ELSEVIER

Intern. J. of Research in Marketing 19 (2002) 287–301

International Journal of Research in Marketing

www.elsevier.com/locate/ijresmar

Retail segmentation using artificial neural networks☆

Derrick S. Boone*, Michelle Roehm

Babcock Graduate School of Management, Wake Forest University, P.O. Box 7659, Winston-Salem, NC 27109-7659, USA

Abstract

Advances in information technology (e.g., scanner data, cookies, and other electronically based data collection methodologies) have enabled researchers to collect unprecedented amounts of individual-level customer data. As a result, customer databases are becoming increasingly larger and more complex, and may tax the capabilities and exacerbate the shortcomings of the techniques currently used to analyze them. To address this challenge, we examine the use of artificial neural networks (ANNs) as an alternative means of segmenting retail databases. In particular, we investigate the Hopfield–Kagmar (HK) clustering algorithm, an ANN technique based on Hopfield networks, and empirically compare it to *K*-means and mixture model clustering algorithms. Our results indicate that ANNs may be more useful to retailers for segmenting markets because they provide more homogeneous segmentation solutions than mixture model and *K*-means clustering algorithms, and are less sensitive to initial starting conditions.

Keywords: Artificial neural networks; Segmentation; Cluster analysis; Retailing

1. Introduction

Retailers have long recognized the importance of tailoring their marketing mixes to suit the specific needs and preferences of different customer groups. However, access to unprecedented amounts of individual-level customer data may make such targeted promotional efforts increasingly difficult to effectively implement. Recent growth in the use of loyalty programs, personal shopping programs, scanners, cookies, and other electronically based data collection methodologies has resulted in an "embarrassment of riches" that may only serve to complicate market segmentation and targeting.

Given the easy access to customer data afforded by information technology, retailer databases are becoming substantially larger and noisier. As a result, how to effectively handle, analyze, and interpret customer information will be one of the key challenges facing retailers who wish to execute segment-specific marketing mixes. According to Shapiro and Varian (1999), those firms that master information about their customers will thrive by delivering customized, highly valued offerings, while those that do not will be at a competitive disadvantage.

In order to more effectively implement targeted marketing mixes, it is vitally important that retailers segment their customer bases such that sufficient commonalities exist within, and sufficient distinctions between, each segment to justify taking the time and expense to create separate marketing mixes. Furthermore, the correct assignment of customers to the resulting segments is critical because improper seg-

☆ This paper was evaluated following the normal evaluation process and accepted for the Special Issue by the Guest Editors.

* Corresponding author. Tel.: +1-336-758-4475; fax: +1-336-758-4514.

E-mail address: derrick.boone@mba.wfu.edu (D.S. Boone).

0167-8116/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-8116(02)00080-0

mentation reduces the effectiveness of a segmentation strategy and squanders marketing resources. Consider, for example, a retailer that has information regarding several million customers in its database and wants to send targeted promotional messages (e.g., brochures, etc.) to the various segments. If customers are sent the wrong brochures because they are assigned to the wrong segments, the marketing efforts directed towards them are likely to be ineffective and for naught. Moreover, the financial costs associated with such mistakes can be nontrivial.

According to the Direct Marketing Association (1999), the cost of printing, handling, and mailing a single page letter and brochure is between US$.93 and US$1.10 per customer. Thus, when segmenting databases that contain millions of customers, the assignment of even a small fraction of customers to incorrect segments can waste tens, or even hundreds, of thousands of marketing dollars. Therefore, retailers must exercise care when selecting a clustering algorithm for segmenting their markets because the resulting solutions are highly dependent upon the algorithm being used. (See Arabie, Hubert, and De Soete (1996), Punj and Stewart (1983), and Wedel and Kamakura (2000) for a review of clustering algorithms.)

For example, prior researchers have noted that the computational burdens associated with hierarchical clustering techniques make them less appropriate than nonhierarchical techniques for segmenting large databases (Milligan & Sokol, 1980; Punj & Stewart, 1983; Wedel & Kamakura, 2000). However, these same authors note that nonhierarchical techniques tend to perform poorly unless "rational" information regarding initial centroid locations (cluster seeds) is provided a priori. Commonly, such information is based on a hierarchical clustering technique and/or managerial expertise. We contend that for large data sets, the computational burdens associated with obtaining hierarchical-based cluster seeds, and the biases and inaccuracies that may be associated with managerially based seed selection, highlight the need for new methods that provide optimal segmentation solutions independent of a priori information.

The central theme of this paper follows from the above notions. We argue that analytical techniques that do not require a priori seed information are needed in order for retailers to more effectively segment and target noisy, data-rich marketplaces. Accordingly, we examine artificial neural networks (ANNs) as an alternative methodology for segmenting retail customer databases. Although ANN computational systems have been applied to a wide variety of problems in engineering, computer science, mathematics, and other areas, management scientists have only recently begun to investigate the potential of ANNs to solve business-related problems in general, and marketing-related problems in particular. (See Krycha and Wagner (1999) for a review of ANN applications in marketing and other business disciplines.)

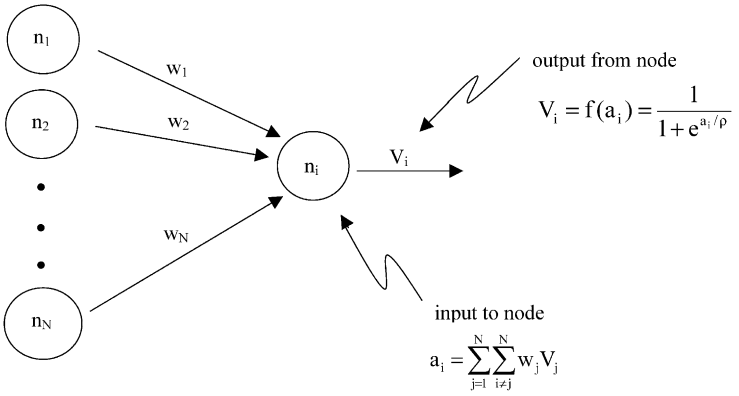
We address this oversight by investigating the segmentation utility of ANNs for retailer databases. More specifically, we focus on an application of Hopfield (1982) networks, an ANN particularly well suited for solving segmentation problems. We begin by providing an overview of ANNs in general. We then review the Hopfield–Kagmar (HK) clustering algorithm and discuss why it should theoretically provide better segmentation solutions than other clustering techniques (Kagmar-Parsi, Gualtieri, Devaney, & Kagmar-Parsi, 1990). Following this discussion, we empirically test HK, *K*-means and mixture model algorithms using a real world retail data set and a variety of artificial data sets (Dempster, Laird, & Rubin, 1977; MacQueen, 1967). Finally, we present a general discussion of our findings and offer observations about limitations and directions for future research.

2. Artificial neural networks (ANNs)

The term "artificial neural network" (ANN) describes a family of analytical models that is based on the physiological properties of animal nervous systems. As with animal nervous systems, ANNs are composed of interconnected "nodes" (neurons) that are capable of processing and transmitting information. How the interconnections are modeled (architecture) determines how information is transmitted in the network and its properties. (See Gurney (1997) for an overview of ANNs in general, and Arabie et al. (1996) for an overview of ANNs related to clustering and classification.)

