Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 313–317
Expert Systems with Applications
www.elsevier.com/locate/eswa

Marketing segmentation using support vector clustering

Jih-Jeng Huang [a], Gwo-Hshiung Tzeng [b,c,*], Chorng-Shyong Ong [a]

[a] Department of Information Management, National Taiwan University, Taipei, Taiwan
[b] Institute of Management of Technology, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan
[c] Department of Business Administration, Kainan University, Taoyuan, Taiwan

Abstract

Marketing segmentation is widely used for targeting a smaller market and is useful for decision makers to reach all customers effectively with one basic marketing mix. Although several clustering algorithms have been proposed to deal with marketing segmentation problems, a soundly method seems to be limited. In this paper, support vector clustering (SVC) is used for marketing segmentation. A case study of a drink company is used to demonstrate the proposed method and compared with the *k*-means and the self-organizing feature map (SOFM) methods. On the basis of the numerical results, we can conclude that SVC outperforms the other methods in marketing segmentation.

Keywords: Marketing segmentation; Clustering algorithms; Support vector clustering (SVC); *k*-means; Self-organizing feature map (SOFM)

1. Introduction

With the various demands and the dynamic environment, market segmentation has become a central concept both in marketing theory and in practice. Market segmentation can be described as the process of partitioning a large market into the smaller groups or the clusters of customers (Weinstein, 1987; Smith, 1956; Kotler & Gordon, 1983; Croft, 1994; Myers, 1996). The similarities within each segment indicate the similar purchase behavior. The information of the segments is useful for decision makers to reach all customers effectively with one basic marketing mix (Anderson & Vincze, 2000).

Several benefits can be obtained from the market segmentation strategy. The most obvious benefit is that the decision makers can use a particular marketing mix to target a smaller market with the greater precision. This benefit allows the decision makers to deploy resources more effectively and efficiently. In addition, market segmentation forges the closer relationships between the customers and the company. Furthermore, the result of market segmentation can be used for the decision makers to determine the particular competitive strategies (i.e. differentiation, low cost, or focus strategy) (Aaker, 2001).

Cluster analysis is a technique employed for partitioning a set of objects into k groups such that each group is homogeneous with respect to certain attributes based on the specific criterion. The purpose of cluster analysis makes it be a popular tool for marketing segmentation. Generally speaking, clustering algorithms can be classified into partitioning methods (e.g. *k*-means), hierarchical methods (e.g. agglomerative approach), density-based methods (e.g. Gaussian mixture models), and grid-based methods (e.g. self-organizing feature maps (SOFM)) (Han & Kamber, 2001).

Although these algorithms have been successfully used in many areas, such as taxonomy, medicine, and business, some issues should be considered for the further applications in practice (Han & Kamber, 2001). First, some methods are restricted to the particular data type (e.g. *k*-means can only be suitable for the interval-based data). Second, some methods are sensitive to the outliers and lead to the poor clustering quality. Third, the clustering method

* Corresponding author. Address: Institute of Management of Technology, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu 300, Taiwan. Tel.: +886 3 5712121x57505; fax: +886 3 5753926.
E-mail addresses: ghtzeng@cc.nctu.edu.tw, u5460637@ms16.hinet.net (G.-H. Tzeng).

0957-4174/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2005.11.028

should have the ability to deal with the high-dimensional data set. Finally, the clustering method should discovery of clusters with arbitrary shape.

In this paper, a nonparametric method, called support vector clustering (SVC), is proposed for marketing segmentation. The reason why we adopt SVC is that it can deal well with the issues above and provide a satisfactory result. In addition, a case study of a drink company is used to demonstrate the proposed method and compared with the *k*-means and the SOFM methods. On the basis of the numerical results, we can conclude that SVC is the appropriate tool for marketing segmentation and outperform to other methods.

The remainder of this paper is organized as follows. We review the two clustering methods, *k*-means and SOFM, in Section 2. Support vector clustering is presented in Section 3. In Section 4, a case study of a drink company is used to demonstrate the proposed method and compared with the *k*-means and the SOFM methods. Discussions are presented in Section 5 and conclusions are in the last section.

2. The review of literature

In this section, two famous clustering methods, *k*-means and SOFM, are presented below. The *k*-means method is the most popular statistical tools used for cluster analysis due to its simplicity and scalability. On the other hand, SOFM is widely used to cluster data set in the field of neural network. Both of *k*-means and SOFM are employed to compare with SVC in this paper.

