PERGAMON

Computers & Operations Research 29 (2002) 1475–1493

computers & operations research

www.elsevier.com/locate/dsw

Integration of self-organizing feature map and K-means algorithm for market segmentation

R.J. Kuo[a,*], L.M. Ho[b], C.M. Hu[c]

[a]Department of Industrial Engineering, National Taipei University of Technology, Taipei, Taiwan 106, ROC

[b]Department of Leisure, Recreation and Tourism Management, Shu-Te University, Kaohsiung County, Taiwan 824, ROC

[c]Division of System Information, Foxconn Industrial PCE Production Group, Hsing-Chu, Taiwan, ROC

Received 1 July 2000; received in revised form 1 February 2001

Abstract

Cluster analysis is a common tool for market segmentation. Conventional research usually employs the multivariate analysis procedures. In recent years, due to their high performance in engineering, artificial neural networks have also been applied in the area of management. Thus, this study aims to compare three clustering methods: (1) the conventional two-stage method, (2) the self-organizing feature maps and (3) our proposed two-stage method, via both simulated and real-world data. The proposed two-stage method is a combination of the self-organizing feature maps and the K-means method. The simulation results indicate that the proposed scheme is slightly better than the conventional two-stage method with respect to the rate of misclassification, and the real-world data on the basis of Wilk's Lambda and discriminant analysis.

Scope and purpose

The general idea of segmentation, or clustering, is to group items that are similar. A commonly used method is the multivariate analysis [4]. These methods consist of hierarchical methods, like Ward's minimum variance method, and the non-hierarchical methods, such as the K-means method. Owing to increase in computer power and decrease in computer costs, artificial neural networks (ANNs), which are distributed and parallel information processing systems successfully applied in the area of engineering, have recently been employed to solve the marketing problems. This study aims to discuss the possibility of integrating ANN and multivariate analysis. A two-stage method, which first uses the self-organizing feature maps to determine the number of clusters and the starting point and then employs the K-means method to find the final solution, is proposed. This method provides the marketing analysts a more sophisticated way to

* Corresponding author. Tel.: + 886-2-2771-2171; fax: + 886-2-2731-7168.

E-mail address: rjkuo@ntut.edu.tw (R.J. Kuo).

PII: S0305-0548(01)00043-0

analyze the consumer behavior and determine the marking strategy. A case study is also employed to demonstrate the validity of the proposed method.

Keywords: Cluster analysis; Market segmentation; Self-organizing feature maps; *K*-means

1. Introduction

Market segmentation has become one of the fundamental concepts of marketing since first presented by Wendal Smith in 1956 [1]. Faced with heterogeneous markets, a firm could increase the expected profits by following the market segmentation strategy. For many decades, a number of market segmentation methods have been proposed for solving this problem. However, the advancement of market segmentation research requires a narrowing of the gap between the academically oriented research on segmentation and its real-world application [2].

The general idea of segmentation is to group items that are similar. These items can be people, species of plants, parts [3], or signals. A commonly used method is the multivariate analysis [4]. Owing to increase in computer power and decrease in computer costs, artificial neural networks (ANNs) which are distributed, and parallel information processing systems successfully applied in the area of engineering, have recently been employed to solve the marketing problems.

From a manager's point of view, ANNs could be applied to solve the problems which have been conventionally tackled by statistical methods, such as sales forecasting [5], bankruptcy prediction, stock market forecasting [6], the travelling salesman problem, and grouping or clustering. In recent years, ANNs have also been applied to market segmentation [7].

Therefore, this study aims to discuss the possibility of integrating ANN and multivariate analysis. A two-stage method was suggested by Punj and Steward [4]. This method consists of a hierarchical method, like Ward's minimum variance method, followed by a non-hierarchical method, such as the *K*-means method. In the present study, a modified two-stage method, which first uses the self-organizing feature maps to determine the number of clusters and the starting point and then employs the *K*-means method to find the final solution, is proposed. Both numerical simulations and real-world problem data are used to validate the feasibility of the proposed method. The simulation results show that the proposed two-stage method is slightly more accurate than the conventional two-stage method (Ward's minimum variance method followed by the *K*-means method) with respect to the rate of misclassification. Similarly, the real-world problem results also indicate that the proposed method is more superior on the basis of both Wilk's Lambda and discriminant analysis. In addition, the results of using self-organizing feature maps alone are found to be the worst compared with both the modified and conventional two-stage methods.

The rest of this paper is organized as follows. Section 2 discusses the general idea of market segmentation and applications of ANNs in marketing segmentation, while the proposed two-stage method is described in Section 3. Section 4 presents the simulation algorithm and results. The real-world problem results are detailed in Section 5. Finally, the concluding remarks are made in Section 6.

2. Background

2.1. Market segmentation

Mass marketing will no longer exist in the coming century. The marketing strategy has moved from mass marketing to target marketing. Owing to multiple requests from different customers, it is impossible to satisfy every customer's needs. Therefore, in order to enhance the customers' satisfaction, the firm has to divide the market into some sub-markets and select one with the largest profits. This idea, market segmentation, was first presented by Wendel Smith [1]. It is the first procedure of target marketing, following positioning and targeting. For the detailed discussion of market segmentation, refer to [8,9].

Kotler [10] claimed that to assure different market segments, three stages, survey stage, analysis stage, and profiling stage, are involved. Wind [2] also discussed similar steps. Kolter pointed out that an effective market segmentation should possess the following five characteristics: (1) measurability, (2) substantiality, (3) accessibility, (4) differentiability, and (5) actionability. Though there are many different clustering methods available, a good clustering result should fit the following six criteria [11]: (1) homogeneous within, (2) heterogeneous between, (3) substantiality, (4) operational, (5) availability and reliability, and (6) enough similarity within each group.

