The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1467-8047.htm

Received 18 July 2012
Accepted 19 March 2013

A comparison of academic libraries: an analysis using a self-organizing map

Damien Ennis
Department of Mathematics, Truckee Meadows Community College, Reno, Nevada, USA
Ann Medaille
Mathewson-IGT Knowledge Center, University of Nevada, Reno, Reno, Nevada, USA
Theodore Lambert
Department of Mathematics, Truckee Meadows Community College, Reno, Nevada, USA, and
Richard Kelley and Frederick C. Harris Jr
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, Nevada, USA

Abstract
Purpose – This paper aims to analyze the relationship among measures of resource and service usage and other features of academic libraries in the USA and Canada.
Design/methodology/approach – Through the use of a self-organizing map, academic library data were clustered and visualized. Analysis of the library data was conducted through the computation of a "library performance metric" that was applied to the resulting map.
Findings – Two areas of high-performing academic libraries emerged on the map. One area included libraries with large numbers of resources, while another area included libraries that had low resources but gave greater numbers of presentations to groups, offered greater numbers of public service hours, and had greater numbers of staffed service points.
Research limitations/implications – The metrics chosen as a measure of library performance offer only a partial picture of how libraries are being used. Future research might involve the use of a self-organizing map to cluster library data within certain parameters and the identification of high-performing libraries within these clusters.
Practical implications – This study suggests that libraries can improve their performance not only by acquiring greater resources but also by putting greater emphasis on the services that they provide to their users.
Originality/value – This paper demonstrates how a self-organizing map can be used in the analysis of large data sets to facilitate library comparisons.
Keywords Self-organizing map, Machine learning, Neural network, Academic libraries, Cluster analysis, Performance
Paper type Research paper

Introduction
Data comparisons among libraries can provide valuable information for making choices regarding resource allocations and service provisions. In its *Standards for Libraries in Higher Education*, the Association of College and Research Libraries (ACRL) (2011b) recommends that academic libraries use external comparisons with their peers for benchmarking purposes in order to identify strengths and weaknesses

Performance Measurement and Metrics
Vol. 14 No. 2, 2013
pp. 118-131

1467-8047
DOI 10.1108/PMM-07-2012-0026

and "to develop a more informed picture of institutional standing within the higher education marketplace" (p. 22). In "Determining quality in academic libraries," Pritchard (1996) writes that "The ability to make unambiguous and meaningful comparisons is an important issue in assessment [...]" (p. 584), and in a discussion of library criteria, Knightly (1979) includes comparison with other organizations as one of seven types of measurement, noting that comparisons can reveal both strengths and areas in need of improvement. While library comparison data needs to be understood within the context of individual library environments, it can inform decision making when used in combination with other data.

Cluster analysis is one method that can be used to compare library data. Lorr (1983) defines clustering as "the grouping of entities into subsets on the basis of their similarity across of set of attributes" (p. 11), and cluster analysis can be especially useful for revealing patterns and relationships within complex data sets. In the field of library and information science, cluster analysis has been used to study term indexing, web searching, journal citations, and user behavior. Cluster analysis has been less frequently used to study libraries as a whole; however, in a study of the multiple dimensions that comprise academic library effectiveness, Micikas and McDonald (1994) used cluster analysis to distinguish among five different library groups. Of these groups, the authors identified a cluster of "highly effective" libraries that were located at institutions that had good financial support, limited enrollments, specialized curricula, and a large ratio of books per student (Micikas and McDonald 1994, pp. 76-78).

One clustering technique that has not yet been used to evaluate academic library data is a machine-learning algorithm called the self-organizing map (SOM). Widely employed in the field of computer science, an SOM is used for data analysis, classification, and visualization, and is frequently used as a means of deriving information from large data sets. SOMs have at times outperformed other data analysis tools because of the large amounts of data that they can process simultaneously (Hamel and Sun, 2005), and the information produced by SOMs has been found to be reliable over time (du Jardin and Séverin, 2011). The use of an SOM makes it possible to detect and visualize relationships among complex data that would be difficult to discern without the machine-learning algorithm.

This study used an SOM to identify data points that could be correlated with high resource and service usage in academic libraries. To choose the metrics for analysis, a number of different studies were consulted that describe the value of measures such as circulation, attendance, weekly public service hours, building and e-resource usage, reference transactions, and attendance at instruction sessions, among others (Dugan *et al.*, 2009; Whitmore, 2002; Van House *et al.*, 1990; Poll, 2003; Weiner, 2005; Association of College and Research Libraries, 2010). When correlated with other data, these metrics have provided valuable insights. For example, in analyzing the connection between traditional and newer measures of academic libraries, Weiner (2005) found that a significant relationship existed among service metrics (numbers of reference transactions and instructional presentations, and attendance at instructional presentations) and more traditional library measures such as budget, staff, and clientele. Whitmore (2002) found that at certain types of institutions a positive relationship existed between library resources and students' gains in critical thinking, but a negative relationship existed between library services and undergraduates' library use. Emmons and Wilkinson (2011) looked for correlations among academic library measures of staff, collection, circulation, and services (number of reference questions and percent of students receiving instruction), and institutional measures of retention and graduation. They found that a significant relationship existed between library staff and both retention and graduation rates. Because of the prevalence of service and usage factors in recent library correlation studies, several of these oft-cited metrics were selected for analysis using an SOM.

