Workshop on Artificial Intelligence

The application of an Extended Self-Organizing Map Networks to Market Segmentation

M. Y. Kiang [*]
Information Systems Department
College of Business Administration
California State University at Long Beach
Long Beach, CA 90840
Tel: (562)985-8944
Fax: (562)985-4080
email: mkiang@csulb.edu

A. Kumar
Department of Marketing
College of Business
Arizona State University
Tel:(480)-965-5544
Fax:(480)-965-8000
email:ajith.kumar@asu.edu

R.T. Chi
Information Systems Department
College of Business Administration
California State University at Long Beach
Long Beach, CA 90840
Tel: (562)985-4238
Fax: (562)985-4080
email: rchi@csulb.edu

[*] Corresponding author

The application of an Extended Self-Organizing Map Networks to Market Segmentation

Abstract

This paper presents the technique of an extended SOM networks and shows how it can be applied as a data analysis tool to market segmentation. The Self-Organizing Map (SOM) network, an unsupervised learning neural network, is a categorization network developed by Kohonen. In this research, we implemented an extended version of SOM networks that further groups SOM output map into user specified number of clusters. We then compared the extended SOM network and K-means analysis, a popular clustering technique, in the context of market segmentation problems. The a priori motivation for considering a neural networks alternative is that SOM networks, like most of the neural network models, do not assume the multivariate normality of data and hence, may be more robust. Test results indicate that the extended SOM networks perform better than K-means analysis when the data are skewed.

Keywords: Artificial Intelligence Application, SOM Neural Network, Clustering Analysis, Market Segmentation

1. Introduction

Market segmentation refers to “the subdividing of a market into distinct subsets of customers where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix [Kotler 1980]. The reason for segmenting market is that consumers are often numerous, geographically dispersed, and heterogeneous, and therefore seek varying benefits from the products they buy. Consumers within a segment are expected to have homogeneous buying preferences whereas those in different segments tend to behave differently. By properly identifying the benefit segment of a firm’s product, the marketing manager can target the sales effort at specific groups of consumers rather than at the total population. The identification of consumer segments is of critical importance for key strategic issues in marketing involving the assessment of a firm’s opportunities and threats. The marketing manager must evaluate the potential of the firm’s products in the target segment and ultimately select the most promising strategy for the segment. The value of performing market segmentation analysis includes: to acquire a better understanding of the market in order to properly position its products on the marketplace; to identify the correct benefit segment when introducing new products; to find new opportunities for existing products; and to increase competitive advantage through product differentiation. Although it was introduced into the academic marketing literature in the fifties, market segmentation continues to be an important focal point of ongoing research and marketing practices (e.g., Chaturvedi et al. 1997). Most of the academic research in market segmentation has been in the development of new techniques and methodologies for segmenting markets. The common thread running through these diverse streams of research is the attempt to classify consumers, deterministically or probabilistically, into a finite number of groups such that consumers within a group/segment are relatively homogeneous with respect to the variables used in the classification.

One compelling reason to explore alternative approaches in market segmentation is the possibility of conflicting of assumption when using statistical methods. Segmentation through K-means analysis invokes assumption of multivariate normality of data whereas the SOM framework does not require the assumption of homogeneity, primarily because it is not based upon any statistical model. In commercial market research studies, the data (especially those involving product ratings, customer satisfaction assessments) tend to be markedly skewed [Sharma et al., 1989], clearly suggesting nonnormality. A study by Deakin [1976] suggests that financial ratios are not normally distributed; rather, they are positively skewed. Violations of the normality assumptions may lead to a biased and overly optimistic prediction of the performance of rules in the population, and thus limit the usefulness of the model.

The Self-Organizing Map (SOM) network, a variation of neural computing networks, is a categorization network developed by Kohonen [1984, 1989, 1995]. The main function of SOM networks is dimension reduction that maps the input data from an n-dimensional space to a lower dimensional (usually one or two-dimensional) plot while maintaining the original topological relations.

