Educational Applications for Students

M.Tech. MTP Stage1 Report

Submitted in partial fulfillment of the requirements for the degree of

Master of Technology

by

Pavan Kotha
Roll No: 113050010

under the guidance of

Prof. Deepak B Phatak

Department of Computer Science and Engineering
Indian Institute of Technology, Bombay
Mumbai
Abstract

Education today is influenced by technology evolution on one side and requirements of society on other side. The main mission of our educational research is to solve the problems of society and give better education to everyone. To satisfy the increasing demand for technical education, computers and web are being harnessed. Currently there are many e-learning platforms which provide online education. No current e-learning platform[15],[17] is concentrating on improving the performance of student. This survey provides some key insights of improving the performance of students and also suggests customized references according to student learning style. Slow learners and those with learning gaps can benefit from this personalized resource recommendation.
Contents

1 Enhancing Learning Through Online Materials .................................................. 5
   1.1 Problem Based Learning .................................................................................. 5
   1.2 Opportunities For Improved Learning .......................................................... 5
      1.2.1 Active Learning ...................................................................................... 6
      1.2.2 Explanations with voice and graphics ...................................................... 6
      1.2.3 Simulations ............................................................................................. 6
      1.2.4 Probe student thinking on multiple scales ............................................... 6
      1.2.5 Timely assessment of progress .................................................................. 7
      1.2.6 Peer interaction ....................................................................................... 7
      1.2.7 Feedback to instructors .......................................................................... 7
   1.3 Challenges ....................................................................................................... 7
   1.4 Do Diagrams Enhance Learning Always? ....................................................... 10
   1.5 Tutorial Systems ............................................................................................ 13

2 E-learning Content Visualization Module ........................................................... 15
   2.1 Introduction .................................................................................................... 15
   2.2 Mobile Device Recognition Methods ............................................................ 15
   2.3 Content Adaptation Approaches ..................................................................... 16
      2.3.1 Server Based Approach ...................................................................... 16
      2.3.2 Proxy Based Approach .................................................................... 16
      2.3.3 Client Based Approach ................................................................... 17

3 Human and Technological Issues in E-learning .................................................. 18
   3.1 Human Issues ............................................................................................... 18
      3.1.1 Learners Types ..................................................................................... 18
      3.1.2 Social and Cultural Factors ................................................................. 19
   3.2 Technology Issues ......................................................................................... 19
   3.3 E-Learning Environment .............................................................................. 19
      3.3.1 Functional Environment .................................................................. 19
      3.3.2 Mobile Environment .................................................................. 19

4 Predicting Student Performance ...................................................................... 21
   4.1 Decision Trees ............................................................................................. 21
   4.2 Decision Tree Learning Algorithm ............................................................... 23
4.3 Performance Prediction of Engineering Students using Decision Trees . . . 25
  4.3.1 Application to CS101 COURSE . . . . . . . . . . . . . . . . . . . . 27
4.4 Combination of classifiers . . . . . . . . . . . . . . . . . . . . . . . . . . 28
4.5 Understanding Learners Learning Style . . . . . . . . . . . . . . . . . . . 29
4.6 Sequential Pattern Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . 30

5 Usability Testing 32
  5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
  5.2 Current Approaches to Usability Evaluation . . . . . . . . . . . . . . . . 32
  5.3 Aspects of User Satisfaction . . . . . . . . . . . . . . . . . . . . . . . . . . 33
  5.4 SUMI Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
  5.5 SUMI Questionnaires . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

6 Conclusion And Future Work 34
## List of Figures

1.1 Video Controls explaining equilibrium conditions and the effect of different choices of co-ordinate axes ................................................. 8  
1.2 Hints ......................................................................................... 9  
1.3 Scaffolding ........................................................................... 10  
1.4 compareexercise ..................................................................... 11  
1.5 guidedsimulation showing motion of body ................................ 12  
1.6 Did I get This Exercise ............................................................... 13  
1.7 Molecular Diagrams ................................................................. 14  

4.1 Training Dataset for playing tennis ............................................. 22  
4.2 Decision Tree for dataset shown above ....................................... 23  
4.3 Partial Decision Tree for dataset shown above ......................... 25  
4.4 Attributes ................................................................................ 26  
4.5 Decision tree for three class prediction ..................................... 27  
4.6 Grade Distribution .................................................................. 28  
4.7 Sequence Database .................................................................. 30  

6.1 Structure of Proposed System ................................................... 35
Acknowledgments

I would like to thank my guide, Prof. DB. Phatak for the consistent directions towards my work.
Chapter 1
Enhancing Learning Through Online Materials

Paper teaching materials are not meeting the instructional needs, hence e-materials which can integrate multimedia, visual and stereo features etc has become prominent. The electronic teaching material can interpret focal and difficult points in teaching through appropriate teaching media like graph, image, audio, video etc. This chapter deals with advantages of incorporating multimedia into electronic teaching materials to meet the needs of students and also the challenges online learning materials are facing in promoting learning when problem solving is tied to drawing and solving equations.

The online course should be divided into many modules. Each module should contain some text and variety of exercises and simulations. Student Learning should be supported through “Learn by Doing” activities which offer hints and feedback. These materials support to learn effectively. In traditional classroom, principal activity outside of class is solving homework problems. The feedback for these homework problems is weak. Students are not graded homework immediately. So the defects of students are not recognized immediately. Students would benefit if they are instantaneously assessed. The following section deals with interactive online course materials which address above deficiencies.

1.1 Problem Based Learning

According to Socrates “Education is not filling of vessel but kindling of flame”. The way to learn is how to think and solve problem. Hence problem based learning should be introduced in electronic teaching materials. The learner’s characteristics like learning needs, learner’s features, learner’s psychological profile, cognition level should also be considered.