The current study sought to answer the following question: can an SOM cluster analysis of complex academic library data be used to reveal meaningful relationships among resource and service measures and other library factors – relationships that might otherwise be overlooked? To answer this question, this study looked at three commonly reported measures of resource and service usage (circulation, attendance at instruction sessions, and reference transactions) and used an SOM cluster analysis to determine whether correlations could be found with features related to library expenditures, personnel, materials, and service offerings – data that is consistently tracked by most academic libraries. First, an SOM was used to cluster library data. Then the output was analyzed for the purpose of: seeing which libraries clustered together on the basis of their combined features; determining which clusters could be identified as "high-performing"; and identifying the distinguishing characteristics of the high-performing library clusters. It is important to note that while an SOM can be used to cluster data and facilitate the discovery of correlations among the data, it cannot provide an explanation as to why those correlations exist. However, an SOM mapping and cluster analysis can provide a useful starting place for more detailed evaluations.

In the remainder of this paper, the authors will first provide an overview of SOMs and a description of the SOM design. Next, the SOM that was used in this study will be described in relation to the library features that were selected for analysis and the metrics that were used in the examination of the map output. Finally, the results and implications of the library map analysis will be explained and discussed.

The SOM

Applications of an SOM

Machine learning is an area of artificial intelligence in which the algorithms that are used improve their performance with feedback. Machine learning techniques are commonly used to automate processes that would ordinarily be difficult or time-consuming for humans. Some machine learning algorithms are trained on known data and then applied to unknown data; other algorithms adjust themselves in an unsupervised fashion. One specific type of machine learning technique is a neural network, a computer program that provides an algorithmic modeling of the neurons of a biological organism.

Developed by Teuvo Kohonen (2001), an SOM is a type of neural network that is often used for problems that involve detecting correlations, reducing dimensionality, finding hidden patterns, and classifying data. An SOM synthesizes data that are comprised of many variables and produces as its output a simplified view of the data that consists of a regular grid or lattice of nodes. The strength of an SOM is that input data items with similar features are placed in close proximity on the grid. Often, this proximity on the grid elucidates an interesting property of the data set, so an SOM allows one to see at a glance which data items are related in a way that otherwise would not be easily seen. In this paper, the relationships that are of primary interest are the features of academic libraries, although in other applications one could easily choose other variables, such as health patient data or specific features of a collection of web pages. Because an SOM has no domain specificity, it is readily applicable to large data sets, including the library data examined in this paper.

Unlike some other types of neural networks, an SOM is an unsupervised, competitive learning algorithm. As an unsupervised algorithm, an SOM contains no preexisting information in regard to "correct" outputs; instead, the algorithm is used to find similarities between different inputs, and these similarities are visualized through the resulting map.

SOMs have been used in a number of different fields, including biology, medicine, education, and business. In the field of molecular biology, SOMs have been used to analyze molecular structures for the purpose of identifying similarities and differences among protein homologs (e.g. Hamel and Sun, 2005), of comparing newly discovered sequences with existing sequences (e.g. Ahmad *et al.*, 2008), and of classifying DNA sequences (e.g. Naenna *et al.*, 2003). SOMs have also been used to classify patients with diseases. For example, Chen *et al.* (2008) used an SOM to classify and visualize common characteristics in lung cancer patients for the purpose of predicting risk of disease, while Astel *et al.* (2010) used an SOM to analyze the relationships between clinical tests and disease symptoms. In the field of education, SOMs have been used to classify types of e-learners in order to recommend appropriate online courses (Tai *et al.*, 2008) and to cluster schools offering a Master of Business Administration degree in order to help students make decisions about which graduate schools to attend (Kiang and Fisher, 2008). In the field of finance, an SOM has been used to improve a model that predicts corporate bankruptcy (du Jardin and Séverin, 2011).

In the field of information science, the clustering ability of SOMs has been used to make web searching more user-oriented and to make the results of web searches more relevant. For example, SOMs have been used to analyze and cluster health subject terms in order to facilitate greater usability of web directories (Zhang *et al.*, 2009; Zhang and An, 2010), to model user profiles in order to create more personalized web searches (Ding and Patra, 2007), and to construct webpage hierarchies in order to create multilingual web directories (Yang *et al.*, 2011). He *et al.* (2003) used an SOM to create a citation-based method of retrieving scholarly publications from the web, while Subramanyam Rallabandi and Sett (2007) used an SOM to construct an image retrieval system on the basis of color, texture, and shape descriptors of images. Petrilis and Halatsis (2007) used an SOM to develop a two-level web site clustering technique that mined both webpage content and context. SOMs have also been used in information science in the analysis of journal content. Linton *et al.* (2009) analyzed the abstracts of papers in management journals and An *et al.* (2011) analyzed the subject terms of papers in library and information science journals to determine which journals had the most similar content for the purpose of aiding both authors and librarians in their selection of literature.

