Assessing the Stability of Market Segment – A fuzzy Clustering Approach

D.M. Sezhiyan

ABSTRACT

Market segmentation is the process of grouping potential consumers of like-minded decision makers for rendering the company's market offerings. Most of the market segmentation research results portrays only the profile of the segments and the market size. The size of the market is always a question mark for its consistency, because the consumers' decision to prefer a brand is fuzzy in nature. This paper addresses the limitation of crisp cluster over fuzzy segmentation and moreover emphasizes the method to measure the level of stability of the resulted segments. To empirically investigate, a study was conducted on the preferences of motor bikes among college going boys in southern part of Tamilnadu. The result of the research reveals that the segmented market was fuzzy in nature which reflects the real market situation. The resulted fuzzy segments were labelled as 'economic class', 'elation class' and 'robust class' and later it was compared with the crisp clusters to assess the stability of each segment. Marketing managers can choose the segment based on its stability and this research paper adds value to literature by providing a formula for arriving a stability score for each segment.

KEYWORDS: Market Segmentation, Stableness of Segment, Fuzzy Clustering

JEL CLASSIFICATION: M31

BIOGRAPHICAL NOTE: D.M. Sezhiyan, Assistant Professor, Department of Management Studies, National Institute of Technology Tiruchirappalli, India. Prior to his academic career, he was with KSDL and Emami. He has a decade of marketing experience in Financial Services and FMCG Industry. His area of teaching is strategic management and marketing. He can be reached at sezhiyan@nitt.edu/ dm.sezhiyan@gmail.com

INTRODUCTION

Market Segmentation is one of the most central concepts in contemporary marketing (Kotler, 1997). To build competitive advantage both in short term and in the longer term, companies need to segment their target group. Market segment can aid in better achieving strategic and tactical objectives (Lamb et al., 1992). Market definition presupposes the homogeneity of substitutive relations for an aggregate of actual or potential customers. If substitutive relations are heterogeneous, the total demand should be divided internally into more homogeneous submarkets. Otherwise one runs of obtaining measurement artefacts of substitutive relations. The problem of dividing a total demand into submarkets may be subsumed under the concept of market segmentation (Frank et al. 1972, Wind, 1978). Market segmentation solutions should necessarily be effective, promising to gain profit and it should pave a way for designing market strategy. To be useful, the solutions of market segments must rate positively on six key criteria: (i) Identity, (ii) Substantiality, (iii) Accessibility, (iv) Stability, (v) Responsiveness, (vi) Actionability.

Identity refers to the extent that we can really identify segments (e.g. how well our solutions look; can we really see differences between segments; are the segments well
Assessing the Stability of Market Segments – A Fuzzy Clustering Approach

Substantiality refers to the size of the segmented market concern (i.e.) are our identified segments large enough to warrant separate market targeting. Accessibility refers to the extent, to which we reach the customers in our segments; Responsiveness refers to the extent that the different market segments respond uniquely to the marketing efforts directed at them. Finally, actionability refers to the extent to which the identified market segmentation provides direction of marketing efforts. Among all the six criteria, however, the prime aim of this research paper is to focus on assessing the stability of market segmentation.

LITERATURE REVIEW

The market segmentation literature proffers many procedures to segment the market and to assess the stability of segments as well. The goal of this paper is to address the measurement of segments’ stability by means of fuzzy clustering. The fuzzy segmentation holds good potential for assessing the stability of segments because it is easy and one pass algorithm. Moreover it replicates the factual information of the segments which is compatible to compare with hard cluster solution or traditional clustering methods. The petite scrutiny on the mentioned literature review of segmentation method substantiates that the fuzzy segmentation method has a good potential to assess the stability of different segments.

The methods used to identify the segments will be an important aspect in obtaining a successful segmentation. Michel Wedel and Kamakura (1998), describe four classes of segmentation methods:

1. A priori descriptive methods (e.g. through cross-tabs, OLAP cubes).
2. A priori predictive methods (e.g. discriminant analysis).
3. Post hoc descriptive methods (e.g. clustering methods and mixture models).
4. Post hoc predictive methods (e.g. cluster wise regression, CHAID, CART, mixture regression modes).

The most important distinction between descriptive and predictive methods is that the descriptive methods are defined in general as those methods that analyze a set of variables, without a distinction between dependent and independent variables. Predictive methods are defined in general as methods that analyze a set of variables, whereby one variable is designated as the dependent variable and the other variables are designed as the independent variables.

The market literature recommends defining market boundaries based on the view of actual or potential customers and the methods used for structuring markets may be made up by two stages (i) If heterogeneity of customers is relevant, customers are grouped into segments with high internal homogeneity and (ii) an appropriate presentation of the relations between brands is determined for each customer segment.