2.2. Artificial neural networks in market segmentation

An artificial neural network (ANN) is a system which has been derived through models of neurophysiology. In general, it consists of a collection of simple non-linear computing elements whose inputs and outputs are tied together to form the network.

The learning algorithms of ANNs can be divided into two categories: supervised and unsupervised [12]. In supervised learning, the network has its output compared with a known answer, and receives feedback about any errors. This is sometimes called learning with a teacher; the teacher tells the network what the correct answer is. For a supervised ANN, both inputs and outputs are necessary for training the network, while an unsupervised ANN requires only the inputs. The network must discover for itself patterns, features, and correlations in the input data, and code for them in the output. The units and connections must thus display some degree of self-organization. The most widely applied unsupervised learning scheme is Kohonen's feature maps. Some research has shown that the two learning algorithms mentioned above can be combined and it is called the hybrid learning.

A number of studies have successfully demonstrated the learning capability of ANNs and their applications in the area of engineering. Though most investigations are interested in the performance of ANNs as compared with the conventional statistical methods [33], these networks have been employed recently to solve the management problems. The results are very promising. Proctor [13] indicated that ANN is an alternative expert system for the solution of marketing decision problems. Two examples shown are the sales forecasting and new product evaluation. Similarly, Venugopal and Baets [14] presented the possible applications of ANNs in marketing management. Three examples, retail sales forecasting, direct marketing and target marketing, were employed to demonstrate the capability of ANNs. Besides, more works on this can be found in [15–17].

Regarding the market segmentation, Venugopal and Baets [14] presented a network with six inputs and three outputs. The inputs are six attributes of market including demographic information, socio-economic information, geographic location, purchase behavior information, consumption behavior information, and attitude to product, while the three outputs represent three segments. Either the Adaptive Resonance Theory Models (ART-1 and ART-2) or self-organizing feature maps can be used for clustering. Fish [18] also suggested that ANNs could be utilized for market segmentation. Bigus [19] suggested that ANNs can be employed as a tool for data mining and presented a network with three different dimensions of data, population (sex, age, and marriage), economic information (salary and family income), and geographic information (states, cities, and level of civilization). Balakrishnan [20] compared self-organizing feature maps with the K-means method. The results reveal that the K-means method has a higher rate of classification through the Monte Carlo algorithm. Two years later, Balakrishnan [7] employed the frequency-sensitive competitive learning algorithm (FSCL) and the K-means method for clustering the simulated data and real-world problem data. Also, the combination of these two methods was presented. Neither the simulated nor the real-world problem data can determine which method alone is better. However, the combination of the two methods seems to provide a better managerial explanation for the brand choice data.

3. Methodology

Though a number of clustering methods have been proposed to solve the market segmentation problem, each carries its own advantages and shortcomings [4]. Even with the recent application of computational intelligence techniques, e.g., artificial neural networks, in the domain of cluster-

Table 1
A comparison of different clustering methods

Characteristics Clustering method	Advantages	Disadvantages
Artificial neural networks (Self-organizing feature maps)	Natural start Can handle large amounts of data	Longer computational time Difficult to set up the training parameters and different parameters setup gives different results
Hierarchical methods (Ward's minimum variance method)	Can determine the number of clusters	Cannot handle large amounts of data Easily affected by the outliers No recovery
Non-hierarchical methods (K-means method)	Can have higher accuracy if the starting point and the number of clusters are provided Can handle large amounts of data	Cannot determine the number of clusters Select randomly the starting point and the number of clusters. It may select two centers, which belong to the same group.

ing, the problem still exists. Table 1 lists three clustering methods, ANNs, Ward's minimum variance method (hierarchical clustering), and the *K*-means method (non-hierarchical clustering), and their corresponding characteristics. It reveals that further improvement is still necessary. Therefore, the integration of these methods becomes desirable. One example proposed by Punj [4] is to combine Ward's minimum variance method and the *K*-means method. This forms a new clustering method called the two-stage clustering method. The main reason for such integration is that Ward's minimum variance method can provide the number of clusters which the *K*-means method requires. That way, the number of clusters need not be assumed by the researcher. Besides, the starting point is not randomly selected. This becomes more practical in meeting the requirements of industries.

Thus, this study aims to examine in detail three clustering methods through both simulated data and real-world problem data. One is the conventional two-stage method (Ward's minimum variance method and *K*-means method) proposed by Punj [4], another is the artificial neural network (self-organizing feature maps), and the other is the proposed two-stage method (self-organizing feature maps and *K*-means method). Though integration of unsupervised neural network with the *K*-means method was conceptually presented in [7], no detailed comparison has yet been made through simulated data.

In the following subsections, the related clustering methods will be discussed.

3.1. Multivariate clustering methods

The multivariate clustering algorithms are a class of data reduction techniques. Technically, it can be divided into two categories: (1) hierarchical methods and (2) non-hierarchical methods. The hierarchical clustering methods can be further classified into two types: (1) agglomerative hierarchical methods and (2) divisive hierarchical methods. Agglomerative hierarchical algorithm starts with n clusters, where n is the number of observations. The distance between observations is calculated. The two closest points are merged into a cluster. The process continues until all observations are in one cluster. Then the decision rule is used to determine the number of clusters [21]. On the other hand, the divisive hierarchical method involves the opposite processes. There are several methods for hierarchical clustering, like the linkage methods which include Ward's minimum variance method [22], single linkage method, average linkage method, complete linkage method, and so on.

