



An extended self-organizing map network for market segmentation—a telecommunication example

Melody Y. Kiang^{a,*}, Michael Y. Hu^b, Dorothy M. Fisher^c

^a *Information Systems Department, College of Business Administration, California State University, 1250 Bellflower Blvd., Long Beach, CA 90840-8506, United States*

^b *Graduate School of Management, Kent State University, United States*

^c *Information Systems Department, School of Business and Public Administration, California State University, Dominguez Hills, United States*

Received 22 May 2003; received in revised form 2 September 2004; accepted 9 September 2004

Available online 11 November 2004

Abstract

Kohonen's self-organizing map (SOM) network is an unsupervised learning neural network that maps an n -dimensional input data to a lower dimensional output map while maintaining the original topological relations. The extended SOM network further groups the nodes on the output map into a user specified number of clusters. In this research effort, we applied this extended version of SOM networks to a consumer data set from American Telephone and Telegraph Company (AT&T). Results using the AT&T data indicate that the extended SOM network performs better than the two-step procedure that combines factor analysis and K-means cluster analysis in uncovering market segments.

© 2004 Elsevier B.V. All rights reserved.

Keywords: SOM neural network; Extended SOM network; Factor analysis; K-means cluster analysis; Market segmentation

1. Introduction

Market segmentation refers to the process of forming groups of consumers, whereby the groups are homogeneous in terms of demand elasticities and are accessible via marketing strategies [13]. A marketing manager can select and effectively execute segment-specific marketing mixes. The value of

performing market segmentation analysis includes positioning the product in the marketplace properly, identifying the appropriate segments for target marketing, finding opportunities in existing markets, and gaining competitive advantage through product differentiation. The bottom line is to increase profitability by enabling firms to target consumers more effectively. Although it was introduced into the academic marketing literature in the 1950s, market segmentation continues to be an important focal point of ongoing research and marketing practices [4]. Most of the academic research in market segmentation has been in

* Corresponding author. Tel.: +1 5629858944; fax: +1 5629854080.

E-mail address: mkiang@csulb.edu (M.Y. Kiang).

the development of new techniques and methodologies for segmenting markets. The common thread running through these diverse streams of research is the attempt to segment consumers, deterministically or probabilistically, into a finite number of segments that are homogenous within and heterogeneous between with respect to consumer demand and demographics.

It should be noted that the usefulness of market segmentation hinges upon the accuracy of market segmentation. Relatively low accuracy in forecasting segment memberships indicates a high portion of unintended members in each segment. Misplacements will result in ineffective marketing programs that are designed to stimulate sales as well as in potential negative impact on revenue generation from the unintended segment members.

From a methodological point of view, a clustering procedure is often employed to form segments of consumers with similar preferences. When the number of dimensions underlying preferences is large, a researcher may first use a dimension reduction technique such as principal component or factor analysis to reduce the dimensions to a manageable set before subjecting the output factors to a clustering routine. Thus, a two-step approach is typically used [29].

Statistical dimension reduction routines such as factor analysis evolve around the Pearson's product moment correlation coefficient. Multivariate normality and linearity among the variables are key assumptions underlying the use of correlation coefficient. Clustering algorithms are in general heuristics in that they assign objects into clusters based on some distance measures between an object and the centroid of the cluster. Clustering algorithms are not statistical in the sense that they do not rely on any distributional assumptions. Violation of statistical assumptions in this two-step procedure will likely come about in the data reduction stage. In commercial market research applications, measures such as consumer product ratings, customer satisfaction assessments tend to be markedly skewed [26]. Violations of the normality assumption may lead to bias and incorrect assignment of consumers to the resulting segments and eventually lead to ineffective marketing strategies.

Kohonen's [10–12] self-organizing map (SOM) network, a variation of neural computing networks, is a nonparametric approach that makes no assumptions about the underlying population distribution and is

independent of prior information. Similar to principal component analysis and factor analysis, the main function of an SOM network is dimension reduction; that is, it maps an n -dimensional input space to a lower dimensional (usually one- or two-dimensional) output map while maintaining the original topological relations and, thus, enables decision maker to visualize the relationships among inputs. While Kohonen's Self-Organizing networks have been successfully applied as a classification tool to various problem domains, including speech recognition [15,32], image data compression [17], image or character recognition [2,25], robot control [23,28], and medical diagnosis [27], its potential as a robust substitute for clustering tools remains relatively unexplored. Given the non-parametric feature of SOM, it can be expected that SOM will yield superior results to the factor/cluster procedure for market segmentation. Recently, neural network applications for clustering and prediction in marketing [3,7] showed promising results as compared to the traditional statistical approaches.