2.1. *k*-means method

The *k*-means method (Anderberg, 1973) was proposed to overcome the scaling and the merged problems of the hierarchical clustering methods. The characteristics of simplicity and scalability make it be widely used in the field of statistics. The procedures of the *k*-means method can be summarized as follows.

Step 1: Randomly select k initial cluster centroids, where k is the number of the clusters.
Step 2: Assigned each object to the cluster to which it is the closest based on the distance between the object and the cluster mean.
Step 3: Calculate the new mean for each cluster and reassign each object to the cluster.
Step 4: Stop if the criterion converges. Otherwise go back to Step 2.

However, several deficiencies of the *k*-means method have been proposed as follows (Han & Kamber, 2001). First, since the *k*-means method can only be applied only when the means of clusters are defined, it cannot work well when data with qualitative attributes are involved. Second, the *k*-means method is not suitable for discovering clusters with nonconvex shapes or clusters of very different size. Finally, the *k*-means method is very sensitive to noise and outlier data. Such data can substantially influence the mean value and produce the wrong results.

2.2. Self-organizing feature maps

SOFM (Kohonen, 1989, 1990) is an unsupervised competitive learning method and widely used to deal with the clustering problem. SOFM is a two-layer network structure which is composed of the input layer and the output layer. The main characteristics of SOFM are its lateral connections between neurons in the output layer and the mechanism of winner-takes-all. We can depict the architecture of the SOFM network as shown in Fig. 1.

The procedures of SOFM can be described as follows:

Step 1: Set at random the initial synaptic weights between $[0,1]$.
Step 2: Calculate the winner-takes all neuron j^* at iteration p using the criterion:

$$j^*(p) = \min_j \|x - w_j(p)\|, \quad j = 1, \ldots, m \tag{1}$$

where $\|\cdot\|$ denotes the Euclidean norm, and m denotes the number of neurons in the output layer.
Step 3: Update all neurons' weights using the following equation:

$$w_{ij}(p+1) = \begin{cases} w_{ij}(p) + \alpha[x_i - w_{ij}(p)], & j \in \Lambda_j(p) \\ w_{ij}(p), & j \notin \Lambda_j(p) \end{cases} \tag{2}$$

where α denotes the learning rate parameter and $\Lambda_j(p)$ is the neighbourhood function centered around the winner-takes-all neuron j^* at iteration p. Note that the neighbourhood function is a function of the distance between j and j^*. Typical functions are Gaussian and Mexican functions.
Step 4: Go back to Step 2 and continue until the criterion is satisfied.

Recently, SOFM has been widely used in various applications such as image segmentation (Kim & Chen, 2001), texture segmentation (Zhiling, Guerriero, & De Sario,

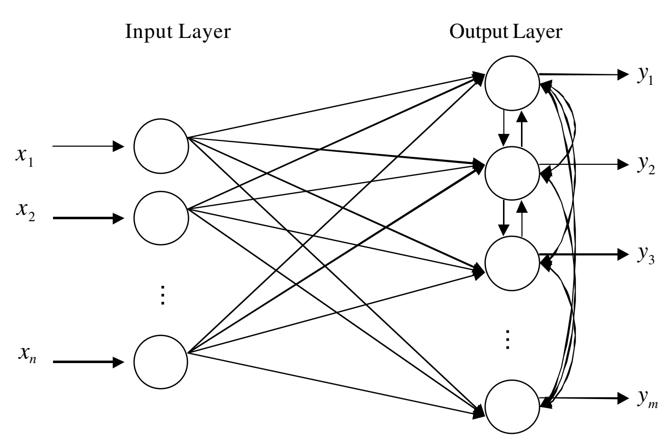

Fig. 1. The architecture of SOFM.

1996), and market segmentation (Bloom, 2005). Next, we describe the contents of SVC in Section 3.