Punj and Steward [4] suggested that the integration of the hierarchical method with the non-hierarchical one is a feasible solution for clustering. Empirical studies of the clustering algorithm performance suggest that one of the iterative partitioning methods is preferable to the hierarchical methods. This holds true only when a non-random starting point can be specified. In addition, iterative partitioning methods require prior specification of the number of clusters desired, while hierarchical methods do not need such specification. Thus, the researcher is confronted with determining both an initial starting point and the number of clusters in order to use the non-hierarchical methods, like the *K*-means method. Therefore, first, the hierarchical methods, which have demonstrated superior performance, such as Ward's minimum variance method, can be applied to obtain a rough solution. Such a solution can suggest the number of clusters and also provide the starting point. Then the non-hierarchical methods, like the *K*-means method, can use the information to obtain the final clustering results. In this study, Ward's

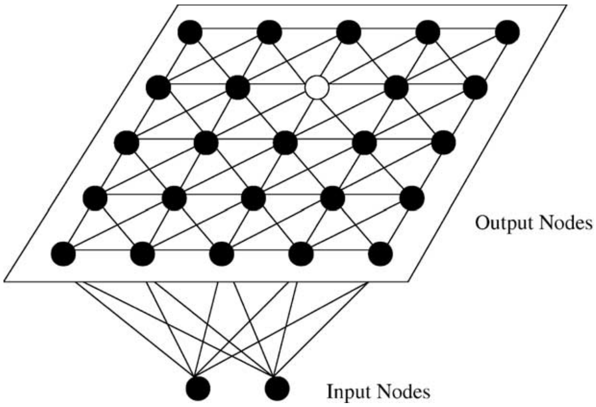

Fig. 1. The Kohonen's feature maps.

minimum variance method is used to determine the initial information for the K-means method while the K-means method will determine the final clusters.

3.2. Self-organizing feature maps

For supervised learning, both inputs and outputs are necessary for training the network, while the unsupervised learning needs only the inputs. The network must discover for itself patterns, features, and correlations in the input data and code for them in the output. The units and connections must thus display some degree of self-organization. The most widely used unsupervised learning scheme is the self-organizing feature maps developed by Kohonen as shown in Fig. 1 [23].

The learning rule is

$$\Delta w_{i,j} = \eta \Lambda(i, i^*)(\xi_j - w_{i,j}) \tag{3}$$

for all i and j. The neighborhood function $\Lambda(i, i^*)$ is 1 for $i = i^*$ and falls off with distance $|r_i - r_{i^*}|$ between unit i and i^* in the output array. A typical choice for $\Lambda(i, i^*)$ is

$$\Lambda(i, i^*) = \exp\left(\frac{|r_i - r_{i^*}|^2}{2\sigma^2}\right), \tag{4}$$

where σ is a width parameter that is gradually decreased.

3.3. Proposed two-stage method

As suggested by Punj and Steward [4], the integration of hierarchical and non-hierarchical methods can provide a better solution. The reason is that hierarchical methods, like Ward's minimum variance method, can determine the candidate number of clusters and starting point that non-hierarchical methods, like the K-means method, need, while non-hierarchical methods can provide better performance with the specified information. In this study, we propose that the computational intelligence techniques, self-organizing feature maps, can replace Ward's minimum variance method or the average linkage method. The main reason is that the first stage of the conventional two-stage clustering methods always involves the hierarchical methods. One of their shortcomings is non-recovery. Once an observation has been assigned to a cluster, it should not be moved at all. However, self-organizing feature maps are a kind of learning algorithm, which can continually update, or reassign, the observation to the closest cluster. From the final output array, the researcher can easily determine the candidate number of clusters as well the starting point. On the other hand, self-organizing feature maps can always converge very fast.

4. Simulation

Clustering methods have been presented and evaluated exclusively in the literature [4,24]. Though these techniques are optimal for some specific distributional assumption or dimensionality, further study is still necessary for determining their robustness to data which do not satisfy the assumed structure. However, in the real-world problems, it is quite difficult to determine which clustering method is the best, since the true, or real, clustering solutions are unknown. Thus, the bulk of the validation literature tries to solve this problem through the Monte Carlo framework. One of the main advantages is that the researcher can use the analytical data with a known structure.

A number of different schemes for generating artificial data have been presented in the literature, like [25–27,21]. The present study adopts the algorithm proposed by Milligan [21]. It has been used in several studies that examine the properties of the clustering algorithm [20,7]. Then the simulated data with known cluster solutions can be applied for validation purposes.

4.1. Algorithm for generating simulated data

Several factors can affect the quality of cluster recovery, e.g., the number of clusters in the data, the number of dimensions used to describe the data, and the level of errors in the data [27]. The simulated data sets for this study contain either 2, 3, 4, or 5 distinct non-overlapping clusters with a roughly equal distribution of points in each cluster. The number of dimensions is varied so that all points in a data set are described by a four-, six-, or eight-dimensional space, while the three levels of error are no error, low error and high error. A full factorial design is taken. Therefore, the result is a $4 \times 3 \times 3$ design with three replications per cell. Totally there are 108 data sets. Each set contains 100 data points, while only 50 data points were used in [7]. The simulated data are generated as truncated, multivariate normal distributions in Euclidean space. The data sets would have reasonably distinct and separate clusters as defined by internal cohesion and external isolation [28]. The algorithm is summarized in the following [21]:

Step 0: Initialize the random number generator, constants, and array.
Step 1: Determine the number of replications, the number of dimensions, and the levels of error.
Step 2: Create non-overlapping cluster boundaries for the first dimension of the variable space.
Step 3: Determine the cluster boundaries for the remaining dimensions of the variable space and a random ordering of the clusters in each dimension.