Balakrishnan et al. [1] compared several unsupervised neural networks with K-means analysis. Due to the lack of an extended grouping function such as the one implemented in our extended SOM network, the two-layer Kohonen network implemented in their study is designed so that the number of nodes in the output layer (Kohonen layer) corresponds to the number of desired clusters. It is a different kind of Kohonen network that does not provide a two-dimensional map for users to visualize the relationships among data points. We have found that the majority of the studies that apply Kohonen network to clustering have implemented this type of network. The performance of these neural networks is examined with respect to changes in the number of attributes, the number of clusters, and the amount of error in the data. Their results show that the K-means procedure always outperforms neural networks especially when the number of clusters increases from two to five. Unlike the compromising approach of Balakrishnan et al., our extended SOM method preserves the dimension reduction function of the original SOM and further groups the nodes on the output map into a user specified number of clusters. The extended SOM will be discussed and presented in detail in a later section. Key features specific to the expended SOM will be pointed out. How these added features serve to enhance

the performance of SOM in dimension reduction and clustering will also be examined.

In this study, we will first employ the extended SOM to group consumers into segments using their attitudes toward long-distance communications. The resulting segments will be cross validated using consumer usage of four primary modes of communications (long-distance phone calls, letters, cards, and personal visits) and demographic factors. The key demographic factors identified by American Telephone and Telegraph Company (AT&T) are number of friends (FRIENDS) and relatives (RELATIVES), number of moves made in the past 5 years (MOVES), number of people over 16 years of age currently living in the household (OVER16), and marital status of the head of household (MSTATUS). The first two factors correspond to the size of community of interest. All of these demographic factors have been found by the company to be strongly correlated with the usage rate of long-distance phone calling. Results from using the extended SOM will be compared with those from the factor/cluster procedure.

The balance of the paper is organized as follows: Section 2 presents the basic concepts of SOM networks and illustrates their use as a dimension-reduction tool (analogous to factor analysis). This is followed by a discussion of the extended grouping capability integrated into the original SOM networks. Section 3 describes the experimental procedures and results from AT&T data set. In Section 4, we compare the performance of the extended SOM with that of the factor score-based approach, both qualitatively and quantitatively. The segments will be validated using consumer usage and demographic factors. The paper concludes with a summary of our findings.

2. Self-organizing map (SOM) networks

The self-organizing map (SOM) network is a neural network based method for dimension reduction. SOM can learn from complex, multidimensional data and transform them into a map of fewer dimensions, such as a two-dimensional plot. The two-dimensional plot provides an easy-to-use graphical user interface to help the decision-maker visualize the similarities between consumer preference patterns. In the AT&T data set, there are 68 customer attitude variables. It would be difficult to visually classify

consumers based on all these attributes because the grouping must be done in a 68-dimensional space. By mapping the information contained in the 68-variable set into a two-dimensional plot, one can visually group customers with similar preferences into clusters. These relationships can then be translated into an appropriate type of structure that genuinely represents the underlying relationships between market segments. Hence, SOM networks can be used to build a decision support system for marketing management.

Kohonen [10–12] developed the SOM network between 1979 and 1982 based on the earlier work of Willshaw and Malsburg [31]. It is designed to capture topologies and hierarchical structures of higher dimensional input spaces. Unlike most neural network applications, the SOM performs unsupervised training, i.e., during the learning (training) stage, SOM processes the input units in the network and adjusts their weights primarily based on the lateral feedback connections. The nodes in the network converge to form clusters to represent groups of nodes 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 [11]. 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 the Kohonen 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 subareas of the Kohonen layer to different categories of information and, therefore, the location of the processing element in a network becomes specific to a certain characteristic feature in the set of input data.

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 such that:

$$\|x_v - w_i\| = \min\|x_v - w_i\|, \quad i = 1, \dots, N, \quad (1)$$

where $\|\cdot\|$ indicates euclidean distance which is the most common way of measuring distance between

vectors. 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:

$$\Delta w_i = \alpha(x_v - w_i^{\text{old}}), \quad \text{for } i \in N_r, \quad (2)$$

where α is the learning coefficient, x_v 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 $r=1$ (see Fig. 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 it is usually reduced as training progresses.

Previous research has proposed various ways to improve the learning of SOM networks [5,16,22]. In this research, we implemented a Gaussian-type neighborhood adaptation function $h(t,r)$, similar to the one used by Mitra and Pal [18].

$$h(t,r) = \frac{\alpha(1 - r*f)}{\left[1 + \left(\frac{t}{\text{cdenom}}\right)^2\right]} \quad (3)$$

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 b . To determine the number of training cycles to be run at each stage, we use the

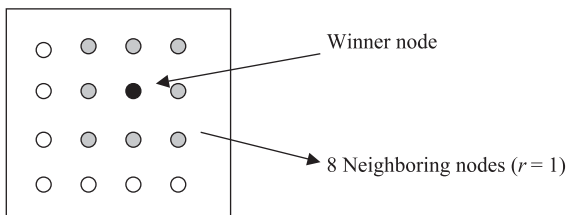


Fig. 1. A 4×4 Kohonen Layer and definition of neighboring nodes with radial distance ($r=1$).

index of disorder D proposed by Mitra and Pal [18]. Essentially, D measures the “improvement” in the “state” of the network at discrete time intervals. When this index falls below a certain threshold ($D < \text{convergence coefficient } \delta$), the next stage of training begins with a reduced r value. The reader may refer to Mitra and Pal [18] for the detailed explanation of the algorithm.