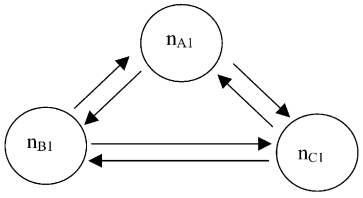
The most basic ANN is a perceptron. Perceptrons were first mathematically described by McCulloch and Pitts (1943) and contain a single input layer, consisting of multiple nodes, and a single output layer,

(a) McCulloch and Pitts (1943) Artificial Neuron



n_j and w_j are the nodes and weights, respectively, for the $j = 1, 2, \ldots, j \neq i, N$ nodes in the network

(b) Three Node Hopfield (1982) Artificial Neural Network



n_{pi} = node responsible for processing input information and generating output information regarding Customer *i*'s membership in Segment *p*. In this example, *i* = Customer 1 and there are three segments (A, B, and C). Arrows indicate multidirectional flow of information between nodes.

Fig. 1.

consisting of a single node (see Fig. 1a). As modeled, a node receives inputs from upstream nodes, weights them, and if the weighted summation exceeds some threshold value, transmits a signal in a single, downstream direction. That is, the activation energy of node i, a_i, is described by:

$$a_i = \sum_{j=1}^{N}\sum_{i\neq j}^{N} w_j V_j \qquad (1)$$

where w_j and V_j are the weights and outputs (inputs to node i), respectively, for node $j = 1, 2, \ldots, j \neq i, N$. If node i were modeled to mimic an animal neuron, its output would be: $V_i = 1$ (on) or $V_i = 0$ (off) by setting:

$$V_i = 1 \quad \text{if } a_i > \theta$$

$$V_i = 0 \quad \text{if } a_i \leq \theta$$

where θ is some fixed threshold activation level. For mathematical manipulation, however, a continuous output function is preferable to a step function.

Consequently, the output for node i is often modeled as a logistic function:

$$V_i = f(a_i) = \frac{1}{1 + e^{a_{pi}/\rho}} \quad (2)$$

where ρ determines the steepness of the curve, with progressively smaller values approximating a step function.

Given its architecture, a perceptron is capable of processing multiple input variables (e.g., number of credit cards held, total dollars spent, total number of orders, etc.) to generate a simple 0/1 output (e.g., exclusion/inclusion in "heavy credit card user" segment). However, by modifying the architecture of the network, more complex inputs and relationships can be examined. Most commonly, including one or more "hidden layers," with variable numbers of nodes between the input and output layers, and modifying the number of connections and directions of information flow between nodes, are used to affect different network properties and capabilities. For example, feed forward networks with back propagation, in which the output of the network acts as feedback and reenters the network as an input, afford more sensitive modeling of real world data by iteratively adjusting the weights until an optimal solution is found. Similarly, Kohonen (1989) networks, in which the nodes are highly interconnected, allow the nodes to self-organize themselves (adjust weights) such that spatial relationships (e.g., distance or similarity measures) are preserved when going from higher (e.g., multiple input variables) to lower (e.g., cluster membership) dimensional spaces.

To date, ANNs have not been fully investigated as potential tools for segmenting markets. This is especially surprising in light of the widespread usage of segmentation in marketing settings, the magnitude of the body of research on the topics of cluster analysis and related segmentation techniques—over 1600 articles, according to Wedel and Kamakura (2000)—and the recognized limitations, such as sensitivity to initial starting conditions and data complexities, of extant algorithms.

In the few reported studies of ANN-based market segmentation, results are mixed. Dasgupta, Dispensa, and Ghose (1994), in one of the first studies investigating the segmentation utility of ANNs, report that a back propagation ANN compares favorably with traditional multiple discriminant analysis and logistic regression in determining customer segments. Mangiameli, Chen, and West (1996) compare Kohonen (1989) self-organizing maps (SOMs) to traditional hierarchical clustering techniques and find SOMs to be superior. Contrarily, Balakrishnan, Cooper, Jacob, and Lewis (1994) compared SOMs to K-means and found them to be inferior. Balakrishnan, Cooper, Jacob, and Lewis (1996) investigate another ANN based on frequency-sensitive competitive learning and again find that the ANN did not perform well, but suggest its results could be used as seeds for K-means clustering (Krishnamurthi, Ahalt, Melton, & Chen, 1990).

In a more recent study, Natter (1999) found that a feedforward ANN with back propagation that incorporated both clustering and discriminant analysis produced good cluster solutions. Hruschka and Natter (1999) similarly report that the feedforward ANN they tested obtained solutions comparable to those of K-means. Importantly, these authors note that weaknesses in the architectures, rather than a failure of ANNs in general, may account for the mixed successes of ANNs to date.

It thus appears that ANNs can effectively be used to segment markets, but the choice of network architecture determines the degree of success observed. Along these lines, we investigate an algorithm based on Hopfield (1982) networks, an ANN particularly well suited for segmentation applications.

3. Hopfield–Kagmar (HK) clustering

Fig. 1b depicts a Hopfield (1982) network. Unlike other ANNs, each node in a Hopfield network is connected to all other nodes, but not to itself, and information can flow from node to node in multiple directions. Thus, all information in a Hopfield network is propagated within the net, making it highly useful for combinatorial optimization applications in which the goal is to minimize some objective function.

Given its architecture, a Hopfield network has certain inherent advantages over other ANNs in segmentation applications. First, unlike multiple layer perceptrons (MLPs) and Kohonen networks, the weights associated with each input do not change as the network evolves. Instead, an optimal solution is obtained by minimizing an objective cost function

(e.g., within-segment variation). Second, the threshold activation levels are fixed for each node, not globally. Thus, each node is allowed to behave somewhat differently when obtaining a solution. Third, whereas MLP and Kohonen networks have to be "trained," no training is required for Hopfield networks because all information is propagated within and inherent to the network.

Kagmar-Parsi et al. (1990) exploit these properties and develop a Hopfield-based clustering algorithm that they empirically find to be superior to MLP and Kohonen networks (Kagmar-Parsi, Kagmar-Parsi, & Sciortino, 1996). An overview of their algorithm, from a market segmentation perspective, is provided in Table 1.

Table 1
Overview of the Hopfield–Kagmar (HK) clustering algorithm

Phase one: network initialization

(1) Create $K \times N$ matrix, where K = number of desired segments, specified a priori, and N = number of customers in the data set.
(2) Randomly assign $0 \le V_{pi} \le 1$ to each element of the matrix, where V_{pi} = strength of membership of Customer i in Segment p.
(3) Randomly select small subset of customers to use in determining initial segment centroids using a V_{pi}-based weighted average.

