Knowledge Mining for Adaptive Multimedia Web-based Educational Platform

LEYLA Zhuhadar

Doctoral Candidate at University of Louisville, KY USA OLFA Nasraoui

Main Advisor: Assosiate Professor and Chair of E-Commerce at Dept. of Computer Science and Computer Engineering, University of Louisville, KY USA ROBERT Wyatt

> Advisor: Assistant Professor of Biology and the Director of Distance Learning at Western Kentucky Universit, KY USA

AIED2007-Artificial Intelligence in Education 13th International Conference on Artificial Intelligence in Education

Abstract. This workshop introduces an Adaptive Web-Based Educational platform that maximizes the usefulness of the online information that online students retrieve from the web. It shows in a data driven format that information needs to be personalized and adapted to the needs of individual students. Therefore, educational materials need to be tailored to fit these needs: learning styles, prior knowledge of individual students, and recommendations. This approach offers several techniques to present the learning material for different types of learners and for different learning styles. User models (user profiles) are created using a combination of clustering techniques and association rules mining. These models represent the learning style, and learning sequence, which can help improve the learning experience on the website for new users.

Introduction

At the department of Distance Learning at Western Kentucky University, an interactive web environment was developed, where teachers, researchers, and knowledge seekers can discover information about their distance learning students. Every single access to our platform, from each individual student to different types of learning material, such as *text, audio, podcasting*, and *video* lectures were traced and recorded in log files. The audio and video lectures were presented through the latest technology, Podcasting and VODcasting to enhance the learning *mobility*. By tracking the behavior of each online student and knowing which lectures he/she has selected, the sequence of lectures that were selected, the type of the selection (text, audio, or video), and the method used (online or offline), we can build a user model (user profile), which is a system representation of how the learner relates to the conceptual structure of the application.

Our approach provides answers to four A's: Automatic Synchronization, Accessibility, Availability, and Adaptivity. Automatic Synchronization distinguishes both Podcasting and VODcasting from the traditional multimedia (audio and video) on the web: Most likely Pod/VODcasting will not replace traditional multimedia on the web, but will rather become a more flexible extension of it, offering more diversity to a

considerably larger audience. The key element of this intelligent technology is the automatic feed, which allows online students to subscribe to this feed only once and then the updated lectures, audio recordings of textbooks, texts, recent audio or video interviews, etc. will automatically be transferred to students' MP3 devices. Accessibility means offering learners with different needs alternative ways to navigate through the information. For instance, the inclusion of closed caption text was embedded into our system to help different ways of information delivery (hearing impaired students need the caption as an alternative to sound). Using this method, we can provide students with disabilities (hearing impaired) with alternative ways of accessing online course materials. Availability means enabling online students to access lectures any way and anyhow, i.e. through the internet via streaming the media online to a browser or an MP3 device, streaming the media offline which allows students to "read" or "review" texts while walking or driving. Adaptivity refers to learner preferences regarding different learning styles. Some learners prefer learning by reading (text), others by listening (audio) and yet others prefer a visual learning style (video). What is innovative about our system is the four A's can be encapsulated into the personalization aspect, which includes all aspects of the learning situation, such as personal preferences of student learning style and needs. We will call this system, which is illustrated in Figure [1], the *My Way adaptive e-learning platform.*

Data-Driven Adaptive E-Learning Platform

1. System Framework

Our Approach is based on a dynamic website that provides online lectures in three different learning styles: text, audio, and video. Our personalization strategy relies on data that consists of website logs of distance learning at Western Kentucky University. Website logs record every user's activity on the website. This includes the resources that have been accessed, along with a time stamp, and other information. The pre-processing of weblogs which includes cleaning them from irrelevant information and transforming them often consumes 80% to 95% of the effort and resources needed for web usage mining (Edelstein, 2001) [1]. In this platform, all the activities of the students were traced, including the sequence of browsing the lectures, and the type of media that has been chosen.



