ONLINE LEARNING BEHAVIORS FOR RADIOLOGY INTERNS BASED ON ASSOCIATION RULES AND CLUSTERING TECHNIQUE

Hsing-Shun Chen[1,2] and Chuen-He Liou[1]
[1]*Graduate Institute of Health Allied Education,*
National Taipei University of Nursing and Health Sciences, Taipei, Taiwan
[2]*Department of Radiology, Far Eastern Memorial Hospital, Taipei, Taiwan*

ABSTRACT

In a hospital, clinical teachers must also care for patients, so there is less time for the teaching of clinical courses, or for discussing clinical cases with interns. However, electronic learning (e-learning) can complement clinical skills education for interns in a blended-learning process. Students discuss and interact with classmates in an e-learning collaborative environment. E-learning can assist clinical training and provides a collaborative environment, but every student has individual learning preferences on the e-learning platform. A typical platform, such as a learning management system (LMS) does not provide individual learning activities for every student. This paper clusters students into two groups: active and inactive groups. In each group, students' learning behavior patterns, i.e., the association rules for activities, are derived from the transaction data for the LMS. The cluster to which a student belongs defines the online learning behaviors, from the activity association rules. The method then provides individual preferred activities. Teachers instruct students in accordance with their aptitude, as derived from the learning behavior pattern. The cluster analysis shows that students in active group often view teaching videos after completing feedback. Students in the inactive group often view teaching materials after adding posts on a forum.

KEYWORDS

Internship, clinical skill education, e-learning, blended learning, clustering, association rules

1. INTRODUCTION

In a hospital, clinical teachers must devote time to patients, so they have less time to teach clinical courses, or to discuss clinical cases with interns (Prideaux et al., 2000, Ramani and Leinster, 2008). However, e-learning can be a complement to clinical skill education for interns (Ruiz et al., 2006). Most medical students feel that e-learning has a positive impact on the acquisition of clinical skill and knowledge. It is an integrated, blended approach (Gormley et al., 2009). Students who use an e-learning platform as a complement to in-classroom education obtain higher scores in the final examination (Seluakumaran et al., 2011). Blended learning, which is clinical learning combined with e-learning, results in a greater acquisition of knowledge in radiology internship (Mahnken et al., 2011). An e-learning course in radiology gives greater knowledge acquisition than a standard lecture-based course and is a cost-effective alternative to standard lecture-based teaching (Hadley et al., 2010).

Medical education in radiology includes undergraduate, postgraduate and continuing education. The existing e-learning platforms for radiology are on-line neuroradiology education resources (NeuroRAD) (Sparacia et al., 2007), the American Association of Physics in Medicine (AAPM) and the Radiological Society of North America (RSNA) (Brambilla et al., 2011). These platforms were developed for physicists and radiologists in continuing education, but few platforms have been developed for undergraduate radiology internship. Radiology clinical training includes image diagnosis, nuclear medicine, radiation oncology, quality assurance and safety examination, which require discussion and interaction with a teacher and classmates. In an e-learning collaborative environment, radiology interns can interact with their classmates at anytime, to construct their own knowledge base(Brambilla et al., 2011). Clinical teachers provide the radiological content, such as image banks and special clinical cases (Perriss et al., 2006). Radiology interns in

ISBN: 978-989-8704-08-5 © 2014

different areas discuss their cases with others by posting on a forum. This study establishes an e-learning platform to assist clinical skill training.

E-learning can assist clinical training and provides a collaborative environment, but every student has individual learning preferences on the e-learning platform. A typical platform, such as the learning management system (LMS), does not provide individual learning activities for every student. In order to provide an individual learning environment, data mining technology is applied to educational systems (Romero and Ventura, 2007). This paper clusters students into two groups, based on their activity preferences: active and inactive groups. Students in the active group, often view the course, complete feedback, add posts, update posts and view discussions on the forum. All of the activity association rules for the two groups were retrieved, in order to determine students' learning behavior patterns, for each cluster. For example, active group students often view teaching videos for the lessons, after viewing the course and completing feedback. They also prefer to write messages to other classmates to communicate emotions.

