December 2006

A Visual Data Mining Approach to Understanding Students Using Computer-Based Learning Technology

Antonina Durfee
Appalachian State University

Scott Schneberger
Appalachian State University

Donald Amoroso
Appalachian State University

A Visual Data Mining Approach to Understanding Students Using Computer-Based Learning Technology

Antonina Durfee
Appalachian State University
durfeea@apstate.edu

Scott Schneberger
Appalachian State University
schnebergers@appstate.edu

Donald L. Amoroso
Appalachian State University
amoroso@appstate.edu

ABSTRACT

Educators are increasingly using online computer-based training and assessment software—especially with large classes or in distance education settings. This technology is often criticized, however, for hampering personalized interaction with students. This paper introduces a unique approach for analyzing student characteristics influencing their adoption and use of computer-based educational technology so that instructors can better meet student learning needs. Using visual, self-organizing mapping, our data mining approach clustered students based on input data from thirty-six survey questions posed to over 400 students with experience using computer based training and assessment. The data mining technique provided clear descriptions of four different student clusters. Based on the unique characteristics of the four clusters, instructors could optimize classroom resources as well as provide individualized support once specific students are matched to their respective cluster group. In this manner, continual computer-based assessments of students can be used to maximize computer-based learning and evaluation.

Keywords

CBT, CBA, training, assessment, educational technology, hybrid classes, data mining, SOM, Self-Organizing Map

INTRODUCTION

In recent years, researchers, teachers, and technologists have attempted to develop more effective computer-based education systems. Educational institutions are increasingly turning to distance education with various computer-based methods of course content delivery and computer-based assessment (CBA) tools. The computer based course tools are adapted for entry-level courses like introductory computing, or for large section, or for administering computer literacy or proficiency exams. Electronic, online, or computer-based learning can provide a number of advantages—such as time and place convenience for students and instructors, standardized delivery, self-paced learning, economies of scale in terms of classrooms and instructors, automated feedback to students and instructors, a variety of available content (Siegfried and Kennedy, 1995). IT can assist an instructor in extending availability beyond class time and office hours, establishing links among classmates, and accomplishing administrative activities (Benbunan-Fich, Hiltz, 2002).

Online education systems and large classes (more than 80 students) are often criticized for generality and a lack of personal contact. Instructors not getting adequate feedback from students while explaining a difficult concept can not elaborate on the topic to ensure better understanding. Information about the characteristics of a particular student class and of individual students can allow an instructor to tailor course delivery for maximum learning benefits. But what is the best method to gauge student characteristics to maximize computer-based training and assessment benefits?

Data mining methods seek knowledge from seemingly hidden data relationships and patterns where data dimensionality, complexity, and quantity are prohibitively large for common manual analysis. The commonly used techniques for data mining are statistical methods and machine learning methods. Assessing the value of discovered patterns, their usability, and their validity toward the investigated problem requires domain knowledge expertise, however, and thus may be difficult to automate. But the ease of interpreting visualized and validated patterns often makes data mining attractive to finding complex characteristics and patterns.

This paper explores using a visual data mining approach—the self-organizing map or SOM—for understanding student characteristics and patterns that can be used to guide instructors in optimizing learning resources. This study is the second in a series of studies on the factors that affect student learning using computer-based training and assessment software. While the first study indicated the significant factors affecting that learning, this study constructs a data mining artifact to help instructors better tailor computer-based training and assessment to meet class and individual needs. Our aim is to provide an instructor with computer-based data mining tools for assessing the learning needs of students using computer-based training and assessment systems.

LITERATURE REVIEW

Previous literature relative to this study's research issue is in two major areas: the relationships between various technical and individual characteristics and academic performance, and the use of data mining methods to learn more about students using educational technology. Two types of classroom information technologies are reviewed here: using technology to improve student learning (computer-based training or CBT), and using technology to improve student performance evaluation (computer-based assessment, or CBA).

Individual Characteristics and Technology-Enabled Learning

Many researchers have studied the relationships among student individual characteristics and academic performance. Arias and Walker (2004) found strong negative relationships between class size and student performance calculated as aggregate exam points while teaching economics. The results suggested to them that student ethics and proximity to an instructor in small classes help students understand economic concepts better. They included several measures of student academic abilities, i.e., SAT, SAT verbal and SAT math, GPA, and demographic data (such as year of study, age, and gender) as explanatory variables and class size as the control variable. Bostrom, et al., (1990) argued that individual differences are important for end-user training. Two studies in particular examined factors that influence computer training and skill gaining (Leidner and Jarvenpaa, 2001; Willet, 2002).

Ricketts and Wilks (2002) suggested that well-designed CBA can benefit students by improving their performances in assessments in the introduction of statistics in biology. Noyes, Garland, and Robbins (2004) studied paper-based and computer-based assessments, comparing the test performances of undergraduate students taking each test type. Given the identical multiple choice questions, students who used CBA achieved better results than those taking paper-based tests, and students with higher scores were found to benefit the most from CBA. Finally, CBA helped to improve long term recall of key concepts and resulted in higher scores than conventional exams, and students with computer experience had no additional advantages versus less experienced students (Bocij and Greasley, 1999).