1.2 Opportunities For Improved Learning

The following are some of the benefits in e-learning which aim at improvement in learning.
1.2.1 Active Learning

Active Learning in a large classroom is difficult. But computer-based materials which can provide short review questions and give hints and feedback can promote higher levels of cognitive activity for students. By appropriately devising online materials, students can be actively engaged throughout the process by giving frequently small questions and checking their progress. In this way, students can assimilate new information and check their understanding by tests with feedback.

1.2.2 Explanations with voice and graphics

The combination of voice and graphics offers advantages over textbooks as words are linked to relevant diagrams. An instructor can provide good explanation involving voice and graphics. Students can repeat selected portions of lecture multiple times if they do not understand the concept by replaying a video file. The disadvantage is that video files do not have an altered way of explanation based on student queries. Generally, in any online course (e.g., courses in Coursera), the video lectures are 10 to 15 min duration only. The video lectures are short because researchers say that people concentration cannot stay on a topic more than 15 min continuously.

1.2.3 Simulations

Neither a static textbook nor an instructor can offer dynamic simulations, particularly in the simulations controlled by user parameters. Online learning materials can include this aspect of simulations. E.g., in Statics, it can be shown clearly in online materials, how motion of object varies with applied force.

1.2.4 Probe student thinking on multiple scales

Progress in learning can not only be assessed by big problems, such as found in textbooks. Frequent short questions on fine-grained issues (e.g., simply with yes or no answers, or multiple choice questions or conceptual questions that require written explanation which can immediately be compared by the student to an expert answer) are also appropriate. In online environment, this is feasible than with traditional homework. We can also pose conditional questions which depend on the answers to the previous questions. Example: Coursera, an online learning platform also uses these techniques in teaching. Many videos contain simple multiple choice questions (on the topic taught in that video) embedded within them. Some videos also contain reflection exercises where the student is asked to think about something, write down a response, and then share it later in the forums. These exercises concentrate on fine-grained issues. Apart from these exercises, there are quizzes which concentrate at coarse-grained issues.
1.2.5 Timely assessment of progress

With little feedback and the delay in returning homework, makes student unaware that they are lagging behind in subject. Computer-based learning materials helps the students to recognize progress instantaneously. Students can then choose to repeat selected exercises to get more practice.

1.2.6 Peer interaction

Using discussion forums, students can interact with other students to clarify their doubts on subjects. Example: Piazza is a forum which has many good features for making peer interaction better.

1.2.7 Feedback to instructors

We can log student online interactions and then data-mining techniques can be used to track progress and also the detailed paths taken by individual students, then find the topics where students are finding difficulty. Instructor can then use class time more productively to discuss those challenging problems.

1.3 Challenges

Despite these benefits, there are many challenges that are being faced by computer-based online materials for learning. The materials should engage the student in as many kinds of interactions as possible. For eg. in Statics course, we can ask students to choose from several forces, to choose points where forces act, or to move a slider to an appropriate point where support should be. Students can also be asked to enter free-form explanations, the computer can only save this input for later inspection by the instructor. Student can view an expert answer after submitting. Another challenge is good user interface design. If students want to engage with the system without any external intervention, it is important to signal to users where they are and what is expected at each instant. The following are some examples which addresses above elements.

Complex explanations are difficult to understand with written text and diagrams. In these scenarios, both aural and visual ways of explanations can be combined so that diagrams and voice go in synchrony and makes the users attention focused. By using video controls, user has ability to pause, stop, rewind, repeat. Figure 1.1 explains Video Controls explaining equilibrium conditions.

Online materials can also help the students by giving hints and feedback on wrong answers. These are called as “Learn By Doing” exercises. Figure 1.2 asks the student to solve for equilibrium conditions and determine the unknown support on their own and offers scaffolding, if they need help. So users can solve the problem on their own and
Figure 1.1: Video Controls explaining equilibrium conditions and the effect of different choices of co-ordinate axes

[1]

enter the answer. User is taken through a sequence of steps and expected to perform one step at a time

In problem shown in Figure 1.2, user is taken through a sequence of steps to solve the problem, with one step at a time. Figure 1.3 shows 4th step of scaffolding given to student in writing algebraic equation with drop down menus. The ultimate goal to follow step by step procedure and giving hints to student at each step is to make student solve problem on his own.

Figure 1.4 shows how student can compare his answer with an expert answer. In case of dynamic simulations, these learning materials can explain the concepts in a better way than static textbook, because motion of body depends on applied force. Static
textbook can only show single picture whereas these materials show motion of body depending on user entered parameters. Figure 1.5 shows how student can observe the motion of body based on the forces he applied.

Student progress can be assessed with Did I Get This Exercise. Figure 1.6 shows how student is asked to choose points where forces are acting. If the student chooses wrong answer, feedback to student explains the principles of free body diagrams.

Andrew Ng, founder of Coursera also followed many of above principles when teaching his Machine Learning course because the above way of teaching clearly explains the fundamentals of the subject and also tests his level of understanding. This section explained the benefit in integrating multimedia, visual and stereo features in online materials. So the enhancements need to be included in online materials are active learning, integration of simulation, feedback during problem solving, fine grained assessment, timely tracking of progress.
1.4 Do Diagrams Enhance Learning Always?

In the previous section, we have included many diagrams for better explanation of concepts. But the question is Do Diagrams enhance the learning always? Generally, there are large learning gains when e-materials include diagrams and verbal explanations, but diagrams do not always lead to improved outcomes. This section describes when diagrams can enhance learning.

