Using Data Mining to Discover the Correlation between Web Learning Portfolios and Achievements

Chien-Ming Chen¹, Cheng-Hao Ma¹, Bin-Shyan Jong¹, Yen-Teh Hsia¹ and Tsong-Wuu Lin²
Chung Yuan Christian University¹, Soochow University²
calculusxp@cg.ice.cycu.edu.tw, col@cg.ice.cycu.edu.tw, bsjong@ice.cycu.edu.tw, hsia@ice.cycu.edu.tw, twlin@cis.scu.edu.tw

Abstract - Internet learning is different from traditional classroom teaching. This is because there is no actual contact between teachers and students; therefore, it is difficult for teachers to keep track of students’ learning conditions. By analyzing the learning portfolios of students studying through online learning, teachers can observe their learning activities. This study conducted a cluster analysis on each attribute of students’ learning portfolios. By observing the activities of individuals in each cluster and investigating their activities at every stage, the behavioral styles were determined and the appropriate warning messages were sent to students in certain clusters. The learning portfolio analysis system proposed in this study provides statistical and data mining techniques, focusing on students’ assignment grades, test results, and records of online learning portfolios, for explorative analysis. It also provides a remedial mechanism for teachers to send out adaptive warning messages according to students’ learning portfolios. The evaluation in this study demonstrates that the sent warning messages can affect students’ learning behaviors and achievements.

Index Terms - Learning portfolio, cluster analysis, adaptive warning message

INTRODUCTION

Due to the great popularity of the personal computer and Internet in the modern world, online learning has become a significant trend. Students are able to participate in online learning using PCs, handheld devices, or mobile phones via the Internet from any location and at any time. Students can also determine their learning status through online evaluations, thus allowing them or their teachers to arrange remedial instruction or learning processes regarding the unfamiliar parts of the curriculum.

However, online learning is different from the traditional classroom instruction. Since teachers and students do not interact face to face, the teachers cannot ascertain the learning status of their students effectively. Moreover, since the students have more freedom and experience less pressure in the online learning environment than in the traditional learning environment, the possibility exists that they become inefficient learners and their attitude toward learning becomes negative. If teachers do not identify their students’ learning situations immediately, there is a great chance that their learning will be ineffective. This study will analyze students’ learning portfolios to aid teachers in immediately determining their students’ learning status and behaviors promote their learning effects.

Learning portfolios generally refer to any documents created during the learning process, for example, notes, assignments, discussion contents, and online learning activities. In addition, students’ behaviors, for instance, the time they spend on reading learning materials, the time they spend online, their logon frequency, their history of completing assignments, and records of their online conversations with others on the same learning platform, can be recorded in a database (DB) of the platform. Thus, the learning portfolios created during the course of online learning should include even more detailed raw data. Therefore, we have built a learning portfolios analysis system to investigate the correlation between students’ learning behaviors and learning achievements in order to enable teachers to exercise a greater degree of control over their students’ overall and personal learning situations. We also examine how teachers can provide immediate guidance to promote the students’ learning effects.

This study proposes a learning portfolio analysis system based on statistical and data mining techniques that focuses on students’ assignment grades, test results, and records of online learning portfolios in order to conduct an explorative analysis and help teachers to identify the learning variations in students’ learning processes. According to the results of the system analysis, teachers should adjust their teaching methods accordingly in order to provide learning assistance and promote learning effects in student whose negative behavior is affected by low learning performance.

LEARNING PORTFOLIOS

Learning portfolios provide students with a specific method to evaluate their own learning situations [1]. Therefore, all kinds of student activities are included in the portfolios, such as records of their interaction with others, assignments, test papers, discussion content, and online learning records. By analyzing such data, students can understand their own learning status, and teachers can determine the effectiveness of their lectures. Analyzing learning portfolios is important.
not only because it provides the opportunity to observe students’ learning effects but also because it improves the principles of teaching and learning [1][2][3].

In the traditional classroom environment, teachers can identify which portfolio they require clearly and easily. For example, a teacher can take a roll call to immediately obtain a student’s attendance status or conduct evaluations and hold discussions to understand the level of familiarity that a student has with a certain concept. These learning portfolios are defined by the teachers themselves, and they clearly understand the function of each portfolio. On the other hand, students’ behaviors are recorded through a learning platform in online learning. Unless the raw data is analyzed and interpreted appropriately, teachers will not easily understand the significant attributes of students’ learning portfolios and evaluate their learning effects. Thus, research has been conducted to build a system wherein learning activities can be observed in the online learning environment [4].