Phase two: network implementation

(4) Each of the $n = K \times N$ nodes in the network receives input information from other nodes in the network. a_{pi}, the input to node$_{pi}$, is a function of V_{pi}, the strength with which Customer i is assigned to Segment p, V_{qi}, the strength with which Customer i is assigned to all segments other than p, and R_{pi}, the squared Euclidean distance of Customer i from the centroid of Segment p.
(5) At time t, the first node in the network updates its output, V_{pi}, following a two-stage process:
(a) Calculate a_{pi} in accordance with the mathematical operations specified in Eq. (4).
(b) Use a_{pi} to generate an "updated" V_{pi} in accordance with the transfer function specified in Eq. (2).
(6) Recalculate the centroid of Segment p, based on the "updated" V_{pi} weighted average, and recalculate R_{pi}, the distance of Customer i from the (recalculated) centroid of Segment p.
(7) At time $t+1$, the second node in the network updates V_{pi}, and recalculates the centroid of Segment p and R_{pi} as specified in steps 5 and 6 above.
(8) At time $t+2$, $t+3$, $t+4$,. . ., all remaining nodes in the network iteratively and sequentially receive and process input information from other nodes in the network, update V_{pi}, and recalculate segment centroids and R_{pi} as specified in steps 5 and 6 above.
(9) The network continues to iterate until the outputs for each node no longer change. That is, when V_{pi} (time t) = V_{pi} (time $t+\delta t$) $\forall$ V_{pi}, an optimal solution has been found.

In their research, Kagmar-Parsi et al. (1990) propose that the following equation can be used to control the evolution of a Hopfield (1982) network:[1]

$$S = \frac{A}{2}\sum_{i=1}^{N}\sum_{p=1}^{K}\sum_{q\neq p}^{K} V_{pi}V_{qi} + \frac{B}{2}\sum_{i=1}^{N}\left(\sum_{p=1}^{K} V_{pi} - 1\right)^2 + \frac{C}{2}\sum_{i=1}^{N}\sum_{p=1}^{K} R_{pi}V_{pi}^2 \quad (3)$$

In Eq. (3), V_{pi} ($0 \le V_{pi} \le 1$) represents the output of node$_{pi}$ and reflects the strength with which Customer i is assigned to Segment p. V_{qi}, which is the output of node$_{qi}$, similarly reflects the strength with which Customer i is assigned to segments other than p. Finally, R_{pi}, the square of the Euclidean distance of Customer i from the centroid of Segment p, reflects the degree of within-segment homogeneity. A, B, and C are positive constants.

An optimal solution to a segmentation problem can be obtained by minimizing Eq. (3). Given that $V_{pi} = 1$ and $V_{qi} = 0$ $\forall$ $V_{q\neq p}$ denote Customer i's inclusion and exclusion in Segments p and q, respectively, the first term in the equation will have a higher value when a customer is assigned even partially to more than one segment. Similarly, the second term in the equation will have a higher value when a customer is not assigned to any segment. Finally, the third term will have a higher value when the summed R_{pi} (squared Euclidean distances of the customers from the segment centroids) are not minimized. Thus, Eq. (3), when minimized, results in a segmentation solution that promotes the assignment of customers to segments such that within-segment variation is as low as possible.

Importantly, it should be noted that the minimization of Eq. (3) does not *necessarily* require that seg-

[1] In order to be technically correct, a constraint term that ensures each segment contains at least one customer, $\sum_{p=1}^{K}\theta \times\left(1-\sum_{i=1}^{N} V_{pi}\right)$, where $\theta(x)=0$ for $x \le 0$ and $\theta(x)=1$ for $x>0$ is the step function, should be included in Eq. (3). However, because this term is not continuous and therefore nondifferentiable, Kagmar-Parsi et al. (1990) omitted it and offered as a better strategy the rejection of solutions that violate this constraint. Furthermore, Kagmar-Parsi and Kagmar-Parsi (1992) report that in several thousand trials, the solutions never violated this constraint and its exclusion appears to be of little consequence.

ment memberships be unique ($V_{pi}=0$ or 1), as is the case with hierarchical, nonhierarchical (e.g., *K*-means), and nonparametric clustering algorithms. Additionally, the minimization of Eq. (3) does not require that $\sum_{p=1}^{k} V_{pi} = 1$, as is the case with fuzzy clustering algorithms. Instead, the HK algorithm's only requirement is that when it stops, the terminal V_{pi} are assigned in such a manner that the first two terms and the within-segment variation (third term) in Eq. (3) are minimized.

To obtain a solution, the network is first started in some initial state in which each customer is randomly and partially assigned to all possible segments ($0 \leq V_{pi} \leq 1$). The network then evolves toward an optimal solution via an iterative process. Each node sequentially revises its V_{pi} from the initial state to one in which each customer is ideally assigned to exactly one segment ($V_{pi}=1$ or $0 \ \forall \ V_{pi}$), such that within-segment variation is minimized. The process by which each node sequentially and iteratively "updates" its output (V_{pi}) at time t can be described by the $n=K\times N$ coupled, ordinary differential equations obtained from $-\partial S/\partial V_{pi}$:

$$\frac{\mathrm{d}a_{pi}}{\mathrm{d}t} = -A\sum_{q\neq p}^{K} V_{qi} - B\left(\sum_{q=1}^{K} V_{qi} - 1\right) - CR_{pi}V_{pi} \quad (4)$$

A better understanding of the updating process can be derived from the following discussion (see Fig. 2.) The inputs to node$_{pi}$, a_{pi} (segment strengths of membership and distances from segment centroids), will change over time as the input strengths of membership and distances to centroids change as the network evolves. Input–output transformation (i.e., updated strength of membership V_{pi}) is accomplished via a logistic transfer function as described in Eq. (2). The iterative process specified by Eq. (4) progresses, with all nodes sequentially updating over time, until all V_{pi} become stable (i.e., the change in V_{pi} after node$_{pi}$ updates is insignificant). The resulting solution ideally contains homogeneous segments (i.e., minimum within-segment variation) in which each customer is uniquely assigned to a segment.[2]

[2] Technically speaking, because an analog (logistic) transfer function is being used, the V_{pi} only approach 0 or 1, and accordingly, if $V_{pi}<\eta_0$, then $V_{pi}=0$, and $V_{pi}\geq 1-\eta_0$, then $V_{pi}=1$, where η_0 is some arbitrarily small positive constant.

Based on the above discussion, HK conceptually resembles *K*-means, fuzzy clustering methods, and mixture model algorithms (Dempster et al., 1977; MacQueen, 1967; Wedel & Steenkamp, 1989, 1991). HK, however, differs from existing segmentation techniques in several keys ways. First, like *K*-means, mixture model, and other nonhierarchical procedures, HK follows an iterative process, based on the minimization of within-segment variation, to adjust segment membership. Unlike *K*-means, however, which reassigns segment membership in toto, HK *partially* reassigns segment membership. Consequently, the potential skewing effect of an outlying customer on any one segment is reduced because all customers are partially assigned to all segments. Only when HK terminates does segment membership become unambiguous.

Second, HK is more flexible in its initial segment assignments because, unlike mixture model and fuzzy clustering methods, HK does not require prior segment memberships to sum to one. This property makes HK less sensitive to initial starting conditions (poorly specified seeds), and less likely to return suboptimal solutions when analyzing relatively unstructured data sets. Third, unlike mixture models that use a maximum likelihood estimation approach, HK uses a neural network approach to determine posterior segment membership. As a result, HK can sequentially adjust segment membership and within-segment variation (error) after processing each customer, whereas mixture models adjust cluster membership using pooled error across all *N* customers.