The need of diagram in a particular learning situation depends on learning objective, design of visual representation and cognitive processing of learner. Pictures only help if they are relevant to current instruction and exclude interesting but extraneous information. A series of studies was conducted in chemistry classrooms to test whether molecular-level diagrams (Figure 1.7) would enhance conceptual understanding of chemical equilibrium. Some participants are given materials consisting text+diagrams and some other are given materials with only text and test was conducted for both sets of people. The result showed that students who learnt through diagrams+text did not outperform students who learnt this concept with only text. The reason for this is, students who learnt using text+diagrams mapped the features of diagrams into text and did not read
the text. The self-explanation here is negatively correlated with performance because of extraneous information i.e. here diagrams are unnecessary. They perceived information incorrectly as molecular level diagrams have extraneous information.

The current models of multimedia learning do not specify what features of diagrams lead to enhanced learning. Researchers suggested that three factors that determine the effectiveness of diagram are a) the specific learning objectives, b) how the diagram makes key information salient and c) the learners' cognitive processing (student must select meaningful information from the diagram for processing) and prior knowledge.

**a) Specific Learning Objective** The diagram should convey correct mental model of the system. For eg., if students want to construct mental model of a mechanical system

\[
A = \frac{W + L(\frac{d}{b} + 1)}{2}
\]

\[
R = \frac{W - L(\frac{d}{b} - 1)}{2}
\]
Figure 1.5: guided simulation showing motion of body

(such as a bicycle pump), then learner must identify the parts of the system and the relationships between these parts. The mental model here consists of process of air being pushed through the pump. A series of diagrams enhances the development of the mental model because misinterpretation might happen if only single picture used.

b) Design of diagram

Diagram should guide to important information. Every representation makes certain aspects salient while suppressing other details. For instance, black and white line drawings make the parts and overall shape of an object salient but suppress details about color and texture. If the parts of a mechanism are important and the color is irrelevant, then black and white line drawings may be more effective than color.

c) Cognitive Processing of Learner

Student must select meaningful information from the diagram for processing. Processing ability and prior knowledge of the student influence the information extracted from a diagram. For eg. in learning about a bicycle pump, the student must understand that pushing down the handle of the pump causes air to flow
out of the open valve at the bottom because air is an entity that has mass and decreasing the volume of air results in greater pressure on the valve. The diagrams created using above 3 principles improve learning.

1.5 Tutorial Systems

This section describes types of tutorial systems which need to be incorporated in e-learning. General and Domain Specific are two types of tutorial system.

a) General Tutorial Systems These are also called as Help-type systems. These are designed to support teaching activities. Help type systems in the software product appear as specific command menu. Their purpose is to assist the user when he needs information
about using the application.

i) **Text Type Help**: Information is present as text only. It displays the explanation about option chosen by the user. Read me files also come under this category.

ii) **Hyper link help**: Displays the information of selected terms by clicking the term.

iii) **Help with related topics**: It is used to access information related to data present on main screen.

iv) **What’s This type**: It offers help regarding buttons and commands on screen.

v) **Offline Tutorial Help**: These contain tutorial fragments that can be accessed without internet connection. These contain short tutorials to guide the student in step by step manner.

vi) **Online Tutorials**: These contain tutorials which can be accessed using internet connection.

b) **Domain Specific Learning Systems** These systems are designed based on its working domain. Example: Learning Software for Economics uses graphics to explain financial concepts. Learning Software for Mathematics provides contents according to age groups. For 12-15 years children it explains basic algebra and for age > 15 it explains advanced algebra.
Chapter 2

E-learning Content Visualization Module

Till now, we have seen what content to be put up in e-learning materials. In this chapter, we see how e-learning content is visualized, and the existing methods for mobile device recognition, and different content adaptation approaches.

2.1 Introduction

The evolution of technology has led to wide variety of devices like desktop computers, notebook computers, tablets, personal digital assistants (PDAs), cell phones, mobile phones etc. Each device differs from other in web browser they use, supported markup languages (HTML, XHTML, cHTML, WML etc), supported script languages, file formats, screen resolution etc. These different characteristics of the devices shows the need to develop an adaptation method for correct visualization of electronic content.

The solution for problem of correct content visualization on different devices is very important for e-Learning because same education content need not be developed again.

Example: The existing courses in the e-Learning Shell Software Platform are available only for desktop computers. They cannot be used with mobile devices. Hence there is need for developing new content visualization subsystem which delivers content to users that use any device like PC’s, tablets, mobile devices etc.

2.2 Mobile Device Recognition Methods

First device has to be recognised for content adaptation. Currently request header field in http protocol is used by servers to identify the device. Alternative methods to identify device are:

HTTP USER AGENT HEADER: Browsers use http protocol to transfer information on web. The server decides the kind of information to send depending on device profile. Http request sends Accept Header which indicates the types of data the browser can
accept. The client also sends User Agent Header for identifying client device and contains information about browser, operating system and hardware information. The drawback is the information in user agent header is not sufficient as number of different kind of devices are growing up.

**Composite Capabilities/Preferences Profiles (CC/PP):** World Wide Web Consortium has standard way for specifying CC/PP. The standard says, the devices should transmit their configuration details and abilities like screen resolution, audio characteristics, frequency band etc to web servers. CC/PP is universal profile that describes the devices characteristics.

**WAP User Agent Profile:** The device is identified using user agent profile. When mobile device sends request to server, it also sends an URL address to its mobile profile by adding X-Wap-Profile Header in request. This header indicates the server where to find the device profile. The content server extracts the necessary information for the client from device profile repository and can store it, so that, it can be used later. Information is present in XML, hence using XSLT information can be transformed to the type which device can recognize.