By analyzing various types of learning portfolios, teachers might aim to determine the cause-and-effect relationships in order to diagnose their students’ learning effects or other unknown aspects. For example, they aim to understand the following: “What elements or behaviors would lead to ineffective learning?” and “Could midterm examination scores affect students’ online behaviors?”

**DATA MINING**

Data mining [5] through tools such as KDD (knowledge discovery in database) is a useful technique for searching for valuable hidden information in a DB. Techniques related to data mining include logical analysis, statistics, and AI technology. The purposes of data mining are to analyze data to obtain the relevant information and determine the relations among them and to build correlative models in all kinds of domains. Presently, the use of this technique is very popular. In an online learning platform, it can assist teachers in analyzing all kinds of learning portfolios, presenting students’ online learning records, seeking out students with certain learning behaviors, and providing guidance to promote their learning achievements.

**CLUSTER ANALYSIS**

In general, a cluster analysis [5] is applied to classify a large amount of data in order to determine similar clusters based on certain variables and separate them from the clusters that are not similar. It is also used to search for hidden characteristic attributes and some special data attributes.

**ADAPTIVE LEARNING**

“E-Learning” provides students with the opportunity to learn online any time and anywhere. Hence, the information services provided by learning platforms have become one of the most important factors that affect students’ learning behaviors and achievements. One study indicates that an uninteresting curriculum can result in failure for most students learning through the internet [6]. The curriculum design is generally such that it satisfies most students’ needs, but not each individual learner’s needs. For example, without creative classroom presentation, all students are unable to pay full attention during a lecture or course. Hence, expecting the same learning results from each student will not achieve fruitful outcomes, and many students’ self-confidence would be reduced. To resolve the abovementioned issue, some scholars propose Adaptive Educational Hypermedia (AEH) [7][8], which focuses on individualized learning experiences. Considerable research has been devoted to develop a general Adaptive Educational Hypermedia System (AEHS). An AEHS contains four important components: Knowledge Space (KS), User model (UM), Observations (OBS), and Adaptation Model (AM) [9]. KS mainly includes the following two parts: (1) educational resources and relative quality information and (2) a controllable graphics framework and relational learning objectives in the knowledge field. The UM describes a learner’s relative background, while the OBS comprise learners’ activity records and monitoring mechanisms, which is what this study is interested in. The AM defines learners’ behavioral rules in the AEHS. These rules are formed on the basis of Concept Selection Rules and are used to supplement the other components to form the AEHS.

**LEARNING PORTFOLIOS ANALYSIS SYSTEM**

The instruction provided through online learning is quite different from that provided in traditional classes. While in the former, students learn through the Internet and acquire knowledge independently, in the latter, teachers provide instruction and solve students’ questions. Thus, the level of difficulty of the evaluations that teachers perform on their students’ learning portfolios is different in the two situations. Since there is no actual contact between teachers and students, the teachers encounter problems in controlling the learning status of each student. This study adopts a learning portfolio diagnostic system [10] that teachers can use to analyze such students’ learning portfolios. Through this system’s interface, teachers can observe students’ activities online. If students experience a bad learning situation or achieve low scores, teachers can detect it immediately and begin the remedial process.

Students’ online activities are directly recorded in a DB; however, this results in the DB containing a large amount of unorganized data. Therefore, an important issue is how teachers can obtain significant information from such a huge data set. The diagnostic system adopted in this study saves portfolios in its portfolio DB after reorganizing and filtering them appropriately. This could make it easy for teachers to determine the attributes of students’ online activities and provide a time scale for the learning portfolios. Thus, teachers will be able to obtain more precise learning portfolios for each student. Furthermore, it provides an interface for teachers to import additional data related to the courses—such as the scores of students’ written papers.
or projects—and data relating to students’ actions that reveal their learning styles into the portfolio DB.

After determining the attributes of online activities, how should we define students’ learning behaviors and good learning situations? For example, should these be defined based on how frequently students log on or how frequently they participate in discussions? Badly defined learning behaviors would result in an inaccurate reflection of students’ learning status and would not be able to detect students’ behaviors efficiently. Therefore, constructing an appropriate definition for learning behaviors is a very important issue. For this reason, this system provides various methods through which teachers can define students’ learning behaviors by combining hierarchical logical descriptions and using field information from the learning portfolios DB.