Finally, unlike *K*-means and mixture models, HK does not require a priori rational information in order to perform well (Punj & Stewart, 1983; Wedel & Kamakura, 2000). That is, when used with HK, rational seeds provide no better starting information than random seeds. This occurs because HK initially randomly and partially assigns customers to all segments and updates segment membership after processing each customer. As a result, rational seed locations vary after the first node updates just as if random seeds had been used. Thus, HK is capable of returning optimal solutions even when used with random seeds.

Taken together, the unique properties of HK discussed above theoretically provide several advantages over extant clustering algorithms. Specifically, HK is less sensitive to initial starting conditions (centroid

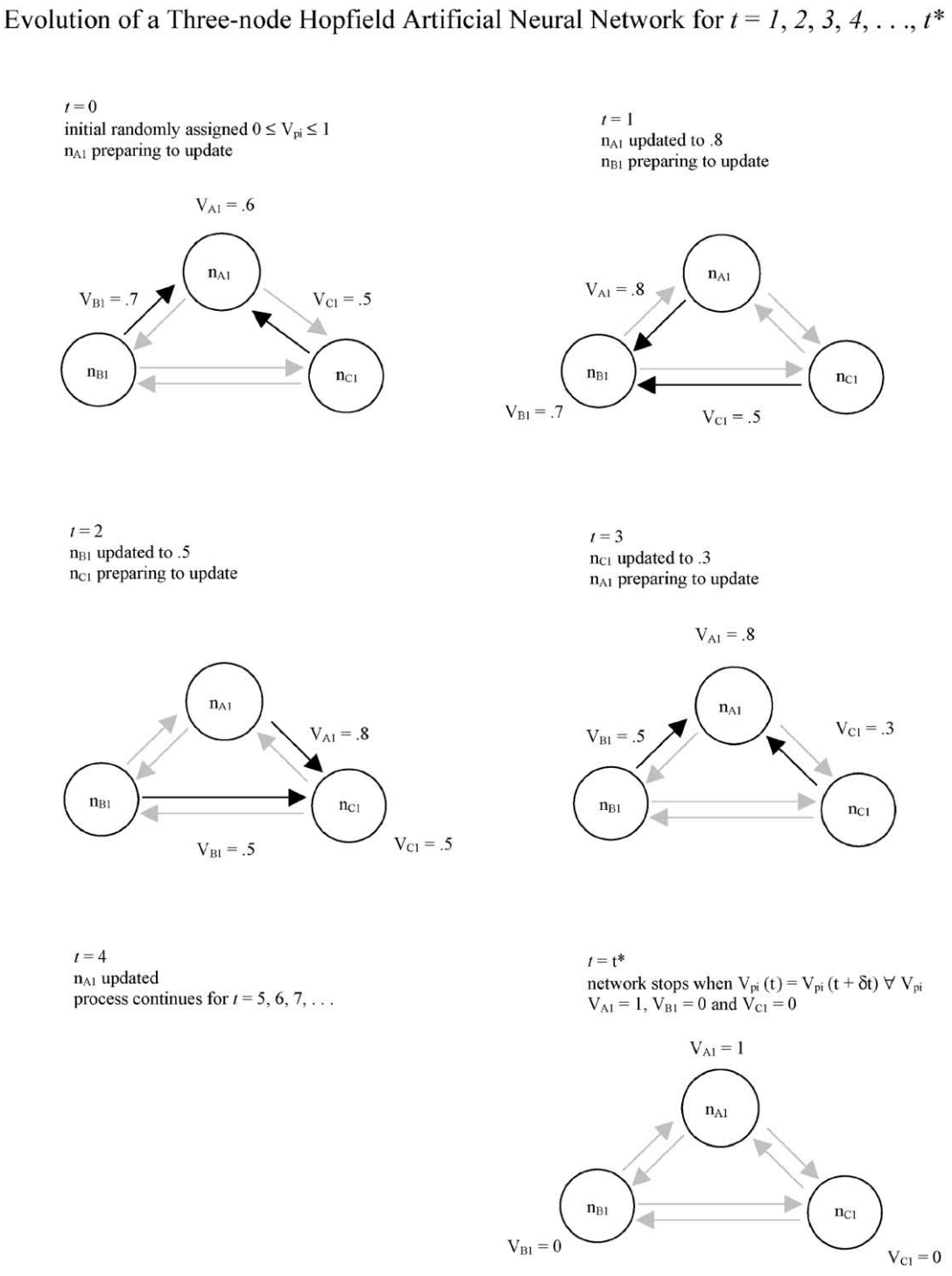

Evolution of a Three-node Hopfield Artificial Neural Network for t = 1, 2, 3, 4, . . ., t*
t = 0
initial randomly assigned 0 ≤ Vpi ≤ 1
nA1 preparing to update
VA1 = .6
nA1
VB1 = .7
VC1 = .5
nB1
nC1
t = 1
nA1 updated to .8
nB1 preparing to update
nA1
VA1 = .8
nB1
nC1
VB1 = .7
VC1 = .5
t = 2
nB1 updated to .5
nC1 preparing to update
nA1
VA1 = .8
nB1
nC1
VB1 = .5
VC1 = .5
t = 3
nC1 updated to .3
nA1 preparing to update
VA1 = .8
nA1
VB1 = .5
VC1 = .3
nB1
nC1
t = 4
nA1 updated
process continues for t = 5, 6, 7, . . .
t = t*
network stops when Vpi (t) = Vpi (t + δt) ∀ Vpi
VA1 = 1, VB1 = 0 and VC1 = 0
VA1 = 1
nA1
nB1
nC1
VB1 = 0
VC1 = 0

Fig. 2.

locations) and can more discriminately assign and reassign segment memberships. Thus, HK does not require rational a priori input information regarding cluster seeds and is more likely to return optimal (i.e., more homogenous) solutions when analyzing relatively unstructured data.

4. Empirical study

In testing the HK algorithm, we used a two-pronged approach. First, we compared HK and traditional clustering algorithms using a real world data set. Second, because the efficacy with which a clustering algorithm returns a market's true structure can only be ascertained when the true cluster structure is known unambiguously, HK and traditional clustering algorithms were compared using a variety of artificial data sets.

4.1. Real world data set

Our real world data set consisted of 4317 customer-level observations on six purchase behavior variables, obtained from a major retailer database. Segmentation variables and descriptive statistics for the data set are presented in Table 2.

When analyzing real world data sets, the number of segments that are actually present is not known with certainty. Therefore, researchers often rely on one or more ''stopping rules'' for determining the most appropriate number of segments to market to. According to Milligan and Cooper (1985), the Pseudo-*F*-test and Cubic Clustering Criterion are among the most accurate stopping rules for recovering the true number of clusters in a data set (Calinski & Harabasz, 1974; Sarle, 1983). Therefore, based on these criteria, the optimal number of segments specified a priori for the real world data set was determined to be five.