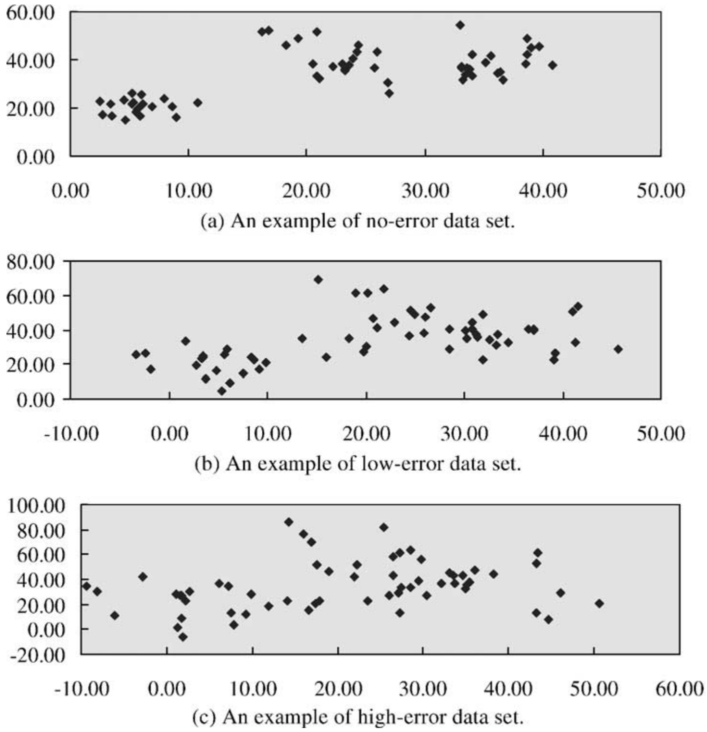

Fig. 2. Three examples of data sets of different error level.

Step 4: Generate the points within cluster.
Step 5: Generate outliers to each cluster.
Step 6: Error perturbation of the coordinates.
Step 7: Generate additional noise dimensions.
Step 8: Standardize the dimensions.
Step 9: Generate the random noise data sets.

Fig. 2 displays one example of the data sets with no error, low error and high error in two dimensions.

4.2. Simulation results

This section will examine the performance of three clustering methods, the conventional two-stage method, the self-organizing feature maps, and the proposed two-stage method. SPSS [29] is utilized for Ward's minimum variance method and the K-means method. Self-organizing feature maps are programed using Visual C + + language. The number of learning epochs and the training rate are set as 1000 and 0.5, respectively.

The results of the three methods are compared with respect to the overall cluster recovery and sensitivity to cluster characteristics. The percent of the 100 observations misclassified by the

Table 2
The simulated results of ANOVA under different factors

Control variable	Self-organizing feature maps	Conventional two-stage method	Proposed two-stage method
Main effect	0.001[a]	0.000[a]	0.000[a]
Clustering number (CN)	0.232	0.002[a]	0.000[a]
Dimension (D)	0.218	0.108	0.094
Error perturbation level (EPL)	0.000[a]	0.000[a]	0.000[a]
EPL × CN	0.492	0.001[a]	0.084
EPL × D	0.315	0.239	0.000[a]
CN × D	0.826	0.67	0.298
EPL × CN × D	0.841	0.015[a]	0.485

[a]Means significant difference ($\alpha = 0.05$).

Table 3
The mean misclassification rates under different levels for each factor (%)

	Level	Self-organizing feature maps	Conventional two-stage method	Proposed two-stage method
Total average		1.7778	0.5833	0.5463
Error	No	0	0	0
Perturbation	Low	0.58	0.31	0.18
	High	4.67	1.44	1.47
Clustering	2	2.59	1.19	1.48
Number	3	2.74	0.70	0.41
	4	1.15	0.3	0.30
	5	0.52	0.14	0
Data	4	2.72	0.86	0.81
Dimension	6	1.86	0.33	0.28
	8	0.75	0.56	0.55

different methods indicates cluster recovery performance. Also, SPSS is employed to perform the analyses of variance (ANOVAs) for testing the hypotheses about the effects of the data characteristics on the clustering results. Tables 2 and 3 show the simulation results.

Hypothesis 1: The percentage of misclassifications does not differ across the number of clusters in the data set.

This hypothesis was confirmed by both the conventional and proposed two-stage clustering methods and not by the self-organizing feature maps. As the number of clusters increases, cluster recovery becomes better. The reason is that in the data generation algorithm, the first dimension is non-overlapping. The larger the number of clusters is, the longer the length of the first dimension is. In addition, the lengths of the other dimensions are based on the first dimension. Thus, this may enlarge the distance between two clusters.

Hypothesis 2: The percentage of misclassifications does not differ across the number of dimensions (or attributes) of each observation.

Table 2 reveals that all the three clustering methods have no significant difference regarding the percentage of misclassifications. For the self-organizing feature maps, the percentage of misclassification with eight dimensions is the lowest among the three levels, while the percentage of misclassification with six dimensions is the lowest among the three levels for both two-stage clustering methods. In addition, recovery improves as the number of dimensions increases from four to eight in the case of self-organizing feature maps. This pattern is consistent with the results of [7,30].

Hypothesis 3. The percentage of misclassifications does not differ across the levels of error.