To avoid a few nodes ending up representing too much of the input data due to the effect of the initial random weights assigned to them, we incorporate a “conscience” mechanism. This mechanism proposed by DeSieno [6] 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 all the input data equally.

2.1. The extended clustering function

Sometimes it is difficult to visually group the output from SOM especially when the map is highly populated. Hence, a scientific approach that can help the user to group the output from SOM network based on certain objective criteria is needed. To automate the grouping process to complement Kohonen SOM networks, Murtagh [20] proposed an agglomerative contiguity-constrained clustering method. This method groups the output from SOM using a minimal distance criterion to merge the neighboring nodes. The rationale is that the SOM networks will maintain the original topological relations; therefore, the nodes that are located closely together on the representational grid should have the same 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 data set. Hence, we decided to use the minimal variance criterion for our contiguity-constrained clustering method. The criterion we implemented is a modified version of Murtagh’s [19] that attempts 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 that the result of the merge will minimize the global minimal variance given the number of clusters. The merge process is repeated until a user-specified number of clusters are arrived at or when only one cluster remains. Readers should refer to Kiang and Kumar [9] for a detailed step-by-step procedure used to implement the algorithm.

2.2. The use of extended SOM approach

If a desired output is not known for each training example but some knowledge of the relationship among a data set needs to be extracted, the extended SOM method may be appropriate. The clustering task that groups subjects into clusters of similar elements belongs to this category. An unsupervised learning method is a classification method where the network discovers its own classes. Examples include applications in market segmentation, and group technology in flexible manufacturing systems. The SOM method has also been used a lot as a preprocessor for other statistical or neural networks methods. When functioning as a preprocessor, the unsupervised learning networks have been applied to perform tasks such as dimension reduction, exploratory data analysis, and novelty detection. Kohonen SOM is one of the most popular unsupervised learning algorithms for dimension reduction.

Unlike the more popular neural networks models that implement supervised learning which usually requires extensive trial-and-error processes to determine the best network configuration for a given problem, the standard SOM network structure always has two layers. The main decision is the size of the Kohonen layer. While the different network size should not have an effect on the performance of the network, previous research suggested the selection of the network size in proportion to the number of data points in the data set. Thus, when the result is plotted on the two-dimensional map, the data points will not be excessively cluttered together and clear clusters can be easily identified visually. It is also suggested that the input values should be standardized to avoid one or a few variables dominating the mapping process. As far as result validation is concerned, no satisfactory methods exist for determining the true number and composition of population clusters for any type of

cluster analysis. Therefore, the only way to validate the method is by using simulated data sets with known cluster results.

3. The empirical study

There are three main components identified in the market segmentation literature including: formation of market segments, differences across segments with respect to consumer demand, and identification of the segments using consumer demographics. The factors used in the formation of segments are referred to as the base variables. In our example, these are the 68 attitude items. Criterion variables correspond to the consumer demand/preferences within each segment. Again, these are expected to be homogenous within each segment and heterogeneous among the segments. In order to allow targeting to take place, each segment has to be identified or profiled using demographic factors. As mentioned earlier, AT&T relied on consumer attitudes toward long-distance communications to form segments. The usefulness of these segments depends on to what extent consumer demand for long-distance communications differs by these segments. If no differences in consumer demand are detected, one can question the use of attitudes for the formation of segments. Furthermore, in order to stimulate demand in each of the segments, each attitudinal segment must be uniquely identified using demographic variables. Thus, differences in demographic variables are captured across the attitudinal segments. Consumer demand and demographic factors provide the opportunity to cross-validate the appropriateness of using attitudes for segmentation.

3.1. Data set

Data were obtained from a panel study that the American Telephone and Telegraph Company (AT&T) undertook in early 1980s for the purpose of market segmentation and strategy development. At that time, the study was one of the largest residential consumer projects undertaken by a single company. The out-of-pocket cost to the company totaled around 1 million. The database still provides valuable insights for market researchers [14].

The company maintained a demographically proportional, national panel of 3602 heads of households who participated in a 12-month diary. The AT&T sample was demographically balanced in terms of income, marital status, age, gender, population density, and geographic region. In a pre-diary survey, each panel member was asked to respond to a series of 68 attitude questions related to long-distance communications. The items were derived from a series of focus group interviews and had been tested for reliability and validity. Examples of the items are: ‘Sometimes I make a long-distance call simply to save the hassle of writing a letter’, ‘some people expect long-distance calls from me at a fairly regular intervals’, and ‘it bothers me that I have lost touch with many friends that I was close to at one time’. A six-point Likert scale ranging from ‘agree completely’ (coded as ‘6’) to ‘disagree completely’ (coded as ‘1’) was used for each of these 68 items. A closer examination of the skewness index for each of these 68 items shows that in numerous cases the univariate distributions are highly skewed, providing evidence that the distributions are not close to normal. The appropriateness of using factor analysis hinges upon the bivariate normality assumption since the procedure is based primarily on the Pearson’s product moment correlation. It is anticipated that violation of the normality assumption in the use of factor analysis will substantially affect the results based on the K-means.