### 2.3 Content Adaptation Approaches

Content adaptation is a process of selecting, generating or modification of content (text, images, audio, video), so that it can be presented to different devices. If the web page is accessed using desktop computer then it needs no modification. If the same web page is accessed via mobile device, images must be resized and compressed, text must be formatted and video is presented as text description or as an image depending on available frequency band. Hence content adaptation is needed.

Different Content Adaptation Approaches are:

#### 2.3.1 Server Based Approach

In this approach, web server has modules for content adaptation. The same content is stored in different versions on the server, but only the content, which coincides with the client’s profile is presented. General approach used for content delivery for different devices is to store the information in XML format and then use XSLT to transform to the client preferable markup language.

#### 2.3.2 Proxy Based Approach

In this approach, a proxy server analyses and transforms the content before sending it to the client. Proxy server also caches the adapted content so that it can be used later. Proxy server and the web server should know type of device to send the appropriate content. Eg: AvantGo is service which uses proxy based approach and delivers mobile websites to consumers personal digital assistants (PDAs) and smart phones.
2.3.3 Client Based Approach

In this approach, the necessary transformation is done by a consumer's device. Eg: Opera Software uses client-based adaptation technology called Small-Screen Rendering (SSR) which transforms the content according to consumer device type. Another method is use XHTML and different CSS for each device type.
Chapter 3

Human and Technological Issues in E-learning

Many e-learning systems fail to satisfy the learners needs and requirements due to the absence of human and technological issues such as social and cultural factors, quality components and pedagogy requirements, technological issues such as the learning environment. E-learning systems are complex web applications as they should transfer traditional pedagogy methods and practices to electronic environment.

3.1 Human Issues

The development of a successful learning system is based on the understanding learners needs and behavior and incorporating them in learning process. This will help achieve the pedagogical goals. When developing an e-learning system, we need to classify the learners needs and behaviour. General user requirements for E-learning system in perspective of web-design process are

a) Keep navigation clicks to minimum
b) Keep scrolling to minimum
c) Have contents both for low and high-connection speeds
d) Have a consistent user interface

3.1.1 Learners Types

The learners are classified into following types:

Traditional Learner: Focuses his effort on reading the required material.
Achiever: Focuses his effort on completing the quizzes and review questions.
The Interactive Learner: Focuses his effort on interacting with peers and tutors
The Struggler: Studies less frequently than all other students. Clearly, with above classification we can see some learners are passionate, some are exam based and so on. These different types of learners should be given different kinds of learning resources and different
types of learning support to suit their individual needs.

3.1.2 Social and Cultural Factors

Learner behavior is shaped by social and cultural factors also. **Demographics:** The Web engineer must specify and design the E-Learning application based on the targeted population. **Social characteristics:** The developer should examine the educational system, the literacy level, and the languages spoken in the country.

3.2 Technology Issues

Identifying the technology level of each targeted country helps the Web engineer to decide on the type of technology and resources to use. Countries with advanced technologies and high Web usage are excellent candidates for an E-Learning application. Countries new in the Internet arena have basic technologies, hence need to design E-Learning systems with low bandwidth and capabilities due to poor communications. The e-learning platform should have quality factors like usability, functionality, efficiency, reliability and maintainability.

3.3 E-Learning Environment

E-Learning environment is separated into two sub-environments a) Functional Environment and b) Mobile Environment

3.3.1 Functional Environment

E-Learning system should have following operations: Class announcements, access to course material, assignments and case studies, online quizzes with a timer and a feedback mechanism, a virtual classroom with collaborative study groups. In addition, a Personal Preferences section where the system provides progress report, a homework submission utility, access to classmates public information, and should be able to view other courses enrolled by student. Other facilities to be provided are online help with access to the helpdesk administrator, a calendar for scheduling, a customization utility for specific preferences such as, the change of language, a requirements analyzer tool for suggesting new requirements, and external links to online databases and related material.

3.3.2 Mobile Environment

Mobile education is defined as the dissemination of pedagogical material through the use of wireless networks and devices. With technology revolution e-learning is migrating to mobile environment. Technologies used for providing wireless access are WLAN, WAP, Short
Message Service (SMS), and UMTS. WLAN is usually only available at the institution premises, whereas WAP, SMS and voice-technologies are more widely available.
Chapter 4

Predicting Student Performance

Classification is one type of prediction where predicted variable is binary or categorical (means takes some discrete value) variable. Classification methods like decision trees, Bayesian network etc can be applied on the educational data for predicting the students behavior, performance in examination etc. Classification is nothing but dividing data tuples into different classes. This prediction will help the instructors to identify the weak students and help them to score better marks. This survey only includes decision tree technique in predicting student performance. This chapter explains the decision tree algorithm and how it can be applied in real life scenario to predict the performance of the student.

4.1 Decision Trees

Statistical data is represented in terms of tuples. The data consists of many attributes and there is target attribute which we have to predict. For example, in the dataset shown in Figure 4.1, the target attribute to be predicted is PlayTennis. This attribute shows whether we can play tennis or not on a particular day i.e there are 2 classes, yes or no. This problem is called as classification problem because we are classifying the dataset into 2 classes which says whether we can play tennis or not on a particular day. The PlayTennis attribute is predicted on other attributes like Temperature, Humidity etc. of the particular day. For prediction, here we use Decision Tree technique.

Decision tree is a technique used for classification of instances. It is a method for predicting discrete-valued target functions, in which the learned function is represented by a decision tree. Learned trees can also be represented by if-then rules. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute in the given example. This process is then repeated for the subtree rooted at the new node. Figure 4.2 shows a learned decision tree from dataset shown in Figure 4.1. This decision tree classifies whether morning of a day is suitable for playing tennis or not. For
example, if the morning of a day has these attributes (Outlook = Sunny, Temperature = Hot, Humidity = High, Wind = Strong), then this instance would traverse via leftmost branch of decision tree shown in Figure 4.2 and would therefore be classified as a negative instance (i.e., the tree predicts that PlayTennis = no).