3. Support vector clustering

SVC is proposed by Ben-Hur, Horn, Siegelmann, and Vapnik (2001) to cluster data set based on the theory of support vector machine (SVM). Support vector machines (SVM) was pioneered by Vapnik (1995, 1998) to deal with the problems of pattern classification and nonlinear regression by minimizing the structural risk (Vapnik, 1995, 1998). Based on the perspective of statistical learning theory, the principle of SVM is related to minimizing the Vapnik–Chervonenkis (VC) dimension and the upper bound on the number of test errors. Recently, SVM has been widely used in many areas to handle various classification and curve fitting problems such as pattern recognition (Drucker, Wu, & Vapnik, 1999), text categorization (Rowley, Baluja, & Kanade, 1998), and bioinformatics (Jaakola, Diekhans, & Haussler, 2000). On the basis of SVM, SVC extended SVM to consider the problem of clustering. Now, we can describe the concepts of SVC as follows.

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_n\} \in \mathbb{R}^n$ be the data space. Using a nonlinear transformation Φ to some high-dimensional feature-space, the smallest enclosing sphere of the radius R can be defined as

$$\|\Phi(x_j) - a\|^2 \leqslant R^2, \quad \forall j = 1, \ldots, n \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm and a is the center of the sphere. In order to deal with the problem of the outlier, the slack variables ξ_i are incorporated into Eq. (3) to display the soft constraint:

$$\|\Phi(x_j) - a\|^2 \leqslant R^2 + \xi_i, \quad \xi_i \geqslant 0 \tag{4}$$

The problem above can be solved by introduced the Lagrangian as follows:

$$L = R^2 - \sum_j (R^2 + \xi_i - \|\Phi(x_j - a)\|^2)\beta_j - \sum_j \xi_i \mu_j + C \sum_j \xi_j \tag{5}$$

where β_j, $\mu_j \geqslant 0$ denote the Lagrange multiplies, C is the user-defined constant, and $C\sum_j \xi_j$ is the penalty term. To solve the equation above, we can set the derivative of L with respect to R, a and ξ_j, respectively, as follows:

$$\sum_j \beta_j = 1 \tag{6}$$

$$a = \sum_j \beta_j \Phi(x_j) \tag{7}$$

$$\beta_j = C - \mu_j \tag{8}$$

Next, by adopting the KKT complementary condition (Fletcher, 1987), we can derive

$$\xi_j \mu_j = 0 \tag{9}$$

$$(R^2 + \xi_i - \|\Phi(x_j - a)\|^2)\beta_j = 0 \tag{10}$$

By eliminating the variable R, a and μ_j, we can re-write the Lagrangian into the Wolfe dual form as the following equation:

$$W = \sum \Phi(x_j)^2 \beta_j - \sum \beta_i \beta_j \Phi(x_i) \cdot \Phi(x_j) \tag{11}$$

and subject to

$$0 \leqslant \beta_j \leqslant C \tag{12}$$

Note that the dot product $\Phi(x_i) \cdot \Phi(x_j)$ should be satisfied Mercer's theorem (Cristianini & Taylor, 2000). In this paper, the Gaussian kernel is employed and can be represented as

$$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j) = \mathrm{e}^{-q\|x_i - x_j\|^2} \tag{13}$$

where q denotes the width parameter. Three common types of the inner-product kernels can be described as shown in Table 1.

Now, we can determine the cluster assignment as follows. Let a segment of points y, the clustering rule can be represented as the adjacency matrix:

$$A_{ij} = \begin{cases} 1, & \forall\ y \text{ on the line segment connecting } x_i \text{ and } x_j \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

All data points are checked to assign a specific cluster. In addition, outliers are unclassified since their feature space lie outside the enclosing sphere. Next, we use the customers of a drink company to demonstrate the proposed method.

4. Marketing segmentation: a case study of a drink company

In this section, the 40 potential customers of a drink company is used to demonstrate the proposed method and compared with the k-means and the SOM methods. The data set contains four life style attributes including the degree of socialization (Y_1), leisure (Y_2), knowledge retrieving (Y_3), and achievement (Y_4).

In order to determine the appropriate parameters in SVC, we first use singular value decomposition (SVD) to project the data into the three-dimensional space as shown in Fig. 2.

On the basis of Fig. 2, it can be seen that three clusters should be the rational segments. On the other hand, the agglomerative hierarchical method is also employed to show the clustering tree for determining the appropriate clusters.

From Fig. 3, we can also visually determine the same numbers of the clusters with the SVD method. Next, by setting $C = 1$ we can adjust the parameter q to cluster the data for three segments. In this study, we can derive the $q = 1.5$. Next, using Eqs. (11)–(13) we can obtain the clustering results of the SVC method as shown in Table 2. In addition, we also employ the k-means and the SOM method to compare with the proposed method as shown in Table 2.