**SYSTEM ARCHITECTURE**

This system analyzes raw data obtained from the learning portfolios of “i-learning” (http://i-learning.cycu.edu.tw), which is a famous learning platform in Taiwan [11]. It provides teacher with all kinds of services to conduct distance learning. While students can study learning materials, join course discussions, and interact with their classmates or teachers through this platform, teachers can use this platform to arrange online tests and create discussions regarding the curriculum. In the adopted system, the interface that teachers can use has been developed with the JSP language, which was also used to deal with the backend operation of managing the learning portfolios. The learning portfolios analysis system saves the portfolios in its portfolio DB through reorganization and filtration. Through the integrated analysis conducted by this system, teachers are able to determine students’ learning behaviors and learning situations.

![System Architecture Diagram](image)

**CAPTURING AND FILTERING ONLINE LEARNING PORTFOLIOS AND INPUTTING RELATED INFORMATION**

The online learning portfolios are recorded in the learning platform’s DB and Web server log. However, since the content of the Web log is simplified and system information oriented, it is difficult for the teachers without a computer science background to interpret the information. For this study, we were required to investigate the learning platform’s DB scheme, define students’ learning portfolios, and divide the learning period into specific units. Therefore, it is evident that the raw data must be filtered in order to capture valuable information from the students’ learning portfolios. Moreover, a student can log in or log out too frequently, the content of an article might not be related with the topic under which it is posted, the learning materials study time might be shown as being excessively long or short, students may click on learning material links randomly, etc., which are actions that can result in inaccurate information and hinder the diagnosis of students’ learning situations.
EVALUATING AND DEFINING STUDENTS’ LEARNING STATUS AND BEHAVIORS

Even after capturing, filtering, and integrating all of the students’ learning portfolios, the following concern remains: although we can obtain the values of various attributes, how can we classify the types of students’ learning behaviors? For example, what kinds of behaviors are represented by the attributes of “learns online frequently,” “often participates in discussions,” and “studies online learning materials every week”? The solution lies in analyzing these activity records to define online learning behaviors. The diagnostic system applies many types of clustering algorithms to form clusters based on students’ learning portfolios. It can allocate the students to different clusters. For example, let us consider the attribute of “the time spent studying online learning materials.” When we divide them into two clusters by using the average study time as the threshold value, the students with study times that are less than the average are categorized in the cluster of “spent less time studying online learning materials.” On the other hand, the students with higher study times are categorized into the cluster of “more time spent studying online learning materials.” Moreover, teachers can apply automatic clustering, conditional clustering, logic clustering, or other cluster analysis algorithms to categorize students’ learning status and behaviors more accurately.

- Automatic clustering: This method applies a cluster analysis algorithm to form clusters of the target attributes. This study adopted the EM (Expectation Maximization) [12], FarthestFirst [13], and Kmeans [5] algorithms. Each algorithm produces a different clustering result. Thus, teachers can obtain more references that provide different interpretations of students’ online learning situations, which assist the teachers in evaluating the students’ overall online learning situations.

- Conditional clustering: In this method, clusters of students are formed according to predefined conditions. The teachers can define the conditions of the clusters that they want the students to be divided into.

- Logic clustering: This method uses a serial of logical combinations such as “and” and “or” to form multilevel clusters. These provide teachers with more flexible choices in terms of clustering results. These clustering methods are integrated in the learning portfolios analysis system. It can combine many types of students’ learning activities and observe the learning status of a specific student. Moreover, the learning status and behaviors thus obtained can be combined with other behaviors through multilevel clustering.

COUNSELING MECHANISM

This system provides counseling mechanisms through e-mails and cell phone text messages to specific students. The content of the messages is written either by teachers or by the system according predefined samples. With these mechanisms, teachers can encourage the well-performing students or remind the poor-performing students that they must take studying more seriously. In addition, such mechanisms can help students to understand their learning status, regulate their own behavior and motivate themselves to promote the learning effect, and increase the degree of participation in online learning.

EVALUATION

The purpose of this study was to verify the effectiveness of this system by examining the online behaviors, learning achievements, and effects of warning messages on students from different colleges. Since there are different styles of classroom presentation and learning methods, an investigation of the learning processes in different colleges would enable us to obtain the student learning model of each college. We conducted exams on six courses in all the colleges. Table 1 presents the number of people who were taking the courses. We found students learning ineffectively on line out according to analysis of the system, and then they were divided into an experiment group and a control group. The experiment group was evaluated based on its online performance from the beginning of the semester to one week before the exams. The students who did not perform well received warning messages to let them know that the time of the exam was approaching. The results pertaining to the students’ achievements on the exams are as follows.