The remainder of this paper is organized as follows. Section 2 discusses related studies. Section 3 describes the proposed method. Section 4 presents the experimental results. Section 5 summarizes the findings, states the limitations of this study.

2. RELATED STUDIES

Two data mining techniques are used: clustering technique and association rules. Some data mining applications for e-learning are also described.

2.1 Clustering Technique

Clustering techniques, which are usually used to segment markets (Punj and Stewart, 1983), seek to maximize the variance between groups and to minimize the variance within groups. A number of clustering algorithms have been developed, such as K-means, hierarchical and fuzzy c-means algorithms (Omran et al., 2007). K-means clustering (MacQueen, 1967) is a widely used similarity grouping method that partitions a dataset into k groups. The K-means algorithm assigns instances to clusters, based on the minimum distance principle. An instance is assigned to a cluster based on the minimum distance to the center of the cluster, over all of the k clusters.

2.2 Association Rules

Association rule mining determines the associations between two sets of products in a transaction database. Agrawal et al. (1993) formalized the problem of determining association rules that satisfy the minimum support and the minimum confidence requirements. For example, if a set of purchase transactions includes a set of product items I, an association rule is an implication of the form, $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \Phi$. X is the antecedent (body) and Y is the consequence (head) of the rule. Two measures, support and confidence, are used to determine the quality of an association rule. The support of a rule is the percentage of transactions that contain both X and Y and the confidence of a rule is, the fraction of all transactions that contain X that also contain Y.

2.3 Data Mining Applications for E-learning

Data mining technologies include on-line analytical processing (OLAP), clustering, association rules and classification and visualization (Zaíane, 2002, Talavera and Gaudioso, 2004, Zorrilla et al., 2005, Mostow and Beck, 2006, Romero and Ventura, 2007, Romero et al., 2008). For example, Zorrilla et al. (2005) built a web log data cube for OLAP operation, to analyze the log to obtain the information that allows teachers to evaluate the learning process. Talavera and Gaudioso (2004) clustered users into groups, to determine their behavior patterns and evaluations. Mostow and Beck (2006) developed a listening tool that uses visualization technology to help children to decode words and comprehend stories. Zaíane (2002) used association rules that use a learner's access history to recommend on-line learning activities, or shortcuts on a course website. Romero et al. (2008) used clustering, association rules and classification technologies to discover knowledge from a learning content management system.

3. METHODOLOGY

This section proposes a method to understand learning behavior patterns for interns, which is shown in Figure1.The activity usage count for each student on LMS was calculated. The missing values were filled using proper data preprocessing. Because some activities are used by few or no students, feature selection was applied to some activities. These activities were not taken into account. The continuous values for the activity usage count were then transformed into discrete preference values, to form a student-activity preference matrix.

The K-means clustering method was then used to cluster students into activity preference groups, based on the similarity between students' activity preferences, which were measured using Pearson's correlation coefficient, as shown in Eq.(1). The $\overline{r_{S_i}}$ and $\overline{r_{S_j}}$ denote the average rating score of all activities used by students S_i and S_j respectively. The variable I denotes the mix of the set of activities. The $r_{S_i,A}$ and $r_{S_j,A}$ denote the rating score that students S_i and S_j used activity A.

$$\mathrm{corr}(S_i, S_j) = \frac{\sum_{A \in I}\left(r_{S_i,A} - \overline{r_{S_i}}\right)\left(r_{S_j,A} - \overline{r_{S_j}}\right)}{\sqrt{\sum_{A \in I}\left(r_{S_i,A} - \overline{r_{S_i}}\right)^2 \sum_{A \in I}\left(r_{S_j,A} - \overline{r_{S_j}}\right)^2}} \qquad (1)$$

In each group, students' online learning behavior pattern, the association rules, were derived from the transaction data of the learning management system (LMS). The cluster to which a student belongs defines the online learning behavior pattern, based on the association rules in that cluster. The method shows the preferred activities for every student.