While Kohonen's Self-Organizing networks have been successfully applied as a classification tool to various problem domains, including speech recognition [Zhao & Rowden 1992; Leinonen et al., 1993], image data compression [Manikopoulos 1993], image or character recognition [Bimbo et al., 1993; Sabourin & Mitiche 1993], robot control [Walter & Schulen 1993; Ritter et al., 1989], and medical diagnosis [Vercauteren et al., 1990], its potential as a robust substitute for clustering tool remains relatively unresearched.

Balakrishnan et al. [1994] compared several unsupervised neural networks with K-means analysis. Due to the lack of extended grouping function such as the one implemented in our extended SOM network, the two-layer Kohonen network implemented in their study was designed so that the number of nodes in the output layer (Kohonen layer) corresponds to the number of desired clusters. This is a different kind of Kohonen network that does not provide a two-dimensional map that allows users to visualize the relationships among data points. Most of the studies we have found that apply Kohonen network to clustering have implemented this type of network. The performance of these neural networks was examined with respect to changes in the number of attributes, the number of clusters, and the amount of error in the data. Their result shows that the K-means procedure always outperformed the neural networks especially when the number of clusters increased from two to five. Unlike Balakrishnan's compromising approach, our extended SOM method preserves the dimension reduction function of the original SOM and further groups the output map into the number of clusters specified by the user.

The balance of the paper is organized as follows: Section 2 presents the basic concepts of SOM network and illustrates its use as a data-reduction tool. This is followed by a discussion of the extended grouping capability integrated to the original SOM networks. Section 3 describes the experimental procedure that was used to generate the data sets and determine the network configurations. In Section 4 we compare the performance of the extended SOM and K-means analysis using the simulated data sets. The paper concludes with a summary of our findings.

2. Self-Organizing Map (SOM) Networks

In this research, we introduce a neural networks based method, the self-organizing map (SOM) network, to identify and represent the relationship among various consumer preferences and market segments. SOM can learn from complex, multi-dimensional data and transform them into a map of a fewer dimensions, such as a 2-dimensional plot. The 2-dimensional plot provides an easy-to-use graphical user interface to help the decision-maker visualize the similarities between consumer preference patterns. For example, suppose there are ten measures of customer attributes that are to be used to segment a market. It would be difficult to visually classify individuals based on all these attributes because the grouping must be done in a 10 dimensional space. By using the information contained in the 10-variable set but mapping the information into a 2-dimensional space, one can visually combine customers with similar attributes. These relationships can then be translated into an appropriate type of structure that genuinely represents the underlying relationships between market segments. The method can be used as the basis for building a decision support system for marketing management.

Unlike most neural networks applications, the SOM performs unsupervised training, that is, during the learning (training) process the processing units in the networks adjust their weights primarily based on the lateral feedback connections. The nodes in the network converge to form clusters to represent groups of entities with similar properties. A two-dimensional map of the input data is created in such a way that the orders of the interrelationships among objects are preserved [Kohonen 1989]. The number and composition of clusters can be visually determined based on the output distribution generated by the training process. Teuve Kohonen developed the SOM network between 1979 and 1982 based on the earlier work of Willshaw and von der Malsburg. The SOM network performs unsupervised learning; its primary use is for tasks such as clustering, pattern recognition, and various optimization problems. It is designed to capture topologies and hierarchical structures of higher dimensional input spaces.

The SOM network typically has two layers of nodes, the input layer and the Kohonen layer. The input layer is fully connected to the Kohonen layer that is also the output layer. The Kohonen layer, the core of the SOM network, functions similar to biological systems in that it can compress the representation of sparse data and spread out dense data using usually a one or two-dimensional map. This is done by assigning different sub-areas of the Kohonen layer to the different categories of information, therefore the location of the processing element in a network becomes specific to a certain characteristic feature in the set of input data. The resulting network resembles the tree structure that can be derived by conventional clustering method.