Figure [1] Recommendation System for Multimedia Web-based Educational Platform Meaningful information from the weblogs was extracted to suit the objective of the adaptive system. The Weblogs were cleaned, and their final transformation is shown in Table [1].

Student ID	File Name	College	Day of Week	Time	Media Type				
Student1	Waters_ENG200_DocLecture1.php	ENGLISH	WEEKEND	DAYTIME	TEXT				
Student2	Plummer_Hist119_DocLecture1.php	HISTORY	WEEKEND	DAYTIME	TEXT				
Student2	Plummer_Hist119_PodLecture1.php	HISTORY	WEEKEND	DAYTIME	AUDIO				
Student3	Plummer_Hist119_Podlecture3.php	HISTORY	WEEKEND	DAYTIME	AUDIO				
Student3	Plummer_Hist119_DocLecture3.php	HISTORY	WEEKEND	NIGHTTIME	TEXT				
Student4	Waters_ENG200_DocLecture1.php	ENGLISH	WEEKDAY	NIGHTTIME	TEXT				
Student115	Waters_ENG200_DocLecture1.php	ENGLISH	WEEKDAY	DAYTIME	TEXT				
	Table [1] Weblog excerpt after cleaning and transforming the data								

We had 931 individual log sessions, with 115 different students from 6 different colleges and 12 different courses. Each student logged to different lectures with a different learning style (Text, Audio, and Video). Table [2] shows the extracted information about each student session, which consists of StudentID, lectures that he/she picked, college name, the day of the week (Weekdays/weekends) time of day (Day time/Night time) and media style (Text/Audio/Video).

# of Students in total	115										
# of College	6 (English	6 (English, History, Chemistry, Business, Health & Human Services, College of Education)									
# of courses	12										
	English	History	Chemistry	Business	Health & Human Services	College of Education					
	Eng200	Hist11 9	Chm100	BA580	Swrk320	LM22					
	Eng200			BA592	CFS111						
]		AC450							
				MG333							

				Internation al Trade					
Day of Browsing the lecture	(Weekdays	(Weekdays, Weekends)							
Time of Browsing the lecture	(Day Time, Night Time)								
Media type of lecture	Text Audio Video Subscribe								
Table [2] Features and attributes of the collected data									

The percentages of each learning style for each college were calculated over all the courses, and are shown in Table [3].

	Text	Audio	Video	Sub- scribe	#of Sessions	#of Student s	% Text	% Audio	% Video	% Sub- scribe
English	248	148	17	15	428	40	57%	35%	4%	4%
History	26	22	0	2	50	15	52%	44%	0%	4%
Chemistry	47	27	0	7	81	18	58%	33%	0%	9%
Health & Human Services	16	13	0	7	36	17	44%	36%	0%	20%
Business	170	130	0	10	310	22	55%	42%	0%	3%
College of Education	12	12	0	2	26	3	46%	46%	0%	8%
Total	519	352	17	43	931	115	57%	35%	4%	4%
%	55%	37%	3%	5%						
Table [3] Data	distribut	ion and	the perce	entages of	each lea	rning s	tyle for e	each Co	llege

2. Implementing a Recommendation System for Distance Learning

Our adaptive multimedia web-based educational platform relies on a recommender system that can recommend lectures to a learner based on their previous navigation or access to lectures, and based on navigation made by other "similar" learners. We have attempted to extract the user models by applying two data mining methodologies: first data clustering, then association rule mining. The Group User Model (Bollen, 2000) [2] is the collective knowledge of a group of users on a given domain transformed from hyperlink structure. After this group user model is formed, it is used to improve and recommend hyperlinks to individual users rather than a group of users. The web sessions go through data mining. During this stage, user activity models (such as user profiles) were created using either clustering techniques or association rule mining. For example, association rule models can be used to infer the learning experience of new online students.