SOM design

An SOM provides a method of mapping data from a higher-dimensional space to a lower-dimensional output space (e.g. mapping data with ten features onto a two-dimensional map). The most commonly used SOM configuration is two-dimensional because it facilitates the easy visualization of information. Unlike supervised neural networks, where neural connections are often segmented into layers, an SOM contains neurons (that is, nodes or connection points) that are connected to their immediate neighbors in a grid pattern. The neurons are traditionally arranged in either a hexagonal or rectangular lattice. Figure 1 shows two input data points which are

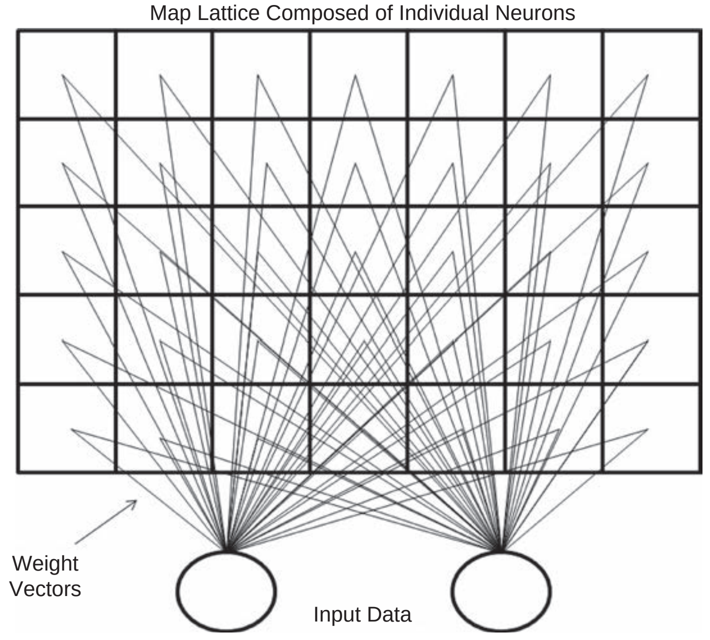


Figure 1.
An illustration showing the dimensionality of the SOM data

mapped to a rectangular lattice, with each square representing a neuron on the map. Each neuron's weight vector (i.e. the collection of numerical data associated with a neuron) is modified by the collected numerical data from the input points. Figure 2 shows how the numerical distance between data points is represented on a two-dimensional map. Input data points that are located close to each other in the input space are mapped to nearby neurons on the output map.

Through a weight vector, each neuron in an SOM is connected to the input layer. The dimensionality of the weight vector matches the number of features of the input data. Each neuron's weights are set to an initial random value within a designated range. The n-dimensional weight vector is defined as follows:

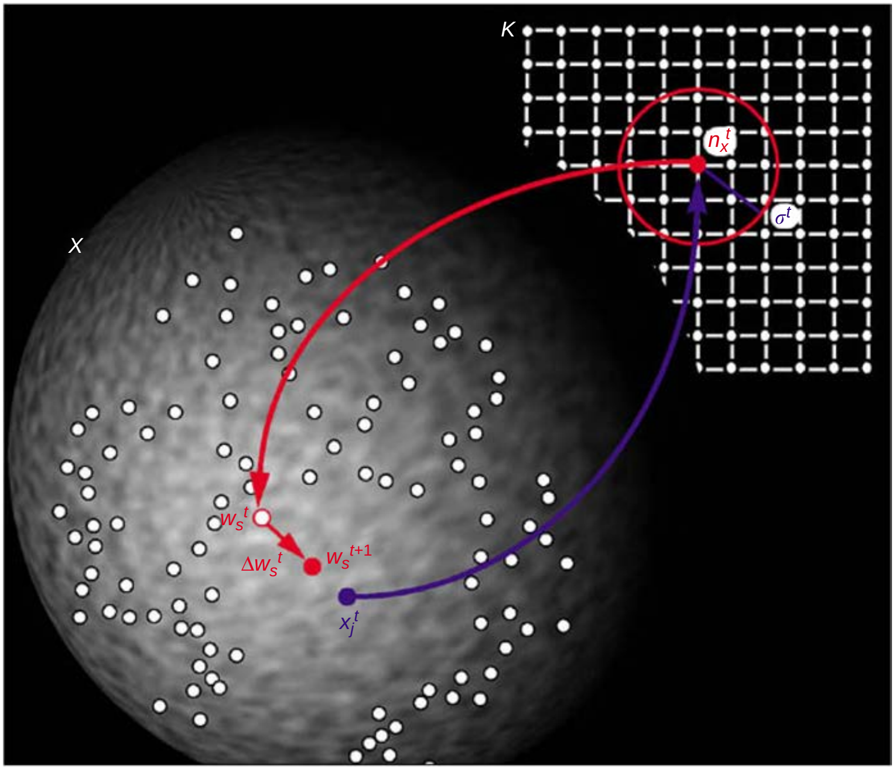
$$w_i = [w_1, w_2, \ldots, w_n]$$

The procedures required to apply an SOM can be divided into three parts: data gathering and normalization; training an SOM using the data; and extracting information from the trained SOM.