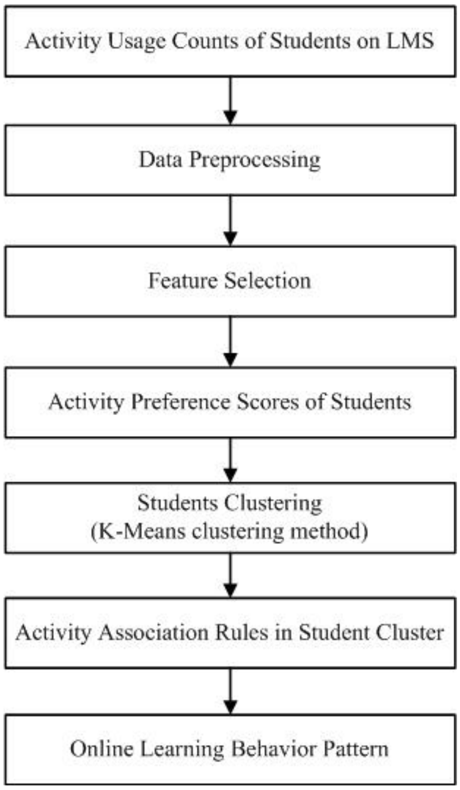

Figure1. An overview of the proposed method

ISBN: 978-989-8704-08-5 © 2014

3.1 Data Preprocessing

The students and activities transaction data from the learning management system (LMS) were firstly pre-processed. The linear method was used to fill the missing values. Feature selection was then applied. There are many learning activities on the learning management system, but some activities are never or seldom used. The number of the selected activities is 18. They include viewing the course, folder, discussion, resource, video, completing feedback, adding and updating posts on a forum, writing messages and updating user profiles.

3.2 Activity Preference Scores of Students

The usage counts for the selected activities are continuous values. Eq. (2) (Lin et al., 2003) is used to transform the continuous values to discrete values, i.e., -1, 0, or 1.

$$Z = \frac{X - \bar{x}}{\sigma_x} \tag{2}$$

where X is the activity usage count, $\bar{x}$ and σ_x are, respectively, the mean value and the standard deviation of the activity usage count and Z is a semantic variable.

All of the continuous usage count values were normalized to discrete preference scores (PS) and 0.3 is selected to cluster students into suitable groups by using Eq. (2) with $Z < -0.3$, $-0.3 \leqq Z \leqq 0.3$, and $Z > 0.3$, to respectively represent inactive, neutral and active preferences. The preference score (PS) is the degree of preference that a student demonstrates for an activity, which is defined as in Eq. (3). The preference score is 1 if the usage count, $X > \bar{x} + 0.3\sigma_x$, and the preference score is -1, if the usage count, $X < \bar{x} - 0.3\sigma_x$. 0 represents a neutral preference.

$$\text{Preferece Score (PS)} = \begin{cases} 1, & \text{when } X > \bar{x} + 0.3\sigma_x \\ -1, & \text{when } X < \bar{x} - 0.3\sigma_x \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The student-activity preference score matrix is shown in the following Table 1.

Table 1. Student–activity preference score matrix

Student ID	View courses	Complete feedback	Add post on forum	Update post on forum	View discussion on forum	
1	-1	0	1	0	1	…
2	0	-1	0	-1	1	…
3	1	0	-1	1	0	…
4	0	1	0	1	-1	…
⋮	⋮	⋮	⋮	⋮	⋮	…

4. EXPERIMENTAL RESULTS

4.1 Experiment Setup and Dataset

A dataset obtained from the learning management system (LMS) of a hospital was used for the experiment. The hospital is a medical center in northern Taiwan. This study was approved by the hospital ethics committee. Clinical teachers added learning resources (course, teaching material videos, or questionnaires) and activities (forum, chat room). Students viewed the courses, talked in chat rooms, discussed on forums and completed feedback on the LMS. The details are shown in Figure2.

The experiment dataset was extracted from the learning management system of the hospital, for the first semester of 2013. There are 10,637 items of student activity transaction data on the LMS. The dependent variable, support and confidence thresholds for the association rules were set to 0.5, 0.2 and 0.6. These thresholds were set based on our observation on the online learning behaviors of the students. The most popular activities are viewing the course, completing feedback, adding and updating posts and viewing discussions on a forum.