Each respondent was also asked to provide for the four modes (CALL, LETTER, CARD, VISIT) the number of times during a typical month he/she would communicate for nonbusiness reasons with relatives and friends who live over 50 miles away from their home. The key demographic factors are as mentioned previously: FRIENDS, RELATIVES, MOVES, OVER16, and MSTATUS. Appendix A provides summary statistics of the variables used in the segmentation analysis.

3.2. The method of factor analysis and K-means

In this section, we used Version 11.0 for Windows of the SPSS Statistical package to perform the two-step procedure that uses the factor scores from factor analysis as inputs to K-means cluster analysis. The first step was to factor analyze the data set and reduce the 68 attitude variables to a smaller and more

interpretable set of factors. Specifically, principal factor analysis with varimax rotation was used to condense the customers’ responses to the 68 questions into 14 attitude factor scores. The resulting factor scores represent the following major attitudinal dimensions uncovered in the AT&T study:

- Personal emotional satisfaction with long-distance calling
- Lack of expressive skills on long-distance calls
- Long-distance calling is better than writing
- Long-distance calling is seen as obligatory
- Cost-consciousness about long-distance call usage
- Long-distance call seen as a medium for exchange function information
- Compulsive planning
- Insensitivity to communication situations
- Long-distance calls seen as disturbing
- Demonstrative personality
- Emotionally reserved life style
- Long-distance call seen as a luxury for special occasions
- Long-distance calls show others you care
- Long-distance call rates are reasonable

The second step was to use the K-means clustering algorithm to group the 3602 respondents into six segments with the 14 factor scores. Respondents grouped into a segment using K-means cluster analysis had a unique pattern of responses to the attitude items that were different from the patterns of all other segments.

With the 14 factor scores, K-means identified six market segments. Labeling of the clusters was based on the average factor scores in each segment. We then examined the average scores of attitude items within each segment. Based on the average attitude scores, we label the six segments as: (1) routine communicators; (2) emotional communicators; (3) anxious worriers; (4) budgeters; (5) detached communicators; and (6) functionalists. For example, routine communicators ‘see long-distance communications as an everyday necessary tool for frequent communication’; emotional communicators ‘derive great personal emotional gratification from long-distance usage’; anxious worriers ‘associate long-distance communications with disturbing experiences’; budgeters ‘are very cost-conscious’; detached communicators ‘do not have strong need

to communicate caring to others’; and functionalists ‘do not relate to long-distance communications emotionally’. It should be noted that other studies conducted by AT&T repeatedly identified the same six attitudinal segments. Thus, these segments are shown to be quite stable across different studies and time.

For a long time, AT&T marketing strategy has been targeted toward the emotional communicators. The long-distance communication service via phone calling is available practically in all households in the U.S. Usage rate among the emotional communicators can be stimulated by simply reminding them to communicate. The advertising campaign ‘Reach-Out and Reach-Out and Touch Someone’ reminds these consumers to stay in touch with their friends and relatives via the phone.

We are able to replicate AT&T results using factor/K-means approach. The company experimented with the number of segments and found that six segments yielded the optimal results. The 14 factors and the resulting six segments we have identified using its data set are largely the same as what AT&T has found. Because these six segments have specific operational meanings for marketing strategies, we employ the extended SOM to form six segments and will then compare the SOM results with those derived from factor/k-means. It should be noted that the optimal number of segments using SOM may not be six. Thus, in the comparative study section, we have run both K-means and extended SOM methods for different number of clusters and compared their results to find the best number of clusters.

3.3. The extended SOM clustering method

First, we performed a normality test on the 68 attitude items in the AT&T data set. All 68 variables exhibit significant deviations from normality as indicated by the Shapiro–Wilk statistic [24,30]. This strongly suggests that SOM as a nonparametric approach that makes no assumptions about the underlying population distribution may be more appropriate than parametric models to describe this data set.

An implementation of the extended SOMs in C++ was used to reduce the 68 attribute variables to a two-dimensional plot. After some preliminary runs, it shows that the network size has no significant effect on the performance of the network. We fixed the

Kohonen layer to a 7×7 network. The resulting two-dimensional plot depicting the six market segments is presented in Fig. 2 where the segment boundaries are delineated. A close examination of the 68 items within each of the six segments leads us to conclude that we have uncovered the same six segments. Emotional communicators constitute the segment with the largest market potential and budgeters, the smallest market potential. The two segments indeed fell on the two opposite corners of the SOM output map as shown in Fig. 2.

The labels of the six segments allow us to provide some tentative interpretations of the two dimensions in Fig. 2. The x -axis seems to correspond to the level of psychological involvement that a consumer experiences when engaged in long-distance communications. The y -axis seems to imply the personal value a consumer ascribes to long-distance communications.