The training process involves making adjustments to the map on the basis of the input provided, with the final result being a map that places similar input points at adjacent locations. An SOM is trained in an iterative process, and, for each iteration, an input vector is selected at random from the training data set and presented to the network. A winning neuron is selected whose weight vector most closely matches the input vector. Euclidian vector distance is often used to choose the best match. This is defined as:

$$\text{winning neuron } c = \arg\ \min \|x - w_i\|$$

After the winning neuron has been selected, the weight vectors of the winning neuron and the neurons that reside in its neighborhood are adjusted to more closely match the



Source: From Bachelier (2003), available through a creative commons license

Figure 2.
An SOM schematic showing the relationship between input data and output neurons

current input. The closer a neuron is to the winning neuron, the larger the change to its current weight values. This is represented by the following equation:

$$w_i(t+1) = w_i(t) + \alpha(t)h_{ci}(t)[x(t) - w_i(t)]$$

where

- $x(t)$ is the input vector randomly drawn from the input set at time t;
- $\alpha(t)$ is the learning rate function; and
- $h_{ci}(t)$ is the neighborhood function centered on the winning neuron at time t.

Both the learning rate (i.e. the size of the change to the weight vectors) and the neighborhood radius decrease over the number of iterations. A Gaussian function is often used to represent this decrease. The Gaussian neighborhood function is defined as follows:

$$h_{ci}(t) = \exp\left[\frac{-\|r_c - r_i\|^2}{2\sigma(t)^2}\right]$$

where

- $\sigma(t)$ is the width of the Gaussian kernel; and
- $-||r_c - r_i||^2$ is the distance between the winning neuron c and the neuron i with r_c and r_i representing the two-dimensional positions of neurons c and i on the SOM grid.

A new vector is then chosen from the training data, and the steps above are repeated until changes to the weight values in a given iteration are below a specific threshold.

Methodology

Library data were collected from US and Canadian academic libraries that are members of the ACRL by using the ACRL*Metrics* data portal. Fifteen library features for the fiscal year 2010 were selected for analysis in an SOM and are listed with their definitions in Table I. These features consist of a combination of resources such as collections, staffing, and expenditures, and activities such as giving presentations to groups and staffing service desks. These particular features were selected because they were consistently reported among the majority of libraries in the data portal and they covered a broad range of library features.

Next, an SOM was constructed according to the specifications described in the previous section. Prior to running the data through the SOM, the data were examined for outliers that suggested mistakes or inaccuracies. These outliers were removed, leaving data from a total of 1,395 libraries. The data were then normalized according to the number of full-time equivalent (FTE) students enrolled at the institution. Finally, the data were run through the SOM and fed to a 44 × 44 neuron map, a size which was

Feature	Definition (derived from Association of College and Research Libraries (ACRL), 2011a)
Volumes in library	Total number of physical units that have been cataloged, classified, and made ready for use
Total serials	Total number of unique serial titles
Monographs	Annual monograph expenditures
Current serials	Annual serial expenditures
Other library materials	Annual expenditures on items other than monographs and serials such as backfiles of serials, charts and maps, audiovisual materials, and manuscripts
Miscellaneous expenditures	Annual expenditures on items other than library materials such as expenditures for bibliographic utilities, literature searching, and security devices
Salaries and wages of professional staff	All salaries and wages of professional staff, excluding fringe benefits
Salaries and wages of support staff	All salaries and wages of support staff, excluding fringe benefits
Salaries and wages of student assistants	All student wages, regardless of budgetary sources of funds
Professional staff	Number of FTE staff that the library considers professional, such as librarians, computer experts, systems analysts, and/or budget officers
Support staff	Number of FTE staff that are not included in the count of professional staff, excluding maintenance and custodial staff
Student assistants	Number of FTE student assistants employed by the library
Staffed service points	Number of staffed public service points in main and branch libraries
Weekly public service hours	Total hours that the library is open per typical full-service week
Presentations to groups	Total number of presentations made as part of bibliographic instruction programs and through other planned class presentations, orientation sessions, and tours

Table I.
Library features used in the SOM

selected because it was large enough to accommodate the data and resulted in a graphic that facilitated the analysis.

The resulting SOM visualization was then analyzed for common features among the clusters. The analysis was conducted by computing a "library performance metric" (LPM) that was based on three features that represent usage of the library: total number of reference transactions, total participants in group presentations, and total circulation transactions (Table II).

The three features that comprise this metric were selected because they have consistently been employed to measure library usage in a number of studies (e.g. Weiner, 2005; Whitmore, 2002; Emmons and Wilkinson, 2011, etc.). While other usage metrics – such as the number of database logins, web site visits, or full-text paper downloads – would have provided both interesting and valuable information as well, these data were either tracked too inconsistently or were not provided by a significant number of libraries. The three LPM values were combined so that they had equal weight in the final score and were normalized according to the number of students (FTE) at each institution.

