The Fuzzy Clustering on Market Segment

Tsuen-Ho Hsu
Professor, Department of Business Administration
I-SHOU University
1, Section 1, Hsueh-Cheng Rd.,
Ta-Hsu Hsiang, Kaohsiung
County, Taiwan, R.O.C.
Tel : 886-7-6563711 ext.5911
E-mail : Thhsu@csa500.isu.edu.tw
Kao-Ming Chu
Master of Business Administration, Soochow University
Tel : 886-7-6313071
Fax: 886-7-6316034
E-mail : golden86@ms23.hinet.net
Hei-Chun Chan
Instructor of Business Department, Wen Tzao College
Tel : 886-7-3426031 ext 266
Fax: 886-7-6316034
E-mail : golden86@ms23.hinet.net

ABSTRACT

In any study of market segmentation, researchers often use clustering analysis as a tool. The analysis often is in a crisp partition form. But in practice, the sample are usually not well distributed, therefore the form may not be precisely defined. That is, one sample can belong to two or more groups. But, due to the fact that the requirements on the consumers and on the market are very high and the many real-market problems are fuzzy by nature and not random, the probability applications have not been very satisfactory in a lot of case.

In this study, we adopt the fuzzy cluster method and attempt to combine a new compactness and separation validity Function to build market segmentation in order to address the fuzziness among the group boundaries. Then we can use membership grade to describe each group. Therefore, the real market situation is clearly presented.

Through membership grade, we depict the reality of the market, which lies between integers and real number. Buyers' mindset are both rational and complicate, so their purchasing decisions are not predictable and are affected by many factors. The structural stability of the market can be tested by the loyalty of buyers who pertain to different clusters. Marketing strategies will also have effects on the movement of group for housing buyer.

Keywords: Fuzzy Cluster Method, Market Segmentation, Consumer Behavior

I. INTRODUCTION

In any study of market segmentation, researchers often use clustering analysis as a tool. The analysis often is in a crisp partition form. But in fact, the samples are usually not well-distributed; therefore the form may not be precisely defined. That is, one sample can belong to two or more groups. But, due to the fact that the requirements of the consumers and on the market are very high and many real-market problems are fuzzy by nature and not random, the probability applications have not been very satisfactory in a lot of cases. According to Punj and Stewart (1983), the traditional method, crisp partition, cannot fit the real product market, and so is the boundary between customers and competitors. This indicates that different segments overlap each other. In this study, we adopt the fuzzy cluster method and attempt to combine a new compactness and separation validity function to build market segmentation in order to address the fuzziness among the group boundaries. Then we can use membership grade to describe each group. Therefore, the real market situation is clearly presented. To this end, we used the Taiwan real estate industry as an example. It can offer the real estate industries to understand real consumer's constructure and find the new market niche. While seeking a large market and the various product benefits, a company also needs to properly analyze its different market segments in order to survive and succeed in a competitive environment.

II. THE FUZZY CLUSTERING ANALYSIS METHOD

2.1 Objective Function

Because consumers belong to different market segments, and fuzzy clustering can denote more information than hard clustering (Zimmermann, 1991), in this study we use fuzzy clustering theory to solve the segmentation problem. A sample is no longer said to "belong to" or "not belong to" a certain market segment. Instead, a sample can belong to more than one market segment with a tendency to be identified with a single, stronger segment. We also use the membership grade to analyze the characteristics of the market segments. That is, the fuzzy clustering method means more flexible modeling — by extending the zero-one membership to the membership in the interval [0,1], more flexibility is introduced.

The fuzzy clustering method can be divided into two varieties (Yang, 1993). One uses fuzzy relation to perform fuzzy clustering; the other is based on objective function to determine fuzzy clustering. The grouping results on the fuzzy relation are separate segments. Therefore, in this study, we use the objective function to get soft segmentation. The objective function of fuzzy clustering derives from Fuzzy C-Means (FCM). The FCM clustering algorithm was first presented by Dunn (1974). Bezdek (1981) further developed the FCM clustering algorithm. Subsequent revisions came from Roubens (1982), Goth (1989), Gu and Dubuisson (1990), and Xie and Beni (1991). However, Bezdek's FCM remains the most commonly used.

The effective validation is not taken into consideration in Bezdek's FCM. We add the compactness and separation validity function presented by Xie and Beni (1991) to revise Bezdek's FCM. This allows us to determine the proper number of segments and the segmentation validity.

The measurement criteria of the objective function are the minimum distance between each sample X_k and the center of each group V_i. In this way, we can determine the minimum objective value. This paper uses Bezdek's (1981) FCM objective function,

$$Min \quad J_m(U,V) = \sum_{i=1}^{c} \sum_{k=1}^{n} (\mu_{ik})^m d^2(X_k, V_i)$$

subject to: $0 \le \mu_{ik} \le 1$,

$\sum_{i=1}^{c} \mu_{ik} = 1$, $1 \le k \le n$,

$0 < \sum_{k=1}^{n} u_{ik} < n$, $1 \le i \le c$, $c \ge 2$, $m \geqq 1$.