4. The comparative study

In market segmentation studies, the accurate assignment of respondents to clusters/segments is critical. Kiang and Kumar [9] used simulated data where true cluster memberships are known and found that SOM networks provide more accurate recovery of underlying cluster structures when the input data are skewed. Because true cluster memberships are unknown for the real-world data set, it is difficult to determine which clustering method is best for the real-life problem. The accuracy of membership assignment can be measured using the pooled within cluster variance or the sum of squares (SSE). Given the fact that only two output dimensions are used in the extended SOM and 14 factor

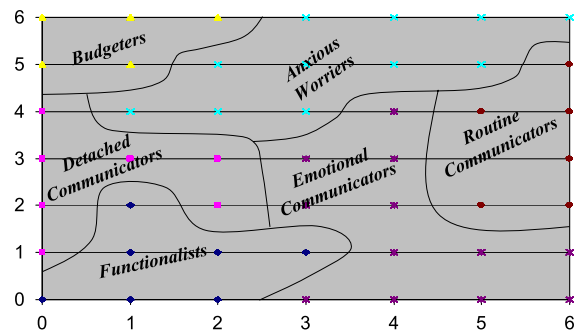


Fig. 2. The 7×7 SOM with six labeled clusters.

Table 1
Total within cluster variance for 2–7 clusters

Number of clusters	K-means	Extended SOM	Difference
2	628,249	609,537	18,712 (3.07%)
3	622,185	588,167	34,018 (5.78%)
4	601,514	577,351	24,163 (4.19%)
5	594,669	569,766	24,903 (4.37%)
6	591,141	562,241	28,900 (5.14%)
7	594,929	556,686	38,243 (6.87%)

scores are used for clustering, in order to establish a more valid basis for comparison, all 68 items are used in the calculation of total within group variance. Because the true number of segments is not known, we ran both methods on the same data set six times for different number of clusters. Table 1 shows that the Extended SOM is slightly better than the factor score-based procedure in all scenarios. We also noticed that, in general, the total variance decreases monotonically as the number of clusters increases. However, for K-means analysis, the total variance of seven clusters is higher than that of six clusters. The finding again suggests there are six segments and it also concurs with the result presented in original AT&T report.

The salient feature of the extended SOM network is its ability to reduce the input space to a one- or two-dimensional output map while maintaining the original topological relations. In other words, the data points that were close in the higher dimensional space should remain close in the reduced lower dimensional map. The two-dimensional plot provides an easy-to-use graphical interface and helps the decision-maker visualize the relationships among inputs. On the other hand, by using factor scores, we are able to interpret and label the six segments.

At the individual consumer level, the correspondence between the two approaches is captured in Table 2. It should be noted that when going down each column of SOM, the largest frequencies occur in the diagonal terms, again indicating on the whole, the two procedures yield a large percent of consumers belonging to the same six segments. The sum of the six diagonal entries has 1873 observations, roughly 52% of the 3602 observations. Conversely, 48% of the respondents have been classified into different segments by the two procedures. While there is some indication of aggregate level stability, at the individual level, there remains substantial amount of variation.

A more sophisticated approach for measuring the similarity between the clustering results of SOM and K-means analysis is the Rand Index [8]. The Rand Index is a way of computing the agreement between two partitions of the same data set. This measure varies between zero and one, with one indicating perfect agreement between the two partitions, and zero indicating the opposite. Interested readers should read [9,21] for a detailed explanation of the method. Table 3 summarizes the results for comparing the two methods based on Rand Index. It shows that, as the number of clusters increases, so does the agreement between the two methods.

Next, we address the issue of relationship between segment compositions with the consumer demand factors. As mentioned previously, in the survey, respondents were asked during a typical month the number of times each was engaged in the four modes of long-distance communication: phone calling (CALL), letter (LETTER), card (CARD), and personal visit (VISIT). With SOM and K-means as independent variables and CALL, LETTER, CARD,

Table 2
Cross-tabulation of SOM's clusters and K-means' clusters

	Clusters	SOM						Total
		1	2	3	4	5	6	
K-MEANS	1	289 (8.02%)	57 (1.58%)	17 (0.47%)	54 (1.5%)	211 (5.86%)	1 (0.03%)	629 (17.46%)
	2	31 (0.86%)	205 (5.69%)	21 (0.58%)	279 (7.75%)	85 (2.36%)	3 (0.08%)	624 (17.32%)
	3	154 (4.28%)	87 (2.42%)	159 (4.41%)	13 (0.36%)	7 (0.19%)	0 (0.00%)	420 (11.66%)
	4	4 (0.11%)	1 (0.03%)	120 (3.33%)	444 (12.33%)	10 (0.28%)	8 (0.22%)	587 (16.30%)
	5	14 (0.39%)	121 (3.36%)	0 (0.00%)	13 (0.36%)	372 (9.94%)	51 (1.42%)	557 (15.46%)
	6	0 (0.00%)	10 (0.28%)	0 (0.00%)	80 (2.22%)	277 (7.69%)	418 (11.60%)	785 (21.79%)
	Total	492 (13.66%)	481 (13.35%)	317 (8.80%)	883 (24.51%)	948 (26.32%)	481 (13.35%)	3602 (100%)

$\chi^2=4845.58$; p -value=0.0001.

