Proceedings of the 2007 IEEE Symposium on
Computational Intelligence and Data Mining (CIDM 2007)

Selecting the Right Peer Schools for AACSB Accreditation - A Data Mining Application

Melody Y. Kiang
California State University, Long Beach
1250 Bellflower Blvd.,
Long Beach, CA 90840

Steven A. Fisher
California State University, Long Beach
1250 Bellflower Blvd.,
Long Beach, CA 90840

Dorothy M. Fisher
California State University, Dominguez Hills
1000 E. Victoria Street
Carson, CA 90747-0005

Robert T. Chi
California State University, Long Beach
1250 Bellflower Blvd.,
Long Beach, CA 90840

Abstract—**For a business school, the selection of its peer schools is an important component of its International Association for Management Education (AACSB) (re)accreditation process. A school typically compares itself with other institutions having similar structural and identity-based attributes. The identification of peer schools is critical and can have a significant impact on a business school's accreditation efforts. For many schools the selection of comparable peer schools is a judgmental process. This study offers an alternative means for selection; a quantitative technique called Kohonen's Self-Organizing Map (SOM) network for clustering. SOM as a software agent uses visualization to present information to the school in choosing its peer schools.**

I. INTRODUCTION

AACSB accreditation is critical to the success of a business school. It gives the school prestige by placing it in an elite group of accredited institutions that meet strict quality standards. However, more tangible benefits of accreditation include improved curriculum and operations, enhanced fund-raising and the ability to attract and retain quality students and faculty. The accreditation process is an arduous and costly process that requires a school not just to meet AACSB standards but also to reflect on its own mission, operations, and direction.

A key dimension of the (re)accreditation process is the selection of its comparable peer schools. A school typically compares itself to other institutions with similar structural attributes such as size, degree-granting type, resources, etc. Comparisons with peer schools are necessary for understanding and evaluating the business school's current performance and future goals. AACSB defines comparable peers as "…schools who are considered to be similar in mission and are assumed to be appropriate for performance comparison" [5] (http://www.aacsb.edu/accreditation/glossary.asp). AACSB requires a business school to identify a minimum of six comparable schools.

The selection of comparable peers is a critical process and can have a significant impact on a school's accreditation efforts. For many schools the selection of comparable peers is a judgment process. However, this can be formidable task given the large number of business schools and the many relevant attributes. We offer an alternative means for selection; a quantitative technique called Kohonen's SOM networks, an unsupervised learning neural network for clustering, to assist schools in identifying their "AACSB Comparable Peers".

The Self-Organizing Map (SOM) network, a variation of neural computing networks, is a categorization network developed by Kohonen [2,3,4]. The SOM network was originally designed for solving problems that involve tasks such as clustering, visualization, and abstraction. The main function of SOM networks is to map the input data from an n-dimensional space to a lower dimensional (usually one or two-dimensional) plot while maintaining the original topological relations. The physical location of points on the map shows the relative similarity between the points in the multi-dimensional space.

In this research, we apply the clustering and visualization capabilities of SOM to plot the 229 AACSB accredited schools into a two-dimensional map. The map will assist a candidate school to properly identify its peer schools for comparison during the AACSB (re)accreditation process. Thus, SOM as a software agent uses visualization to present information to assist a school in choosing its peer schools.

The balance of the paper is organized as follows: Section two presents the basic concepts of SOM network and illustrates its use as a data-reduction tool. It is followed by a discussion of the extended grouping capability. Section three describes the data sets and the experimental design. The paper concludes with a summary of our findings.

II. SELF-ORGANIZING MAP (SOM) NETWORKS

Unlike other neural network approaches, the SOM network performs unsupervised training; that is, during the learning process the processing units in the network adjust their weights primarily based on the lateral feedback connections. The more common approach to neural networks required supervised training of the network (i.e., the network is fed with a set of training cases and the generated output is compared with the known correct output). Deviations from the correct output result in adjustment of the processing units' weights. On the other hand, unsupervised learning does not require the knowledge of target values. The nodes in the network converge to form clusters to represent groups of entities with similar properties. The number and composition of clusters can be visually determined based on the output distribution generated by the training process.