Decision trees can be represented as a disjunction of conjunctions of constraints on the attribute values of instances. Each path from the tree root to a leaf corresponds to a conjunction of attribute tests, and the tree itself to a disjunction of these conjunctions. For example, the decision tree shown in Figure 4.2 corresponds to the expression (Outlook = Sunny AND Humidity = Normal) OR (Outlook = Overcast) OR (Outlook = Rain AND Wind = Weak). This expression says whether to play tennis or not on a particular day which is dependent on attributes of the day. The attributes we considered are climatic conditions.

<table>
<thead>
<tr>
<th>Day</th>
<th>Outlook</th>
<th>Temperature</th>
<th>Humidity</th>
<th>Wind</th>
<th>PlayTennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>Weak</td>
<td>No</td>
</tr>
<tr>
<td>D2</td>
<td>Sunny</td>
<td>Hot</td>
<td>High</td>
<td>Strong</td>
<td>No</td>
</tr>
<tr>
<td>D3</td>
<td>Overcast</td>
<td>Hot</td>
<td>High</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D4</td>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D5</td>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D6</td>
<td>Rain</td>
<td>Cool</td>
<td>Normal</td>
<td>Strong</td>
<td>No</td>
</tr>
<tr>
<td>D7</td>
<td>Overcast</td>
<td>Cool</td>
<td>Normal</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D8</td>
<td>Sunny</td>
<td>Mild</td>
<td>High</td>
<td>Weak</td>
<td>No</td>
</tr>
<tr>
<td>D9</td>
<td>Sunny</td>
<td>Cool</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D10</td>
<td>Rain</td>
<td>Mild</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D11</td>
<td>Sunny</td>
<td>Mild</td>
<td>Normal</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D12</td>
<td>Overcast</td>
<td>Mild</td>
<td>High</td>
<td>Strong</td>
<td>Yes</td>
</tr>
<tr>
<td>D13</td>
<td>Overcast</td>
<td>Hot</td>
<td>Normal</td>
<td>Weak</td>
<td>Yes</td>
</tr>
<tr>
<td>D14</td>
<td>Rain</td>
<td>Mild</td>
<td>High</td>
<td>Strong</td>
<td>No</td>
</tr>
</tbody>
</table>

Figure 4.1: Training Dataset for playing tennis
[Tom Mitchell TextBook]

Fig 4.1 shows data set which has target attribute PlayTennis. It can have values yes or no for different mornings. Whether a person has to play Tennis or not depends on attributes like Outlook, Temperature etc. Given a new tuple we predict whether he can play Tennis or not by traversing decision tree.
4.2 Decision Tree Learning Algorithm

Basic algorithm in decision tree is called as ID3. It learns decision trees by constructing it top-down fashion. Each attribute is evaluated using a statistical test to determine how well it alone classifies the training examples. The best attribute is selected and used as the test at the root node of the tree. A descendant of the root node is then created for each possible value of this attribute, and the training examples are splitted to the branch corresponding to the example’s value for this attribute. The entire process is then repeated using the training examples associated with each descendant node to select the best attribute to test at that point in the tree.

The attribute to be tested at a node in the tree for classifying examples depends on a statistical property called information gain of attribute. Information gain measures how well a given attribute separates the training examples according to their target classification. ID3 uses this information gain measure to select among the candidate attributes at each step while growing the tree.

Given a collection S, and if the target attribute has c different values, then the entropy of S relative to this c-wise classification is defined as $\text{Entropy}(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$, where $p_i$ is proportion of S belonging to class i. Information gain, $\text{Gain}(S, A)$ of an attribute A relative to a collection of examples S, is defined as $\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$ where Values(A) is the set of all possible values for attribute A and $S_v$ is the subset of S for which attribute A has value v.
In the given dataset, S is a collection of 14 examples. Attribute Wind has the values Weak or Strong. There are 9 positive (i.e., play tennis = yes) and 5 negative examples denoted as \([9+, 5-]\). Of these 14 examples, 6 of the positive and 2 of the negative examples have Wind = Weak. The information gain from attribute Wind is calculated as follows:

Values(Wind) = Weak, Strong  
\[S = [9+, 5-]\]  
\[S_{\text{weak}} = [6+, 2-]\]  
\[S_{\text{strong}} = [3+, 3-]\]  

\[
\text{Gain}(S, \text{Wind}) = \text{Entropy}(S) - \sum_{v \in \{\text{Weak, Strong}\}} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v)
\]

\[
= \text{Entropy}(S) - \frac{8}{14} \cdot \text{Entropy}(S_{\text{weak}}) - \frac{6}{14} \cdot \text{Entropy}(S_{\text{strong}})
\]

\[
= 0.940 - \frac{8}{14} \cdot 0.811 - \frac{6}{14} \cdot 1.00
\]

\[
= 0.048
\]

Information gain is used to select the best attribute at each step in constructing the tree. So information gain is calculated for each attribute. The information gain values for all four attributes are:

Gain(S, Outlook) = 0.246  
Gain(S, Humidity) = 0.151  
Gain(S, Wind) = 0.048  
Gain(S, Temperature) = 0.029

where S denotes the collection of training examples from dataset shown in Fig 4.1.