<table>
<thead>
<tr>
<th>Course name</th>
<th>Total students</th>
<th>Experiment group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>System program</td>
<td>143</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>Economics</td>
<td>61</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Educational test and evaluation</td>
<td>68</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Developmental Psychology</td>
<td>80</td>
<td>29</td>
<td>44</td>
</tr>
<tr>
<td>Statistics</td>
<td>68</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>Curriculum design</td>
<td>46</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

System program: The list of control group is created the same way as for experiment group, without sending messages. Table 2 presents the ANOVA (Analysis Of Variance) test results, with a confidence level of 95%, for the students’ achievements obtained from the system.

<table>
<thead>
<tr>
<th>Team</th>
<th>Number of Students</th>
<th>Summation</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>16</td>
<td>946</td>
<td>59.125</td>
<td>359.1833</td>
</tr>
<tr>
<td>Control group</td>
<td>22</td>
<td>1002</td>
<td>45.54545</td>
<td>284.5455</td>
</tr>
</tbody>
</table>

As shown in Table 2, there is a significant difference in exam performance between the experiment and control groups (F = 5.41169 > F = 4.113165). The students in the...
experiment group were more active and cautious while learning online, and they also demonstrated a certain amount of progress due to the warnings they received. On the other hand, those in the control group used self-learning methods and were unable to realize their weaknesses or extent of inactivity without the guidance of the warning messages. Thus, there was a difference between the learning achievements of the students in the control and experiment groups.

**Economics:** The students in the control group were those who had demonstrated average performances online, had comparatively low exam results, and did not receive warning messages. Table 3 contains the ANOVA results for the economics exam performances.

**Economics Statistical Results of Test Performance**

<table>
<thead>
<tr>
<th>Team</th>
<th>Number of Students</th>
<th>Summation</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>10</td>
<td>557</td>
<td>55.7</td>
<td>357.3444</td>
</tr>
<tr>
<td>Control group</td>
<td>14</td>
<td>596</td>
<td>42.57143</td>
<td>90.10989</td>
</tr>
</tbody>
</table>

Based on the statistical tests, there is a significant difference in exam performance between the students who did and did not receive warning messages. Further, they demonstrated that under supervision, the students who received warnings performed better than those who did not receive warnings in terms of their learning achievements.

**Educational testing and evaluation:** The students in the control group were randomly selected from among those who did not receive messages. Since there was no midterm exam held for this class, we first observed the condition of this class and then defined the score system for the online performance. The performance online is the results were as follows: duration of login—20%, days of login—15%, number of postings for discussions and course interactions—10% each, number of clicks for the relevant material—15%, and reading time for the relevant material—30%. Table 4 contains the educational exam results and ANOVA results for the online performances.

**Educational Test and Evaluation Statistical Tests of On-line Performance**

<table>
<thead>
<tr>
<th>Team</th>
<th>Number of Students</th>
<th>Summation</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental group</td>
<td>12</td>
<td>556</td>
<td>46.33333</td>
<td>324.9697</td>
</tr>
<tr>
<td>Control group</td>
<td>14</td>
<td>636</td>
<td>53</td>
<td>340.3636</td>
</tr>
</tbody>
</table>

Based on the statistical results, there is no significant difference between the experiment and control groups. The students who received warning signals in the experiment group did not lag behind the students in the other group.

**Statistics:** We used gender as a criterion to analyze this class. We examined whether there exists a gender difference with respect to online learning and the effectiveness of warning messages. In the experiment group, we randomly selected female students and observed whether they required warning messages during the online course. This experiment, which involved sending adaptive messages, was based on the students’ online learning conditions and midterm exam performances. The students in the control group were randomly selected male students. Table 6 presents the ANOVA results for the exam performances.

**Curriculum design:** The students in the control group were randomly selected from among those who did not receive warning messages. Table 7 presents the ANOVA results for the online performances in the curriculum design course.

**Developmental Psychology:** The students in the control group were randomly selected from those who did not receive warning messages. Table 5 presents the ANOVA results for the online performances in the developmental psychology course.