The data has been divided into 6 categories corresponding to the different departments: English, History, Chemistry, Business, Health & Human Services, and College of Education. Then it was analyzed both on a department level, as well as globally. For creating User Models (user profiles), we clustered students based on the similarity between their cumulative access sessions (a record of all lectures viewed by a particular student throughout all their sessions). This would allow us to discover lecture access profiles. Each student's sessions were combined into one long transaction vector with one attribute (or dimension) per lecture, and where the visit to a lecture was represented by a 1 in the corresponding dimension, while a 0 was used for lectures that

have not been visited. We used the K-means algorithm (McQueen, 1967) [3] for clustering similar students.

3. Results of Clustering Student Cumulative Sessions

Clustering was performed on the cumulative student sessions from each department separately. The numbers of clusters varied from one department to another (English=15, Business=13, History=4, Chemistry=5, Health Human services=2, Education=1). Each cluster centroid is an average of the cumulative sessions assigned to (i.e. closest to) that cluster. The weights are listed in decreasing order. We can see that the cluster centroids summarize the different assortments or combinations of lectures viewed by most students in a particular group of students. Such assortments are interesting because even when interpreted after discovery, can be seen to reflect intuitive relationships. For example cluster 5 in the English department consists of assortments that combine a lecture on drama (taught by one instructor) with other lectures by a different instructor that are focused on character presentation, reading poetry out loudly, and sonnet rhythms and schemes. This cluster fits a group user profile interested in learning about drama that involves poetry, and in all likelihood certain types of sonnets. What is really interesting about such an output of clustering is that it is automatically discovered by data mining, i.e., by a purely data-driven approach, and not inferred by Human reasoning.

4. Pattern Discovery with Association Rules

In order to discover patterns such as trends and relationships within the web usage data, we applied association rule mining. Association rule discovery is a classical data mining problem (Agrawal, 1993) that serves to discover patterns in users' behavior. We used the *Apriori* Association Rule mining algorithm (Agrawal, 1993) [4] with minimum support 5% and confidence 75%.

Experiment 1: Mining Association Rules from Single Transaction Records with All Attributes

The first experiment was performed on all students' sessions to find association rules that relate all the attributes (such as Day, Time, Media Type, and Academic Department). We used a minimum support of 5% and minimum confidence of 75%. From the discovered association rules we may conclude that:

- Media type (Text/Audio) was correlated with day_time: Most of the students Read/Listen to the lectures during the day.
- Weekdays were correlated with Day time: The students who accessed their lectures during the weekdays tended to access them during the day as opposed to the evening hours.
- English courses were more accessed during weekdays, whereas Chemistry courses were more accessed during Weekends.
- Business courses were accessed during weekends more than weekdays and during the day_time, and with a preference toward Audio lectures (podcast) format.

Experiment 2: Mining Association Rules from Cumulative Session Records with Lecture Attributes

In this experiment, we extracted association rules between lectures, by using as input data, the *cumulative* sessions of each student (into a long binary transaction vector consisting of *all* lectures accessed by the *same* student, as was done in the clustering phase). Consequently, we were able to discover the lecture access patterns that can be used to compute lecture recommendations for a given student who has accessed certain lectures. We started by dividing our sessions into six separate subsets, one for each department, and then mined association rules from each subset.

Recommendation in the Adaptive Web-based Educational Platform

In this section we explain how the results of data mining, as described in the previous sections, can form the basis for automated recommendations. From the association rules that link "all" attributes, we may conclude that different areas of study have different user behavior, and different preferences. Moreover, different students have different learning styles, different preferences of time access, and browsing access. From the association rules that relate the lectures accessed by the same student, we notice that students follow different paths through the website, thus choosing various combinations or assortments of lectures. From the college-based clustering of the cumulative transaction vectors of each student, we notice that students within the same college tend to be divided into several groups based on the combination of lectures that they have chosen, and that this division is not only based on course number or instructor. Rather it truly shows variability in the menu selections (if we regard a set of chosen lectures as one menu). We can build a recommender system by using any or all of the above models, i.e.:

- **A.** Association rules between all attributes (such as Day, Time, media Type, Academic Department).
- **B.** Association rules between lectures (accessed by the same student).
- **C.** Cluster centroids/profiles of lecture assortments that are often chosen by a "group" of similar students in the same college.

