topic ID="4" Category="Electrical Machines">
    <Title>Circuit Models and Characteristics of DC Machines</Title>
    <SupportedPresentations>
      <Reading>topic4r</Reading>
      <Listening>topic4l</Listening>
      <Watching>topic4w</Watching>
    </SupportedPresentations>
    <Level>1</Level> <Time>60</Time>
    <Count>3</Count> <Question>5</Question>
    <RelatedConcepts>
      <Concept ID="1" Url="concept1">Position</Concept>
      <Concept ID="3" Url="concept3">Moment</Concept>
    </RelatedConcepts>
    <RelatedTopics>
      <RTopic ID="11" Url="dcbrushless">Brushless DC Motor</RTopic>
    </RelatedTopics>
    <Pre-ConditionTopics>
      <Pre-Topic ID="1" Url="dcseries">DC Series Motor</Pre-Topic>
    </Pre-ConditionTopics>
    <RelatedAnimations>
      <RAnm>anm1</RAnm>
    </RelatedAnimations>
    <RelatedSimulations>
      <RSim>simulation1</RSim>
    </RelatedSimulations>
  </Topic>
</TopicList>

Fig. 2. An XML specification document about a topic in the domain model

3 Process of User Modeling


In this paper, rule-based inference and classification approaches were adopted to decide knowledge state of the users about the domain model network and to develop the UMC. The UMC records the activities such as material pages clicked, how long pages viewed, and the number of clicks pressed. These records are separately logged for each element in the user’s domain model. These data are generally used for statistical purposes in the most of web based educational systems [1]. Contrary to the systems, the data were obtained in the proposed system converted useful info using rule-based inference mechanism by the UMC. The influence of logged data over the adaptation is seen in the user model depending on rules and relations among the elements in the domain model network [7]. Therefore when the user clicks a link the logged data are processed by the rule-based inference model and the NB classifier was then applied. The results are seen on the umpteen concepts and topics depending on the relations among the concepts and topics. The operation process of rule-based inference mechanism is detailed in the “Rule-based inference mechanism” subsection. In the classification process, the useful info offered by the rule-based inference mechanism is used to classify users according to their knowledge levels for each topic by the UMC. It uses NB classifier for this task. The operation process of NB classifier is detailed at the end of “Classification of users by the UMC” subsection.

The UMC decides users’ knowledge status and online topic in the domain model network. The UMC determines knowledge status of a user for each concept in the domain model in accordance with degrees of the study time and the number. Similarly, the UMC determines knowledge status of the user for each topic in the domain model in accordance with degrees of the study time and period, knowledge status about the concepts related to the topic of the user. So, the developed application for each element views and updates the knowledge status of user from concepts to topics in the domain model. All of the parameters are represented as the features of NB classifier.
3.1 Rule-based inference mechanism

A rule-based inference model was developed to convert the gathered data about the user to useful info. This task was executed by the UMC. Consideration the expert domain in the rule-based inference model was taken into consideration according to the recommendations. Four features are considered to determine the knowledge status of user for topics in this process. The considered features are to be the degree of study time (P), the degree of study count (Q), the degree of study time to concepts related to the topic (R) and the knowledge state about the concepts related to the topic (S). The degrees for the features were assigned as “poor”, “insufficient”, “average”, and “advanced”. These degrees were calculated using rules that declared by the domain model expert.

i. Calculating the degree of study time for a topic (A)

Required study time for each topic was declared by the expert of domain in the "Topics Specification XML Document". The UMC calculates the degree of study time for each topic. The study time of user for each topic material is denoted as "x". "t" is the time that a user must study. The determined rules for “A” by the expert are:

\[
A(x,t)=\begin{cases} 
\text{advanced} & \text{if } x < \frac{t}{2} \\
\text{average} & \text{if } \frac{t}{2} < x < \frac{t}{4} \\
\text{insufficient} & \text{if } \frac{t}{4} < x < t \\
\text{poor} & \text{if } x > t
\end{cases}
\]

ii. Calculating the degree of study count for a topic (Q)