The SOM network typically has two layers of nodes, the input layer and the Kohonen layer. The input layer is fully connected to a two-dimensional Kohonen layer. During the training process, input data are fed to the network through the processing elements (nodes) in the input layer. As the training process proceeds, the nodes adjust their weight values according to the topological relations in the input data. The node with the minimum distance is the winner and adjusts its weights to be closer to the value of the input pattern.

The network undergoes a self-organization process through a number of training cycles, starting with randomly chosen weights for the nodes in Kohonen layer. During each training cycle, every input vector is considered in turn and the winner node is determined. The weight vectors of the winning node and the nodes in the neighborhood are updated using a weight adaptation function. The learning algorithm we implemented for network training is similar to the one implemented by Kiang [1]. Reader may refer to [1] for detailed algorithm.

The output from SOM networks is a two-dimensional map (Kohonen layer). Each node on the map may represent zero to many input data. The input data that are similar in higher dimension should be close to each other on the output map. We can consider each node on the output map as one group and cluster the input data accordingly. However, this type of Kohonen network usually has many nodes in the output layer. For example, a network of size 10x10 will have total 100 nodes in the output layer. When the number of nodes on the map is more than the number of clusters we desire, additional procedure to further group the nodes into fewer number of clusters is required In this study, we applied the extended SOM network method developed by [1] to automate the segmentation process to complement the usage of the Kohonen SOM networks. The method groups the output from SOM based on a *minimal variance criterion* to merge the neighboring nodes together. We start with each node in the map representing one group, and calculate the centroid of each group. Then we try to merge two neighboring groups so the result of the merge will maintain the global minimal variance for that number of clusters. The merge process is repeated until a user specified number of clusters has derived or when only one cluster remains. Readers should refer to [1] for the discussion of the detailed process.

III. EXPERIMENTAL DESIGN AND RESULTS

In this study, we have chosen eleven attributes as input parameters to train the SOM network. The eleven attributes are the important structural and identity-based characteristic identified from the school profile data collected at AACSB. They are Degree Offered (Undergraduate/Masters/Doctoral), Private/Public and Commuter/Residential, Carnegie Classification, Endowment, ratio of Budget to Full Time Equivalent Faculty, Total Full Time Equivalent Faculty, MBA Degree Confirmed, ratio of Full Time Faculty Doctorate to Full Time Faculty, ratio of Full Time Equivalent Faculty to Full Time Faculty, the GMAT score, and MBA tuition. The data sources for this study are from AACSB. Table 1 presents the encoding scheme used for some of the attributes:

TABLE 1
ENCODING SCHEME

Highest Degree Offered		Private/Public & Commuter/Residential		Carnegie Classification	
Under-graduate	1	Public & Commuter	1	Bachelor/Specialized Institution	1
Masters	2	Private & Commuter	2	Master's I	2
Doctoral	3	Public & Residential	3	Master's II	3
		Private & Residential	4	Doctoral – Intensive	4
				Doctoral –Extensive	5

After removing school with missing data, there are a total of 229 schools (input vectors) with 11 attributes each. The input values were preprocessed to reduce the impact of off-scaled attributes on the result of the output. The most commonly used input pre-processing function in SOM is implemented:

$$\text{Input}_{i,\,new} = (\text{Input}_{i,\,old} - \text{Mean}_i)/\text{Standard\_Deviation}_i$$

After some preliminary runs, it shows that the network size has no significant effect on the performance of the network. Therefore we used network sizes of 11x11 for our experiments. We implemented the algorithm in C++ programming language. The output map of the 229 schools is shown in Figure 1.