Compeau and Higgins (1995) concentrated on studying self efficacy—the conviction that one can control his/her outcomes and do what is necessary to produce a certain result—and its importance in user acceptance and use of information technology. Learning style defined through demographic variables were found to have an effect on teaching and learning processes (Bostrom, et al., 1990). Student major as a predictor was mentioned in McGray (2000). There is also literature on the effectiveness of technical support for computer assisted learning. Bocij and Greasley (1999) concluded that students with computer experience had no additional advantages versus less-experienced students.

Data Mining Approaches

Data mining approaches reside on the shoulders of two giants of data analysis: statistics and artificial intelligence. Data mining helps convert data into knowledge by uncovering interesting, unobvious, or “golden” patterns to enhance the performance of an organization or a product (Fayyad, 1996). Data mining approaches are widely used in finance (Tan, 2002), marketing (Eklund, et al., 2002), logistics (Piatetsky-Shapiro, 2000), engineering, natural science research, bioinformatics (Fayyad and Uthurusamy, 2002; Hand, 2001), medicine (Goldman, et al., 1998) and text analysis (Tkatch, 1997).

Two common statistical data mining techniques are clustering and classification. *Clustering* identifies collections of similar data objects different from other objects; they are grouped based on the principle of maximizing intraclass similarity and minimizing interclass similarity. *Classification* finds patterns or a set of models in “training” data that describe and distinguish data cases or concepts. Classification constructs a model to predict the class of objects whose class type is known. The derived models may be presented as a set of association rules, decision trees, mathematical formulae, or neural networks.

Neural networks transform nonlinear, multidimensional input variables into another multidimensional output variable (Safer and Wilamowski, 1999). A derived measure of performance indicates how well a neural network has “learned” the relationships in the data. One popular neural network data mining technique using clustering is the self-organizing map

(SOM)—a good method for visualizing multidimensional data that has been used in pattern recognition, image analysis, and process monitoring (Vesanto, 1999).

Using machine learning, SOM creates a two-dimensional map from *n*-dimensional input training data. The map shows a landscape of standard shapes (such as hexagons) with borders between the shapes that define data clusters based on probability density (Kohonen, 1997). These clusters consist of input variables with similar characteristics but different intra-cluster characteristics that provide density. Overall, these clusters can provide visual clues about variable characteristics. Visualizing data can help when analyzing data for intuitive patterns—especially with exploratory research involving high numbers of variables—such as with this study.

SOM analysis is done in two steps: the training process, and the mapping process. When processing the data for neural training, node vectors are initially set to random values and then adjusted based on the input data. The SOM algorithm traverses each node using Euclidean distance formulas to find similarity between the input vector and the map's node weight vectors. The algorithm then identifies the node that produces the smallest weight vector distance—called the Best Matching Unit. In essence, it's a "winner take all" process where a node with its weight vector closest to the vector of inputs is declared the winner and its weights are adjusted to make it closer to the input vector—and its neighbors' weights are also changed relative to how close the neighbor is. This process is then repeated for each input vector, over and over, for a large number of cycles.[1] Resulting "feature planes" represent the values in a single vector column for identifying cluster characteristics. The end result of the entire data mining process is a map of output nodes called a U-matrix.

Much education-related, published research using data mining has focused on using the World Wide Web such as Chen, et al., (2001) using data mining tools to assist instructors in changing their pedagogical strategies and interventions by analyzing large volumes of Web-access logs. The authors are not aware of published research based on using visual, self-organizing maps to analyze the use of educational technology.

RESEARCH METHODOLOGY

We used a four-step methodology: data collection (student performance in hybrid class and CB assessments and survey responses), factor analysis, clustering analysis, and result visualization using a self-organizing map.

Data Collection: Survey Instrument

This paper built on a previous exploratory study on computer-based training and assessments (CBA) (Schneberger, et al., 2006). A wide range of variables were chosen from the initial study based on the model shown in Figure 1.

