LEARNER CLUSTERING TO FACILITATE INCREASED EFFECTIVENESS OF WEB-BASED LEARNING

Richard Czarnecki*
Michelle LaFleche**

INTRODUCTION

Web-based instruction can provide the connectivity and data access that is not available in the more traditional, on-site instructional systems. Data can be collected, aggregated and analyzed real-time in order to evaluate learner progress and overall system effectiveness. While there have been studies to evaluate web-based instruction such as using web page access history information (e.g. web site statistics), we believe there is additional potential in using learner instructional performance data to improve online courseware and educational systems. The Destinations ™ educational technology system (product of NCS Learn) enables the collection of a wide array of learner performance information that we believe could be used to further enhance the learning experience. We have the ability to use the multitude of learner performance data tracked by this system in order to 1) further individualize the learning process as well as 2) evaluate the overall effectiveness of instruction. The purpose of this study was to explore potential ways of applying the results of learner data analysis to enhance the system's functionality for our learners.

METHODOLOGY

There are many methods in which technology can be applied to data to help further refine the learning process, as well as to evaluate the overall effectiveness of instruction. One such method is to cluster learners into meaningful groups to detect similar learning patterns, and then enhance the system based on those patterns. Our plan was to use K-means clustering to first identify main groups of learners who performed similarly on the system's initial mathematics placement assessment. We would then examine the resulting clusters in relation to the learners' corresponding curriculum performance data to determine if any patterns could be established.

Our objective for using learner initial placement information was two-fold: 1) to determine if learners were accurately placed into the mathematics curriculum by the system, and 2) to identify learners who seemed to require additional instructional intervention. Three sets of data were selected to represent the diversity of the learners, as shown in Table 1.

Table 1: Data Sets

	Description	# of Learners
Data Set #1	adult learning center	76
Data Set #2	public high school	62
Data Set #3	GED prep	52

Senior Systems Engineer, NCS Learn
***Educational Systems Technologist, NCS Learn*

First, a k-means cluster analysis was performed on each individual data set. We chose to identify four clusters because we felt this number would allow us to obtain a manageable number of groups, but yet still provide enough detail for us to detect any significant emerging patterns between the groups. Next, we identified variables that would provide information regarding how learners were progressing in the curriculum after their initial placement. These variables were 1) math activity average score on first attempt, and 2) overall math activity average score. An activity is an instructional element lasting approximately 15-20 minutes that covers a specific skill concept. In order to have "mastered" an activity, learners are required to obtain a score of 85%. Up to three attempts on a given activity are permitted. We felt the first attempt of a given activity would reflect a learner's current knowledge state of a specific skill, while the overall activity average score provided us with a more holistic view of how the learner was progressing. A one-way ANOVA was then performed based on the four cluster memberships using these variables.

RESULTS

Table 2 below shows the results of the cluster analysis for each of the three data sets. Also shown are the one-way ANOVA results of the variables first attempt average activity score and the overall average activity score for each cluster. The comparison of clusters against these variables yielded statistically significant results.

Table 2: Data Set Analysis Results

	Cluster Centers		**ANOVA Results**	
	Initial Math Level	**# of Learners**	**1st Try—Math Activity Avg. Score**	**Math Activity Avg. Score**
Data Set #1				
Cluster 1	1.0	4	58%	56%
Cluster 2	4.0	57	77%	76%
Cluster 3	6.0	13	84%	84%
Cluster 4	9.0	2	81%	81%
Significance			.047	.033
Data Set #2				
Cluster 1	2.0	17	56%	51%
Cluster 2	3.0	30	85%	85%
Cluster 3	4.0	13	65%	67%
Cluster 4	10.0	2	64%	67%
Significance			.000	.000
Data Set #3				
Cluster 1	4.0	5	78%	81%
Cluster 2	6.0	16	90%	90%
Cluster 3	9.0	25	71%	74%
Cluster 4	10.0	6	90%	92%
Significance			.038	.046

INTERPRETATION

Based on the analysis results, site demographics appeared to have an impact on the clusters' initial math levels (see Table 2 above). For example, the public high school learners were using the system mainly for supplemental instruction instead of primary instruction. Therefore, they may have lacked the motivation to perform well. Or, they may simply not have possessed the skill level required to place higher. On the other hand, the GED-prep group, which consisted of learners who were participating voluntarily, tended to place high overall in the curriculum.

The instructional philosophy of the system is to initially place learners in their appropriate skill level, focusing on those skills with which they need the most assistance. Therefore, it is not expected that the learner is to master each activity with a score of 90-100%. This type of high average score may indicate that the learner is placed too low and therefore already comprehends the material. Conversely, a very low average activity score may indicate that the learner is placed too high and may not have the necessary prerequisite skills.

The majority of learners appeared to be placed correctly with average activity scores of >70% after initial placement. There were a few cases of high initial placement but low (<70%) average activity performance. In addition, some learners placed very low (e.g. level 1.0 or 2.0), and yet still performed poorly in activities (<60%). We used the average activity score information to help determine if learners may have been misplaced in the curriculum. For example, we would be able to see if learners had possibly been placed too high based on higher placement but low average activity scores, or too low based on higher activity scores. We focused on cases showing the possibility of too high of a placement rather than too low. This is due to the fact that the system has built-in mechanisms that move high-performing learners through curriculum more quickly. On the other hand, learners who are placed too high and consequently begin to struggle through the curriculum tend to become frustrated which may have a further negative effect on their performance.

There are several ways in which to use this type of information to further individualize the learning process. For example, to address learners who are placed low in the curriculum and are performing poorly in activities, the system may signal the instructor that these learners need extra assistance. Also, clustering to identify groups could be automated to provide detailed learner performance information for each cluster. Learners performing below a defined threshold for each cluster would then be identified for additional instructional intervention. In addition, reports could be generated that grouped like-performing learners together by skill area in which they are working. This would provide the instructor with information that could be used to direct some kind of small group instruction or collaborative learning project for these specific learners.