A user's study count for each material of topic is denoted as "u". "c" is the count determined by the domain expert for each material that a user must study. The determined rules for “Q” by the expert are:

\[
Q(u,c)=\begin{cases} 
\text{advanced} & \text{if } u < \frac{c}{2} \\
\text{average} & \text{if } \frac{c}{2} < u < \frac{c}{4} \\
\text{insufficient} & \text{if } \frac{c}{4} < u < c \\
\text{poor} & \text{if } u > c
\end{cases}
\]

iii. Calculating the degree of study time for concepts related to the topic (R)

The UMC finds the concepts which are related to topics using "Topics Specification XML Document". The UMC then learns the degree of study time of user about these concepts from “User Domain Model”. Total degree is denoted as "d" to the concepts. "n" is the number of concepts related to the topic. The determined rules for the “R” by the expert are:

\[
R(d,n)=\begin{cases} 
\text{poor} & \text{if } d/n \leq 1 \\
\text{insufficient} & \text{if } 1 < d/n \leq 2 \\
\text{average} & \text{if } 2 < d/n \leq 3 \\
\text{advanced} & \text{if } d/n > 3
\end{cases}
\]

iv. Calculating the degree the number of studied concepts related to the topic (S)

The UMC finds the number of concepts which are related to the topics and earns the number of the studied concepts of user from “User Domain Model”. The "p" is the number of concepts related to the topic. “y” is the number of studied concepts, which are related to the topics, of a user. The determined rules for S by the expert are:

\[
S(y,p)=\begin{cases} 
\text{poor} & \text{if } y < \frac{p}{4} \\
\text{average} & \text{if } \frac{p}{4} < y < \frac{p}{2} \\
\text{insufficient} & \text{if } \frac{p}{2} < y < \frac{p}{4} \\
\text{advanced} & \text{if } y > \frac{p}{2}
\end{cases}
\]

3.2 Classification of users by UMC

According to the degrees of four features, calculated by the rule-based inference mechanism, the NB algorithm classifies the user as “beginner”, “intermediate” or “advanced” for each topic in the domain model network. The classification was followed as:

For each target value \( v_j \):
\[
\hat{p}(v_j) \leftarrow \text{estimate } p(v_j) \tag{1}
\]

For each feature degree \( f_i \) of each feature \( f \)
\[
f : \hat{p}(f_i | v_j) \leftarrow \text{estimate } p(f_i | v_j) \tag{2}
\]

Classify new instance \( z \):
\[
\nu_{\tilde{z}} = \arg \max_{v_i \in v} \hat{P}(v_j) \prod_{f_i \in z} \hat{P}(f_i | v_j) \tag{3}
\]

UMC is tested for a sample case. In the sample case the user’s knowledge state for "k" topic is, which is

Table 1. The conditional probability distribution for the topics

<table>
<thead>
<tr>
<th>Features</th>
<th>Degrees</th>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Poor</td>
<td>11/45</td>
<td>3/32</td>
<td>7/26</td>
</tr>
<tr>
<td></td>
<td>Insufficient</td>
<td>12/45</td>
<td>6/32</td>
<td>4/26</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>12/45</td>
<td>8/32</td>
<td>5/26</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>10/45</td>
<td>15/32</td>
<td>10/26</td>
</tr>
<tr>
<td>Q</td>
<td>Poor</td>
<td>10/45</td>
<td>7/32</td>
<td>4/26</td>
</tr>
<tr>
<td></td>
<td>Insufficient</td>
<td>20/45</td>
<td>4/32</td>
<td>5/26</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>8/45</td>
<td>13/32</td>
<td>7/26</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>7/45</td>
<td>8/32</td>
<td>10/26</td>
</tr>
<tr>
<td>R</td>
<td>Poor</td>
<td>19/45</td>
<td>2/32</td>
<td>0/26</td>
</tr>
<tr>
<td></td>
<td>Insufficient</td>
<td>18/45</td>
<td>8/32</td>
<td>0/26</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>6/45</td>
<td>14/32</td>
<td>11/26</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>2/45</td>
<td>8/32</td>
<td>15/26</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>20/45</td>
<td>2/32</td>
<td>0/26</td>
</tr>
<tr>
<td></td>
<td>Insufficient</td>
<td>15/45</td>
<td>13/32</td>
<td>0/26</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>7/45</td>
<td>12/32</td>
<td>9/26</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>3/45</td>
<td>5/32</td>
<td>17/26</td>
</tr>
</tbody>
</table>