Table 2
Segmentation variables and descriptive statistics for real world data set ($N=4317$)

Variable	Mean	Standard deviation	Maximum	Minimum
Number of credit cards used	1.16	0.65	5	0
Account age (months)	39.34	11.22	69	19
Days since first purchase	698.05	211.74	1095	184
Days since last purchase	546.65	255.47	1004	2
Total number of orders	1.81	1.19	9	1
Total dollars spent	172.84	162.07	1468.70	9.00

4.2. Artificially generated data sets

Whether or not a clustering algorithm recovers a market's true structure cannot be known with certainty when analyzing real world data sets. Accordingly, artificial data sets, in which the cluster structure is known a priori, must be generated and used in order to unambiguously compare how well a given segmentation technique captures the true structure of a market.

Following Milligan and Cooper (1985) and Carmone, Kara, and Maxwell (1999), data sets containing four clusters and 2000 observations were generated for three factors. The three factors were varied experimentally using a full factorial research design with three replications. The factors, described below, were selected because they represent data complexities likely to be present in real world data sets, have been previously reported to affect cluster recovery, and represent realistic conditions under which to test the efficacy of various clustering algorithms (Milligan, 1980; Milligan, Soon, & Sokol, 1983).

For example, segmentation studies are often based on large, complex, relatively unstructured data sets that may contain noisy variables, outliers, and unequally sized clusters. Accordingly, our first factor was number of variables (dimensions). This factor was designed to have two levels, each of which represents a relatively large numbers of noise-free and noisy variables. For the first level, there were eight noise-free and one noisy variable. The data were generated such that no cluster overlap existed on the first dimension (variable). For the second level, there were eight noise-free and two noisy variables. Again, no cluster overlap was permitted on the first dimension (variable). Moderate (one noisy variable) and more severe (two noisy variables) levels of noise were included because the presence of noisy variables has been reported to greatly impact cluster recovery (Milligan, 1980). The coordinates for the noisy vari-

ables were random draws taken from a uniform distribution, with the range of the noisy variables set equal to range of the first variable.

The second factor, cluster density, was varied along two levels so that unequally sized clusters would be present in the data. The two levels of density were: (1) equal size clusters, and (2) one large cluster containing 60% of the data points, with the remaining points distributed equally across the remaining clusters. According to Milligan et al. (1983), cluster density can also impact the recovery of the true cluster structure.

The third factor, percentage of outliers in each cluster, also had two levels: (1) no outliers, and (2) 20% outliers. These outliers consisted of data points that exceeded $\pm$ 1.5 standard deviations of the cluster mean on at least one variable, and were generated in addition to the initial 2000 observations. Consequently, data sets contained either 2000 observations (no outliers) or 2400 observations (0.20 outliers).

A $2 \times 2 \times 2$ research design and procedure developed by Milligan (1985) were used to generate eight data sets that contained "natural" clusters exhibiting external isolation and internal cohesion (Cormack, 1971). Each data set was replicated three times, for a total of 24 data sets. Elongated clusters were present in the data. All clusters were embedded in a K-dimensional (9–10 variables) space, with the data points for each cluster drawn from a truncated multivariate normal distribution ($\pm$ 1.5 standard deviations of the cluster mean on each dimension) with uncorrelated variables.

Our general analytical plan was to compare the segmentation solutions returned by HK to those returned by traditional cluster analysis techniques.

4.3. Comparative cluster analysis techniques

Prior researchers have noted that there exist over 50 clustering methods that might be applied to market segmentation problems (Milligan & Cooper, 1985; Wedel & Kamakura, 2000). However, Arabie et al. (1996) and others note that no one clustering technique is generally superior across data sets (Punj & Stewart, 1983; Wedel & Kamakura, 2000). Furthermore, Vriens, Wedel, and Wilms (1996) report that in a test of 13 conjoint segmentation methods, "no single method can be designated as universally preferable" (p. 84), an indication that there is apparently no one "gold standard" for comparative purposes. Accordingly, K-means and a mixture model were selected as the comparative clustering algorithms.

K-means was selected because it has been used as the comparative standard in other, similar studies (Balakrishnan et al., 1994, 1996; Hruschka & Natter, 1999; Natter, 1999) and is currently the most widely used and most popular segmentation technique (Chaturvedi, Carroll, Green, & Rotondo, 1997; Wedel & Kamakura, 2000). A mixture model was selected as the second comparative approach given its statistical basis and growing popularity among researchers (Wedel & Kamakura, 2000).

4.4. Dependent variables

The primary goal of this research is to determine how well HK compares to K-means and mixture models in producing homogeneous market segments and recovering a market's "true" structure. When analyzing real world data, however, a market's true structure cannot be known unambiguously because segment memberships are generally not known a priori. Given that homogenous solutions are generally preferred over less homogenous solutions, the degree of within-segment variation can be used to assess the appropriateness of a given segmentation scheme. Accordingly, total within-segment variation serves as the dependent variable for our real world data set.

Because a market's true structure can be distorted when based on the minimization of within-segment variation (over-fitting the data), an additional assessment of segmentation accuracy is necessary to fully test the efficacy of HK. For example, a "light credit card user" may be incorrectly assigned to the "heavy credit card user" segment because she's an outlier in the number of cards she holds. In order to ascertain if her segment assignment is in fact incorrect, her true segment membership must be known a priori. Accordingly, artificial data sets must be used because only then can "true" segment memberships be known unequivocally and used as a standard for testing segmentation methods.

With artificial data sets, a measure of the accuracy of the segmentation solutions returned by each technique (relative to the referenced, known segment memberships) is the Hubert–Arabie adjusted Rand Index (HARI) (see Carmone, Kara, and Maxwell,

1999; Hubert and Arabie, 1985). The HARI, corrected for chance, is based on the pairwise comparison of all N customers to see which customers are segmented together in the referenced solution versus those segmented together in the solutions being evaluated. Milligan and Cooper (1986), in a Monte Carlo study comparing several other evaluative indices, found that the HARI was the preferred external criterion measure for determining the efficacy of different clustering techniques. Accordingly, the HARI, which reflects segment structure recovery, and total within-segment variation, which reflects segment homogeneity, both serve as the dependent variables for our artificial data sets.

4.5. Procedure

JMP 4.0.2 software (1989–2000 SAS Institute) was used to perform the K-means (MacQueen, 1967) and mixture model analyses for all data sets. For the mixture model, a mixture of normals was used (MacLachlan & Krishnan, 1997). The HK analyses for all data sets were run in a UNIX environment using a C-program provided by Kagmar-Parsi and modified by the authors (Kagmar-Parsi et al., 1990). In executing the HK algorithm, the constants in Eq. (4) were set to $A=B=1$ and $C=0.9/R_{\text{avg}}$ (Kagmar-Parsi et al., 1990). C was rescaled as a function of the average squared within-segment variation, R_{avg}, to ensure good solutions because as R_{avg} gets smaller, C becomes larger and is given more weight in Eq. (4). As a result, reducing within-segment variation becomes more important in finding an optimal solution as the network evolves.