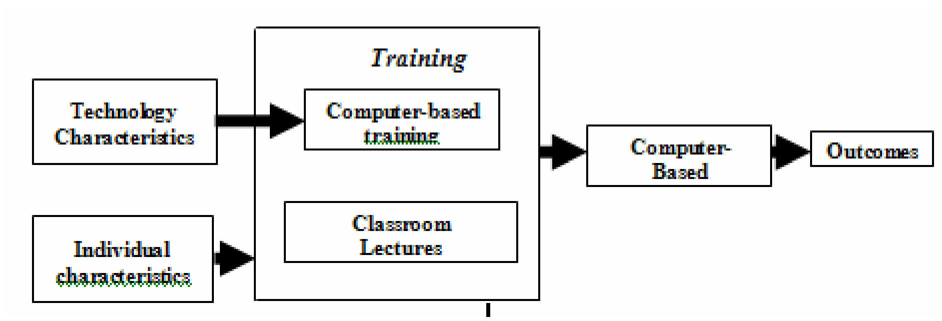


Figure 1. Initial Model for Hybrid Educational Technology Use

Data collection combined a survey methodology with direct performance measurement. Thirty-six questions were posed to over 550 students with direct and current experience using CBT and CBA for course credit. Additionally, the subjective or perceptive survey data was matched with measured, objective course performance scores. The combination of two collection approaches (perceptional and measured) allowed us to search for relationships among the subjective data, among the objective data, and between the subjective and objective data. All students used the same CBT/CBA software linked to the same course textbook. Six course sections were small (<35 students), and four were large (>100 students). All sections were taught individually by two professors; one had five small sections one semester then three large sections the following semester, the other had one large and one small section the same semester. Both professors followed identical syllabi over both semesters—each taught the same topics in the same sequence using the same textbook, lecture slides, and exams. The students were university undergraduate students of a wide range of ages and of all academic years from all schools across

[1] wikipedia.com, Self-Organizing_Map, 2/13/2006.

campus. It was a required course for some students, but not for all. The specific results of the survey analysis can be found in Schneberger, et al., 2006.

Factor Analysis

Factor analysis is an interdependence technique which defines the underlying structure among variables in the analysis (Hair, 2006). Highly interrelated (correlated) variables form factors which represent a data dimension. We used factor analysis to refine the structure of the survey data, consolidating or collapsing the initial list of variables into a smaller group of highly cohesive yet unique factors we could use for optimal data mining. The overall sample size met the general minimum criteria for factor analysis—having at least five times the number of observations than the number of variables. We used SPSS 14.0 to reduce data using Principal Component Analysis.

Data Mining Using a Self-Organizing Map (SOM)

SOM identifies borders defining unique clusters of input data based on common characteristics; we used the output of the factor analysis to improve the SOM clustering. SOM analysis ends when the average quantization error is small enough to present diminishing marginal returns from continued analysis. The SOM maps were created using SOM_PAK, a SOM software package developed at the Helsinki University of Technology and freely available at http://www.cis.hut.fi/research/som_pak/. The U-Matrix map is visualized using Nenet v1.1a (Neural Networks Tool) application designed to illustrate the visualization abilities of SOM. The demo version of Nenet is available at http://koti.mbnet.fi/~phodju/nenet/Nenet/l.

DATA ANALYSIS

Using the results of the earlier, exploratory study on CBT and CBA variables affecting learning outcomes, the initial data analysis step was to use principal component analysis with Varimax and Kaiser normalization rotation to cluster the initial 28 input variables into their orthogonal factors. The eight significant resulting factors are shown in Tables 1 and 2 explaining 65.25% of cumulative variance and a sufficient level of sampling adequacy of .778 as measured by the Kaiser-Meyer-Olkin measure. “SAM” is the brand name of the computer-based training and assessment software.

Question	SAM Value	Computer Skills	SAM Usability	CBA Help	SAM Tech Support	Preparation	Internet Experience	Emotional Support
	1	2	3	4	5	6	7	8
Q32	0.833							
Q19	0.803							
Q20	0.797							
Q31	0.788							
Q22	0.711							
Q21	0.681							
Q17	0.632							
Q12		0.830						
Q13		0.805						
Q14		0.744						
Q11		0.677						
Q27		0.659						
Q16			0.749					
Q15			0.741					
Q18			0.697					
Q29				0.836				
Q30				0.794				
Q28				0.689				
Q24					0.894			
Q25					0.885			
Q23					0.448			
Q33						0.797		
Q26						0.579		
Q8							0.787	
Q9							0.571	
Q35								0.757
Q34								0.690

Table 1. Factor Components Analysis

Factors	Questions	Title	Interpretation
1	Q17,19-22, 31,32	SAM value	SAM clarity, usefulness, and ability to prepare them for exams
2	Q 10-14, 27	Computer skills	Initial computer skills and those acquired in class
3	Q 15-16, 18	SAM usability	Ease in learning and navigating SAM
4	Q 28-30	CBA help	Help from instructor, peers, and tech support
5	Q 23-25	SAM tech support	Accuracy and helpfulness of SAM tech support
6	Q 26, 33	Preparation	Difficulty preparing for class and CBA
7	Q 8, 9	Internet experience	Number of years and experience on the Internet
8	Q 34, 35	Emotional support	Moral support outside of class

Table 2. Factor Interpretation

The next step was clustering and visualizing the converted dataset with aggregated factor scores using SOM data mining. Table 3 shows the training parameters used with the SOM algorithm. Appendix 1 shows the feature planes for the eight factors, where colder (bluer, or darker) and warmer (reddish, or grayer) color codes show higher values of factors.