Where μ_{ik} is membership grade that X_k belong to V_i, and m is the parameter. The higher the value of m, the fuzzier the segmentation.

$$d(X_K, V_i) = \|X_k - V_i\| = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2\right]^{1/2}$$

d is the distance from X_k to V_i.

2.2 Compactness and Separation Validity Function

Zimmermann (1991) notes that segmentation validation depends on whether the segmentation results can reflect the data and the segmentation structure. This problem can be converged to determine the correct segmentation number, *C*. Bezdek (1974), Dunn (1976), Windham (1980, 1982), Roubens (1982), Rousseeuw (1987), and Trauwaert (1988), present different validation methods. But, Xie and Beni (1991) note that it is difficult to use the aforementioned validation methods to present geometric characteristics among the data. They present the compactness and the separation validity function to improve the segmentation effectiveness.

In this paper we use Fukuyama and Sugeno(1989) method to measure the effectiveness of fuzzy clustering. Its purpose is to solve the cluster numbering, *C*. Fukuyama and Sugeno's compactness and separation validity function are as follows:

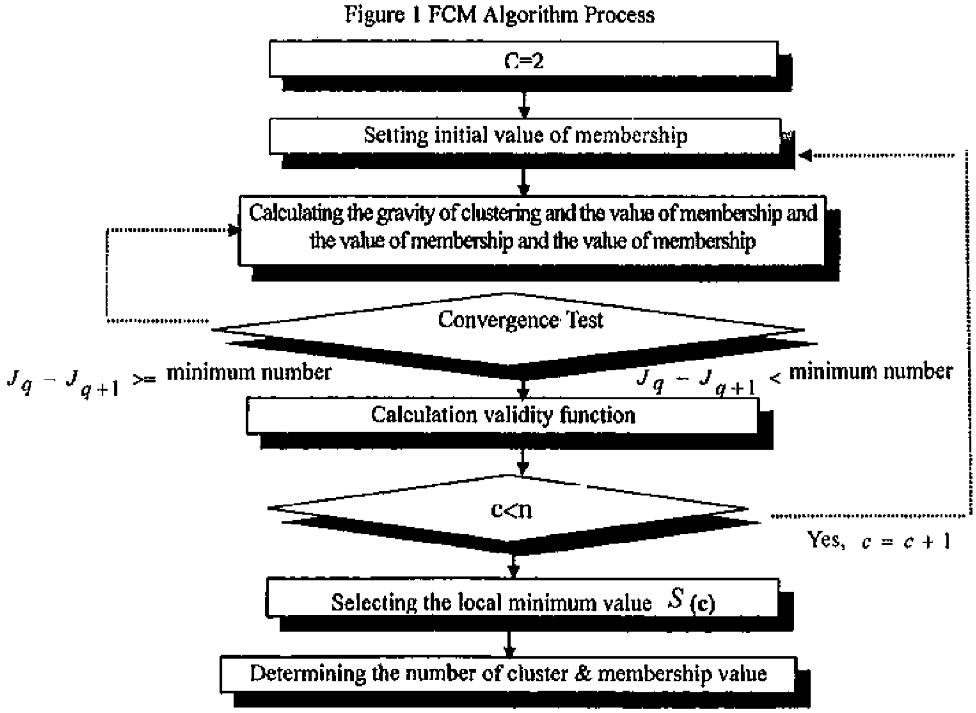
$$S(c) = \sum_{k=1}^{n}\sum_{i=1}^{c} (u_{ik})^m \left(\|x_k - v_i\|^2 - \|v_i - \bar{x}\|^2\right)$$

The smaller the value of S (c), the better the compactness and separation between the clustering groups of in-cluster samples. Therefore, the goal should be to minimize the value of S (c). At the same time, we can also determine the minimal objective function of FCM.

The process of the fuzzy clustering analysis algorithm is an iteration calculation process. That is, an iteration is to revise the cluster numbering. We use Matlab language to develop the fuzzy clustering algorithm process.

The algorithm process contains eight steps and is shown below:

Figure 1 FCM Algorithm Process



III. CASE STUDY

3.1 Questionnaire Design and Sampling Method

3.1.1 Questionnaire design

The questionnaire includes three parts: 1) 36 questions about house buying decision criterion. 2) 11 questions about house buying motivation. 3) 12 questions about demographic data.

3.1.2 Sampling method

In this study we visit the potential buyers of three consumer housing selling sites located in Kaohsiung city from April 1 to May 31, 1999. After the buyers visit the housing product, the staff ask one of each visit group to fill a questionnaire of house buying-demand and behavior. 350 questionnaires are provided during two months. Of the 335 questionnaires returned, 312 are usable with an effective response rate of 89%.