Research often uses clustering analysis as a tool for segmenting the market. Usually most researchers use Hierarchical and Non- Hierarchical clustering procedures which typically results in crisp partitioning form. To explain more, the partition is characterized by the fact that pairs of clusters do not contain the same objects. Most clustering algorithms build up such partitioning methods where one sample or member cannot belong to two or more groups and another important aspect to ponder is that the diverse substitution indicators may also serve as segmentation criteria. The similarity of customers with regard to the substitution indicator forms the basics of the construction of homogeneous submarkets. These pairwise similarities between customers can be transferred into simpler classification types. (i) Hard or Crisp Partition, (ii) Overlapping Classification, (iii) Fuzzy Partition.

As mentioned earlier, the crisp partitioning is one which the one sample or member cannot belong to two or more groups where as the overlap clustering solutions will produce the similar kind of results as of fuzzy clustering solutions but it has different from the hierarchical clustering algorithm. Hierarchical clustering algorithms are among the most commonly used clustering analysis
D.M. Sezhiyan

in market segmentation researches. These approaches will discard much of details related to the levels of nesting that happen for specific clusters. The overlap clustering will obtain a partition of the set of entities being clustered. That is to say, the objects are segregated into mutually exclusive (and exhaustive) subsets. For a single partition, there is no nesting of subsets or clusters, so that the representation of structure through a partition is necessarily non-hierarchical. The methods of overlap clustering can be allude to following ways like BK method of Jardine and Sibson (1968), as well that of Peay (1974) and ADCLUS (ADditive CLUStering) method by Phipps A.J et al. 1981). About fuzzy partitioning, according to Zadeh (1965), the set theory concept of a partition may be generalized to a fuzzy partition. Fuzzy set theory replaces the hard membership or non-membership of an object to a subset by gradual membership indicating the nearness of an object to the class. Contrary to classical crisp partitions, the membership values may not only take the two values 0 or 1, but all values in the interval between 0 and 1. To substantiate the mentioned conceptual issue Cattell (1978), proves the boundary between groups is fuzzy in the real market, and so is the boundary between customers and competitors. Same like overlap clustering, the fuzzy clustering method can also be divided into two types (Yang 1993). One uses a fuzzy relation to perform fuzzy clustering and the other uses objective function to determine fuzzy clustering. The objective function of fuzzy clustering holds good potential for soft segmentation. The fuzzy clustering methods can convey more information than hard clustering (Zimmermann 1991, Wedel and Steen Kamp, 1991), In fuzzy clustering theory, the 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 segment, and has a tendency to be identified with a single stronger segment and also possible to measure the membership grade to analyze the characteristics of market segments.

The FCM clustering algorithm was first presented by Dunn (1974), and it was further developed by following contributors, Backer (1978), Jim Bezdek et al. (1981), Trauwaert (1985), Gath and Geva (1989), Gu and Dubuisson (1990) and Xie and Beni (1991). However, Jim Bezdek’s FCM remains as the most commonly used. FCM Partitions a collection of n vectors \( \mathbf{x}_i \), \( i = 1, \ldots, n \) into c fuzzy groups and finds a cluster center in each group such that a cost function of dissimilarity measure is minimized. The major difference between FCM and HCM is that FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by the membership grades between 0 and 1. However, by imposing normalization which stipulates that the summation of degree of belongingness of a data set always is equal to unity:

\[
\sum_{i=1}^{n} u_{ij} = 1, \quad \forall = 1, \ldots, n
\]  

The cost function (or objective function) for FCM is then a generalization of Equation (1):

\[
j(U, C_1, \ldots, C_c) = \sum_{i=1}^{n} J_i = \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij}^m d_{ij}^2
\]

Where \( u_{ij} \) is between 0 and 1; \( C_i \) is the cluster center of fuzzy group \( i \); \( d_{ij} = \| \mathbf{C}_i - \mathbf{x}_j \| \) is the Euclidean distance between \( i \) th cluster and \( j \) th data point; and \( m \) is \( \in [1, \infty) \) is a weighting exponent.

The necessary conditions for Equation (2) to reach a minimum can be found by forming a new objective function \( \tilde{j} \) as follows:

Where \( \lambda_j, j = 1 \) to \( n \), are the Lagrange multipliers for the \( n \) constraints in Equation (1). By differentiating \( j(U, C_1, \ldots, C_c, \tilde{e}_1, \ldots, \tilde{e}_n) \) with respect to all its input arguments, the necessary conditions for Equation (2) to reach its minimum are

\[
\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \lambda_j (\sum_{i=1}^{c} U_{ij} - 1) = 0
\]

\[
\sum_{j=1}^{n} u_{ij}^m \lambda_j (\sum_{i=1}^{c} U_{ij} - 1) = 0
\]

\[
\sum_{j=1}^{n} u_{ij}^m \lambda_j (\sum_{i=1}^{c} U_{ij} - 1) = 0
\]

\[
\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \lambda_j (\sum_{i=1}^{c} U_{ij} - 1) = 0
\]

\[
\sum_{j=1}^{n} u_{ij}^m \lambda_j (\sum_{i=1}^{c} U_{ij} - 1) = 0
\]

\[
\sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \lambda_j (\sum_{i=1}^{c} U_{ij} - 1) = 0
\]
Assessing the Stability of Market Segments - A fuzzy Clustering Approach

\[ c_i = \frac{\sum_{j=1}^{n} u_{ij} x_j}{\sum_{j=1}^{n} u_{ij}} \quad (4) \]

And

\[ u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{d_{ij}}{d_{kj}} \right)^{1/m-1}} \quad (5) \]

The fuzzy C-means algorithm is simply an iterated procedure through the preceding two necessary conditions. In a batch-mode operation, FCM determines the cluster centers \( c_i \) and the membership matrix \( U \) using the following steps:

**Step 1:** Initialize the membership matrix \( U \) with random value between 0 and 1 such that the constraints in equation-1 are satisfied.