The real world data set was standardized and analyzed using HK, K-means, and the mixture model. The initial seeds for all algorithms were based on the centroid locations obtained from hierarchical clustering using the average method (Milligan & Sokol, 1980; Punj & Stewart, 1983). Additional analyses using random seeds were also conducted. For HK, the random seeds were based on a random 5% subsample of the data set (Kagmar-Parsi et al., 1990). Two hundred fifty analyses were conducted for each algorithm when random seeds were used in order to minimize the impact of any poor initial starting conditions on the final segmentation solutions (Wedel & Kamakura, 2000). Two hundred fifty analyses were also conducted for HK when used with rational seeds because, as previously discussed, such solutions vary due to the initial random assignment of V_{pi} and sequential membership reassignment. ANOVAs comparing the total within-segment variation for each algorithm, seed-type, and their interaction were then conducted.

Each of the 24 artificially generated data sets was standardized and analyzed using HK, K-means, and the mixture model. The initial seeds for K-means and the mixture model were obtained using the average hierarchical clustering method. The initial seeds for HK were obtained from a random 5% subsample of the data set. The HARI for each of the solutions was then calculated using the known cluster memberships resulting from the experimental design (Carmone, Kara, and Maxwell, 1999; Hubert & Arabie, 1985; Milligan, 1985). ANOVAs comparing the HARI for each algorithm, experimental design factor, and their interaction were then conducted. The total within-segment variation for each solution was also determined and similarly compared across algorithm, experimental design factor, and their interaction.

5. Results

The results of our analyses for the real world data set, which consisted of 4317 customer-level observations of six purchase behavior variables, are presented in Tables 3 and 4.

Table 3
Total within-segment variation for real world data set using standardized variables ($N=4317$)

Seed-type	Clustering algorithm		
	HK	K-means	Normal mixtures
Rational (based on hierarchical clustering using average method)	19.29	19.66	29.21
Random (mean across 250 analyses)	19.23	20.47	27.31

The segmentation solutions returned by K-means and normal mixtures, when used with rational seeds, do not change across repeated analyses. Therefore, the total within-segment variation for rationally seeded K-means and normal mixtures reported above is for a single analysis.

Table 4
Segment means by algorithm and seed-type for real world data set

Segment	Number of credit cards used	Account age (months)	Days since first purchase	Days since last purchase	Total number of orders	Total dollars spent	Segment size
(a) HK random seeds (minimum variation solution)							
One	1.6	39.8	797.2	319.2	4.0	425.67	603
Two	1.2	54.8	824.0	659.4	1.6	137.02	648
Three	0.9	31.7	541.0	452.5	1.4	113.53	1173
Four	1.5	48.0	454.4	335.2	1.6	144.16	645
Five	1.0	33.8	858.3	795.8	1.4	139.85	1248
(b) K-means rational seeds							
One	1.1	56.8	712.1	591.9	1.5	133.00	684
Two	1.0	35.0	480.3	417.8	1.4	115.83	1296
Three	0.9	34.7	839.2	794.7	1.4	138.48	1383
Four	2.0	38.2	787.2	333.8	2.4	207.74	605
Five	1.4	41.5	765.6	322.3	4.7	536.43	351
(c) Normal mixtures rational seeds							
One	1.5	40.5	754.8	273.2	4.3	429.05	404
Two	3.1	45.9	725.3	557.4	1.6	163.77	141
Three	4.0	51.5	871.0	188.5	6.0	787.52	12
Four	1.0	38.9	689.4	577.6	1.5	136.20	3729
Five	1.2	41.6	809.7	483.0	4.9	1044.47	31

The rational seeds for *K*-means and normal mixtures were based on the centroid locations obtained from hierarchical clustering using the average method.

Analysis of within-segment variation, collapsed across seed-type, revealed that there are differences in the abilities of HK, *K*-means (KM) and the mixture model (MM) to recover homogenous solutions ($\bar{x}_{HK} = 19.26$, $\bar{x}_{KM} = 20.07$, $\bar{x}_{MM} = 28.26$; HK versus KM: $F(1,996) = 3.39$, $p = 0.07$; HK versus MM: $F(1,996) = 424.85$, $p < 0.00$). Further analysis revealed that the solutions returned by *K*-means and HK did not differ based on the use of random (RAND) versus rational (i.e., those based on hierarchical clustering (HC) algorithms) seeds (HK_{RAND} versus HK_{HC}: $F(1,996) = 0.57$, $p = 0.45$; KM_{RAND} versus KM_{HC}: $F(1,996) = 0.86$, $p = 0.35$). However, the solutions returned by the mixture model were found to be more homogenous when rational seeds were used (MM_{RAND} versus MM_{HC}: $F(1,996) = 4.77$, $p = 0.03$) (see Milligan and Sokol, 1980; Punj and Stewart, 1983).

The overall minimum variation solution (17.85) was returned by HK using random seeds. When random seeds were used across algorithms, HK was significantly better than *K*-means or the mixture model in returning minimum variance solutions (HK_{RAND} versus KM_{RAND}: $F(1,996) = 253.92$, $p < 0.00$; HK_{RAND} versus MM_{RAND}: $F(1,996) = 10828.51$, $p < 0.00$). When rational seeds were used, HK and *K*-means returned solutions with similar variation (HK_{HC} versus KM_{HC}: $F(1,996) = 0.18$, $p = 0.67$). However, rationally seeded HK was found to be significantly better than the rationally seeded mixture model (HK_{HC} versus MM_{HC}: $F(1,996) = 130.07$, $p < 0.00$). No differences were found between HK, used with random seeds and *K*-means, used with rational seeds (HK_{RAND} versus KM_{HC}: $F(1,996) = 0.24$, $p = 0.62$).

Finally, examination of Table 4 reveals that the segmentation results returned by each of the three clustering algorithms are dissimilar. These differences occur for several reasons. First, the real world data set is large, complex, noisy, and contains poorly defined, overlapping segments. Second, the mixture model and HK partially assign segment membership whereas *K*-means assigns membership in toto. Third, the mixture model adjusts segment membership

based on pooled error while HK adjusts segment membership after each node updates. Accordingly, the mixture model is less capable of strongly revising posterior segment membership in messy data environments and more likely to return a suboptimal solution. Similarly, *K*-means is more likely to return a suboptimal solution under such conditions because it is highly susceptible to poor initial seed locations. As a consequence, the solutions returned by each algorithm are likely to vary, with HK returning a more homogeneous solution.

The above findings thus suggest that when analyzing complex, real world data sets, *randomly* seeded HK recovers, on average, segmentation solutions that are as good as or better than those returned by *K*-means and mixture models using computationally intensive *rational* seeds. Furthermore, the optimal solution across multiple runs was returned by randomly seeded HK, and not by rationally or randomly seeded *K*-means or mixture models (Wedel & Kamakura, 2000). However, to fully test HK and explore boundary conditions under which HK may under- or out-perform the comparative clustering algorithms, the underlying structure of the data set must be known. Accordingly, we also analyzed artificial data sets containing known cluster structures (large number of observations and variables, outliers, noisy variables, and unequally sized clusters) using HK, *K*-means, and the mixture model. Our results are presented in Table 5.