For all of the three clustering methods, the main effect of the level of error on the data is significant. For the proposed two-stage clustering method, the level of error is significant at 0.01 level ($N = 108$, $p < 0.001$). The mean misclassification of observations increases from 0.0% to 0.18% to 1.47% for the no error, low error, and high error data sets, respectively. A similar pattern is seen for the conventional and the proposed two-stage clustering methods. The performance of the self-organizing feature maps is the worst among the three clustering methods.

In summary, the misclassification rate of the self-organizing feature maps is the highest. On the other hand, both two-stage clustering methods can provide better performance. The mean misclassification rates for conventional and proposed two-stage clustering methods are 0.5833% and 0.5463%, respectively. The paired-sample t test indicates that there is no significant difference. However, Table 3 shows that the proposed two-stage clustering method is worse than the conventional two-stage clustering method only in the case when the number of clusters is two. The former is better than the latter for the rest of the cases. It may be reasonable to select the proposed two-stage clustering method. However, further comparison will be made in the following section through the real-world data.

5. Real-world problem results and discussion

Both the proposed and conventional two-stage clustering methods are excellent for clustering analysis as shown in Section 4. An advanced comparison of the two methods was made using the actual data collection for 3C (computer, communication, and consumer electronics) market. In this study, a 3C store is defined as the store which sells products including computers, communication equipment, and consumer electronics only. The results can be used to determine the different clusters for the 3C market.

5.1. Questionnaire design and attributes

5.1.1. Questionnaire design

This study uses the benefit-looking segmentation model. The main purpose of this study is to examine how the 3C market can provide different benefits to the consumers according to different segments. With this goal, a questionnaire was designed through reviewing the newspaper/magazine/academic papers and interviewing 3C market managers. The questionnaire was pre-tested by consumers via personal interviews to avoid unnecessary misunderstanding.

5.1.2. Attributes

The attributes of the questionnaire are divided into three dimensions: (1) customer experience attributes, (2) benefit attributes that the 3C stores can provide, and (3) demographic information attributes. The first part, which contains customer experience attributes, aims to find out the visiting customers' satisfaction with the existing 3C stores located in Kaohsiung City, which is the second largest city in Taiwan. The attributes include customers, the degree of customers' satisfaction of these stores, customers' companies, main products purchased, frequency of purchase and time spent on each visit. In the second part, 20 different benefit attributes are covered. We aim to determine the importance of each attribute to the customers. Five ration scales based on Likert's summated rating scale are used. These five scales are not very important, not important, normal, important, and very important. The third part collects the personal information of the respondents, such as sex, education, living area, average family income and occupation.

5.2. Population and sampling

The present study targets the customers who have visited the 3C stores located in Kaohsiung City. The stores actually sell only 3C products. According to the annual city reports, there are a total of seven 3C stores in the city.

This study collects data through personal interviews and adopts the quota sampling defined as follows. Totally there are seven 3C stores located in Kaohsiung City. Each store is assigned 35 copies of the questionnaire. The interviews are conducted during weekends. Totally, 245 questionnaires are collected. Since all the questionnaires are completed through personal interviews, there is no failure situation. However, five copies contained incomplete answers, and are excluded. Therefore, only 240 questionnaires are counted. The return rate is 97.8%.

After the standard procedures, such as 'verify, check, delete and coding', the data are analyzed using SPSS. The statistics results show that 65% of the respondents are males. Most of the respondents' age is between 21 and 30 years (34.2%), while the education of 92.5% of the respondents ranges from high school to college. Living areas covered evenly and 32.5% of the average family income is from NT$ 25000 to 35000. Among the respondents, 29.2% are involved in business while 22.5% are students.

5.3. Factor analysis

Twenty benefit-looking variables as shown in Table 4 are employed for factor analysis. First, the Bartlett–Ball test is used to determine whether the same variance exists for each variable. As a result, the chi-square value and P value of 1620.4490 and 0.00, respectively, are obtained which imply that the data are suitable for factor analysis. Moreover, the KMO (Kaiser–Meyer–Oklin) analysis is made to check whether the sampling numbers and sampling objects are suitable or not. The evaluation result indicates that the KMO value is 0.82765. As Kaiser (SPSS/PC + , 1988) stated, if the KMO value is more than 0.5, the factor analysis is acceptable. The bigger the KMO value is, the better the result is. Thus, the factor analysis of the present study can provide a good result. Next, these 20 benefit-looking variables are used for factor analysis.

First, the principal component analysis is used to extract the main structure of the factors. The standard cluster criterion is based on the Eigenvlaue which is greater than 1. Totally six factors are

Table 4
The computational results for 20 benefit-looking variables

Factors	Benefit-looking variable (Question)	Factor loading	Eigenvalue	Cumulated explanation variance (%)	Cronbach's α
1	17. Attitude of shop assistants	0.77116	6.04059	30.2	0.8112
	19. Mark the market and selling prices	0.76818			
	18. Store's reputation	0.68067			
	15. Price marked clearly	0.53031			
	20. Store with products of its own brand	0.50654			
	16. Lowest price guaranteed	0.50189			
	8. Large and clear store space	0.47149			
2	3. Self-service shopping	0.81576	1.84685	39.4	0.5080
	1. Convenient transportation and easy parking	0.60277			
3	6. A wide selection of products	0.76885	1.43100	46.6	0.6882
	5. Professional knowledge of shop assistants	0.60554			
	4. Negotiable price	0.53264			
	7. Good after-service	0.48856			
4	14. Frequent lucky draw activities	0.75970	1.17653	52.5	0.5776
	11. Frequent promotion activities	0.60668			
5	12. Selling products of famous brands	0.73908	1.06593	57.8	0.6750
	13. Recreation area like coffee bar or fast food shops	0.61787			
	10. Free internet and computer games provided	0.58210			
	2. Located in the neighborhood	0.48457			
6	9. Credit card accepted and without extra charge	0.83323	1.02389	62.9	—

obtained and the cumulated explanation variance is 62.98%. Then, the orthogonal rotation is implemented through Varimax's in order to explain the contents and meaning of each factor. The ones with factor loading of more than 0.45 are selected. The results are listed in Table 4. These six factors are named business reputation, shopping convenience, a wide selection of products, promotion activities, supplementary service, and credit card service. The corresponding factor loading for each factor is also listed in Table 4.