Since the purpose of clustering is to partition a set of objects into k groups such that each cluster is as

Table 1
Three common types of the kernels

Type of SVM	$K(x_i, x_i)$	Parameter
Polynomial	$(1 + x_i' x_j)^p$	where p denotes the power and is specified by user
Radial-basis function	$\exp(-\frac{1}{2\sigma^2}\|x_i - x_j\|^2)$	where σ denotes the width and is specified by user
Multilayer perceptron	$\tanh\left(\beta x_i' x_j + b\right)$	where β, b denote the coefficients and are specified by user

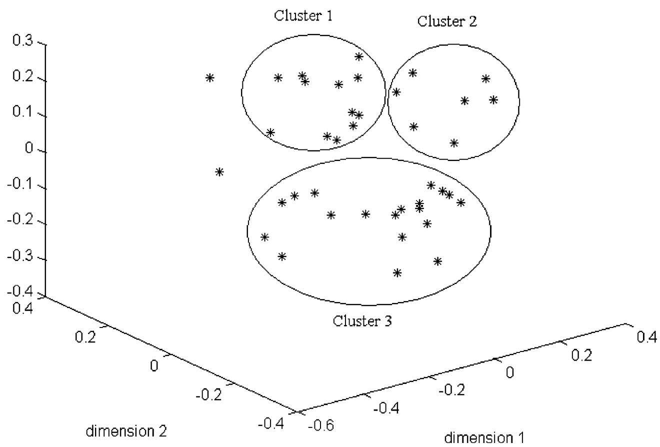

Fig. 2. The data mapping using the three-dimensional space.

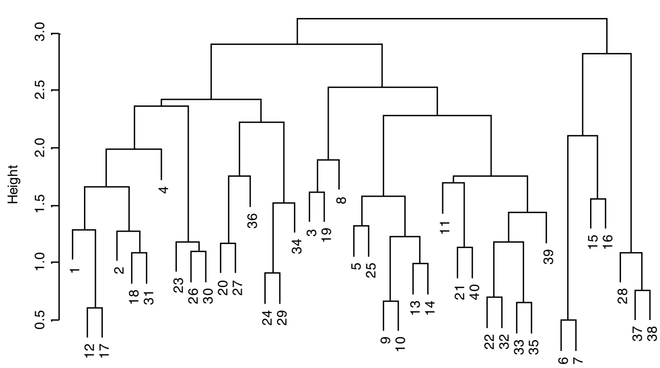

Fig. 3. The clustering tree.

heterogeneous as possible and the data within-cluster is as homogeneous as possible, we use the mean and the standard error to compare the performance of the various methods above.

On the basis of table, it can be seen that SVC can well separate each cluster by the significantly different mean of the all factors. Form the index of standard error, in the other hand, we can also conclude that SVC can outperform to other methods for grouping the clusters more homogeneously. Next, we provide the depth discussions about the comparison of the methods above in Section 5.

5. Discussions

Marketing segmentation involves clustering a whole market into several meaningful segments. It is a clear that different people have different needs. In order to meet these various needs, market has to be divided into smaller segments in order marketers to have the ability to plan good marketing and positioning of its product.

Recently, many artificial intelligence tools including neural network and fuzzy-based methods are introduced to deal with the clustering problems. However, as mentioned previously, several issues should be considered so that the clustering method can widely used in the real-life problems. In this paper, SVC is employed for marketing segmentation by considering the issues previously.

In this paper, a case study of a drink company is used to demonstrate the proposed method. First, we adopt the SVD and the agglomerative hierarchical methods to determine the appropriate numbers of the clusters. Next, we adjust the parameters to derive the results of marketing segmentation using SVC. Compared the results with the *k*-means and the SOFM methods using the mean and the standard error, we can conclude that SVC outperform to other methods in our case study.