According to the information gain measure, the Outlook attribute provides the best prediction of the target attribute, PlayTennis, over the training examples. So, Outlook is selected as the decision attribute for the root node, and branches are created below the root for each of its possible values it takes i.e., Sunny, Overcast, and Rain. The resulting partial decision tree is shown in Figure 4.3. The training examples are split to the branch corresponding to the example’s value for this attribute. Every example for which Outlook = Overcast is a positive example of PlayTennis. Therefore, this node of the tree becomes a leaf node with the classification PlayTennis = Yes. In contrast, the descendants corresponding to Outlook = Sunny and Outlook = Rain do not have all positive or all negative. So the decision tree procedure is repeated for these nodes.

The process of selecting a new attribute and partitioning the training examples is now repeated for each descendant node, this time using only the training examples associated with that node. Attributes that have been incorporated higher in the tree are excluded, so that any given attribute can appear at most once along any path through the tree. This process continues for each new leaf node until either of two conditions is met: (1) every attribute has already been included along this path through the tree, or (2) the training examples associated with this leaf node all have the same target attribute value. Final decision tree is shown in Figure 4.2.
4.3 Performance Prediction of Engineering Students using Decision Trees

Decision tree algorithms are applied on engineering students past performance data to generate the model and this model can be used to predict the students performance in first year engineering exams. This will enable to identify the students who are likely to fail in advance so that they can improve their performance.

This section describes the model that predicts the academic performance of the engineering students using decision trees. The features selected for creating the model are based on students past performance, because past performance of a student is indicative of his future performance. The model is built by analyzing the data tuples from training data having a set of attributes shown in Figure 4.4. The training data here consists of previous students past performance and their grades in engineering first year. For each tuple in the training data, the value of target class label attribute is known. Using this data we build the model using decision tree technique. Next, test data is used to check
the accuracy of the model. If the accuracy of the model is acceptable then the model can be used to classify the data tuples for which the class label is not known (i.e., we predict freshies grades).

<table>
<thead>
<tr>
<th>Sr No.</th>
<th>Name</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Branch</td>
<td>COMP, IT, FLX, ETC, MECH</td>
</tr>
<tr>
<td>2</td>
<td>HSCPercent</td>
<td>Distinction(above 75%), Firstclass(60%-75%),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HSecondclass(50%-60%), Secondclass(less than 50%)</td>
</tr>
<tr>
<td>3</td>
<td>HSCMaths</td>
<td>Real</td>
</tr>
<tr>
<td>4</td>
<td>HSCPCM</td>
<td>Real</td>
</tr>
<tr>
<td>5</td>
<td>HSCCET</td>
<td>Real</td>
</tr>
<tr>
<td>6</td>
<td>SSCpercent</td>
<td>Distinction(above 75%), Firstclass(60%-75%),</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HSecondclass(50%-60%), Secondclass(less than 50%)</td>
</tr>
<tr>
<td>7</td>
<td>SSCMaths</td>
<td>Real</td>
</tr>
<tr>
<td>8</td>
<td>SSCScience</td>
<td>Real</td>
</tr>
<tr>
<td>9</td>
<td>Category</td>
<td>Open, OBC, SC, ST, Others</td>
</tr>
<tr>
<td>10</td>
<td>Gender</td>
<td>Male, Female</td>
</tr>
<tr>
<td>11</td>
<td>Livinglocation</td>
<td>Rural, Urban</td>
</tr>
<tr>
<td>12</td>
<td>SSCBoard</td>
<td>State, CBSE</td>
</tr>
<tr>
<td>13</td>
<td>Atype</td>
<td>CAP, MGMT</td>
</tr>
<tr>
<td>14</td>
<td>Father_Occupation</td>
<td>PublicSectorjob, Privatejob,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>business, Farmer, Teacher, Other</td>
</tr>
<tr>
<td>15</td>
<td>Mother_Occupation</td>
<td>PublicSectorjob, Privatejob,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>business, Farmer, Teacher, Other</td>
</tr>
<tr>
<td>16</td>
<td>SSCMedium</td>
<td>English, Regional</td>
</tr>
<tr>
<td>17</td>
<td>Fersult (target variable)</td>
<td>PASS, FAIL, ATKT</td>
</tr>
</tbody>
</table>

**Figure 4.4: Attributes**

[13]

HSCPercent - The percentage of marks obtained by student in Higher secondary class.
HSCMaths - Marks in HSC Mathematics
HSCPCM - Sum of Physics, Chemistry and Mathematics marks in HSC exam (out of 300).
HSCCET - Marks obtained in common entrance test.
SSCpercent - The percentage of marks obtained by student in Higher secondary class.
SSCMaths - The Percentage of marks in SSC Mathematics.
SSCSci - The Percentage of marks obtained in SSC Science.
Atype - The admission type which may be through central process or through Management
quota of institute.
SSCMedium - Medium of the student at secondary school level.
FEresult - Result of student in First Year of Engineering. This can take the values PASS, FAIL, or ATKT. In general, if a student fails in up to three theory and two practical subjects of current academic year, he/she is awarded ATKT and promoted to next class provided they do not have backlog of previous year.

The decision tree obtained is shown in Figure 4.5. The study has shown that the developed model helped First Year Engineering students to improve their grades in final exam.

![Decision Tree](image)

Figure 4.5: Decision tree for three class prediction

4.3.1 Application to CS101 COURSE

In a college like IIT we cannot use this because all are intelligents. But we can use the decision tree technique to predict the grade of student in CS101 COURSE so that student can improve their performance in endsem and improve their grade. Because in a course like CS101, showing marks of one student to other is not good and student cannot predict their
grade as course has huge strength and they also do not know marks of all other people. The attributes are quiz1, quiz2, midsem, quiz3, project marks. Through historical information we can build the decision tree. So for a student who is yet to write endsem this year can approximately know his grade before writing endsem, which is predicted with the marks of exams conducted till then. Based on his predicted grade, we can also provide him customized references so that he read those materials and can improve in endsem. This model will work per professor because each professor paper difficulty level, exam criterion grading strategy etc are different.