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 by the winner-take-all function. The weight vectors of the winning node and the nodes in the neighborhood are updated using a weight adaptation function based on the following Kohonen rule:

$$\Delta w_i = \alpha (X - w_i^{old}), \text{ for } i \in N_r,$$

Where α is the learning coefficient, X is the input vector, and N_r is the collection of all nodes in the neighborhood of radial distance r. For a two dimensional Kohonen layer, there could be up to a total of eight neighboring nodes when $r = 1$. The process will adjust the weights of the winning node along with its neighbor nodes closer to the value of the input pattern. The neighborhood size (r) can change, and is usually reduced as training progresses.

Previous research has proposed various ways to improve the learning of SOM networks [Cottrell and Fort 1986; Ritter and Schulten 1986; Lo and Bavarian 1991]. In this research, we implemented a Gaussian type neighborhood adaptation function $h(t, r)$, similar to the one used by Mitra and Pal [1994]:

$$h(t,r) = \frac{\alpha(1 - r * f)}{\left[1 + \left(\frac{t}{cdenom}\right)^2\right]}$$

This function decreases in both spatial and time domains. In the spatial domain, its value is the largest when node *i* is the winner node and it gradually decreases with increasing distance from *i*. Parameter α determines the initial value of $|h|$ while the parameter f ($0 < f < 1/r$) determines the rate of decrease of $|h|$ in the spatial domain. In the time domain, t controls the value of $|h|$ whereas the parameter *cdenom* determines the rate of its decay.

The training is conducted in many stages; at each stage, we reduce r by one. Note that r affects the number of nodes in the set N_i. To determine the number of training cycles to be run at each stage, we use the index of disorder D proposed by [Mitra and Pal 1994]. Essentially, D measures the "improvement" in the "state" of the network at discrete time intervals. When this index falls below a certain threshold ($D <$ *convergence coefficient* δ), the next stage of training begins with a reduced r value. The reader may refer to [Mitra and Pal 1994] for the detailed algorithm.

To avoid a few nodes end up representing too much of the input data due to the effect of the initial random weight values assigned to them, we incorporate a “conscience” mechanism, proposed by DeSieno [1988], that prevents the nodes with higher winning frequency from winning repeatedly and makes the nodes with lower winning frequency more likely to win. The purpose of this mechanism is to give each node in the Kohonen layer an opportunity to represent approximately equal information about the input data.

The Extended Clustering Function

Sometimes it is hard to visually group the output from SOM especially when the map is highly populated. Hence, a more scientific approach that can help the user to group the output from SOM network based on certain objective criteria is needed. To automate the segmentation process to complement the usage of the Kohonen SOM networks, Murtagh [1995] proposed an agglomerative contiguity-constrained clustering method. The method groups the output from SOM based on a minimal distance criterion to merge the neighboring nodes together. The rationale is that the SOM networks will maintain the original topological relations; therefore the nodes that are closely located on the representational grid should have similar cluster centers. Murtagh also stated that a minimal variance criterion might be used in place of the minimal distance one. To test our cases, we have implemented both approaches. After a few preliminary runs, we found that the minimal variance criterion we implemented consistently outperformed the minimal distance approach using our sample cases. Hence, we decided to use the minimal variance criterion for our contiguity-constrained clustering method. The criterion we implemented is modified from Murtagh's [1985] and tries to minimize the overall within cluster variance at each step of the process. 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 is derived or when only one cluster remains. Readers should refer to Kiang & Kumar [2001] for a detailed step-by-step procedure to implement the algorithm.

3. The Experimental Design

Because this study is a preliminary investigation of the performance of SOM *versus* K-means in a segmentation context we decided to first focus attention on artificially constructed data. Knowledge of the correct cluster membership for each observation is essential for valid comparisons with respect to the accuracy with which different techniques recover the true cluster structures. We selected a few conditions of variation that would be most insightful in identifying situations in which SOM might be preferable to K-means analysis and vice-versa.