Table 3
Measuring the agreement between K-means and extended SOM methods

Number of clusters	Rand index
2	0.501
3	0.584
4	0.652
5	0.694
6	0.734
7	0.759

and VISIT as dependent variables, one-way analysis of variance (ANOVA) was conducted to capture the differences in usage rates in each of the four modes across the six segments formed either through SOM or K-means. Results are presented in Table 4.

Statistically significant differences are detected for each of the four modes along SOM and K-means. For SOM, the largest F -stat. is associated with CALLS (F -stat.=77.12, p -value=0.0001), similarly for K-means (F -stat.=75.04; p -value=0.0001). It can be concluded that the six attitudinal segments, formed either by SOM or by K-means, lead to differences in consumer demand in long-distance communication usage. Thus, a different set of marketing mix can be identified and used to target each of these six segments to stimulate overall demand.

As for using the demographic factors to profile the segments, we conducted one-way analysis of variance for the continuous demographic factors (FRIENDS, RELATIVES, MOVES, and OVER16) and rely on cross-tabulation for categorical variable (MSTATUS) to establish relationship between SOM, K-means, and

Table 4
ANOVA results of usage by SOM clusters and K-means clusters

	Clusters						F -stat.	p -value
	1	2	3	4	5	6		
<i>A. SOM</i>								
Calls	2.57	2.61	1.18	3.28	5.06	5.40	77.12	0.0001
Letter	3.91	1.67	2.38	2.48	4.01	3.64	22.80	0.0001
Card	2.48	1.41	1.48	1.66	2.56	2.57	15.07	0.0001
Visit	1.27	1.02	1.06	1.50	1.71	1.97	9.91	0.0001
<i>B. K-means</i>								
Calls	3.21	2.81	1.53	2.88	5.33	5.22	75.04	0.0001
Letter	4.86	1.29	2.28	2.88	2.78	4.02	43.72	0.0001
Card	3.04	1.02	1.34	2.01	2.10	2.62	30.26	0.0001
Visit	1.26	1.23	1.09	1.57	1.64	1.91	8.65	0.0001

Table 5
ANOVA results of demographics by SOM clusters and K-means clusters

	Clusters						F -stat.	p -value
	1	2	3	4	5	6		
<i>A. SOM</i>								
Friends	4.75	4.32	3.76	4.59	5.32	5.49	18.23	0.0001
Relatives	5.19	4.91	4.52	5.41	5.91	5.96	17.94	0.0001
Moves	0.31	0.21	0.29	0.41	0.51	0.36	8.21	0.0001
Over 16	2.44	2.28	2.33	2.38	2.22	2.19	4.93	0.0002
<i>K-means</i>								
Friends	5.11	4.14	4.05	4.58	5.11	5.49	19.07	0.0001
Relatives	5.75	4.98	4.79	5.16	5.58	6.02	16.10	0.0001
Moves	0.39	0.32	0.26	0.41	0.35	0.48	3.76	0.0021
Over 16	2.37	2.29	2.38	2.42	2.19	2.01	4.86	0.0002

these demographic factors. The results are presented in Table 5.

Again, the ANOVA results show that significant differences are detected for FRIENDS, RELATIVES, MOVES, and OVER16 along both SOM and K-means. Chi-square statistics indicate significant relationship between SOM ($\chi^2=50.42$; p -value=0.0001), K-means ($\chi^2=48.84$; p -value=0.0001), and MSTATUS. The cross-tab results are not presented in Table 5. It is comforting to note that these demographic factors are useful for segment membership identification.

Discriminant analysis was also conducted using the five demographic factors to profile the differences among the six segments formed by SOM or K-means. Discriminant analysis forms linear combinations of the five demographic factors and uses them to discriminate among the six groups. A stepwise discriminant procedure indicated that all five factors are to be retained and used in both cases. Furthermore, the SOM procedure with all five demographic factors yielded a slightly better correct classification rate of 28.97% as compared to 25.59% for K-means. Both classification rates are substantially higher than chance classification (16.67%), indicating that the demographic factors are indeed useful in identifying the six segments. Thus, it is shown that SOM provides at least as good a classification as K-means.

5. Conclusion

In this study, we first identify the importance of market segmentation in market management and

strategy development. The first and primary component in market segmentation is the formation of groups. Accurate assignment of segment membership is the key to successful market segmentation programs. Traditional approach relies heavily on factor analysis for dimension reduction and cluster analysis for grouping. The imposition of linearity and normality assumptions may lead to less-desirable segment memberships. This paper examines the viability of using the extended SOM method for dimension reduction and grouping.

The extended SOM network is a contiguity-constraint based clustering method integrated with the original SOM networks. The extended SOM network is applied to perform clustering analysis on a real-world data set. Based on the total within cluster variance, the extended SOM method out-

performed the two-step factor/cluster procedure. Moreover, the extended SOM network provides the user with a visual rendering of the market segments. Our procedure is a robust substitute for the current approach of market segmentation with K-means analysis and also for other problem domains that requires clustering.

The appropriateness of the six segments was further validated using consumer usage and demographic factors. Both SOM and K-means segments are found to be appropriate. SOM segments yielded slightly better demographic profiles than K-means segments. Because the SOM segments are more clearly delineated using the demographic factors, implications here are that SOM-based segments provide a viable option for marketers rather than K-means segments for targeting.