Depending on the type of model used, we can form the following types of recommender systems:

- **A.** *Recommenders of Type A* will tend to recommend a certain *Media Type* based on day, time, or department (or combination thereof) whenever a student session's attributes match the antecedent of any of the rules of model type A. The recommended Media Type corresponds to the one in the consequent part of the matching rules.
- **B.** *Recommenders of Type B* will recommend a set of lectures to a student if the lectures that were already visited by this student match some the antecedent of the rules in model B. The recommended lectures will be the top ranking lectures of the consequents of the matching rules.
- **C.** *Recommenders of Type C* will recommend a set of lectures to a student if the lectures that were already visited by this student match some the cluster centroids/profiles in model C. The recommended lectures will be the top ranking lectures when accumulated throughout all the matching and closest profiles.

Cluster-based collaborative filtering recommendation (Type C)

	Waters_ ENG200_ Lecture1.ph p	Waters_ ENG200_ Lecture2.ph p	Waters_ ENG200_ Lecture31.p hp	Waters_ ENG200_ Lecture4.ph p	Waters_ ENG200_ Lecture5.ph p	Waters_ ENG200_ Lecture6.ph p	Waters_ ENG200_ Lecture7.ph p
cluster0	0.75	1	0.875	1	1	1	0.875
cluster1	0	1	0	0	0	0	1
cluster2	1	1	1	0.5	1	0	0.5
cluster3	1	0.25	0.125	0	0	0.125	0
cluster4	0	0.5	0.5	0	1	1	1
cluster5	1	0	0	0	0	0	0
cluster6	1	1	1	1	0	0	0
cluster7	1	1	0	0	1	0	0
cluster8	0	0	0	0	0	0	0
cluster9	0	0	0	0	0	0	0
cluster10	0	1	0	0	0	0	0
cluster11	1	1	0	0	1	0	0
cluster12	0.3333	0.5	0.3333	0.3333	0.3333	0.3333	0.3333
cluster13	0	0	0	0	0	1	0
cluster14	0	0	0	0.1667	0	0	0
	Olmsted_ ENG200_ Lecture1.ph p	Olmsted_ ENG200_ Lecture2.ph p	Olmsted_ ENG200_ Lecture3.ph p	Olmsted_ ENG200_ Lecture4.ph p	Olmsted_ ENG200_ Lecture5.ph p	Olmsted_ ENG200_ Lecture6.ph p	Olmsted_ ENG200_ Lecture7.ph p
cluste0	0	0	0	0	0	0	0
cluste1	0	0	0	0	0	0	0
cluster2	0	0	0	0	0	0	0
cluster3	0	0	0	0	0	0	0
cluster4	0	0	0	0	0	0	0
cluster5	1	1	0	0	1	0	0
cluster6	0	0	0	0	0	0	0
cluster7	0	0	0	0	1	0	0
cluster8	1	0.5	0	1	0.5	0	0.5
cluster9	0	0	1	0	0	0	0
cluster10	0	0	0	0	0	0	0
cluster11	1	1	0	0	0	0	0
cluster12	1	1	1	1	1	1	1
cluster13	0	0	0	0	0	0	0
cluster14	0.1667	0.1667	0	0.3333	0.1667	0.1667	0.1667
clusters	Olmsted_ ENG200_ Lecture8.ph	Olmsted_ ENG200_ Lecture9.ph	Olmsted_ ENG200_ Lecture10.p	Olmsted_ ENG200_ Lecture11.p hp	Olmsted_ ENG200_ Lecture12.p	Olmsted_ ENG200_ Lecture13.p hp	Olmsted_ ENG200_ Lecture14.p
cluste0	0	0	0	0	0	0	. 0
cluste1	0	0	0	0	0	0	0
cluster2	0	0	0	0	0	0	0
cluster3	0	0	0	0	0	0	0
cluster4	0	0	0	0	0	0	0
cluster5	0	1	0	0	0	0	0
cluster6	0	0	0	0	0	0	0
cluster/	0	0	0	0	0	0	0
cluster0	0	1	0	1	1	0	0
cluster10	0	0	0	0	0	0	0
-	, j				ļ	ļ	-