Results and analysis

The library SOM is displayed in Figure 3. Data were clustered according to the features listed in Table I, and each library is represented numerically by its LPM. Figure 3 shows how the libraries clustered according to the 15 features, with libraries with similar characteristics grouping together at different locations on the map. Once the libraries had clustered, the high-LPM libraries were identified through the application of the LPM label, and the features of the high-LPM libraries were then analyzed to determine what led to their locations on the map. Thus, the LPM was used only as a label on the visualized map and had no bearing on the actual cluster position within the map.

Analysis of the SOM revealed a number of different characteristics. One characteristic of the map is that many resource variables (i.e. greater expenditures, greater numbers of materials, and higher numbers of staff; calculated as a ratio to student FTE) increase when descending on the map. Thus, those libraries with greater material and staff expenditures and greater numbers of staff tended to collect along the bottom of the map, while those libraries with lower resources tended to collect along the top part of the map. Moving horizontally on the map revealed variations in terms of other features, including number of presentations to groups and number of public service hours. Those libraries that offered greater numbers of presentations and public service hours per student FTE generally collected on the right side of the map, while those with lower numbers collected on the left.

Metric	Definition (derived from ACRL, 2011a)
Reference transactions	Total number of reference transactions, both in person and through virtual means
Participants in group presentations	Total number of participants in presentations made as part of bibliographic instruction programs and through other planned class presentations, orientation sessions, and tours
Circulation transactions	Total number of items lent, including renewals

Table II. Items used to compute the LPM

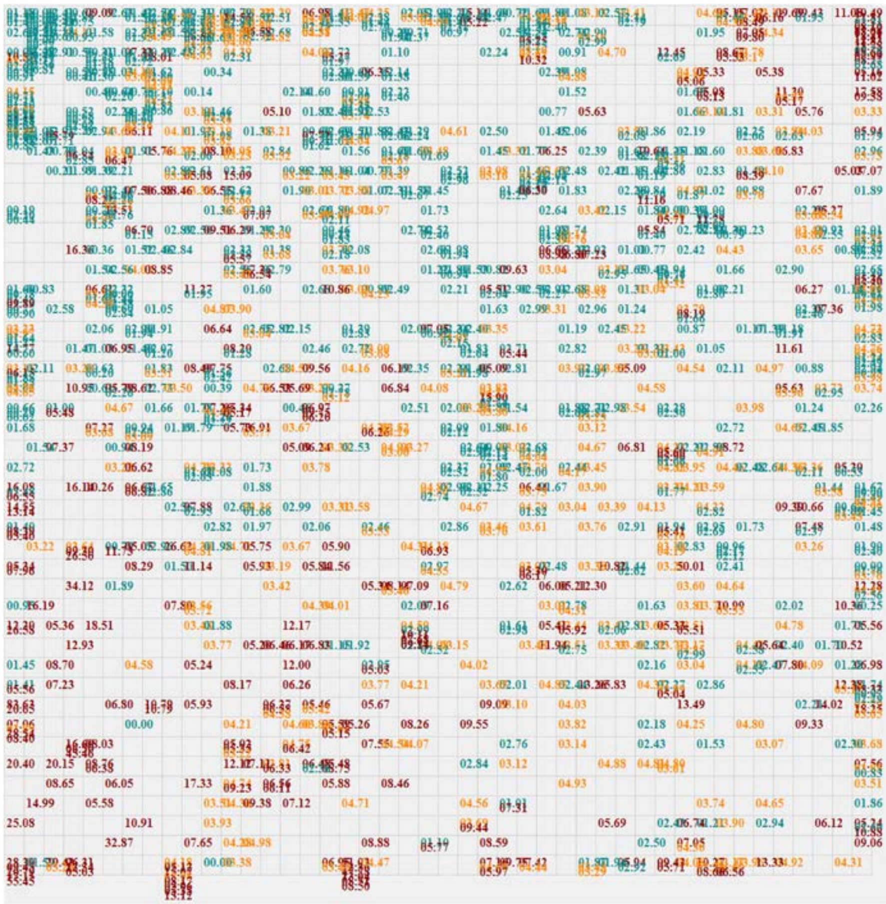


Figure 3.
SOM output with libraries represented by their colored LPM labels

Notes: Red=high, orange=middle, green=low

In Figure 3 the LPM is represented on the SOM by color. Those libraries with a high LPM are colored in red, a medium LPM in orange, and a low LPM in green (in the print version of this paper the high-to-low LPM range is designated from black to light gray). Figure 4 shows the same SOM with the neurons colored according to the average LPM of the libraries assigned to that neuron, with the lowest-scoring libraries in dark green and the highest-scoring libraries in dark red (again, in the print version the high-to-low LPM range is designated from black to light gray). In showing the LPM average of all libraries that were placed at a particular neuron on the map, Figure 4 provides a clearer visualization of the different clusters. In looking at the color locations on the map, it appears that most of the low LPM libraries (green) appear in the upper portion of the map, while a greater number of the middle and high LPM libraries (orange and red, respectively) appear in the lower portion.