Let the population of interest contain C clusters/segments. We assume that each member of the population belongs exclusively to one of the C segments. The assumption of exclusive (as opposed to probabilistic or fuzzy) segment membership is made in order to be consistent with existing segmentation research practice wherein some type of K-means clustering procedure that yields a hard/crisp partition of the data is frequently used to derive segments. Let π_c denote the probability that a respondent sampled at random from the population belongs to the *c*th segment. Let $\mathbf{x}$ denote the (px1) vector of observations used as input into the factor analysis and $p(\mathbf{x};\boldsymbol{\theta})$ its probability distribution in the population, where $\boldsymbol{\theta}$ denotes a vector containing the parameters of the distribution. In marketing research applications, $\mathbf{x}$ would typically represent ratings on attributes of products that are presumed to drive consumer judgments and/or preferences. Let $p_c(\mathbf{x};\boldsymbol{\theta}_c)$ denote the probability distribution of $\mathbf{x}$ in the *c*th segment (i.e., the conditional distribution of $\mathbf{x}$ given that it comes from the *c*th segment). Then the statistical model for $\mathbf{x}$ in the population can be expressed as

$$p(\mathbf{x};\boldsymbol{\theta}) = \sum_{c=1}^{C} \pi_c\, p_c(\mathbf{x};\boldsymbol{\theta}_c).$$

Random samples of observations from $p(\mathbf{x};\boldsymbol{\theta})$ can be generated in two steps. First, generate an observation from a multinomial distribution with C categories. If the observation belongs to, say, the *k*th category of the multinomial, then generate an observation for $\mathbf{x}$ from $p_k(\mathbf{x};\boldsymbol{\theta}_k)$. This method of sample generation facilitates assessments of accuracy of cluster membership assignments, since the true cluster membership of each sample observation is known. We vary the number of variables (p) between 8 and 12 but use a sample size of 800 for all data sets. The sample size is typical of what would be found in many marketing research studies (between 500 and 1000).

Given the assumptions underlying the application of factor analytic techniques typically used in market segmentation studies and the high levels of skewness commonly found in the marketing data, particularly those obtained in the form of evaluative ratings of brands or products we vary the level of skewness in the sample data. Like other neural networks approaches, an appealing feature of SOM is there is no inherent requirement that the input data be multivariate normal, in contrast to traditional statistical methods.

We included data with both equal and unequal cluster sizes. Prior research suggests that SOM networks tend to perform better when the underlying cluster sizes are approximately equal [Balakrishnan et al., 1994]. On the other hand, unequal clusters tend to arise quite frequently in market segmentation studies. We also generated data with different number of clusters as recommended by Balakrishnan et al. [1994]. 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 11, 15, and 17 in our first set of experiments but fixed the Kohonen layer to an 11x11 network through out the second set of experiments. In summary, we examined eight scenarios in detail. The first set of study contains four scenarios: normal/highly skewed crossed with equal/unequal cluster sizes for two factors, three clusters problems. The second set of study contains the other four scenarios: moderately skewed data with equal/unequal cluster sizes for three factors, four and five clusters problems.

As the first step in this study, we implemented the algorithm in C++. This language was selected for its object-oriented approach and its generality to other object-oriented algorithms.

The code was verified by independently testing specific components and comparing computer-generated results with hand calculations. Each processing unit in the network and each input pattern were implemented using objects. All calculations are performed through message passing between objects in the program. Small networks were used to verify the overall program, again, by comparing computer-generated results with hand calculations. The network training for the experiments was performed on a cluster of IBM RS/6000 mini-computers. Version 6.11 of the SAS Statistical package was used to run the K-means and other related statistics. For K-means analysis, we used Proc FASTCLUS (in SAS), with the default procedure for generating starting solutions. The data sets were generated using SAS (to generate multinomial distributions corresponding to the segments) and PRELIS 2 [Joreskog & Sorbom 1989] (to generate multivariate observations corresponding to an underlying factor model).