Table 2
The comparison of *k*-means, SOM, and SVC

Method	Cluster	Mean				Standard error			
		Y_1	Y_2	Y_3	Y_4	Y_1	Y_2	Y_3	Y_4
k-means	1	−0.0438	−0.4490	0.9360	−0.2109	0.8930	0.8014	0.5160	1.0119
	2	0.4879	−0.0412	−0.6132	0.2629	0.7945	0.8986	0.7365	0.9684
	3	−1.3470	1.3208	−0.6564	−0.2764	0.4703	0.6190	0.6944	0.9965
SOM	1	−0.6470	0.8298	−0.2578	0.5274	0.8399	0.7816	0.7521	1.0335
	2	0.1457	−0.1997	1.0638	−0.4366	1.0181	0.8121	0.4353	0.8797
	3	0.5511	−0.6939	−0.7863	−0.1545	0.7804	0.7488	0.6888	0.8576
SVC	1	−0.2717	0.4909	−0.4394	0.9963	0.6375	0.6205	0.5362	0.5597
	2	1.2022	−0.3802	−0.7913	0.5449	0.3778	0.5101	0.5311	0.5675
	3	0.0227	−0.8701	0.7858	−0.2742	0.8105	0.7990	0.6579	0.5004

In addition, several advantages for SVC used in marketing segmentation can be described as follows. First, SVC can deal well with different types of attributes due to it can map the data to the appropriate feature space. In addition, SVC can generate cluster boundaries of arbitrary shape, where other methods are usually limited to the hyper-ellipsoid shape. Furthermore, by incorporating the slack variables, SVC can soundly deal with the problem of the outlier. Finally, SVC is good at handling the high-dimensional data set.

6. Conclusions

Marketing segmentation receives much attention in practice for planning the particular marketing strategies. In this paper, SVC is used for considering the marketing segmentation problems. First, two approaches, the SVD and the agglomerative hierarchical methods, are employed to derive the numbers of the segments. Next, the parameters can be adjusted according to the information above. Finally, the results of marketing segmentation can be obtained. From the numerical results, it can be seen that the proposed method can outperform to the k-means and the SOFM methods. In addition, SVC can also provide other four extra advantages for the clustering problems.

References

Aaker, D. A. (2001). *Strategic market management*. New York: John Wiley and Son.

Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.

Anderson, C., & Vincze, J. W. (2000). *Strategic marketing management*. New York: Houghton Mifflin.

Ben-Hur, A., Horn, D., Siegelmann, H. T., & Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research, 2*(2), 125–137.

Bloom, J. Z. (2005). Market segmentation: A neural network application. *Annals of Tourism Research, 32*(1), 93–111.

Cristianini, N., & Taylor, J. S. (2000). *An introduction to support vector machines and other Kernel-based learning methods*. Cambridge: Cambridge University Press.

Croft, M. J. (1994). *Market segmentation: A step-by-step guide to profitable new business*. London, New York: Routledge.

Drucker, H., Wu, D., & Vapnik, V. (1999). Support vector machines for spam categorization. *IEEE Transactions on Neural Network, 10*(5), 1048–1054.

Fletcher, R. (1987). *Practical methods of optimization*. Chichester: Wiley-Interscience.

Han, J., & Kamber, M. (2001). *Data mining: Concept and techniques*. San Francisco: Morgan Kaufmann Publishers.

Jaakola, T., Diekhans, M., & Haussler, D. (2000). A discriminative framework for detecting remote protein homologies. *Journal of Computational Biology, 7*(1), 95–114.

Kim, J., & Chen, T. (2001). Multiple feature clustering for image sequence segmentation. *Pattern Recognition Letters, 22*(11), 1207–1217.

Kohonen, T. (1989). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 59–69.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*, 1465–1480.

Kotler, P., & Gordon, M. (1983). *Principles of market*. Canada: Prentice Hall.

Myers, J. H. (1996). *Segmentation and positioning for strategic marketing decisions*. Chicago: American Marketing Association.

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(1), 23–38.

Smith, W. R. (1956). Product differentiation and market segmentation as an alternative marketing strategy. *Journal of Marketing, 21*(1), 3–8.

Vapnik, V. (1995). *The nature of statistical learning theory. Support vector domain description*. Springer-Verlag.

Vapnik, V. (1998). *Statistical learning theory*. New York: John Wiley and Sons, Inc.

Weinstein, A. (1987). *Market segmentation: Using Niche marketing to exploit new markets*. Chicago: Probus.

Zhiling, W., Guerriero, A., & De Sario, M. (1996). Comparison of several approaches for the segmentation of texture images. *Pattern Recognition Letters, 17*(5), 509–521.