Number of trials:	100	Network Size	5x7
Training length of 1st part:	1750	Training rate of 1st part:	0.5
Network radius of 1st part:	7	Training rate of the 2nd part:	17500
Training rate of the 2nd part:	0.05	Neighborhood radius of 2nd part:	1

Table 3. The SOM Training Parameters

To plot the U-matrix and identify the resulting clusters of students, we combined the individual feature planes from Appendix A as shown in Figure 2. By analyzing the shading of the hexagons and the borders between them, we identified similarities as well as differences forming clusters. The resulting clusters are shown in Figure 3 as a SOM U-matrix map with highlighted borders. The interpretation of the clusters is based on the factor descriptions.

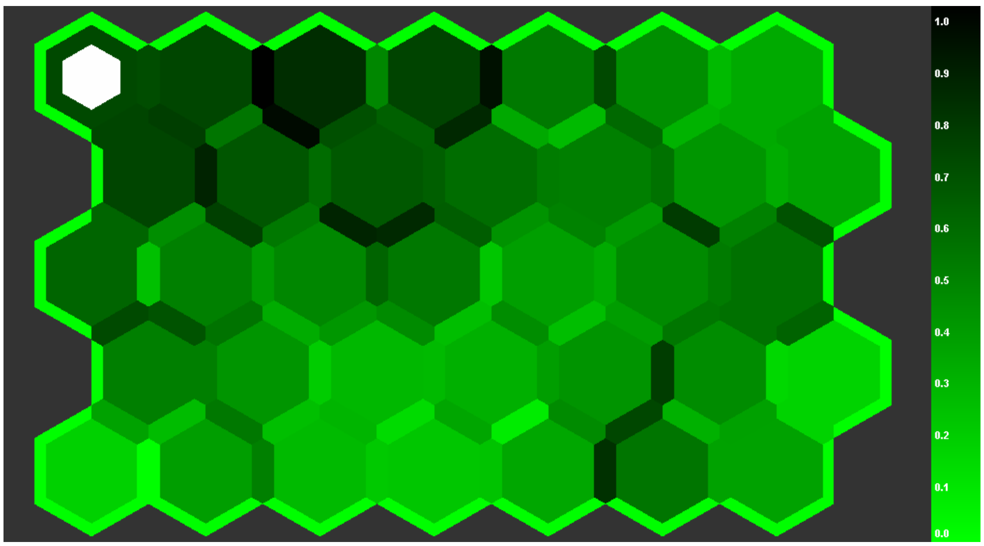


Figure 2. SOM U-Matrix Map

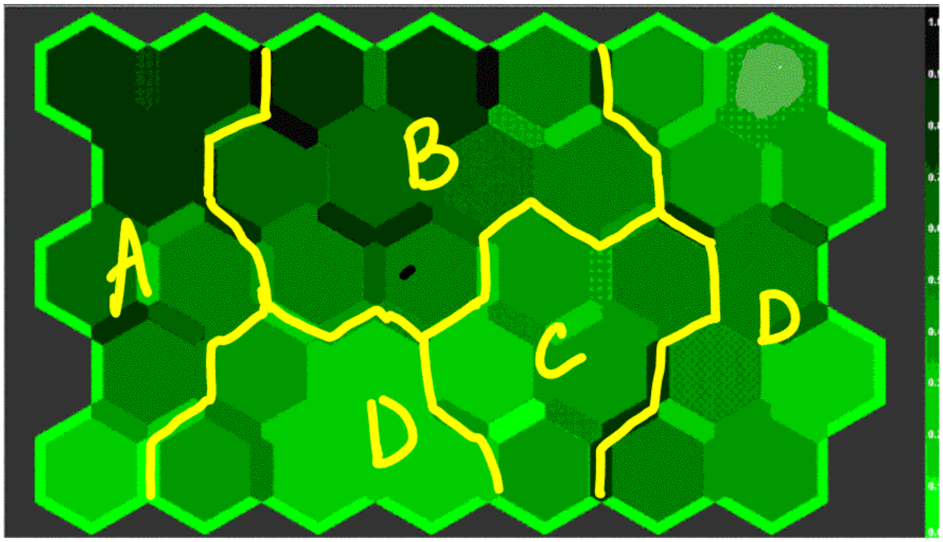


Figure 3. SOM U-Matrix Map With Highlighted Borders

ANALYTICAL RESULTS AND DISCUSSION

Group **A** is located at the left hand side of the map shown in Figure 3. This group represents students who value and use computer-based training often; they already had above average computer skills, had above average access to broadband Internet access, they thought the CBT/CBA software was very easy to navigate, and therefore did not highly need or value associated technical support. Students in this group received minimal moral support outside the classroom and used minimal help while performing assignments. While these students did not consider themselves to be hardworking or persistent, they did place great value on computer-based training and used it to their advantage. In summary, this could be considered an ideal group for exploiting the advantages of CBT/CBA. When working with students in group A, instructors can maximize course learning by maximizing computer-based training opportunities.