4. THE COMPARATIVE STUDY

In market segmentation studies of the type, the accurate assignments of respondents to clusters/segments are critical. Accuracy of respondent assignments to clusters can be quantified in different ways. In this study, we applied the trace criterion to evaluating the outcomes from the two clustering methods. In trace criterion, the cluster membership assignments made by the clustering algorithm were cross-tabulated with the known cluster membership (i.e., the multinomial category for that observation). The rows and columns of this square table were permuted so as to maximize the sum of the joint frequencies in the diagonal cells of the table. This sum, expressed as a percentage of the total sample size, provides a good measure of clustering accuracy.

Table 1 summarizes the results from our first set of experiments based on trace criterion. The results reveal some interesting contrasts. When the data are generated from multivariate normal distributions (linear), the K-means analysis approach performs marginally better than the extended SOM method in recovering the correct segment membership. We also noticed that, in this setting, the extended SOM method yields better results when the underlying segment sizes are approximately equal. When the sample data are skewed, the extended SOM approach clearly outperforms the K-means approach. However, the performance of the extended SOM method is not affected by the cluster sizes while the K-means approach is much worsened when the underlying cluster sizes are not equal.

Table 1. Accuracy of Cluster Recovery for Two Factors, Three Clusters Problems

Data Type	Cluster Sizes	K-means Analysis	Extended SOM
Normal	Equal	86.26	83.8
Normal	Unequal	86.14	82.15
Skewed	Equal	59.38	88.07
Skewed	Unequal	48.51	89.07

*Averaged across network sizes 11, 15, & 17

We then followed up by conducting a second study on four new samples of size 800. The four samples were: (1) 4 clusters with equal segment sizes, (2) 4 clusters with unequal segment sizes, (3) 5 clusters with equal segment sizes, and (4) 5 clusters with unequal segment sizes. Each sample contained twelve variables with a three-factor model as the data generating mechanism. In all four samples the data were generated with moderate levels of skewness, compared to the skewed data generated in the first experimental design.

Table 2 presents the results for cluster recovery. The extended SOM method recovers the true cluster structure more accurately than K-mean analysis across all scenarios. The results are consistent with the results of the first study, given that the data in the four samples were less skewed compared to the skewed data in the first study (i.e., the performance of the traditional approach improves as the data become less skewed).

Table 2. Accuracy of Cluster Recovery for Three Factors, Four & Five Clusters Problems

Number of Clusters	Cluster Sizes	K-means Analysis	Extended SOM
4	Equal	82.89	94.52
4	Unequal	78.01	94.14
5	Equal	85.13	95.01
5	Unequal	83.26	93.27

5. CONCLUSION

In this study, we applied the extended SOM network, a contiguity-constraint based clustering method to integrate with the original SOM networks, to perform clustering tasks. The results show that the extended SOM method either yields comparable performance or outperforms K-means analysis especially when the sample data is skewed. This is in contrary to the results reported in previous research by Balakrishnan et al., [1994]. We believe this is due to the improved grouping capability implemented in our study. The extended SOM network presents a potential robust substitute to the current approach for market segmentation with K-means analysis and also to other problem domains that requires clustering.

REFERENCES

Balakrishnan, P.V., M.C. Cooper, V.S. Jacob, and P.A. Lewis, “A Study of the Classification Capabilities of Neural Networks using Unsupervised Learning: A Comparison with K-Means Clustering,” *Psychometrika*, Vol. 59, No. 4, Dec. 1994, 509-525.

Bimbo A.D., L..Landi, and S. Santini, "Three-dimensional planar-faced object classification with Kohonen maps," Optical Engineering, Vol. 32, No. 6, pp. 1222-1234, June 1993.

Chaturvedi, A., J. D. Carroll, P. E. Green, and J. A. Rotondo, "A Feature-Based Approach to Market Segmentation Via Overlapping K-Centroids Clustering," *Journal of Marketing Research*, Vol. 34, pp. 370-377, August 1997.

Cottrell, M. and J.C. Fort, "A stochastic model of retinotopy: A self-organizing process," Biol. Cybern., Vol. 53, pp. 405-411, 1986.