3.2 The Factor Analysis of Purchase Housing Decision Criterion

We use 36 benefit variables to perform the factor analysis, and reduce the number of the variables from 36 to seven. They are (1) the factor of life style and quality, (2) the factor of payment and loan, (3) the factor of house planning and design, (4) company credibility and service quality, (5) safety, (6) building's value and re-selling, (7) the factor of information equipment.

3.3 Fuzzy Clustering Analysis

According to FCM algorithm, if we set m=2, c=2~10 and convergence value ε=0.001, then when c=5 the compactness and separation validity value(s=0.2800) is the minimum value, so the number of the optimal segment is 5.

3.4 Segment Structure Analysis

3.4.1 Labeling Segmentation Clustering

FCM just can categorize the degree of each segment group for all samples, but not identify to the specific segment. So we use the nearest distant rule to categorize the samples of each market segment. It is shown in Table 1. We just line up 5 of the samples in our study.

Table 1 The degree of each segment group for all samples

Sample	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5	Amount
001	0.0913	0.5506	0.1435	0.1311	0.0835	1.0000
002	0.1777	0.5039	0.1272	0.1048	0.0864	1.0000
003	0.0434	0.1176	0.6452	0.1020	0.0918	1.0000
004	0.0098	0.0174	0.9231	0.0269	0.0228	1.0000
005	0.0548	0.0819	0.1825	0.0315	0.6493	1.0000
Segment Number (Real Vaules)	**0.377**	**1.2714**	**2.0215**	**0.3963**	**0.9338**	**5**
Segment Number (Integer Vaules)	**0**	**2**	**2**	**0**	**1**	**5**

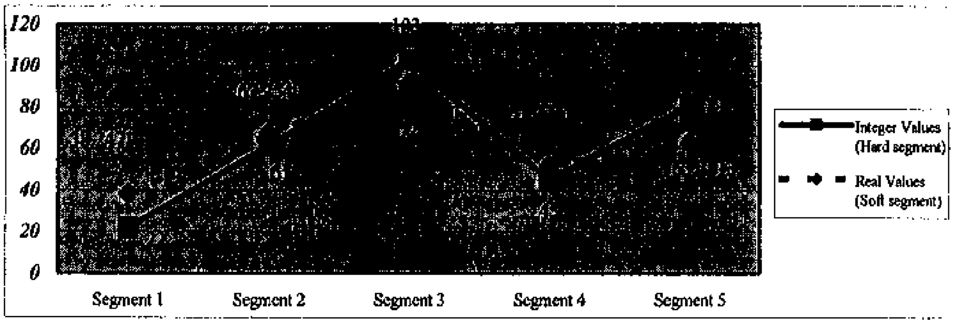
The degrees of each segment group for sample 001 are 0.0913, 0.5506, 0.1435, 0.1311, 0.0835. The strong possibility belongs to segment 2, and the weak possibility belongs to other segments. This sample does not totally belong to segment 2. That is, the sample is toward segment 2 than other segments.

Though FCM we get hard segmentation (integer values) of all sample (312). But we can get their soft segmentation (real values) that is, the aggregation of the values in each column. The result is shown in table 2 and figure2.

Table 2 The integer and real number of each segment

The number of questionnair-takers	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Integer Values(Hard segment)	22	61	103	45	81
Real values(Soft segment)	38.2471	67.2549	91.4726	51.7741	63.2513

Figure 2 The integer and real number of each segment



Our findings are in the following :

1. It is different between the interger value which we gain with hard segment and real value which we gain with soft segment. So in traditional, it often happen higher and lower-account condition, when we want to forecast the size of market. For example using the hard segment we gain intercger value is 22, but using the soft segment is 38.2471. It shows lower-account consition. But it shows opposite condition in segment 3.
2. It exist a gap between interger value and real value. It shows consumer market is a dynamic and uncertainly condition. The values which we gain from har segment and soft segment can present upper limit and lower limit of each segment of size of consumer market. The number of each segment is move between upper limit and lower limit. It shows consumer market is a dynamic market.

4.3 The Analysis for Each Structure Characteristic of Each Market Segment

The findings of FCM for the membership grade (α-cuts) of each questionnaire-taker, it is shown in Table 3 :

Table 3 The α-cuts for Each Segmemt

Segment / α-cuts	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
α=0.5	57.34%	47.42%	53.99%	49.38%	50.41%
α=0.6	39.29%	26.72%	22.78%	39.51%	30.62%
α=0.7	25.74%	15.41%	10.24%	26.24%	25.26%
α=0.8	15.25%	4.34%	5.05%	19.22%	16.29%

The findings of Table 3 are shown below :

1. From Table 3, we find the higher the α-cuts, the higher the customers' loyality is and the more stable the market structure is.
2. The membership grade of more than half of each group is 0.5. Also, more than half of the total questionnaire-takers belong to segment 2 and segment 3, but the α-cuts of these two groups are 4.34% and 5.05% respectively. This shows these two segment markets are unstable.