Frequency the take place of topic in the “beginner”, “intermediate” and “advanced”

<table>
<thead>
<tr>
<th></th>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45/103</td>
<td>32/103</td>
<td>26/103</td>
</tr>
</tbody>
</table>

Classification for the knowledge level of user as “beginner”, “intermediate” or “advanced” for topic “k” according to the conditional probability distribution for the topics as given in Table 1 and the features of user about the topic “k” is calculated from:

\[ k_{NB} = \arg \max_{j \in \{ \text{Beginner}, \text{Intermediate}, \text{Advanced} \}} P(k_j) \]

\[ P(A = \text{Poor} | \text{Beginner-Intermediate-Advanced}) \]
\[ P(Q = \text{Insufficient} | \text{Beginner-Intermediate-Advanced}) \]
\[ P(R = \text{Average} | \text{Beginner-Intermediate-Advanced}) \]
\[ P(S = \text{Advanced} | \text{Beginner-Intermediate-Advanced}) \]

\[ P_{\text{Beginner}} = (11/45)*(20/45)*(6/45)*(3/45)*(45/103) = 0.000422 \]
\[ P_{\text{Intermediate}} = (3/32)*(4/32)*(14/32)*(5/32)*(32/103) = 0.000249 \]

\[ P_{\text{Advanced}} = (7/26)*(5/26)*(11/26)*(17/26)*(26/103) = 0.003615 \]

As a result of the calculations given above, the user knowledge level about the topic “k” is classified as “advanced”.

4 Designing Adaptation Model

Adaptation model is used for adaptive presentation of materials in the domain model according to the user model [6-7]. This model uses both of domain model and user model to execute its task.

Adaptation model for the developed application was given in Fig. 3. The adaptive navigation trees about the topics, concepts, simulations, and help were prepared in that model. For example, to prepare the topics according to knowledge level of user it needs both of “user’s topic model” and “topic specification document”. This model observes the documents and determines the topic which one is suitable for the user. When the goal of user is known it then prepares the related topics to the goal.

User static model stores the preference and the personal information of user. The information is used for the statistical purposes and determination the presentation format of the materials. For example, a user can read, listen or watch the materials. According to the preference, appropriate materials are provided to the user.

Fig. 3. The adaptation model for the developed application
5 Conclusions

The difficulties faced in developing such a system are purifying the gathered data, obtaining useful data, evaluating useful data and providing a powerful adaptation effect in a short time. So the rule-based inference approach has been adopted before processing the gathered data in the machine learning algorithm. Thus, useful data used to classify the process with help of NB classifier. This approach enables accurate and rapid adaptation effect to the developed application. The difficulties faced in the developed application are to represent the domain model for each courses and to prepare the materials. To provide a superior adaptation, elements of the domain model should be represented individually in the domain model. XML specification documents would be also designed for each element class.

The UMC developed used to determine knowledge state of users for each element in the domain model network using the obtained useful information from rule-based inference mechanism in the NB classifier easily and effectively. Users were successfully classified according to the knowledge status of them in the UMC. Finally, a case study was demonstrated to determine the knowledge level of user about a topic in the developed application.

In this paper, a new user modeling approach was presented to determine the knowledge status of user in AEHS. The approach is based on forming the domain model and determining relations among elements of that model to decide the knowledge status of user from domain model independently. The proposed method provides flexibility to individual learning or studying steps. In comparison to the literature, the proposed approach enables quick and powerful adaptation for matching instructional needs of users.

6 References