Table 5

Algorithm	Cluster density		Outliers		Variables and noisy variables	
	Equal	0.60	None	0.20	8+1 Noisy	8+2 Noisy
(a) Average HARI by clustering algorithm and data complexity for N= 72 artificial data sets						
HK	0.92	0.74	0.93	0.72	0.84	0.82
K-means	0.80	0.87	0.92	0.75	0.82	0.85
Normal mixtures	0.66	0.73	0.92	0.48	0.68	0.72
(b) Average total within-segment variation by clustering algorithm and data complexity for N= 72 artificial data sets						
HK	21.82	24.76	20.29	26.31	21.39	25.20
K-means	23.87	29.13	20.34	32.66	26.24	26.77
Normal mixtures	43.95	43.64	20.40	67.20	42.17	45.42

Analysis of the HARI for the artificial data sets, collapsed across cluster structures, indicated that there are differences in the abilities of the three clustering algorithms to recover a market's true structure (see Table 5(a)). HK, used with random seeds, was found to be as accurate as *K*-means, used with rational seeds ($\bar{x}_{\text{HK-RAND}} = \bar{x}_{\text{KM-HC}} = 0.83$; $F(1,69) = 0.02$, $p = 0.89$). Both HK and *K*-means afforded a statistically significant 13% improvement in cluster recovery accuracy over the rationally seeded mixture model ($\bar{x}_{\text{MM-HC}} = 0.70$; HK versus MM: $F(1,69) = 5.56$, $p = 0.02$).

Further analysis of the HARI revealed that when cluster densities were equal, HK outperformed *K*-means and the mixture model in cluster structure recovery (HK versus *K*-means: $F(1,66) = 2.54$, $p = 0.11$; HK versus MM: $F(1,66) = 11.35$, $p < 0.00$). However, when one large cluster was present, *K*-means outperformed HK in cluster structure recovery (HK versus KM: $F(1,66) = 3.21$, $p = 0.07$).

When no outliers were present, HK, *K*-means, and the mixture model were equally accurate in recovering the true cluster structure (HK versus KM and MM: $F(1,66) = 0.19$, $p = 0.67$). When outliers were present, randomly seeded HK was found to be as accurate as rationally seeded *K*-means (HK versus KM: $F(1,66) = 0.48$, $p = 0.49$), and significantly more accurate than the rationally seeded mixture model (HK versus MM: $F(1,66) = 25.93$, $p < 0.00$), in recovering the cluster structures.

When one noisy variable was present, HK was as accurate as *K*-means (HK versus KM: $F(1,66) = 0.04$, $p = 0.84$), and significantly more accurate than the mixture model (HK versus MM: $F(1,66) = 4.02$, $p = 0.05$), in recovering the true cluster structure. No differences between HK, *K*-means, and the mixture model were found in structure recovery when two noisy variables were present (HK versus KM: $F(1,66) = 0.15$, $p = 0.70$; HK versus MM: $F(1,66) = 1.61$, $p = 0.21$).

In order to assess the efficacy with which HK, *K*-means and the mixture model would return managerially desirable, homogenous market segments in complex data environments, we also analyzed total within-segment variation collapsed across cluster structure (see Table 5(b)). Analysis revealed that HK, used with random seeds, was as effective as rationally seeded *K*-means, and significantly more effective than the rationally seeded mixture

model, in recovering homogenous market segments ($\bar{x}_{\text{HK-RAND}} = 23.30$, $\bar{x}_{\text{KM-HC}} = 26.50$, $\bar{x}_{\text{MM-HC}} = 43.80$; HK versus KM: $F(1,69) = 0.50$, $p = 0.48$; HK versus MM: $F(1,69) = 20.59$, $p < 0.00$).

Further analysis of within-segment variation revealed that when cluster densities were equal, HK was as effective as *K*-means, and significantly more effective than the mixture model, in recovering homogeneous segments (HK versus KM: $F(1,66) = 0.10$, $p = 0.75$: HK versus MM: $F(1,66) = 11.63$, $p < 0.00$). Similar results were also found when one large cluster was present (HK versus KM: $F(1,66) = 0.45$, $p = 0.50$: HK versus MM: $F(1,66) = 8.46$, $p < 0.00$).

When no outliers were present, HK, *K*-means, and the mixture model were equally effective in recovering minimum variation solutions (HK versus KM and MM: $F(1,66) = 0.00$, $p = 0.97$). However, when outliers were present, HK was significantly better than both *K*-means and the mixture model in recovering homogeneous segments (HK versus KM: $F(1,66) = 6.10$, $p = 0.02$; HK versus MM: $F(1,66) = 252.49$, $p < 0.00$). Finally, when one noisy variable was present, HK was as effective as *K*-means (HK versus KM: $F(1,66) = 0.55$, $p = 0.45$), and significantly more effective than the mixture model (HK versus MM: $F(1,66) = 10.21$, $p < 0.00$), in recovering minimum variation solutions. Similar results were also found when two noisy variables were present (HK versus KM: $F(1,66) = 0.06$, $p = 0.81$; HK versus MM: $F(1,66) = 9.67$, $p < 0.00$).

6. General discussion

Our research has been motivated by several simple observations. First, the use of segment-specific marketing mixes is widespread among retailers. Second, increasing access to individual-level customer information has led to the compilation of large databases that are likely to contain hundreds of variables and data complexities such as outliers, noise, and unequally sized clusters. Third, the current analytical techniques used to analyze such databases are strained under such conditions, making the task of recovering managerially desirable homogenous segments more difficult.

Given the nontrivial costs associated with incorrectly segmenting a marketplace, new segmentation techniques that better enable retailers to fully exploit customer information need to be investigated. Wedel and Kamakura (2000) note that empirical generalizations as to the universal superiority of one technique over another are impossible. These and other authors therefore suggest that when segmenting a market, careful consideration should be given to choosing a methodology and segmentation process that optimally matches one's purpose and the nature of available data (Punj & Stewart, 1983; Vriens et al., 1996). For example, the surprisingly poor performance of the mixture model in our research may have been due to the presence of poorly defined, poorly separated segments, numerous local maxima, and other violations of its underlying assumptions (Arabie et al., 1996; Wedel & Kamakura, 2000).

Our research agenda has been to investigate the Hopfield–Kagmar (HK; Kagmar-Parsi et al., 1990) clustering algorithm, an artificial neural network (ANN)-based segmentation technique that appears to effectively segment the types of customer markets likely to be encounter by retailers. The results of our empirical analyses indicate that HK is less sensitive to initial starting conditions (centroid locations), and therefore unlike *K*-means and mixture models, does not require rational seeds in order to recover optimal (i.e., minimum variation) segmentation solutions. Thus, researchers using HK to segment a market can avoid the computational burdens associated with obtaining hierarchical-clustering-based seeds and the biases and inaccuracies that may be associated with subjectively specified seeds.

These findings have important managerial applications because, in practice, the improper selection or application of a segmentation technique can be of significant financial impact. For example, consider a database that contains information on one million customers. Based on our findings, the 13% improvement in segmentation accuracy afforded by HK or *K*-means over mixture models would result in 130,000 customers being more effectively targeted. Given that the costs associated with a targeted marketing campaign are nontrivial, a more accurate segmentation solution could potentially save the firm thousands of dollars that would have otherwise been wasted on inappropriate and ineffective marketing efforts geared towards improperly segmented customers.