Although the diverse features of the libraries resulted in a variety of placements on the map, three areas emerged of particular interest, and these areas are outlined and labeled in Figure 5. First, an area of low-performing libraries appears in the upper left corner of the map. An analysis of their common features reveals that these libraries are

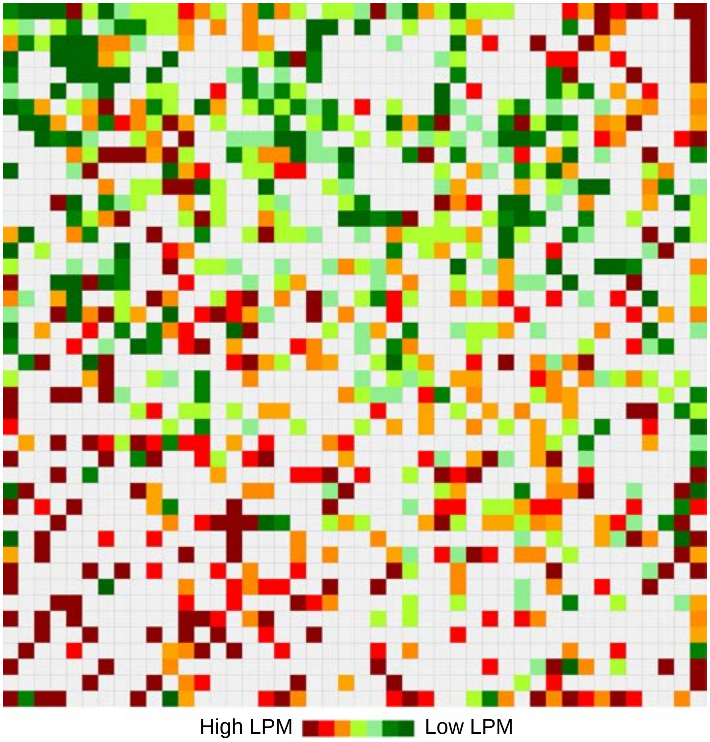


Figure 4.
SOM output with the color dispersion as determined by the average LPM per neuron

generally low in resources such as expenditures and staff; thus, this area has been labeled as lower resource, low performing in Figure 5.

Two high-performing library areas (outlined in Figure 5) emerged in the lower-left and upper-right portions of the map. An analysis of the common features of libraries in the lower-left area reveals that these high performers have greater numbers of resources (i.e. larger budgets, more materials, and higher numbers of staff), and this area has been labeled as higher resource, high performing in Figure 5. This area includes libraries at research universities such as the University of North Carolina, the University of Southern California, and Johns Hopkins University, among others. Thus, for this group of libraries, greater numbers of resources per student FTE can be correlated with better library performance, as measured by the LPM.

The third area of interest in the upper-right area of the map also contains several high-performing libraries (outlined in Figure 5). Unlike the high-resource area in the lower-left, an analysis of the common features of this high-performance area in the upper-right reveals that these libraries are generally low in resources (i.e. smaller budgets, fewer materials, and lower numbers of staff). This area includes libraries at institutions such as Florida Keys Community College and Tougaloo College and has been labeled as lower resource, high performing in Figure 5. An analysis of the features of the libraries in this area reveals that these libraries give greater numbers of presentations to groups, offer a greater number of public service hours, and have greater numbers of staffed service points per student FTE. Thus, it would appear that their high LPM scores can be correlated to these higher service features rather than to greater numbers of resources.