The factors are further evaluated by Cronbach's α reliability parameter, and the results are all larger than 0.5. It implies that the internal consistency of each factor is good.

5.4. Customers market segmentation and evaluation

5.4.1. Cluster analysis

After the factor analysis, the original 20 benefit-looking factors form a six-dimensional structure. Besides, the score of each customer on each factor is calculated, and becomes the basis of market segmentation.

In Section 4, the conventional and proposed two-stage methods have shown no significant difference in clustering. Herewith, the present study uses the self-organizing feature maps and Ward's minimum variance method as the first stage of the two-stage methods.

(1) Ward's minimum variance method

In Ward's method, the increase in total within-groups variance can be employed to decide the number of clusters. The total within-groups variance is found to be in a stable increasing state. Thus, the rule of deciding the number of clusters, the maximum error value rule, cannot be applied. Therefore, the dendrogram is used to determine the criteria. From the dendrogram, one extreme group with four observations is found. This extreme group is deleted and Ward's method is then implemented.

(2) Self-organizing feature maps

The self-organizing feature maps use 240 samples for training. Since there is no rule for determining the best training parameters, they are obtained by trial and error. The stopping criterion is the number of learning epochs. Accordingly, 1000, 1500, and 2000 epochs are tested. In addition, Wilk's Lambda value of each result is calculated. Wilk's Lambda value is the ratio of the within-groups variance (SS_{within}) to the total variance (SS_{total}), and the formula is defined as

$$\text{Wilk's Lambda} = SS_{within}/SS_{total}.$$

A large Wilk's Lambda value implies that there is no difference between within-groups averages. A Wilk's Lambda value closest to zero implies that the source of total variance is from the between-groups variance instead of within-groups variance. Under such a situation, the clustering characteristics, homogeneous within and heterogeneous between, are very significant. Then, the results of Ward's minimum variance method and self-organizing feature maps are used as the initial solution for the K-means method which is then employed to find the final solution.

(3) The determination of segmentation number

The best cluster number selected is the one that has the most significant difference in the variables of demographic information and customer experiences. Thus, the best solution is the one with the most significant difference according to the chi-square test results. This can lead to a better explanation of each segment and the researcher can easily determine the marketing strategy for every segment. Table 5 lists the computational results.

Table 5 indicates that when the number of segments is three with $\alpha = 0.05$, no variable reaches significant difference level. However, as the number of segments is four, both the conventional and

Table 5
The chi-square test results according to the population statistics and customer experience for 3, 4 and 5 segments

		3 Segments		4 Segments		5 Segments	
		Ward's + K-means	SOFM + K-means	Ward's + K-means	SOFM + K-means	Ward's + K-means	SOFM + K-means
Demographic information	Sex	0.392	0.185	0.922	0.001	0.602	0.572
	Age	0.106	0.179	0.848	0.011	0.542	0.203
	Education	0.140	0.110	0.072	0.001	0.005	0.415
	Living area	0.664	0.322	0.251	0.338	0.210	0.972
	Income	0.104	0.201	0.010	0.000	0.006	0.095
	Occupation	0.021	0.550	0.115	0.001	0.026	0.004
Customer experiences	Shopping frequency	0.250	0.110	0.048	0.058	0.785	0.647
	Shopping time	0.652	0.267	0.559	0.226	0.003	0.542

Table 6
Wilk's Lambda values for four different clustering methods and the corresponding F value and P value

Clustering method	Wilk's lambda	F value	P value
Ward's method	0.13233	37.27742	0.000
Ward's method + K-means	0.10534	43.40572	0.000
SOFM	0.10010	45.36322	0.000
SOFM + K-means	0.09906	45.94048	0.000

proposed two-stage clustering methods show the highest difference between the demographic information and customer experiences. The difference in value is smaller than that when the number of segments is equal to five. Therefore, it is reasonable to select four segments, or clusters.

5.4.2. Discussion

There are two criteria for comparing the results. One is the observation of Wilk's Lambda value, the other is the cross validation, which uses the confusion table to observe the error percentage and then evaluate the result of the clustering methods.

(1) Wilk's Lambda value

Wilk's Lambda values for all three clustering methods are calculated in order to observe the difference among them. Table 6 lists Wilk's Lambda values, F values, and P values. The results show that four segments determined by each method have a significant within-groups difference. The smaller the Wilk's Lambda value, the better the segmentation result is. Thus, self-organizing feature maps followed by the K-means method offer the best solution due to its lowest Wilk's Lambda value, 0.09906.