K-means and mixture models, however, may be less appropriate than HK for analyzing such a database because they do not perform as well when used with random seeds. That is, we found that computationally demanding hierarchical-clustering-based seeds, which are impractical for a million-customer data set, *must* be used with *K*-means in order to recover segments as homogenous as those returned by HK using random seeds. Furthermore, we found that HK recovered more homogenous segments than those recovered using mixture models, regardless of the seed-type used. Thus, practitioners using HK would benefit by being able to more effectively identify and target cohesive segments that may be more responsive too different marketing mixes.

Despite the advantages discussed above, our results also indicated that HK did not perform as well as *K*-means in recovering the true structure of a data set when one large cluster was present. However, this shortcoming was offset by HK's ability to more accurately recover the true cluster structure when clusters were equally sized, and to return more homogenous solutions than *K*-means when outliers were present.

7. Limitations and directions for future research

Although this research advances our understanding of the Hopfield–Kagmar (HK) clustering algorithm as an alternative segmentation methodology, it is not without limitations. Given that within-segment variation is the optimization function being minimized, HK may over-fit the data. Thus, in order to more fully assess the performance of HK, it should be evaluated across more real world data sets and validated by comparing HK-based segmentation schemes with actual segmentation outcomes. Also, additional research should be conducted that investigates the sensitivity of the resulting segmentation solutions to the parameters used to execute HK. Together, such research would provide researchers and practitioners with additional insights into HK's robustness, benefits, and limitations.

Clearly, the use of artificial neural networks in general, and in particular Hopfield (1982) networks, in solving marketing-related problems warrants further investigation.

Acknowledgements

The authors would like to thank Heath Holcomb, Behrooz Kagmar-Parsi, Jack Meredith, and Steve Opalanko for their technical assistance and helpful comments. The authors would also like to thank the editor and the three anonymous reviewers whose comments substantially improved the quality of the paper. This research was supported by the Babcock Graduate School of Management, Wake Forest University Research Fellowship Program.

References

Arabie, P., Hubert, J., & De Soete, G. (1996). *Clustering and classification.* River Edge, NJ: World Scientific Publishing.

Balakrishnan, P. V., Cooper, M. C., Jacob, V. S., & Lewis, P. A. (1994). A study of the classification capabilities of neural networks using unsupervised learning: a comparison with *K*-means clustering. *Psychometrika*, *59*(4), 509–525.

Balakrishnan, P. V., Cooper, M. C., Jacob, V. S., & Lewis, P. A. (1996). Comparative performance of the FSCL neural net and *K*-means algorithm for market segmentation. *European Journal of Operational Research*, *93*, 346–357.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*, 1–27.

Carmone, F. J., Kara, A., & Maxwell, S. (1999). HINoV: a new model to improve market segment definition by identifying noisy variables. *Journal of Marketing Research*, *36*(4), 501–509.

Chaturvedi, A., Carroll, J. D., Green, P. E., & Rotondo, J. A. (1997). A feature-based approach to market segmentation via overlapping *K*-centroids clustering. *Journal of Marketing Research*, *34*, 370–377.

Cormack, R. M. (1971). A review of classification. *Journal of the Royal Statistical Society*, *134*(A), 321–367.

Dasgupta, C. G., Dispensa, G. S., & Ghose, S. (1994). Comparing the predictive performance of a neural network model with some traditional market response models. *International Journal of Forecasting*, *10*, 235–244.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM-algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Direct Marketing Association (1999). *Statistical fact book*. New York: Direct Marketing Association.

Gurney, K. (1997). *An introduction to neural networks.* Padstow, UK: T.J. International.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2554–2588.

Hruschka, H., & Natter, M. (1999). Comparing performance of feedforward neural nets and *K*-means for cluster-based market segmentation. *European Journal of Operational Research*, *114*, 346–353.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Kagmar-Parsi, B., Gualtieri, J. A., Devaney, J. E., & Kagmar-Parsi, B. (1990, Spring). Clustering with neural networks. *Biological Cybernetics*, *63*, 201–208.

Kagmar-Parsi, B., & Kagmar-Parsi, B. (1992). Hopfield model and optimization problems. *Neural networks for perception Computation, learning, and architectures, vol. 2*. Boston: Academic Press.

Kagmar-Parsi, B., Kagmar-Parsi, B., & Sciortino Jr., J.C., 1996. Automatic data sorting using neural network techniques. *Naval Research Laboratory,* NRL/FR/5720—96-9803, Washington, DC.

Kohonen, T. (1989). *Self-organization and associative memory.* 3rd ed. New York: Springer.

Krishnamurthi, A. K., Ahalt, S. C., Melton, D. E., & Chen, P. (1990). Neural networks for vector quantization of speech and images. *IEEE Journal on Selected Areas in Communications*, *8*, 1449–1457.

Krycha, K. A., & Wagner, U. (1999). Applications of artificial neural networks in management science: a survey. *Journal of Retailing and Consumer Services*, *6*, 185–203.

MacLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions*. New York: Wiley.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*, 281–297.

Mangiameli, P., Chen, S. K., & West, D. (1996). A comparison of SOM neural network and hierarchical clustering methods. *European Journal of Operational Research*, *93*, 402–417.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbations on fifteen clustering algorithms. *Psychometrika*, *45*, 325–342.

Milligan, G. W. (1985). An algorithm for generating artificial test clusters. *Psychometrika*, *50*, 123–127.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, *50*, 159–179.

Milligan, G. W., & Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research*, *21*, 441–458.

Milligan, G. W., & Sokol, L. M. (1980). A two-stage clustering algorithm with robust recovery characteristics. *Educational and Psychological Measurement*, *40*, 755–759.

Milligan, G. W., Soon, S. C., & Sokol, L. M. (1983). The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PamI-5*(1), 40–47.

Natter, M. (1999). Conditional market segmentation by neural networks: a Monte-Carlo study. *Journal of Retailing and Consumer Services*, *6*, 237–248.

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: review and suggestions for applications. *Journal of Marketing Research*, *20*, 134–148.

Sarle, W.S., 1983. *SAS Technical Report A-108,* The cubic clustering criterion. The SAS Institute, Cary, NC.

Shapiro, C., & Varian, H. R. (1999). Information rules: a strategic guide to the network economy. Boston, MA: Harvard Business School Press.

Vriens, M., Wedel, M., & Wilms, T. (1996). Metric conjoint segmentation methods: a Monte Carlo comparison. *Journal of Marketing Research*, *33*, 73–85.

Wedel, M., & Kamakura, W. (2000). *Market segmentation: conceptual and methodological foundations*. Norwell, MA: Kluwer Academic Publishing.

Wedel, M., & Steenkamp, J. B. (1989). Fuzzy clusterwise regression approach to benefit segmentation. *International Journal of Research in Marketing*, *6*, 241–258.

Wedel, M., & Steenkamp, J. B. (1991). A clusterwise regression model for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research*, *28*, 385–396.