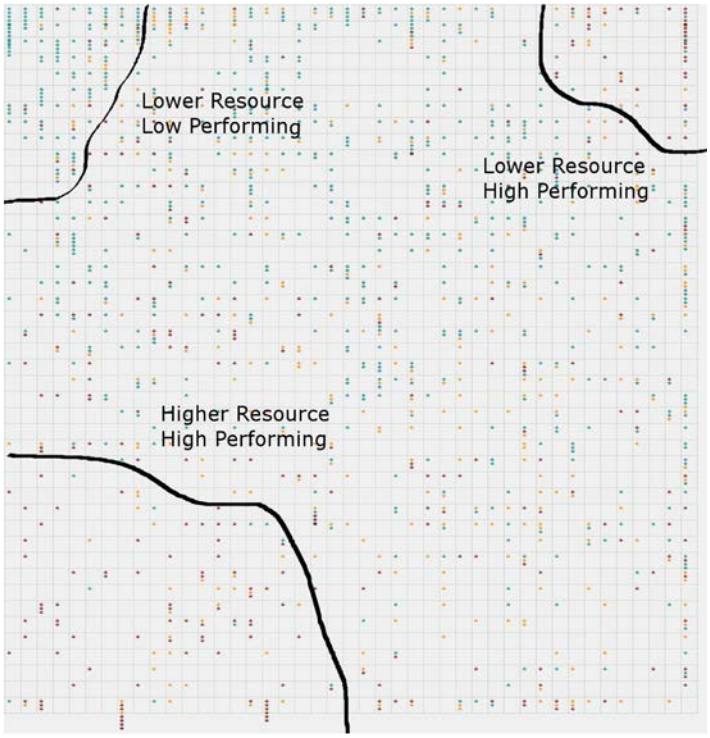


Figure 5.
SOM output showing the three library areas

It is significant, although perhaps unsurprising, that libraries with greater numbers of resources (in the higher resource, high performing area) achieve higher levels of performance, as measured by the LPM. However, the implications of those libraries in the lower resource, high performing area are also significant. In offering greater numbers of these services, the lower resource, high performing libraries may be achieving higher levels of performance by a means other than through greater numbers of resources. It may be, for example, that by offering greater numbers of group presentations, these libraries are educating their users about library resources, which in turn leads to greater numbers of circulation transactions, or they are educating their users about the assistance available in the library, which in turn leads to greater numbers of reference transactions. In addition, by being open for longer hours, libraries are making their collections more readily available for checkout.

Conclusion and suggestions for further study

Data analysis can be a difficult and time-consuming process for humans, but machine-learning techniques automate much of the analysis. The machine learning technique used in this paper, the SOM, provides considerable value in its ability to cluster large library data sets and to facilitate comparisons among libraries. In this study, an SOM has been used to cluster library features such as resources, staff, and expenditures, but it could easily be used to facilitate the analysis of other types of library data as well. While a cluster analysis technique such as the SOM used in this study cannot provide a holistic picture of the state of a single library, it can be used to help one to quickly ascertain data similarities among several features at large numbers of libraries.

In this paper, the SOM used in combination with the LPM labeling has elucidated relationships among certain library features with regard to performance, but it cannot explain why those correlations exist. This is just the starting point for further evaluation. Future research could delve into the demographic characteristics and other features of the high-performing libraries in order to arrive at a more detailed picture of how these libraries achieve high usage rates.

Certainly, the metrics chosen for the LPM used in this paper offer only a partial picture of how libraries are being used. For example, total circulation transactions represent only some use of library collections and do not represent factors such as in-library usage of material, database usage, or paper downloads. In addition, the 15 library features used in this SOM were selected for visualization because of their recurrence as typical points of library comparison, but other library features would provide for a fruitful comparison as well. In addition, a different LPM could be constructed from different outputs, such as building usage or e-resource usage.

For this study, the authors chose to normalize the data on the basis of student FTE, but this choice does not account for all of the factors that could influence the results, such as institutional type, budget, sources of funding, consortial agreements, etc. In this study, the authors felt that the greatest insights would come from using the largest set of library data possible because they were trying to find relationships that were not easily discernible by obvious or pre-existing classifications. However, future work might involve the identification of library groups that cluster within certain parameters, such as size (e.g. small, medium, and large libraries), budget, funding support, institutional mission, etc. In fact, this approach to cluster analysis could aid in the identification of peer groups. Academic libraries typically turn to their institutional administrative offices for an official list of peer institutions (ACRL, 2011b), but libraries on these official lists can differ greatly. Pritchard (1996) writes that "It is possible that the peer institutions used by the administration for strategic planning will not each have a library that functions comparably" (p. 584). An SOM can be used to cluster similar libraries and identify library peer groups; libraries could then look to the high performers within the same cluster to see what kinds of choices these libraries are making.

Although the SOM analysis provided here cannot answer all of the questions prompted by the clustering of the high performers, the analysis does have implications for libraries seeking to improve their performance. In times of decreasing budgets, libraries that want to improve their outputs may be unable to achieve this result through increasing their resource expenditures or hiring more staff. This study suggests that libraries may be able to improve their performance by putting a greater emphasis on the services they offer to their users, whether it be through providing greater numbers of staffed service points and public service hours or providing greater numbers of instruction sessions and outreach opportunities. Future research could analyze the common features among these libraries that are not apparent in the data points examined in this paper.

References

Ahmad, N., Alahakoon, D. and Chau, R. (2008), "Classification of protein sequences using the growing self-organizing map", *4th International Conference on Information and Automation for Sustainability in Colombo*, pp. 167-172.

An, L., Zhang, J. and Yu, C. (2011), "The visual subject analysis of library and information science journals with self-organizing map", *Knowledge Organization*, Vol. 38 No. 4, pp. 299-319.

Association of College and Research Libraries (2010), *Value of Academic Libraries: A Comprehensive Research Review and Report*, researched by Megan Oakleaf Association of College and Research Libraries, Chicago, IL.

Association of College and Research Libraries (2011a), "ACRL statistics questionnaire 2010-2011 instructions", ACRL*Metrics,* available at: www.acrlmetrics.com (accessed June 29, 2012).

Association of College and Research Libraries (2011b), *Standards for Libraries in Higher Education*, Association of College and Research Libraries, Chicago, IL.

Astel, A., Nikolov, M., Christov, V., Simeonova, P. and Simeonov, V. (2010), "Chemometric assessment of clinical data for diabetes mellitus 2 type patients using self-organizing maps", *Journal of Environmental Science and Health. Part A, Toxic/Hazardous Substances & Environmental Engineering*, Vol. 45 No. 5, pp. 560-568.