Table 7
The discriminant analysis confusion table for different clustering methods

		Forecasting result of discriminant analysis				Total
		1	2	3	4	
(a) Ward's minimum variance method						
Clustering method	1	48	0	1	0	49
		97.96%	0%	2.04%	0%	
	2	3	26	4	0	33
		9.09%	78.79%	12.12%	0%	
	3	7	3	53	0	63
		11.11%	4.76%	84.13%	0%	
	4	1	4	5	81	91
		1.1%	4.4%	5.49%	89.01%	
	Total	59	33	63	81	236
	Accurate rate					88.14%
(b)Ward's minimum variance method followed by *K*-means method						
Clustering method	1	62	0	1	0	63
		98.41%	0%	1.59%	0%	
	2	1	42	0	0	43
		2.33%	97.67%	0%	0%	
	3	3	0	78	2	83
		3.61%	0%	93.98%	2.41%	
	4	0	0	0	47	47
		0%	0%	0%	100%	
	Total	66	42	79	49	236
	Accurate rate					97.03%
(c) Self-organizing feature maps						
Clustering method	1	65	0	0	0	65
		100%	0%	0%	0%	
	2	0	54	0	0	54
		0%	100%	0%	0%	
	3	0	0	50	2	52
		0%	0%	96.15%	3.75%	
	4	0	0	0	69	69
		0%	0%	0%	100%	
	Total	65	54	50	71	240
	Accurate rate					99.17%
(d) Self-organizing feature maps followed by *K*-means method						
Clustering method	1	65	0	0	0	65
		100%	0%	0%	0%	
	2	0	54	0	0	54
		0%	100%	0%	0%	
	3	0	0	47	2	49
		0%	0%	95.92%	4.08%	
	4	0	0	0	72	72
		0%	0%	0%	100%	
	Total	65	54	47	74	240
	Accurate rate					99.17%

(2) Examination and analysis

The criterion is to observe the ratio that the discriminant analysis can accurately predict for the segments. Also, the scores of the six factors obtained from the factor analysis are used as inputs to the discriminant analysis. More consistent results indicate a better clustering result. From the confusion table (Table 7), we can see that Ward's method has the worst prediction result. The correct percentage is 88.14%. But if it is followed by the K-means method, the result becomes better with a correct percentage of 97.03%. This improves by almost 9%. On the other hand, self-organizing feature maps followed by the K-means method have a higher accuracy rate, 99.17%. In addition, the first and second groups can be examined and analyzed accurately. Only two observed values are misclassified from group 3 to group 4. As a result, the self-organizing feature maps followed with the K-means method can perform the best.

(3) Composite comparison

Table 8 presents the computational results of the two-stage clustering methods. As seen in Table 8, the conventional two-stage clustering method has a higher reassignment number. There are 44 observations reassigned to another group at the second stage. But the proposed two-stage clustering method shows no significant difference. Segments one and two show no difference, while only three observations are reassigned from segment three to segment four at the second stage. From the performance index, we can see that the proposed two-stage clustering method has accurate rates of 0.09906% and 99.17% in Wilk's Lambda value and discriminant analysis, respectively. Both are better than those of the conventional two-stage clustering method which are 0.10534% and 97.03%, respectively. Besides, the present study also proves that integration of the hierarchical and non-hierarchical methods provides better performance compared with that obtained using the clustering method alone, be it hierarchical or non-hierarchical.

When the K-means method alone is utilized and the number of segmentation is set up as four, the numbers of each segment are 8, 117, 33, and 82. It means that the results are influenced by the

Table 8
A comparison between the conventional two-stage method and the proposed two-stage method

	Conventional two-stage method			Proposed two-stage method		
	Ward's method	Ward's method + K-means	Number of reassignment	SOFM	SOFM + K-means	Number of reassignment
Segment 1	49	63	14	65	65	0
Segment 2	33	43	10	54	54	0
Segment 3	63	83	20	52	49	3
Segment 4	91	47	44	69	72	3
Total	236	236	44	240	240	3
Wilk's Lambda	0.13233	0.10534	0.02699	0.10010	0.09906	0.00104
Accuracy rate of discriminant analysis	88.14%	97.03%	8.89%	99.17%	99.17%	0%

outliers. The group size is not evenly distributed. Since Wilk's Lambda value is 0.16943, which is higher than that of the proposed two-stage method, this indicates that the K-means method alone with random starting point may provide the worst solution.

6. Conclusions

A novel two-stage scheme has been presented for market segmentation. The simulated data has shown that the proposed two-stage clustering method is slightly better than the conventional two-stage method, though the paired-sample t test indicates that there is no significant difference. In addition, the questionnaire survey data reveal that the proposed method is better than the conventional two-stage clustering method according to both Wild's Lambda value and discriminant analysis. The use of self-organizing feature maps alone cannot provide feasible solution.

In this study, the self-organizing feature maps are utilized to determine the number of clusters. However, in some cases, it is quite difficult to determine the cluster number by observing the outcome of network output array, unless the network topology is very clear. Therefore, it may be desirable to apply different unsupervised neural networks, like ART2, for further comparison [31, 32]. Besides, we can observe whether the difference exists or not if fuzzy C-means method replaces the K-means method. Definitely, the simulated data can become more complicated putting more fuzzy data points into each dimension. However, according to the present study, the proposed two-stage method does prove superior in clustering analysis on the basis of both theoretical and practical evaluations.

Acknowledgements

The authors would like to thank the National Science Council, Taiwan, Republic of China for partially supporting this manuscript under Contract No. NSC 88-2416-H-027-001. In addition, the Chinese Management Science Association honored this study with the Master Dissertation Award in July 1998. Their support has been much appreciated.

References

[1] Smith WR. Product differention and market segmentation as an alternative marketing strategy. Journal of Marketing 1956;21:3–8.

[2] Wind YJ. Issues and advance in segmentation research. Journal of Marketing Research 1978;15:137.

[3] Kuo RJ, Chi SS, Den BW. Generalized part family formation through fuzzy self-organizing feature map neural network. Computers in Industrial Engineering 2001;40(1–2):79–99.