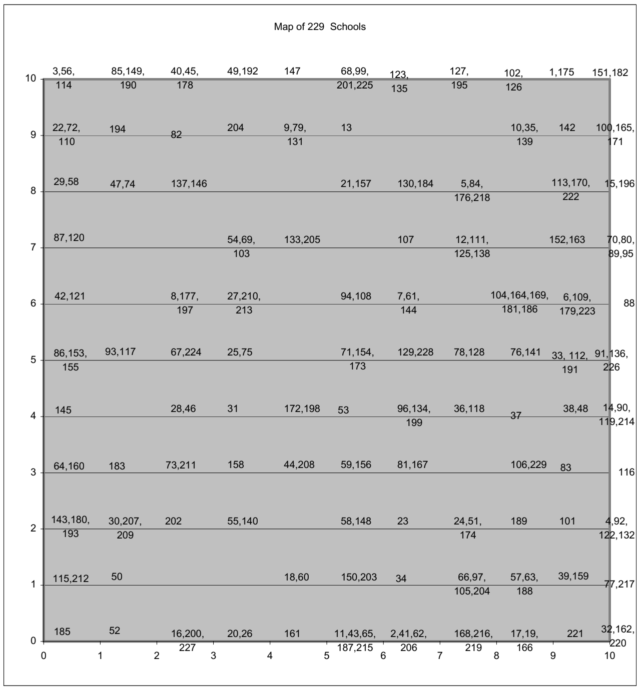

Fig. 1 Output map of 229 schools

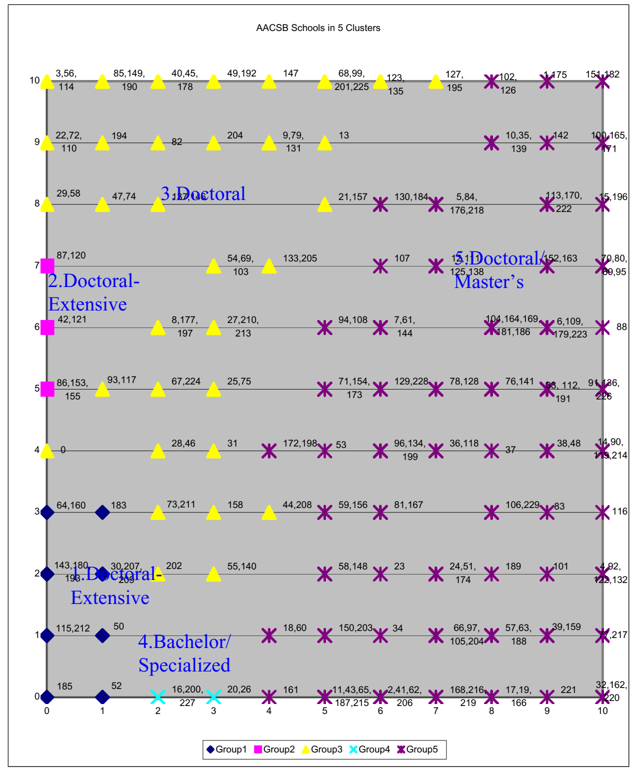

Fig. 2 The output map of the 229 schools grouped into five cluster

We have noticed that the outcome of SOM groups somewhat matches with the Carnegie classification in Table 1, therefore we applied the extended SOM method to further group the 229 schools into five clusters based on their closeness in output map. Figure 2 shows the resulting 2-dimensional plot depicting the five groups.

Although there is no direct match between Carnegie classification and SOM clusters, we tried to label the 5 groups on SOM map according to its closest Carnegie classification. One-way analysis of variance (ANOVA) was conducted to test if there is significant difference in each attribute among the five segments formed through SOM. The results are presented in Table 2. Statistically significant differences are detected for all attributes among all five clusters at $p < 0.0001$. This is a good indication that the extended SOM method has correctly identified five significantly different groups and is an effective decision support tool for clustering and visualization.