Table [4] shows the results of clustering English weblog sessions into 15 clusters; each cell represents the weight of each lecture in the corresponding cluster's centroid.

cluster11	0	0	0	0	0	0	0		
cluster12	1	1	0.6667	0.8333	0.8333	0.6667	0.8333		
cluster13	0	0	0	0	0	0	0		
cluster14	0.1667	0	0.3333	0	0	0	0		
	Table[4]Cluster centroids for the English cumulative sessions (cluster rows were duplicated to accommodate all the lectures)								

Using the cosine similarity in (2), which measures the relative amount of overlap between a student's cumulative session A and a cluster centroid B which can be considered as a prototypical cumulative session of a group of users, thus a group profile, and assuming that A_i represents the presence (i.e. a value of 1) or absence (value of 0) of lecture i in session

$$\cos ine(A,B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} B_i^2}},$$
 (2)

vector A:

we were able to map each student to the closest cluster centroids/group profiles. For example, we illustrate the task of recommendation given as input a small *part* of Student 41's session, i.e. partial session: (*Waters_ENG200_Lecture1.php*,

Waters_ENG200_Lecture2.php, Waters_ENG200_Lecture5.php) by computing the cosine similarity from this student to each cluster's centroid as shown in Table [5]. If we pick the clusters with cosine similarity \geq 0.6 then the matching clusters should be: 0, 2, 3, 7, 11. Table [6] shows the lectures (in shaded color) that should be recommended to student41.

	Cluster0	Cluster1	(Cluster2	Clust	er3	Cluster4		Cluster5	Cluster			
Student41	0.64318	0.40825	C	.8165	0.690	07	0.46291		0.2582	0.57735			
	Cluster7	Cluster8	С	luster9	Cluste	er10	Cluster11		Cluster12	Cluster1	B Clus	ıster14	
Student41	0.86603	0		0	0 0.57735		0.7746		0.18761	0		0	
					Table	[5]							
	Waters_ ENG200_ Lecture1.php	Waters_ ENG200_ Lecture2.p	ohp	Waters_ ENG200 Lecture3) 3.php	Wa EN Lec	ters_ G200_ cture4.php	V E L	Waters_ ENG200_ _ecture5.php	Waters ENG20 Lectur	- 10_ 96.php	Wat ENC Lect	ers_ 3200_ ure7.php
Cluster0	0.75		1		0.875		1		1		1		0.875
Cluster2	1		1		1		0.5		1		0		0.5
Cluster3	1	0	.25		0.125		0		C		0.125		0
Cluster7	1		1		0		0		1		0		0
cluster11	1		1		0		0		1		0		0
	Olmsted_ ENG200_ Lecture1.php	Olmste ENG20 Lecture2.p	ed_)0_ ohp	Olmsted_ ENG200_ Lecture3.phr		Le	Olmsted_ ENG200_ ecture4.php		Olmsted_ ENG200_ Lecture5.php	El Lectu	Olmsted_ ENG200_ Lecture6.php		Olmsted_ ENG200_ cture7.php
Cluster0	0		0		0		0		C)	0		0
Cluster2	0		0		0		0		C)	0		0
Cluster3	0		0		0		0		C)	0		0
Cluster7	0		0		0		0		1		0		0
cluster11	1		1		0		0		C)	0		0
	Olmsted_ ENG200_ Lecture8.php	Olmsted_ ENG200_ Lecture9.p	ohp	Olmsted ENG200 Lecture	l_)_ 10.php	Oln EN Lec	nsted_ G200_ cture11.php	C E L	Olmsted_ ENG200_ _ecture12.php	Olmste ENG20 Lectur	d_ 00_ e13.php	Olm ENC Lect	sted_ 3200_ sure14.php
Cluster0	0		0		0		0		C)	0		0
Cluster2	0		0		0		0		C)	0		0
Cluster3	0		0		0		0		C		0		0
Cluster7	0		0		0		0		C		0		0
cluster11	0		0		0		0		C		0		0
	Table[6]	lectures (in s	haded c	olor) t	hats	should be	re	commende	ed to stu	dent41		