Deakin, E.B., “Distributions of Financial Accounting Ratios: Some Empirical Evidence,” *Accounting Review*, pp. 90-96, Jan. 1976.

DeSieno, D., “Adding a Conscience to Competitive Learning ,” *Proceedings of the International Conference on Neural Networks*, Vol. I, pp. 117-124, IEEE Press, New York, July 1988.

Joreskog, K. G., and D. Sorbom, PRELIS 2 User's Reference Guide, Scientific Software International, 1989.

Kiang, M.Y. and A. Kumar, “An Evaluation of Self-Organizing Map Networks as a Robust Alternative to Factor Analysis in Data Mining Applications,” Information Systems Research, Vol. 12, No. 2, June 2001, pp. 177-194.

Kohonen, T., Cybernetic Systems: Recognition, Learning, Self-Organization, ed. by E.R. Caianiello, G. Musso, Research Studies Press, Ltd., Letchworth, Herfordshire, p. 3, UK 1984.

Kohonen, T., Self-Organization and Associative Memory, Springer-Verlag, second ed., 1989.

Kohonen, T., Self-Organizing Maps, Springer, 1995.

Kotler, P., Marketing Management – Analysis, Planning, and Control, fourth edition, Prentice-Hall Inc. 1980.

Leinonen, L., T. Hiltunen, K. Torkkola, and J. Kangas, "Self-organized acoustic feature map in detection of fricative-vowel coarticulation," Journal of Acoustical Society of America, 93(6), pp. 3468-3474, June 1993.

Lo, Z.-P. and B. Bavarian, "On the rate of convergence in topology preserving neural networks," Biol. Cybern., Vol. 65, pp. 55-63, 1991.

Manikopoulos, C.N., "Finite state vector quantisation with neural network classification of states," IEEE Proceedings-F, Vol. 140, No. 3, pp. 153-161, June 1993.

Mitra, S. and S.K. Pal, "Self-Organizing Neural Network as a Fuzzy Classifier," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 24, No. 3, pp. 385-399, March 1994.

Murtagh, F., Multidimensional Clustering Algorithms, Physica-Verlag, Wurzburg, 1985.

Murtagh, F. “Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering,” Pattern Recognition Letters, Vol. 16, pp. 399-408, April 1995.

Ritter, H. and K. Schulten, "On the stationary state of Kohonen's self-organizing sensory mapping," Biol. Cybern., Vol. 54, pp. 99-106, 1986.

Ritter, H., T. Martinetz, and K. Schulten, "Topology-conserving maps for learning visuo-motor-coordination," Neural Networks, Vol. 2, pp. 159-168, 1989.

Sabourin, M. and A. Mitiche, "Modeling and Classification of Shape Using a Kohonen Associative Memory with Selective Multiresolution," Neural Networks, Vol. 6, pp. 275-283, 1993.

Sharma, S., S. Durvasula, and W. R. Dillon, “Some Results on the Behavior of Alternate Covariance Structure Estimation Procedures in the Presence of Non-Normal Data,” Journal of Marketing Research, Vol. 26, pp. 214-221, May 1989.

Vercauteren, L., G. Sieben, M. Praet, G. Otte, R. Vingerhoeds, L. Boullart, L. Calliauw, and H. Roels, “The Classification of Brain Tumours by a Topological Map,” Proceedings of the International Neural Networks Conference, Paris, 1990, pp. 387-391.

Walter, J.A. and K.J. Schulten, "Implementation of Self-Organizing Neural Networks for Visuo-Motor Control of an Industrial Robot," IEEE Transactions on Neural Networks, Vol. 4, No. 1, pp. 86-95, Jan. 1993.

Zhao, Z. and C.G. Rowden, "Use of Kohonen self-organizing feature maps for HMM parameter smoothing in speech recognition," IEEE Proceedings-F, Vol. 139, No. 6, pp. 385-390, Dec. 1992.