TABLE 2.
TEST OF THE SIGNIFICANCE OF DIFFERENCE AMONG THE FIVE GROUPS

ANOVA				
		df	F	Sig.
Degree Offered	Between Groups	4	161.892	0.000
	Within Groups	224		
	Total	228		
Public/Private & Commuter/Residential	Between Groups	4	9.252	0.000
	Within Groups	224		
	Total	228		
Carnegie Classification	Between Groups	4	84.967	0.000
	Within Groups	224		
	Total	228		
Endowment	Between Groups	4	73.461	0.000
	Within Groups	224		
	Total	228		
Budget/Full_Time_ Equivalent _Faculty	Between Groups	4	138.415	0.000
	Within Groups	224		
	Total	228		
MBA Degree Confirmed	Between Groups	4	48.141	0.000

	Within Groups	224		
	Total	228		
Total Full Time Equivalent Faculty	Between Groups	4	23.822	0.000
	Within Groups	224		
	Total	228		
Full_Time_Faculty_ Doctorate/Total_ Full_Time_Faculty	Between Groups	4	17.907	0.000
	Within Groups	224		
	Total	228		
Total_Full_Time_ Equivalent_Faculty/ Total_Full_Time_Faculty	Between Groups	4	40.266	0.000
	Within Groups	224		
	Total	228		
MBA Tuition	Between Groups	4	61.015	0.000
	Within Groups	224		
	Total	228		
GMAT Score	Between Groups	4	51.639	0.000
	Within Groups	224		
	Total	228		

The extended SOM technique can derive any number of clusters as the user specifies. It is a decision support tool that requires human agent interaction [6]. In order to better identify six peer schools as suggested by AACSB, we printed the clustering results of 5, 10, 20, 40, 60, 80, and 100 clusters. The complete school list with corresponding cluster numbers is available upon request. The various numbers of clusters can help a candidate school to identify any number of peer schools according to its closeness on the map to them. For example, the peer schools of California State University, Long Beach identified by the extended SOM are:

School#	SchoolName	x y	100 clusters	80 clusters	60 clusters	40 clusters	20 clusters	10 clusters	5 clusters
78	Grand Valley State University	7 5	67	59	45	30	17	9	5
128	University of Nebraska at Omaha	7 5	67	59	45	30	17	9	5
76	Georgia Southern University	8 5	68	59	45	30	17	9	5
141	University of Northern Iowa	8 5	68	59	45	30	17	9	5
38	California State University, San Bernardino	9 4	69	59	45	30	17	9	5
48	University of Colorado at Colorado Springs	9 4	69	59	45	30	17	9	5
37	California State University, Long Beach	8 4	70	60	45	30	17	9	5
83	University of Houston-Clear Lake	9 3	70	60	45	30	17	9	5

IV. CONCLUSION AND FUTURE RESEARCH

In this study, we applied the SOM network to group the 229 schools into five segments. We first used a two-dimensional map, the most common Kohonen network, to capture the relationships among the 229 schools. An advantage of the two-dimensional network is that it allows the users to visualize the data distribution on a plot. We further applied the extended SOM method to group the 229 schools into five clusters to compare with the Carnegie classification. The output map of SOM provides a graphical interface to help candidate schools to visualize the different characteristics of the schools thus reduce the task from a multi-dimensional problem to a 2-dimensional map.

REFERENCES

[1] Kiang, M.Y., "Extending the Kohonen Self-Organizing Map Networks for Clustering Analysis," *Journal of Computational Statistics and Data Analysis* (CSDA), 2001.

[2] Kohonen, T., *Cybernetic Systems: Recognition, Learning, Self-Organization* (ed. by E.R. Caianiello, G. Musso), Research Studies Press, Ltd., Letchworth, Herfordshire, UK, 1984.

[3] Kohonen, T., *Self-Organization and Associative Memory*, Springer-Verlag, second ed, 1989.

[4] Kohonen, T., *Self-Organizing Maps*, Springer, 1995.

[5] Fairbank, J. F. & Labianca, G., "Picking the Perfect Peers," BizEd, Novmber/December 2003, pp. 36-41.

[6] Lewis, M., "Designing for Human-Agent Interaction," AI Magazine, 1998.