Using a threshold ≥ 0.5 , we extracted the following lectures :

All Recommended Lectures									
Cluster0	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure3.php	Waters_EN G200_Lect ure4.php	Waters_EN G200_Lect ure5.php	Waters_ENG 200_Lecture 6.php	Waters_ENG 200_Lecture 7.php		
Cluster2	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure3.php	Waters_EN G200_Lect ure5.php					
Cluster3	Waters_EN G200_Lect ure1.php								
Cluster7	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure5.php	Olmsted_ ENG200_ Lecture5.ph p					
cluster11	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure5.php	Olmsted_ ENG200_ Lecture1.ph p	Olmsted_ ENG200_ Lecture2.php				

Since student41 already visited the following lectures in the partial session:

Already visited Lectures							
Waters_ENG200_Lecture1.php	Waters_ENG200_Lecture2.php	Waters_ENG200_Lecture5.php					

We must filter out the lectures that have already been visited. Hence, finally the student will be recommended with the following remaining lectures:

Filtered Recommended Lectures								
Waters_ENG200_Lecture3.php	Waters_ENG200_Lecture4.php	Waters_ENG200_Lecture6.php						
Waters_ENG200_Lecture7.php	Olmsted_ENG200_Lecture1.php	Olmsted_ENG200_Lecture2.php						
Olmsted_ENG200_Lecture5.php								

A more detailed illustration of the recommendation task is depicted in Figure [2].

Evaluation

We can evaluate the quality of our recommendation for student41 by computing the recall (1), precision (2), and F-score (3) measures, as shown below, where the true lectures are the lectures that have really been visited by the student, but not including the input lectures in their partial session. Given Student41's *complete* session was

(Waters_ENG200_Lecture1.php, Waters_ENG200_Lecture2.php,

Waters_ENG200_Lecture5.php, Olmsted_ENG200_Lecture1.php,

Olmsted_ENG200_Lecture2.php), and that his or her partial session was:

(Waters_ENG200_Lecture1.php, Waters_ENG200_Lecture2.php,

Waters_ENG200_Lecture5.php), the *true* lectures (for perfect recommendation) should be (*Olmsted_ENG200_Lecture1.php*, *Olmsted_ENG200_Lecture2.php*).

Hence

 $\{true \ lectures\} = \{\underline{Olmsted_ENG200_Lecture1.php}, \ \underline{Olmsted_ENG200_Lecture2.php}\}, while$

{recommended lectures} = {Waters_ENG200_Lecture3.php,

Waters_ENG200_Lecture4.php, Waters_ENG200_Lecture6.php,

Waters_ENG200_Lecture7.php, <u>Olmsted_ENG200_Lecture1.php</u>,

Olmsted_ENG200_Lecture2.php, Olmsted_ENG200_Lecture5.php}.

where the <u>underlined</u> lectures are the ones that are <u>in both the true and the recommended</u> <u>lecture sets</u>.