Group **B** is located in the top middle part of the map. It contains students who felt they had little computer skills to begin with, but placed the highest value on computer-based training preparing them for computer-based assessments—even though technical support seemed to be of little value. Students in this group relied on collaborative external help and emotional support even though they perceived themselves as hardworking and persistent. Overall, these students had low computer literacy, saw little value in technical support, but placed high value on computer-based training. Instructors dealing with students in this group could provide more background on computer literacy fundamentals, more guidance on using technical support, and take advantage of opportunities to use computer-based training.

Group **C** is the smallest group located in the bottom, middle part of the Figure 3 map. It includes students with the highest computer proficiency and skills, but who find this particular computer-based training and assessment software very hard to navigate—and praise technical support in helping them with it. Students in this group describe themselves as the most persistent and hardworking, but rarely do computer-based training and therefore do not value it highly. At the same time, these students enjoy talking about course content outside the class and rely the most on external emotional support—but do not often get help with computer-based assessments. In summary, they had the strongest computer skills and are very persistent workers, and believed technical support was timely and useful but that this particular CBT/CBA software was very unclear and difficult to navigate. Instructors working with students in this group are dealing with more sophisticated CBT/CBA users, and may, therefore, choose to focus more up front on how to navigate and use the particular CBT/CBA software.

Group **D** is the largest group, located in the right hand side of the map with a second subgroup in between groups A and C. This group represents students who have below average computer skills, find the computer-based training and assessment software very difficult to navigate, and do not use or value its training. They also see themselves as having very little Internet knowledge and experience, rarely discuss the course content outside the classroom, and frequently seek help from friends or an instructor while doing CBT/CBA work. They perceive themselves to be moderately hard working but do not take advantage of training before completing assessments. On the whole, this appears to be the group that needs the greatest amount of attention. Instructors dealing with this group can improve basic computer skills, spend more effort on explaining how to navigate and use the CBT/CBA software, require computer-based training before attempting computer-based assessment, and ensure easy access to help during training and assessment.

IMPLICATIONS

On the surface, it may appear difficult to cleanly label the four SOM groups described above since they might appear to have a mixture of indirectly related characteristics. But the SOM data mining technique clearly found that the students in each group are distinctly related and different from other groups. While focusing on those group similarities and differences may provide a significant insight on students in general who use CBT and CBA, there is a simpler and highly practical use of a U-matrix map such as that shown in Figure 3. After surveying students early in a course, performing the SOM data mining, and matching student IDs to a resulting U-matrix map, an instructor can quickly gauge the needs of an entire student class as well as individual students. Figure 4 shows the Figure 3 cluster map labeled with associated (in this case, coded) student IDs for each respective hexagon.

For example, an instructor could immediately see from Figure 4 that roughly 40% of the students belong to group D, 30% belong to group A, 15% belong to group C, and the remaining 15% to group B. If facing limited time and resources, an instructor could, therefore, seek to optimize time and resources by putting more emphasis on meeting the needs of the largest group—in this case, by stressing student collaboration in training, by providing more background information about computing and the Internet, by spending more effort on explaining how to navigate through the CBT/CBA software, and perhaps by requiring CBT as part of the course performance evaluation. If resources allow, the instructor can then similarly address the needs of the smaller groups. Likewise, knowing which group a particular student belongs to allows an instructor to address individual student needs based on the general characteristics of students in the respective SOM group.

An appealing point about using the SOM data mining method in this manner is that the resulting maps don't show *what* course material students know or don't know—it shows *how* they are likely to best learn the course material. This may be highly beneficial when using computer-based training and assessment since much of that activity may be done outside the classroom—outside of direct observation by an instructor. Moreover, instructors with large classes using computer-based training and assessment may find this data mining technique especially useful in providing targeted, more personal attention to individual students or small groups who might feel isolated in large settings.