[4] Punj G, Steward DW. Cluster analysis in marketing research: review and suggestions for applications. Journal of Marketing Research 1983;20:134–48.

[5] Kuo RJ, Xue KC. A decision support system for sales forecasting through fuzzy neural network with asymmetric fuzzy weights. Journal of Decision Support Systems 1998;24(2):105–26.

[6] Kuo RJ, Chen JH, Hwang YC. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. Fuzzy Sets and Systems 2001;118(1):21–45.

[7] Balakrishnan PVS, Cooper MC, Jacob VS, Lewis PA. Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. EJOR 1996;93:346–57.
[8] Arnbie P, Hubert LJ. Cluster analysis in marketing research. In: Bagozri, RP. editor. Advanced methods of marketing research, Cambridge, MA: Blackwell, 1994. p. 160–89.
[9] Wedel M, Kamakura WA. Market segmentation, conceptual and methodological foundations.. Dordrecht: Kluwer, 1998.
[10] Kolter P. Marketing management: analysis planning, implementation and control, 7th ed.. Englewood, Cliffs, NJ: Prentice-Hall, Inc, 1992.
[11] McCarthy JE. Basic marketing: a managerial approach. 7th ed. Haomewood, IL: Richard D. Irwin Inc., 1981, p. 226.
[12] Hertz J, Krogh A, Palmer RG. Introduction to the theory of neural computation.. Reading, MA: Addison-Wesley Publishing Company, 1991.
[13] Proctor RA. Marketing decision support system: a role for neural networking. Marketing Intelligence and Planning 1992;10(1):21–6.
[14] Venugopal V, Baets W. Neural networks and their applications in marketing management. Journal of Systems Management 1994;45(9):16–21.
[15] Goonalilake S, Treleaven P. Intelligent systems for finance and business.. New York: Wiley, 1995.
[16] Mazanec JA. Positioning analysis with self-organizing maps. Cornell Hotel and Restaurant Administration Quarterly 1995;36:80–95.
[17] Dhar V, Stein R. Intelligent decision support methods—the science of knowledge work.. Englewood Cliffs, NJ: Prentice-Hall, International, Inc, 1996.
[18] Fish KE, Barnes JH, Aiken MW. Artificial neural networks: a new methodology for industrial market segmentation. Industrial Marketing Management 1995;24:431–8.
[19] Bigus JP. Data mining with neural networks.. New York: McGraw-Hill Company, 1996.
[20] Balakrishnan PVS, Cooper MC, Jacob VS, Lewis PA. A study of classification capabilities of neural networks using unsupervised learning: a comparison with K-means clustering. Psychometrika 1994;59(4):509–25.
[21] Milligan GW. An algorithm for generating artificial test clusters. Psychometrika 1985;50(1):123–7.
[22] Ward Jr JE. Hierarchical grouping to optimize an objective function. Journal of American Statistical Association 1963;59:236–44.
[23] Kohonen T. A simple paradigm for the self-organized formation of structured feature maps. In: Amari S, Berlin M. editors. Competition and cooperation in neural nets. Lecture Notes in Biomathematics. Berlin: Springer, 1982.
[24] Vriens M, Wedel M, Wilms T. Metric conjoint segmentation method: a Monte Carlo comparison. Journal of Marketing Research 1996;33:73–85.
[25] Blashield RK, Morey LC. Mixture model test of cluster analysis: accuracy of four agglomerative hierarchical methods. Psychological Bulletin 1976;83:377–88.
[26] Edelbrock C. Comparing the accuracy of hierarchical methodologies. Pattern Recognition 1979;11:235–54.
[27] Milligan GW. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. Psychometrika 1980;45:325–42.
[28] Cormack RM. A review of classification. Journal of the Royal Statistical Society (Series A) 1971;14:279–98.
[29] SPSS/PC + Advanced Statistics, V2.0, SPSS Inc., Chicago, 1988.
[30] Krishnamurthy AK, Ahalt SC, Melton DE, Chen P. Neural networks for vector quantization of speech and images. IEEE Journal on Selected Areas in Communications 1990;8:1449–57.
[31] Wann C-D, Stelios CA. Thomopoulos, A comparative study of self-organizing clustering algorithm dignet and ART2. Neural Networks 1997;10(4):737–53.
[32] Carpenter G, Grossberg AS. ART 2: self-organization of stable category recognition codes for analog input patterns. Applied Optics 1987;26(23).
[33] Vemuri VR, Rogers RD (Eds.) Artificial neural networks: forecasting time series, Los Alamitos, CA:IEEE Computer Society Press, 1994.

Ren-Jieh Kuo received the M.S. degree in Industrial and Manufacturing Systems Engineering from Iowa State University, and the Ph.D. degree in Industrial and Management Systems Engineering from the Pennsylvania State University, University Park, PA. Currently, he is a Professor in the Department of Industrial Engineering, National

Taipei University of Technology, Taiwan, ROC. His research interests include architecture issues of neural networks, fuzzy logic, and genetic algorithms, and their applications in electronic business, logistics and process control.

Li-Ming Ho earned the Ph.D. degree in agriculture economics from National Taiwan University, Taiwan, ROC. Currently, he is the Head and Assistant Professor in the Department of Leisure, Recreation and Tourism Management, Shu-Te University, Taiwan, ROC. His research interests include applications of neural networks in management.

Chen-Ming Hu earned the M.S. degree from the Graduate School of Management Science, I-Shou University, Taiwan, ROC. Currently, he is the system planning engineer of Foxconn Industrial PCE Production Group in Taiwan.