Appendix A. Mean, standard deviation, and skewness of the 68 variables

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Mean	2.97	3.83	4.97	2.43	2.85	3.61	3.02	4.37	4.35	3.74	1.62	4.73
Std. Deviation	1.775	1.930	1.455	1.670	1.747	1.814	1.938	1.620	1.780	1.802	1.206	1.642
Skewness	0.218	0.356	1.513	0.804	0.458	0.124	0.284	0.795	0.837	0.212	2.179	1.132
	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24
Mean	3.74	4.41	4.03	2.43	2.19	3.46	2.57	3.99	2.37	3.98	2.78	4.35
Std. Deviation	1.706	1.511	1.746	1.649	1.569	1.708	1.791	1.881	1.658	1.730	1.700	1.775
Skewness	0.270	0.885	0.565	0.767	1.115	0.103	0.703	0.451	0.863	0.481	0.448	0.745
	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34	A35	A36
Mean	3.45	2.35	2.79	3.63	3.54	4.14	4.69	2.33	3.26	3.63	4.22	4.25
Std. Deviation	1.716	1.540	1.574	1.775	1.675	1.626	1.399	1.433	1.600	1.699	1.644	1.493
Skewness	0.001	0.851	0.437	0.169	0.116	0.606	1.076	0.853	0.038	0.178	0.642	0.697
	A37	A38	A39	A40	A41	A42	A43	A44	A45	A46	A47	A48
Mean	3.84	5.02	4.38	3.91	4.89	5.07	3.35	3.14	3.61	3.78	4.40	3.56
Std. Deviation	1.548	1.259	1.544	1.717	1.161	1.195	1.533	1.573	1.636	1.734	1.730	1.592
Skewness	0.330	1.475	0.748	0.324	1.324	1.506	0.013	0.108	0.147	0.291	0.817	0.200
	A49	A50	A51	A52	A53	A54	A55	A56	A57	A58	A59	A60
Mean	4.25	1.78	4.21	2.66	3.89	3.80	2.97	3.94	3.99	3.79	4.80	2.81
Std. Deviation	1.698	1.203	1.685	1.657	1.774	1.644	1.283	1.637	1.641	1.486	1.570	1.684
Skewness	0.709	1.583	0.674	0.584	0.388	0.353	0.189	0.471	0.479	0.351	1.196	0.393
	A61	A62	A63	A64	A65	A66	A67	A68				
Mean	4.60	4.40	4.20	3.79	3.56	4.34	2.01	3.89				
Std. Deviation	1.324	1.457	1.672	1.843	1.672	1.568	1.399	1.630				
Skewness	1.122	0.763	0.625	0.305	0.211	0.792	1.310	0.389				