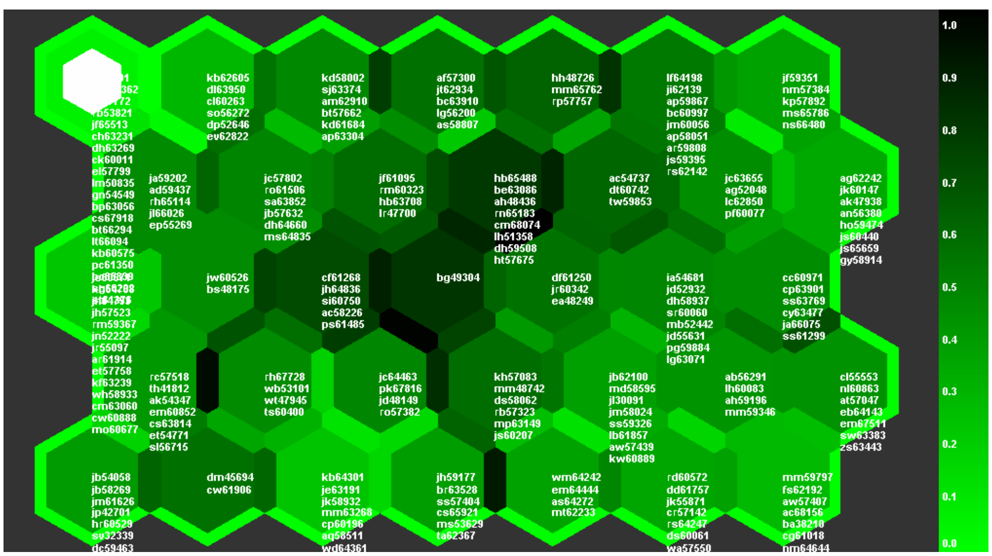


Figure 4. Labeled SOM U-Matrix Map

SUMMARY

Organizations that use computer-based training and assessment tools can potentially reap significant rewards in improving knowledge and skill levels, while minimizing resource expenditures and maximizing student access. But extensive use of computer-based technologies is often criticized in educational circles for a perceived loss of personal interaction between instructors and students. By using a data mining approach such as SOM with self-organizing maps, an instructor can profile students from online or large classes to change the focus or overcome specific pedagogical challenges. The data mining approach provides a personalization and customization tool for clustering student learning and tailoring online training and assessment for use in hybrid classes. In essence students can receive customized learning based upon their matrix location and therefore can learn at much higher rates than if there were only one learning module developed for a class at-large.

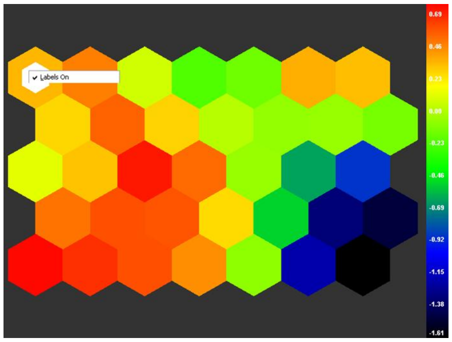
Future research will include the development of predictive models for student learning based upon SOM locations. Corporate training can be enhanced by using the SOM technique prior to corporate investment in training programs.

REFERENCES

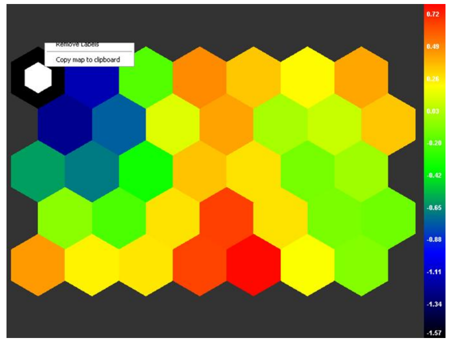
1. Amoroso, D. (2005) Use of online assessment tools to enhance student performance in large classes, *Information Systems Education Journal*, 3, 4, 1-10.
2. Arias, J. J. and Walker, D. M. (2004) Additional evidence on the relationship between class size and Student Performance, *Journal of Economic Education,* 35, 4, 311-326.
3. Benbunan-Fich, R. and Hiltz, S. (2002) Correlates of effectiveness of learning networks, *Proceedings of the 35th Hawaii Conference on System Sciences*, January 6-9, Big Island, HI
4. Bocij, P. and Greasley, A. (1999) Can computer-based testing achieve quality and efficiency in assessment?, *International Journal of Educational Technology*, 1, 1.
5. Bostrom, R., L. Olfman, and M. Sein, (1990) The importance of learning style in end user training, *MIS Quarterly*, March, 101-120.
6. Chen, G., Ou, K., Liu, C. and Liu, B. (2001) Intervention and strategy analysis for web group learning, *Journal of Computer Assisted Learning*, 17, 58-71.
7. Chen, C. and Ying-Ling Hsieh, (2005) Mining learner profile utilizing association rules for common learning misconception diagnosis, *Proceedings of the fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)* 588-592.
8. Compeau, D. and Higgins, C. (1995) Computer self-efficacy: development of a measure and initial test, *MIS Quarterly,* 23, 2, 145-158.
9. Durfee, A., Schneberger, S., and Amoroso, D. (2005) Computer-based assessments of student performance in hybrid classes: does class size really matter?, *Proceedings of Americas Conference on Information Systems (AMCIS'05)* August 2-6, Omaha, NE.
10. Eklund, T., B. Back, H. Vanharanta and A. Visa. (2002) Assessing the feasibility of self-organizing maps for data mining financial information, *Proceedings of the Xth European Conference on Information Systems (ECIS 2002)* Gdansk, Poland.
11. Fayyad, U. (1996) Data mining and knowledge discovery: Making sense out of data, *IEEE Expert*, 20-25.
12. Fayyad, U. and Uthurusamy, R. (2002) Evolving Data Mining into Solutions for Insights, *Communications of the ACM* 45(8) 28-31.
13. Goldman, J. A., Chu, W.W., Parker, D. S. and Goldman, R. M. (1998) Term domain distribution analysis, a data mining tool for text databases: A case history in a thoracic lung cancer text database, *Proceeding of the first International Conference on Discovery Science,* December 14-16, Fukuoka, Japan
14. Hair, J., Black, W., Babin, B., Anderson, R. and Tatham, R. (2006) *Multivariate Data Analysis*, Pearson Prentice Hall, New Jersey, 6th ed.
15. Hand, D., Mannila, H. and Smyth, P. (2001), *Principles of Data Mining,* Boston, USA, A Bradford Book, the MIT Press.
16. Hill, M. (1998) Class size and student performance in introductory accounting courses: further evidence, *Issues in Accounting Education*, 13, 1, 47-64.
17. Kohonen, T. (1997) *Self-Organizing Maps*. Leipzig, Germany, Springer-Verlag.
18. Kozma, R. (2003) Technology and classroom practice: An international study, *Journal of Research on Technology in Education,* 31, 1, 13-20.
19. Leidner, D. and Jarvenpaa, S. (2001) The information age confronts education: case studies on electronic classrooms, *Information Systems Research,* 4, 1, 24-54.
20. Minaei-Bidgoli, B., Kashy D., Kortemeyer G., Punch D. (2003) An application of data mining methods with an educational Web-based, *IEEE/ASEE FIE 2003 Frontier In Education*, - ieeexplore.ieee.org
21. Northcote, M. (2002) Online assessment: friend, foe, or fix? *British Journal of Educational Technology*, 33, 5, 623-625.
22. Noyes, J., Garland, K. and Robbins, L. (2004) Paper-based versus computer-based assessment: is workload another test mode effect?, *British Journal of Educational Technology,* 35, 1 , 111-113.