Recall

Recall (3) is the proportion of true lectures that are recommended out of all true lectures = 2/2 = 100%

$$recall = \frac{\{recommended \ lectures\} \cap \{true \ lectures\}}{\{true \ lectures\}}$$
(3)

Precision

Precision (4) is the proportion of true lectures that are recommended out of all recommended lectures = 2/7 = 28.57%.

$$precision = \frac{\{recommended \ lectures\} \cap \{true \ lectures\}}{\{recommended \ lectures\}}$$
(4)

F-measure

The F-score (5) is the weighted harmonic mean of precision and recall, the traditional F-measure or balanced F-score is $F_1 = 2(1)(0.2857)/(1+0.2857) = 0.44$ or 44%

$$F_{1} = \frac{2 \cdot precision \cdot recall}{(precision + recall)}$$
(5)

Effect of the cluster similarity threshold:

We will illustrate below how the matching threshold affects precision and recall. If we pick the clusters with cosine similarity ≥ 0.7 , then the matching clusters are limited to clusters 2, 7, 11. In this case, it is easy to verify that the recommended lectures will be chosen from

All Recommended Lectures								
Cluster2	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure3.php	Waters_ENG2 00_Lecture5.p hp				
Cluster7	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure5.php	Olmsted_ ENG200_ Lecture5.php				
cluster11	Waters_EN G200_Lect ure1.php	Waters_EN G200_Lect ure2.php	Waters_EN G200_Lect ure5.php	Olmsted_ ENG200_ Lecture1.php	Olmsted_ ENG200_ Lecture2.p hp			

This leads to the filtered recommendations

{recommended lectures} = {Waters_ENG200_Lecture3.php,

Olmsted_ENG200_Lecture1.php, Olmsted_ENG200_Lecture2.php,

Olmsted_ENG200_Lecture5.php}.

Precision now becomes 2/4=50%, while recall remains at 100%. This increases the F_1 to 2(1)(0.5)/(1+0.5) = 0.667 or 66.7%. What we have just illustrated is a general trend where a more sringent matching threshold can increase the precision. A consequence of this increase is typically a decrease in recall, though, this did not occur at threshold 0.7. Recall would however decrease to 0% (as well as precision) if the cluster matching threshold was increased to 0.8, since only clusters 2 and 7 would match the student session, resulting in {recommended lectures} = {Waters_ENG200_Lecture3.php,

Olmsted_ENG200_Lecture5.php }.

We summarize the evaluation metrics in Table [7]. In this case a cluster similarity threshold of 0.7 seems to yield an optimal tradeoff between precision and recall of recommendations.



5. Conclusion and Future Work

The implementation of the proposed web-based education platform takes students' activity into consideration. The main goal was to create a recommender system which was based on discovering patterns from the online students' behavior, and then comparing these patterns to new learners. While the current recommender system did not allow the educator to be involved in the recommendation process, a future improvement would allow the educator to monitor the recommender system's performance metrics, as well as to modify or expand the discovered patterns used as a substrate for the recommendations. The instructor can thus add some input to improve the recommendations. Moreover, keeping track of selected recommender system. We could also compare students who followed the recommendations with those who ignored them, and monitor the time that each one of these groups of students spend to reach the information that they need. Moreover, personalized quizzes could be added for each learning style and a comparison between the results could define which learning style fits a specific topic.

References

[1] Edelstein, H. A. (2001). Pan for gold in the clickstream. *Informationweek*, Retrieved October 31, 2006, from <u>http://www.informationweek.com/828/mining.htm</u>.

 Bollen, J. (2000). Group user models for personalized hyperlink recommendations. Proceedings *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, LNCS* 1892.

[3] McQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of 5th Berkeley Symposium on Mathematics, Statistics and Probability, 1:281-298.

[4] Agrawal, R., & Swami, A. (1993). Mining association rules between sets of items in large databases. Proceedings from 1993 ACM-SIGMOD Int. Conference. *Management of Data*, Washington, D.C, 207–216.