IV. CONCLUSION AND SUGGESTIONS

Through membership grade, we describe the reality of the market, which lies between integer value and real value. Buyers' behavior is both rational and complicate, so their purchasing decisions are not predictable and are affected by many factors. The structural stability of the market depends on membership grade of buyers. Marketing strategies will also have effects on the movement of the groups for housing buyers.

Through the concept of membership grade, marketing planners can measure the market size precisely, market dynamics, and structural aspects of different market segments. Marketing planners (constructors) will have clear picture of various market situations to helpmake marketing strategies.

Through the new fuzzy clustering method developed in this study we can solve the problem of mutual rejection classification by traditional classifying methods. Besides, we can get the cluster number. The method can also be computerlized to help to make it easy to be marketing planners to do market analysis.

References

1. Ajay, M. (1996). "Resource and Market Based Determinants of Performance in the U.S. Banking Industry," Strategic Management Journal, vol. 17, pp. 307-322.
2. Bezdek, J.C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. New York, Plemun Press.
3. Cattell, R.B., (1978). The Scientific Use of Factor Analysis in the Behavioral and Life-Sciences, Plenum Press, New York.
4. Churchill, Jr., G. A. (1995). Marketing Research—Methodological Foundations, 6th ed. New York : The Dryden Press.
5. Day, G. S. (1993). "Continuous Learning about Markets ", California Management Review, Summer,pp.9-31.
6. Dunn, J.C. (1974). "Will-Separated Clusters and the Optimal Fuzzy Partitions," Journal of Cybernetics, 4, pp. 95-104.
7. Dunn, J.C. (1976). Indices of Partition Fuzziness and the Detection of Clusters in Large Data Sets, Fuzzy Automata and Decision Processes Elsevier, New York.
8. Gath, I. And Geva, A.B. (1989). "Unsupervised Optimal Fuzzy Clustering," IEEE Transaction on Pattern Analysis Machine Intelligence, vol. 11, 7, pp. 773-781.
9. Gu, T. and Dubuission, B. (1990). "Similarity of Classes and Fuzzy Clustering," Fuzzy Sets and Systems, vol. 34, pp. 213-221.
10. Jang J. S.(1997). "Data Clustering Algorithms", Neuro-Fuzzy and Soft Computing, A computational Approach to Learning and Machine Intelligence, Prentice-Hall International, Inc., pp. 423-433.
11. Kaufman, L. and Rousseeuw, P. (1989). Finding Groups in Data, Wiley, New York.
12. Rouben, M. (1982). "Fuzzy Clustering Algorithm and Their Clustering Validity," European Journal of Operational Research, vol. 10, pp. 294-301.
13. Roubens, M. (1978). "Pattern Classification Problems and Fuzzy Sets", Fuzzy Sets and Systems, Vol.1, pp.239-253.
14. Roubens, M. (1982). "Fuzzy Clustering Algorithm and Their Clustering Validity", European Journal of Operational Research, Vol.10,pp. 294-301.
15. Rousseeuw, P.J. (1987). "A Graphical Aid of Cluster Analysis," Journal of Computer Applied Math, 20, pp. 53-65.
16. Trauwaert, E. (1988). "On the Meaning of Dunn's partition coefficient for Fuzzy Clusters," Fuzzy Sets and Systems, vol. 25, pp. 217-242.
17. Wedel, M. and Steenkamp, E.M. (1991). "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," Journal of Marketing Research, vol. 28, pp. 385-396.
18. Weinstein, A. (1987). Market Segmentation, Probus Publishing Company.
19. Wind, Y. (1978). "Issues and Advances in Segmentation Research," Journal of Marketing Research, vol. 15, pp. 317-337.
20. Windham, M.P. (1980). "Cluster Validity for Fuzzy Clustering Algorithms," Journal of Fuzzy Sets System, vol. 3, pp. 1-9.
21. Xie, X. and Beni, G. (1991). "A Validity Measure for Fuzzy Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13(8), pp. 841-847.
22. Xie,X. and Beni,G., (1991). "A Validity Measure for Fuzzy Clustering", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.13(8), pp.841-847.
23. Yang, M.S. (1993). "On a Class of Fuzzy Classification Maximum Likelihood procedures," Fuzzy Sets and Systems, vol. 57(3), pp. 365-375.
24. Yospeh, S. (1984). "Slammong Syndicated Data," Advertising Age, vol. 55, May, pp. 47-47.
25. Zimmermann, J.J. (1991). Fuzzy Set Theory and its Applications, Kluwer Academi, Boston.