23. Piatetsky-Shapiro, G. (2000) Knowledge discovery in databases: 10 years after, *SIGKDD Explorations*, 1, 2, 59-61.
24. Ricketts, C. and, Wilks, S. J. (2002) Improving students performance through computer-based assessment: insights from recent research, *Assessment & Education in Higher Education*, 27, 5, 476-479.
25. Riffell, S.K. and Sibley, D.H. (2003) Learning online student perceptions of a hybrid learning format, *Journal of College Science Teaching*, 32, 6.
26. Schneberger, S., Amoroso, D., and Durfee, A. (2006) Computer-based training and assessments: A holistic study. *Proceedings of HICSS-39*, Kauai, HI, January 4-7.
27. Siegfried, J. and Kennedy, P. (1995) Does pedagogy vary with class size in introductory economics?, *AEA Papers and Proceedings*, *Better Learning from Better Management*, 85, 2, 347-351.
28. Van Blerkom, M. L. (1992) Class attendance in undergraduate courses, *Journal of Psychology*, 126, 487-494.
29. Safer, A. and Wilamowski, B. M. (1999) Using neural networks to predict abnormal returns of quarterly earnings, *International Joint Conference on Neural Networks* - *IJCNN'99*, Washington, DC.
30. Tan, R. G. H., van den Berg, J. and van den Bergh, W.-H. (2002) Credit rating classification using self-organizing maps, neural networks in business: Techniques and applications. J. Gupta. Hershey, Idea Group Publishing**:** 140-153.
31. Tkatch, D. (1997) Text mining technology: turning information into knowledge, IBM Corp.
32. Vesanto, J. (1999) SOM-based data visualization methods, *Intelligent Data Analysis*, 3, 3, 111-126.
33. Zaiane, O., and J. Luo, (2001) Towards Evaluation Learners' Behaviour in a Web-based Distance Learning Environment, *Proceeding of the IEEE International Conference on Advanced Learning Techniques (ICALT'01).*

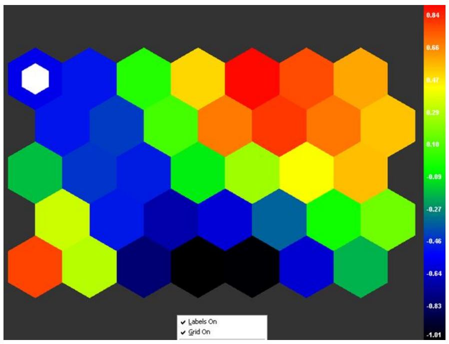
APPENDIX A: SOM FEATURE PLANES FOR FACTORS 1-8