References

- [1] P.V. Balakrishnan, M.C. Cooper, V.S. Jacob, P.A. Lewis, A study of the classification capabilities of neural networks using unsupervised learning: a comparison with K-means clustering, *Psychometrika* 59 (4) (1994 Dec.) 509–525.
- [2] A.D. Bimb, L. Landi, S. Santini, Three-dimensional planar-faced object classification with Kohonen maps, *Optical Engineering* 32 (6) (1993 June) 1222–1234.
- [3] D.S. Boone, M. Roehm, Retail segmentation using artificial neural networks, *International Journal of Research in Marketing* 19 (2002) 287–310.
- [4] A. Chaturvedi, J.D. Carroll, P.E. Green, J.A. Rotondo, A feature-based approach to market segmentation via overlapping K-centroids clustering, *Journal of Marketing Research* 34 (1997 August) 370–377.
- [5] M. Cottrell, J.C. Fort, A stochastic model of retinotopy: a self-organizing process, *Biological Cybernetics* 53 (1986) 405–411.
- [6] D. DeSieno, Adding a Conscience to Competitive Learning, *Proceedings of the International Conference on Neural Networks*, vol. 1, IEEE Press, New York, 1988 (July), pp. 117–124.
- [7] M.Y. Hu, C. Tsoukalas, Explaining consumer choice through neural networks: The Stacked general approach, *Journal of Operational Research* 146 (3) (2003) 650–661.
- [8] L. Hubert, P. Arabie, Comparing partitions, *Journal of Classification* 2 (2/3) (1985) 193–218.
- [9] M.Y. Kiang, A. Kumar, An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications, *Information Systems Research* 12 (2) (2001 June) 177–194.
- [10] T. Kohonen, *Cybernetic Systems: Recognition, Learning, Self-Organization*, in: E.R. Caianiello, G. Musso (Eds.), Research Studies Press, Letchworth, Hertfordshire, UK, 1989.
- [11] T. Kohonen, *Self-Organization and Associative Memory*, 2nd ed., Springer-Verlag, 1989.
- [12] T. Kohonen, *Self-Organizing Maps*, Springer, 1995.
- [13] P. Kotler, *Marketing Management—Analysis, Planning, and Control*, Fourth edition, Prentice-Hall, 1980.
- [14] K. Lee, M. Hu, R. Toh, Are consumer survey results distorted? Systematic impact of behavioral frequency and duration on survey response errors, *Journal of Marketing Research* 37 (1) (2000) 125–134.
- [15] L. Leinonen, T. Hiltunen, K. Torkkola, J. Kangas, Self-organized acoustic feature map in detection of fricative-vowel coarticulation, *Journal of the Acoustical Society of America* 93 (6) (1993 June) 3468–3474.
- [16] Z.-P. Lo, B. Bavarian, On the rate of convergence in topology preserving neural networks, *Biological Cybernetics* 65 (1991) 55–63.
- [17] C.N. Manikopoulos, Finite state vector quantisation with neural network classification of states, *IEEE Proceedings-F* 140 (3) (1993 June) 153–161.
- [18] S. Mitra, S.K. Pal, Self-organizing neural network as a fuzzy classifier, *IEEE Transactions on Systems, Man, and Cybernetics* 24 (3) (1994 March) 385–399.
- [19] F. Murtagh, *Multidimensional Clustering Algorithms*, Physica-Verlag, Wurzburg, 1985.
- [20] F. Murtagh, Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering, *Pattern Recognition Letters* 16 (1995 April) 399–408.
- [21] T.S. Raghu, P.K. Kannan, H.R. Rao, A.B. Whinston, Dynamic profiling of consumers for customized offerings over the Internet: a model and analysis, *Decision Support Systems* 32 (2) (2001 December) 117–134.
- [22] H. Ritter, K. Schulten, On the stationary state of Kohonen's self-organizing sensory mapping, *Biological Cybernetics* 54 (1986) 99–106.
- [23] H. Ritter, T. Martinetz, K. Schulten, Topology-conserving maps for learning visuo-motor-coordination, *Neural Networks* 2 (1989) 159–168.
- [24] J.P. Royston, Approximating the Shapiro–Wilk's W test for non-normality, *Statistics and Computing* 2 (1992) 117–119.
- [25] M. Sabourin, A. Mitiche, Modeling and classification of shape using a Kohonen associative memory with selective multi-resolution, *Neural Networks* 6 (1993) 275–283.
- [26] S. Sharma, S. Durvasula, 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* 26 (1989 May) 214–221.
- [27] L. Vercauteren, G. Sieben, M. Praet, G. Otte, R. Vingerhoeds, L. Boullart, L. Calliauw, H. Roels, The classification of brain tumours by a topological map, *Proceedings of the International Neural Networks Conference, Paris, 1990*, pp. 387–391.
- [28] J.A. Walter, K.J. Schulten, Implementation of self-organizing neural networks for visuo-motor control of an industrial robot, *IEEE Transactions on Neural Networks* 4 (1) (1993 Jan.) 86–95.
- [29] M. Wedel, W. Kamakura, *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, 1999.
- [30] M.B. Wilk, S.S. Shapiro, The joint assessment of normality of several independent samples, *Technometrics* 10 (1968) 825–839.
- [31] D.J. Willshaw, C.v.d. Malsbruge, *Philosophical Transactions of the Royal Society of London, Series B: Biological* 287 (1979) 203.
- [32] Z. Zhao, C.G. Rowden, Use of Kohonen self-organizing feature maps for HMM parameter smoothing in speech recognition, *IEEE Proceedings-F* 139 (6) (1992 Dec.) 385–390.



Melody Y. Kiang is a Professor of Computer Information Systems at California State University, Long Beach. She received her MS in MIS from the University of Wisconsin, Madison, and PhD in MSIS from the University of Texas at Austin. Prior to joining CSULB, she was an Associate Professor at the Arizona State University. Her research interests include the development and applications of artificial intelligence techniques to a variety of business problems. Her research has appeared in *Information Systems Research (ISR)*, *Management Science*, *Journal of Management Information Systems*, *Decision Support Systems*, *IEEE Transactions on SMC*, *EJOR*, and other professional journals. She is an Associate Editor of *Decision Support Systems* and Co-Editor of *Journal of Electronic Commerce Research*.



Dr. Michael Hu holds the Bridgestone Endowed Chair in International Business and he is a Professor of Marketing at Kent State University. He earned his PhD from the University of Minnesota in Management Science/Marketing. He is a dedicated educator and won the University Distinguished Teaching award in 1994. He has published over a hundred and ten journal articles in the areas of artificial neural networks, international business and marketing research. His research has appeared in *Decision Support Systems*, *Annals of Operations Research*, *European Journal of Operational Research*, *Decision Sciences*, *Journal of Marketing Research*, *Journal of International Business Studies*, among many others.

Dorothy M. Fisher is a Professor of Information Systems at the California State University, Dominguez Hills. She received an MA from Duke University and a PhD from Kent State University. Dr. Fisher has had broad consulting experience with private firms as well as educational institutions. Her research emphasizes the applications of statistical and artificial intelligence techniques to management problems. Dr. Fisher has published papers in the *Journal of Computer Information Systems*, the *Journal of Systems Management*, the *Journal of Applied Business Research*, and other academic and professional journals. Currently, she is the Managing Editor of the *Journal of Electronic Commerce Research*.