Figure 2. Knowledge sharing (left) and a discussion on a forum (right) on the e-learning platform

4.2 Student Cluster Identification

It is firstly necessary to identify the characteristics of each student cluster, i.e. the activity preferences for each student cluster. The average usage count for activity X in a cluster is compared to the average usage count $\bar{x}$ plus/minus the standard deviation, σ_x, for all students, to obtain the cluster preference score (PS),using Eq. (2). The preference score for an activity in the cluster ID is $\mathrm{PS}(X^{\mathrm{Cluster\ ID}}_{\mathrm{Activity}})$. For example, $\mathrm{PS}(X^{0}_{\mathrm{View\ course}}) = -1$ and $\mathrm{PS}(X^{1}_{\mathrm{View\ course}}) = 1$. The activity preference scores for two clusters are shown in Table 2. The preference scores for activities such as viewing the course, completing feedback, adding posts, updating posts and viewing discussions on a forum in cluster 0 are $\mathrm{PS}(X^{0}_{\mathrm{View\ course}})$, $\mathrm{PS}(X^{0}_{\mathrm{Complete\ feedback}})$, $\mathrm{PS}(X^{0}_{\mathrm{Add\ post}})$, $\mathrm{PS}(X^{0}_{\mathrm{Update\ post}})$, $\mathrm{PS}(X^{0}_{\mathrm{View\ discussion}})$, which is (-1, -1, -1, -1, -1) and the preference scores for these activities in cluster 1 are $\mathrm{PS}(X^{1}_{\mathrm{View\ course}})$, $\mathrm{PS}(X^{1}_{\mathrm{Complete\ feedback}})$, $\mathrm{PS}(X^{1}_{\mathrm{Add\ post}})$, $\mathrm{PS}(X^{1}_{\mathrm{Update\ post}})$, $\mathrm{PS}(X^{1}_{\mathrm{View\ discussion}})$, which is (1, 1, 1, 1, 1).

Table 2. Activity preference scores for students in the two clusters

Cluster ID	Activity		Cluster	All			Preference Score (PS)
			X	$\bar{x}$	$\bar{x}$ - 0.3 σ_x	$\bar{x}$ + 0.3 σ_x	
0	course	recent	3.0	3.0	2.4	3.6	0
		view	269.3	399.0	319.8	478.2	-1
	feedback	complete	7.0	9.0	7.8	10.2	-1
	folder	view	12.0	10.0	7.9	12.1	0
	forum	add discussion	2.0	2.0	2.0	2.0	0
		add post	38.3	53.0	44.3	61.7	-1
		delete post	1.0	1.0	1.0	1.0	0
		update post	1.0	3.0	2.4	3.6	-1
		view discussion	294.3	402.0	332.7	471.3	-1
1	course	recent	2.5	3.0	2.4	3.6	0
		view	519.3	399.0	319.8	478.2	1
	feedback	complete	14.7	9.0	7.8	10.2	1
	folder	view	13.3	10.0	7.9	12.1	1
	forum	add discussion	3.0	2.0	2.0	2.0	1
		add post	78.5	53.0	44.3	61.7	1
		delete post	1.5	1.0	1.0	1.0	1
		update post	5.5	3.0	2.4	3.6	1
		view discussion	531.3	402.0	332.7	471.3	1

The usage counts for these activities for cluster 1 are twice those for cluster 0. Cluster 0 is an inactive group and cluster 1 is an active group. Students in the inactive group (cluster 0) seldom view the course, complete feedback, add posts, update posts, or view discussions on a forum. However, the active group (cluster 1) students have the opposite learning behaviors. Table 2 shows that the usage counts for adding posts, updating posts and adding discussions for cluster 1 are all larger than those for cluster 0. Students in cluster 1 show confidence in self-expression. The usage counts for recent course, viewing folders and deleting posts for cluster 0 are all close to those for cluster 1. “Recent course” means that students check the latest course status on the LMS.“Viewing folders” means that students download the necessary files, e.g., teaching plans and material, which are placed in folders on the LMS. “Deleting posts” means that students delete their post after adding a post on a forum. In summary, students in both groups download the necessary teaching material files, check the latest course information and delete their posts on a forum.