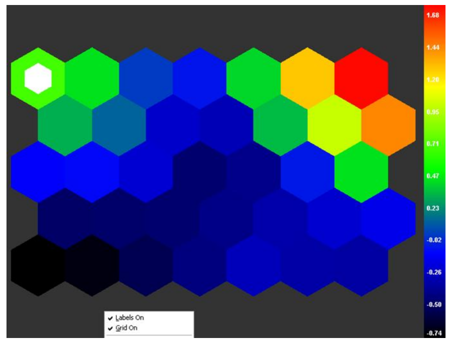
Factor 1



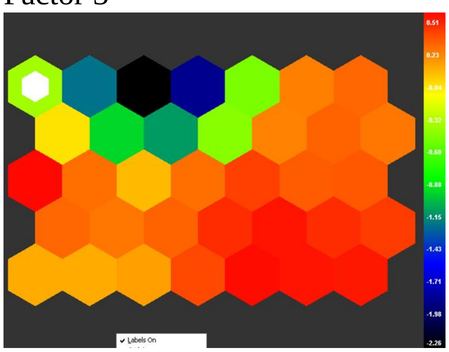
Factor 2



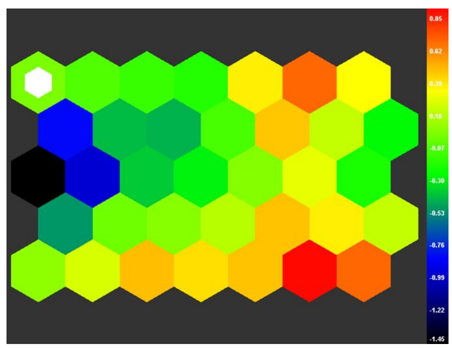
Factor 3



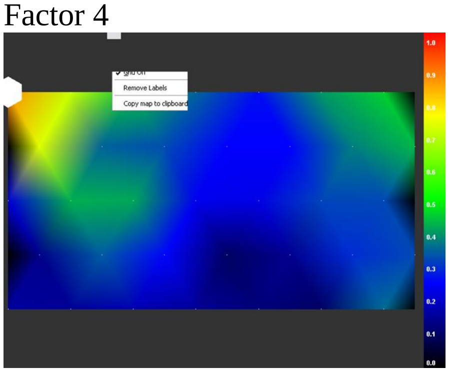
Factor 3



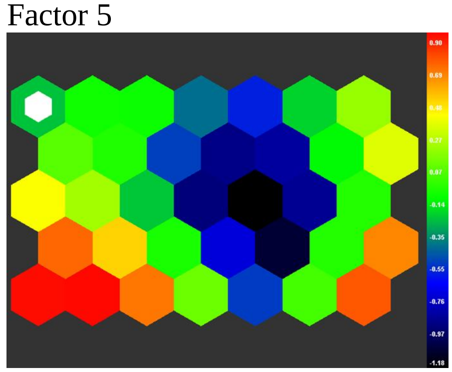
Factor 4



Factor 5



Factor 7



Factor 8

APPENDIX B: SURVEY INSTRUMENT

Demographics

1. English is my primary language (1=Yes, 2=No)
2. My SAM username is ____________
3. My gender (1=Male, 2=Female)
4. My age is __________
5. My academic year (1=Freshman, 2=Sophomore, 3=Junior, 4=Senior)
6. My current GPA: __________
7. My Internet connection type (1=Dial-up, 2=Cable/DSL, 3=T1 or better, 4=Don't know)
8. Number of years using the Internet (<1, 1,2,3,4,5,6,7,8,9,>9)
9. Experience with Internet (1=No, 2=Little, 3=Some, 4=Much, 5=Extensive)

Computing Skills

10. Basic skills like typing a document, etc. (1=strongly disagree, 2=slightly disagree, 3=indifferent, 4=slightly agree, 5=strongly agree)
11. Install programs, etc. (same 5pt. scale)
12. Set up virus checkers, etc. (same 5pt. scale)
13. Install networks, etc. (same 5pt. scale)
14. Install new hardware (same 5pt. scale)

SAM Expertise

15. Learning SAM was easy (same 5pt. scale)
16. Navigating and accomplishing SAM tasks is easy (same 5pt. scale)
17. What SAM tells me is clear and understandable (same 5pt. scale)
18. Overall, I find SAM easy to use (same 5pt. scale)
19. SAM helps me prepare for assessments (same 5pt. scale)
20. SAM easily trains me on MS Office basic functions (same 5pt. scale)
21. SAM decreased the time to learn MS Office functions (same 5pt. scale)
22. SAM improved my ability to use MS Office (same 5pt. scale)
23. I use SAM technical support often (same 5pt. scale)
24. SAM tech support helps me well and timely with SAM problems (same 5pt. scale)
25. SAM tech support is very accessible and knowledgeable (same 5pt. scale)
26. I prepare for classes and assessments well (same 5pt. scale)
28. I receive help from other students while doing assessments (same 5pt. scale)
29. I receive help from SAM/IT tech support while doing assessments (same 5pt. scale)
30. I receive help from my instructor while doing assessments (same 5pt. scale)
31. SAM training reflects what is covered in assessments (same 5pt. scale)
32. SAM training prepares me well for assessments (same 5pt. scale)
33. I use SAM training often (same 5pt. scale)
37. I do assessments (1=on my own computer, 2=in the lab, 3=at a friend's house, 4=elsewhere).

Self-Efficacy

27. I work very hard and persistently in CIS1025 (same 5pt. scale)
34. I am certain I can master the skills in CIS1025 (same 5pt. scale)
35. I often discuss CIS1025 content with friends/family/etc. (same 5pt. scale)
36. I often receive general emotional support from others (same 5pt. scale)