**Curriculum Design Statistical Results of On-line Performance**

As shown in Table 4, there was no significant difference between the experiment and control groups in terms of their online performances. From another perspective, the students in the experiment group who received warnings did not lag behind the students in the other groups. This demonstrates that guidance through the use of warning messages is effective for students who have learning problems. Further, from the statistical results, it is evident that on average, the students in the experiment group did not significantly lag behind those in the control group in terms of completing the online learning activities.

Based on the statistical results, it appears that the use of warning messages is fairly effective for female students, who also demonstrated an overall good performance in the exams. The learning process demonstrates that the online learning condition of females is much better than that of males. After receiving adaptive warning signals, the female students performed much better in terms of completing the online learning activities. From this result, we can tentatively conclude that female students would improve considerably in their learning if placed under system supervision.

Based on the statistical results, there was no significant difference between the experiment and control groups. The students who received warning signals in the experiment group did not lag behind the students in the other group.

**Statistics:** We used gender as a criterion to analyze this class. We examined whether there exists a gender difference with respect to online learning and the effectiveness of warning messages. In the experiment group, we randomly selected female students and observed whether they required warning messages during the online course. This experiment, which involved sending adaptive messages, was based on the students’ online learning conditions and midterm exam performances. The students in the control group were randomly selected male students. Table 6 presents the ANOVA results for the exam performances.

**Curriculum Design Statistical Results of On-line Performance**

As shown in Table 4, there was no significant difference between the experiment and control groups in terms of their online performances. From another perspective, the students in the experiment group who received warnings did not lag behind the students in the other groups. This demonstrates that guidance through the use of warning messages is effective for students who have learning problems. Further, from the statistical results, it is evident that on average, the students in the experiment group did not significantly lag behind those in the control group in terms of completing the online learning activities.
As shown in the statistical results, there is a significant difference in online performance between the students in the experiment and control groups. They also demonstrate that the students in the experiment group did not pay attention to the warning messages. On average, the overall online performances of both the groups were fairly low. However, this study indicates that the reason why students did not pay attention to online learning was because they were enrolled in master’s-level programs. They preferred individual learning; therefore, they did not consider online learning to be very important because it did not place much emphasis on the course content.

CONCLUSIONS

This study proposes a learning portfolios analysis system. This system provides teachers with information that can help them examine and define students’ learning behaviors, motives, and achievements. While observing the online activities of students who have different behavioral features, we guided and assisted them with a warning mechanism. Moreover, we conducted an experiment in six classes on the basis of systematic examinations and observations to test the effectiveness of warning messages. Without a doubt, the students were affected by the warnings to some extent. Many of the results indicate that in terms of learning behaviors and achievements, the students who received the warnings were not likely to lag behind the students who performed better in the class and could even catch up with them. The study on gender differences with respect to the effectiveness of warning messages initially concludes that they are more effective with female learners as compared with male learners. Moreover, the learning method at each college was different. We concluded that the students in the college of business were more active while learning online and responded better to the warning messages than did the students from the other colleges.

FUTURE WORK

Having developed this system, we would like to conduct further explorative experiments in the future. We expect to develop more complete student behavioral evaluation mechanisms by collecting additional information on students and the learning process. Moreover, to further examine the learning process, we will investigate the amount of time students spend reading and analyze in detail the types of material they read. This will be of great help in understanding the content required in learning and the progress that students make.

This study focuses on an investigation of online activities. However, to a certain extent, there is a difference between online activities and students’ actual learning conditions. In order to thoroughly understand the relationship between individual learning styles and online activities, we plan to observe the relationship between thinking styles, learning behaviors, and achievements among all types of students by using a time tracking table. With this, the characteristics of learning can be collected for individual students, which can be used for providing them with customized warning signals and the best guidance.

REFERENCES


AUTHOR INFORMATION

Chien-Ming Chen, Ph.D. Student, Chung Yuan Christian University, calculusxp@cg.ice.cycu.edu.tw

Cheng-Hao Ma, Graduate Student, Chung Yuan Christian University, col@cg.ice.cycu.edu.tw

Bin-Shyan Jong, Professor, Chung Yuan Christian University, bsjong@ice.cycu.edu.tw

Yen-Teh Hsia, Associate Professor, Chung Yuan Christian University, hsia@ice.cycu.edu.tw

Tsong-Wuu Lin, Professor, Soochow University, twlin@cis.scu.edu.tw