4.3 Activity Association Rules for Student Clusters

Table 3 shows the association rules between activities for cluster 0, i.e., from activity (X) to activity (Y). Students fill out the questionnaire, add discussions on a forum and then view discussions and search for the clinical questions on the platform. After adding a discussion on the forum, they view resources to find the answer to a clinical question. The resources are the teaching material files, which are provided by the clinical teacher. It is interesting that students in cluster 0 often remember to logout from the system and they pay more attention to the security of personal information than students in cluster 1.

Table 3. Association rules for activities for cluster 0

Cluster ID	Activity (X)		→	Activity (Y)	
0	course	recent	→	user	update
		view	→	forum	view forum
				resource	view
				user	login, logout, update, view all
	feedback	complete	→	forum	view forum
				resource	view
				user	login, logout, update, view all
	folder	view	→	user	logout, update
	forum	add discussion	→	forum	view forum
				resource	view
				user	login, logout, update, view all
		add post	→	forum	view forum
				resource	view
				user	login, logout, update, view all
		delete post	→	user	logout, update
		view discussion	→	forum	view forum
				resource	view
				user	login, logout, update, view all

Table 4 shows the association rules between activities for cluster 1, i.e., from activity (X) to activity (Y). Students in cluster 1often view a uniform resource locator (URL) and resource after viewing the course and completing feedback. The URL is a web link to teaching videos on YouTube. Students in cluster 1 review the lesson on the platform after it is taught. In addition, students in cluster 1 keep in touch with teachers or classmates by writing messages. Students in cluster 1 also often surf many websites at the same time and are more often compulsorily logged out of the after 5 minutes of no activity.

Table 3 and Table 4 show that students in both groups view other classmate’s recent profiles, which include information on the last login to the system. Students wish to know whether classmates view their recent posts and discussions on a forum, after they add posts and discussions on a forum. Students often check classmates’ latest login time to the system. They also often update their personal photos on the LMS. These are similar social behaviors to those seen on Facebook. It is found that students care about the peer status in the collaborative environment. Peer interaction is important during an internship. Students construct their own knowledge bases and learn from each other (Wilson and Stacey, 2004). In addition, students often observe other students learning status and wish to attract their attention. When they add a new topic for discussion on a forum, they view the posts that other classmates add or update. At the same time, they regularly update their photos and personal information on their own home pages. They attract the attention of peers for personal and social reasons.

Table 4. Association rules for activities for cluster 1

Cluster ID	Activity (X)		→	Activity (Y)	
1	course	recent	→	user	login, update
	course	view	→	message	write
			→	resource	view
				url	view
				user	login, update, view all
	feedback	complete	→	message	write
			→	resource	view
				url	view
				user	login, update, view all
	folder	view	→	user	login, update, view all
	forum	add discussion	→	user	login, update, view all
		add post	→	user	login, update, view all
		delete post	→	user	login, update
		update post	→	user	login, update, view all
		view discussion	→	user	login, update, view all

5. CONCLUSION

After cluster analysis, the average usage counts for viewing the course, completing feedback, adding posts, updating posts and viewing discussions on a forum for the active group are twice those for the inactive group. Therefore, students in the active group often view forums and resources after viewing the course, completing feedback, adding posts and viewing discussions on a forum. The student learning behavior patterns for each cluster can be derived, based on the association rules between activities for the cluster. Next year, when there are new interns, teachers can derive their clusters and association rules, to determine students' behavior, and using their behaviors and the association rules, give proper guidance and resources, to allow students to be taught in accordance with their aptitude. However, this study has some limitations. Some students were not familiar with the e-learning platform. In the future, the e-learning platform will be introduced before the internship.

ACKNOWLEDGEMENT

The research was supported in part by the National Science Council of the Taiwan, under grant NSC 102-2410-H-227-009, and by the Far Eastern Memorial Hospital, under grant FEMH-2014-D-056. The authors would like to thank all of the clinical teachers.