Bachelier, G. (2003), "SOM_Abb_1.jpg", Wikimedia Commons, available at: http://commons.wikimedia.org/wiki/File:SOM_Abb_1.jpg (accessed June 29, 2012).

Chen, S., Zhou, S., Yin, F.F., Marks, L.B. and Das, S.K. (2008), "Using patient data similarities to predict radiation pneumonitis via a self-organizing map", *Physics in Medicine and Biology*, Vol. 53 No. 1, pp. 203-216.

Ding, C. and Patra, J.C. (2007), "User modeling for personalized web search with self-organizing map", *Journal of the American Society for Information Science and Technology*, Vol. 58 No. 4, pp. 494-507.

Dugan, R.E., Hernon, P. and Nitecki, D.A. (2009), *Viewing Library Metrics from Different Perspectives: Inputs, Outputs, and Outcomes*, Libraries Unlimited, Santa Barbara, CA.

du Jardin, P. and Séverin, E. (2011), "Predicting corporate bankruptcy using a self-organizing map: an empirical study to improve the forecasting horizon of a financial failure model", *Decision Support Systems*, Vol. 51 No. 3, pp. 701-711.

Emmons, M. and Wilkinson, F.C. (2011), "The academic library impact on student persistence", *College & Research Libraries*, Vol. 72 No. 2, pp. 128-149.

Hamel, L. and Sun, G. (2005), "Toward protein structure analysis with self-organizing maps", *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1-8.

He, Y., Hui, S.C. and Fong, A.C.M. (2003), "Citation-based retrieval for scholarly publications", *IEEE Intelligent Systems*, Vol. 18 No. 2, pp. 58-65.

Kiang, M.Y. and Fisher, D.M. (2008), "Selecting the right MBA schools – an application of self-organizing map networks", *Expert Systems with Applications*, Vol. 35 No. 3, pp. 946-955.

Knightly, J.J. (1979), "Overcoming the criterion problem in the evaluation of library performance", *Special Libraries*, Vol. 70 No. 4, pp. 173-177.

Kohonen, T. (2001), *Self-Organizing Maps*, 3rd ed., Springer, Berlin.

Linton, J.D., Himel, M. and Embrechts, M.J. (2009), "Mapping the structure of research: business and management as an exemplar", *Serials Review*, Vol. 35 No. 4, pp. 218-227.

Lorr, M. (1983), *Cluster Analysis for Social Scientists*, Jossey-Bass Publishers, San Francisco, CA.

Micikas, L.B. and McDonald, J.A. (1994), *Academic Libraries: The Dimensions of Their Effectiveness*, Greenwood Press, Westport, CT.

Naenna, T., Bress, R.A. and Embrechts, M.J. (2003), "DNA classifications with self-organizing maps (SOMs)", *Proceedings of the IEEE International Workshop on Soft Computing in Industrial Applications in Binghamton, New York, NY*, pp. 151-154.

Petrilis, D. and Halatsis, C. (2007), "Two-level clustering of web sites using self-organizing maps", *Neural Processing Letters*, Vol. 27 No. 1, pp. 85-95.

Poll, R. (2003), "Measuring impact and outcome of libraries", *Performance Measurement and Metrics*, Vol. 4 No. 1, pp. 5-12.

Pritchard, S.M. (1996), "Determining quality in academic libraries", *Library Trends*, Vol. 44 No. 3, pp. 572-594.

Subramanyam Rallabandi, V.P. and Sett, S.K. (2007), "Image retrieval system using R-tree self-organizing map", *Data & Knowledge Engineering*, Vol. 61 No. 3, pp. 524-539.

Tai, D.W.-S., Wu, H.-J. and Li, P.-H. (2008), "Effective e-learning recommendation system based on self-organizing maps and association mining", *The Electronic Library*, Vol. 26 No. 3, pp. 329-344.

Van House, N.A., Weil, B.T. and McClure, C.R. (1990), *Measuring Academic Library Performance: A Practice Approach*, American Library Association, Chicago, IL.

Weiner, S.A. (2005), "Library quality and impact: is there a relationship between new measures and traditional measures?", *The Journal of Academic Librarianship*, Vol. 31 No. 5, pp. 432-437.

Whitmore, E. (2002), "Academic library performance measures and undergraduates' library use and educational outcomes", *Library & Information Science Research*, Vol. 24 No. 2, pp. 107-128.

Yang, H.-C., Hsiao, H.-W. and Lee, C.-H. (2011), "Multilingual document mining and navigation using self-organizing maps", *Information Processing & Management*, Vol. 47 No. 5, pp. 647-666.

Zhang, J. and An, L. (2010), "Visual component plane analysis for the medical subjects based on a transaction log", *The Canadian Journal of Information and Library Science*, Vol. 34 No. 1, pp. 83-111.

Zhang, J., Tang, L.A., Tang, T. and Hong, Y. (2009), "Visual health subject directory analysis based on users' traversal activities", Vol. 60 No. 10, pp. 1977-1994.

Corresponding author
Ann Medaille can be contacted at: amedaille@unr.edu