4.4 Combination of classifiers

In previous model, we used single classifier (decision tree). To improve the accuracy of the model, generally combination of classifiers is used. We can identify students (especially in large classes) who are in risk at early stage so that the professor can advise the students. This section focuses on classification of students to predict their final grade based on features extracted from logged data. We use combination of classifiers to predict the target variable rather than single classifier. Figure 4.6 shows final grade distribution of online PHYSICS course conducted at Michigan State University.

![Graph of distribution of grades in course PHY183](image)

Figure 4.6: Grade Distribution

[10]

We can group the students according to their final grades in several ways. Three of the ways are:

i) 9 possible class labels which are same as students grades.
ii) Group them into three classes, “high” representing grades from 3.5 to 4.0, “middle” representing grades from 2.5 to 3, and “low” representing grades less than 2.5.

iii) Categorize students into 2 class labels. Passed for grades above 2.0 and Failed for grades less than or equal to 2.0

The features that are used for classification are

1. Total number of correct answers. (Success Rate)
2. Getting the problem right on the first try vs. those with high number of tries (Success at the first try) A student who gets all correct answers need not necessarily be in the successful group because they might take an average of 5 tries per problem. This feature helps to identify such people.
3. Total number of tries for doing homework. (Number of attempts before correct answer is derived)
4. Total time that passed from the first attempt to get correct solution. Also the time at which the student got the problem correct relative to the due date.
5. Total time spent on the problem regardless of whether they got the correct answer or not.
6. Reading the supporting material before attempting homework vs. attempting the homework first and then reading upon it.
7. Submitting a lot of attempts in a short amount of time without looking up material in between, versus those giving it one try, reading up, submitting another one, and so on.
8. Giving up on a problem versus students who continued trying up to the deadline.

With these attributes, we can build a model in the same way as discussed earlier. But to improve accuracy of the model, we use combination of classifiers. We use many classifiers because it is not possible to come up with a single classifier that can give good results in all cases. The optimal classifier is dependent on problem domain. If single classifier does not give good performance then combine multiple classifiers to improve classification performance. There are different ways of combining multiple classifiers (CMC).

**Offline CMC:** Out of many classifiers, choose that classifier which has least error rate on given dataset. It has better performance than individual classifier.

**Online CMC:** Each classifier votes to what class, student belongs to. The class getting the maximum votes from the individual classifiers is assigned to student. For example, if many classifiers predict that the student fails, then we assign “fail” label to student.

In the above problem, the objective is to predict the students final grades based on their web-use features, which are extracted from the homework data. We can also include other components like quizzes, midsem, endsem, attendance etc.

### 4.5 Understanding Learners Learning Style

Educational data mining is used to study the data available in the educational field and bring out the hidden knowledge from it. Instructors can first receive a detailed record of a learners behavior and then data mining algorithms can be employed to discover patterns
to characterize learners. By observing how learners behave during their online self-study, we can understand learners learning style and can then make suggestions to learners.

### 4.6 Sequential Pattern Analysis

A sequence is an ordering of events and each event in the ordered list is called item. A sequence \( x \) is a subsequence of another sequence \( y \), if \( x \) can be formed from \( y \) by leaving out some events without disturbing the relative positions of the remaining events. For example, if \( x = \langle C, E, D, B \rangle \) and \( y = \langle C, A, E, D, E, G, C, E, B, G \rangle \), then \( x \) is a subsequence of \( y \). Given a set of sequences \( S \), the support of a sequence \( x \) in \( S \) is the number of sequences containing \( x \), denoted as \( \text{support}(x) \). Given a percentage \( \text{min}_\text{sup} \) as the minimum support threshold, a sequence \( x \) is said to be sequential pattern observed in \( S \) if \( \text{support}(x) \geq \text{min}_\text{sup} \).

Example: Figure 4.7 shows a Sequence Database which contains items \( A,B,C,D,E,F \) and say it has \( \text{min}_\text{sup}=3 \). \( S1, S2, S4, S5 \) contain subsequences \( \langle A, D, B, D \rangle \). So \( s \) is a sequential pattern of length 4. \( \langle A, C \rangle \) is not a sequential pattern as it is supported only by two sequences \( S1 \) and \( S4 \). These sequential patterns are useful to evaluate learners activities and accordingly we can provide customized resources.

<table>
<thead>
<tr>
<th>Sequence_id</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>( A, C, D, F, B, C, D )</td>
</tr>
<tr>
<td>S2</td>
<td>( A, D, E, B, A, D )</td>
</tr>
<tr>
<td>S3</td>
<td>( C, D, C, B, C, D )</td>
</tr>
<tr>
<td>S4</td>
<td>( A, C, D, B, B, D )</td>
</tr>
<tr>
<td>S5</td>
<td>( A, F, D, E, B, D )</td>
</tr>
</tbody>
</table>

Figure 4.7: Sequence Database [11]

To profile learners based on their learning behavior, studying actions are logged by the system to extract patterns from it. The learning logs involve a complex series of low-level events spaced along a time dimension. Through a log parsing mechanism, sequence of temporally ordered learner actions are generated. These sequences of learner actions are then fed into the sequential mining algorithm to discover patterns across the learning logs. With these patterns we can study student learning style and we can evaluate him.
Example: Suppose for each video lecture let us say there is an assignment. If the student doesn’t get good marks in assignment then assume our e-learning system gives personalized references to student to study and then go to next lesson. Then through this pattern analysis we can know whether student is reading those customized references and going to next lesson or not. As the data is logged, we can know detail actions done by the student like how much effort he has put into subject etc.