REFERENCES

Agrawal, R., et al., 1993. Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data.*, Washington, D.C., United States. ACM, pp. 207-216.

Brambilla, C. R., et al. 2011. Collaborative environment for nuclear medicine training. *Radiologia Brasileira,*Vol. 44**,** No. 3**,** pp. 177-182.

Gormley, G. J., et al. 2009. Is there a place for e-learning in clinical skills? A survey of undergraduate medical students' experiences and attitudes. *Medical teacher,*Vol. 31**,** No. 1**,** pp. e6-e12.

Hadley, J., et al. 2010. Effectiveness of an e-learning course in evidence-based medicine for foundation (internship) training. *Journal of the Royal Society of Medicine,*Vol. 103**,** No. 7**,** pp. 288-294.

Lin, Q.-Y., et al. 2003. Mining inter-organizational retailing knowledge for an alliance formed by competitive firms. *Information & Management,*Vol. 40**,** No. 5**,** pp. 431-442.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Statistical Laboratory of the University of California, Berkeley. University of California Press, pp. 281-297.

Mahnken, A. H., et al. 2011. Blended learning in radiology: Is self-determined learning really more effective? *European journal of radiology,*Vol. 78**,** No. 3**,** pp. 384-387.

Mostow, J. & Beck, J. 2006. Some useful tactics to modify, map and mine data from intelligent tutors. *Natural Language Engineering,*Vol. 12**,** No. 02**,** pp. 195-208.

Omran, M. G. H., et al. 2007. An overview of clustering methods. *Intelligent Data Analysis,*Vol. 11**,** No. 6**,** pp. 583-605.

Perriss, R., et al. 2006. Understanding the internet, website design and intranet development: a primer for radiologists. *Clinical radiology,*Vol. 61**,** No. 5**,** pp. 377-389.

Prideaux, D., et al. 2000. Clinical teaching: maintaining an educational role for doctors in the new health care environment. *Medical education,*Vol. 34**,** No. 10**,** pp. 820-826.

Punj, G. & Stewart, D. W. 1983. Cluster analysis in marketing research: review and suggestions for application. *Journal of Marketing Research,*Vol. 20**,** No.**,** pp. 134-148.

Ramani, S. & Leinster, S. 2008. AMEE Guide no. 34: Teaching in the clinical environment. *Medical teacher,*Vol. 30**,** No. 4**,** pp. 347-364.

Romero, C. & Ventura, S. 2007. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications,*Vol. 33**,** No. 1**,** pp. 135-146.

Romero, C., et al. 2008. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education,*Vol. 51**,** No. 1**,** pp. 368-384.

Ruiz, J. G., et al. 2006. The impact of e-learning in medical education. *Academic medicine,*Vol. 81**,** No. 3**,** pp. 207-212.

Seluakumaran, K., et al. 2011. Integrating an open-source course management system (Moodle) into the teaching of a first-year medical physiology course: a case study. *Advances in physiology education,*Vol. 35**,** No. 4**,** pp. 369-377.

Sparacia, G., et al. 2007. Initial Experiences in Radiology e-Learning1. *Radiographics,*Vol. 27**,** No. 2**,** pp. 573-581.

Talavera, L. & Gaudioso, E., 2004. Mining student data to characterize similar behavior groups in unstructured collaboration spaces. *Proceedings of the Artificial Intelligence in Computer Supported Collaborative Learning Workshop at the ECAI 2004*. pp. 17-23.

Wilson, G. & Stacey, E. 2004. Online interaction impacts on learning: Teaching the teachers to teach online. *Australian Journal of Educational Technology,*Vol. 20**,** No. 1**,** pp. 33-48.

Zaíane, O. R., 2002. Building a recommender agent for e-learning systems. *Computers in Education, 2002. Proceedings. International Conference on*. IEEE, pp. 55-59.

Zorrilla, M. E., et al. 2005. Web Usage Mining Project for Improving Web-Based Learning Sites. *In:* MORENO D AZ, R., PICHLER, F. & QUESADA ARENCIBIA, A. (eds.) *Computer Aided Systems Theory – EUROCAST 2005*. Springer Berlin Heidelberg.