Till now, we have seen about what content to be placed in ematerials, prediction of student performance etc. In the next chapter, we discuss about usability testing which is very important to know the satisfaction of users about the e-learning platform which we develop.
Chapter 5

Usability Testing

Software usability is a aspect of Human computer interaction that can benefit from knowledge of the user and their tasks. Instructors often do not have the knowledge to handle usability. One set of methods for determining whether an application enables users to achieve their predetermined goals effectively and efficiently is usability evaluation with end users. This chapter focuses on usability evaluation which is based on SUMI (Software Usability Measurement Inventory) questionnaires.

5.1 Introduction

There are many websites that offer e-learning, but many are not satisfiable. For e-learning website to be usable, we should know what student expect from the site, how they learn, what motivates them and what helps them to achieve their learning goals.

5.2 Current Approaches to Usability Evaluation

Usability Testing is used to verify user interface quality. Usability methods are divided into inspection methods (without end users) and test methods (with end users). Inspection methods are used for identifying usability problems and aim at improving the usability of an interface design by checking it against established standards. These methods include heuristic evaluation, cognitive walkthroughs. In Heuristic evaluation (HE) usability specialists judge whether each element follows established usability principles or not. In cognitive walkthrough (CW), the analyst explores the systems functionalities. Testing with end users provides information about how people use the systems and their problems with interface. Questionnaires need some experience to design and are useful for studying how end users use the system and their preferred features.
5.3 Aspects of User Satisfaction

Efficiency: This shows how quick and effective is software for performing tasks.
Helpfulness: This shows how software communicates in a helpful way and solves operational problems.
Learnability: Ease with which a user can get started and learn new features of the product.

5.4 SUMI Evaluation

SUMI is used for measuring users perception of the software. It assess user satisfaction with software. This tool consists of 50-item paper based questionnaire in which respondents score each item on a three point scale (i.e., agree, undecided, disagree). The following are some sample of questions:
a) Does software responds too slowly to inputs?
b) Whether instructions and prompts are helpful?
c) Whether working with this software is satisfactory?
d) Does presentation of system information is clear and understandable?
The questionnaire is designed to know the aspects of user satisfaction.

5.5 SUMI Questionnaires

SUMI Questionnaires consist of 50 questions for which user selects one of three responses (agree, dont know, disagree). SUMI evaluation package contains SUMISCO software which scores the questionnaires and compares the results to the standardization database. The mean score of the standardization database is 50, and has a standard deviation of 10. A system that gets a score in the range 40 to 60 is acceptable in terms of usability. SUMI questionnaire can also provide information about particular items. This analysis is called Item Consensual Analysis. The result is compared similar to above with standardization database and says whether item is acceptable or not.
Chapter 6

Conclusion And Future Work

Different learners have different learning styles. Different people build process and store knowledge in different ways. So, different people will relate to a particular learning resource in different ways. Human instructors can learn which style of presentation suits which learner and adjust their mode of presentation accordingly. Learners have different backgrounds and previous experience, so different learners may need to focus on different material to achieve the same learning objective. But the current e-Learning systems do not allow for diversification and presents the same sequence of learning modules to every user of the system.

The proposed learning system gives learning support based on individual learning characteristic. Different learning proposals are provided to students in feedback according to learners’ learning rate. A test is provided after completion of each chapter and the system gives corresponding learning proposals according to test result. In the second stage, I will implement the proposed system (Figure 6.1) and also predict the grades of CS101 students by collecting historical data and then give customized references so that they can improve their grades in endsem.

Example: Based on the exam result, we can classify him as dull or average or clever or extra-ordinary. According to his category, we can give him references. For example, if the exam is conducted on random variables topic of probability and the student does not perform well in the exam, then we refer him to read Random Variables topic of Sheldon Ross textbook (very basic book) and we can also refer him set of basic video lectures in this topic. Now the student can read those personalized references suggested to him, and can attempt the test so that he can know his improvement. If the student performs extremely well in the exam, then we can refer him Random Variables topic of Papulios Textbook (tough textbook to understand) and also refer him different set of video tutorials on Random Variables which discuss topic in depth. If the student is dull, then we can impose restriction by not allowing him to go to next chapter until he improves in this chapter by reading suggested references. We impose such a restriction for dull category because student however cannot progress in further topics without having proper basics and there might be a chance that he get scared of the subject if he continues without proper basics. In this manner, according to category of student we can suggest him references.
**General Example 1:** When we buy a book or any item in flipkart, we get many personalized suggestions to buy from flipkart. The way the algorithm runs is, system considers our purchased item as our interest and all resources which are closely related to our interest are given. For example, if we buy a C textbook by Dennis Ritchie then related books on C by Yashwanth Kanetkar, other books on C by Dennis Ritchie etc. are given as suggestions for purchasing.

**General Example 2:** The ads we get in Inbox of gmail are personalized ads. These ads are given to user according to his visited websites, recent searches, clicks and information from emails of inbox.

Personalized resource recommendation can be applied at many granularities, recommendation at problem level, topic level, course level etc. Suppose if a student is unable to solve a problem, then if he clicks on the help button then we can redirect him to a short segment video lecture (recommendation at problem level). So, after listening this the student can solve the problem. Even still, if the student is not able to solve the problem, then we can redirect him to another short video which explains concept in much detailed way.

Finally, these customized suggestions definitely aim at improvement of student.

![Figure 6.1: Structure of Proposed System](image-url)
Bibliography


[17] www.it.iitb.ac.in/ekshiksha/(referred on 31st Aug 2012)