**Step 2:** Calculate \( c \) fuzzy cluster centers \( i = 1..c \) using Equation (4).

**Step 3:** Compute the cost function according to Equation (2). Stop if it's either reaches below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

**Step 4:** Compute a new \( U \) using Equation (5) and go to step 2.

The cluster center can also be first initialized and then the iterative procedure carried out. However, there is no guarantee to ensure that FCM converge to an optimum solution. Such performance depends on the initial cluster centers, thereby allowing us either to use another fast algorithm to determine the initial cluster centers or to run FCM several times, each starting with a different set of initial cluster centers.

As mentioned above, the Fuzzy C-means clustering algorithm uses the minimization of the fuzzy C-means function. There are three input parameters needed to run this function: One - the number of clusters or the partition matrix to be initialized; Two- the fuzziness weight exponent; and Three - the maximum termination tolerance. The last two parameters have their default values, so the user has to define the input about the number of cluster. Fuzzy cluster validity functions like Partition coefficient, Partition entropy, Xie and Beni validity function can be used for initialization purposes. This research has not used any such validity function and initialization, that is to say, the number of cluster was made based on judgemental basis.

The advantage of fuzzy partitioning over the crisp partitioning is that the fuzzy partitioning have fewer values of membership which allows to identify the 'bridges' i.e., objects between the classes. Moreover, the fuzzy membership values may be interpreted to measure the extent of correspondence to typical objects of a class (Kandel, 1982)

The novel thought here is that to compare Hard and Soft segmentation for measuring and there upon to prove the differences in the size of the clusters. The important issue and the point to ponder here is that, the market size obtained from hard segmentations will be different from the soft segmentations. The basic market size differences will help us to evaluate the stability of the different market segments. The smaller differences of market size obtained from the FCM algorithm will indicate that the segmented market will be more stable.

**PURPOSE AND RESEARCH METHODS**

To empirically investigate the stability of market segmentation, the study has carried out to segment the motor bike market based on the preferences of the college going boys in the southern part of Tamilnadu state. To select the most appropriate and meaningful segmentation, the item or questionnaire was designed and developed to suit the basic requirements for segmenting the motor bike market. The following nine major attributes were identified: they are Price, Low maintenance, Fuel efficiency, Young appealing, Latest style and Design, Ruff on road, Longevity of the bike and Overall Performance of the bike. All items are measured on a 7-point Likert-Type scale. To bring face validity for the instrument, a pilot study of fifty in-depth interview were conducted with college students and extensive discussion made with sales personnel of various motor bike companies. Suggestions through feedback regarding the structure, wording and time...
taken to complete the questionnaire were considered, and changes were made to the questionnaire to reflect respondents’ recommendations. Data were collected from the various colleges of different types from southern part of Tamilnadu through personal scheduled interviews. The usable questionnaires were received from the respondents with an average response rate of 76% that is to say 220 samples. The reliability test of Cronbach alpha was performed for all the nine attributes that are portrayed in the below Table 1. The coefficient alpha for the all the nine variables are above 0.9 which indicates the highest reliability and to further stress on the reliability, the value of Cronbach alpha before standardizing was 0.952739 and after standardization, the value of Cronbach alpha increased to 0.952875, which indicates that these nine variables or segment descriptors have high uniformity.

### Table 1: Item analysis for Segmentation descriptors

<table>
<thead>
<tr>
<th>Reliability Item Analysis</th>
<th>Section</th>
<th>Item Values</th>
<th>If This Item is Omitted</th>
<th>Cronbach’s Alpha</th>
<th>Std. Deviation</th>
<th>Total Means</th>
<th>Total Std.Dev</th>
<th>Total Alpha</th>
<th>Cronbach’s Alpha</th>
<th>Std. Deviation</th>
<th>Total Means</th>
<th>Total Std.Dev</th>
<th>Total Alpha</th>
<th>Cronbach’s Alpha</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>4.924581</td>
<td>1.592141</td>
<td>53.03073</td>
<td>12.7715</td>
<td>0.9497</td>
<td>0.7455</td>
<td>0.6504</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low maintenance</td>
<td>5.002793</td>
<td>1.558385</td>
<td>52.95251</td>
<td>12.71936</td>
<td>0.9476</td>
<td>0.801</td>
<td>0.7457</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel efficiency</td>
<td>5.209497</td>
<td>1.553245</td>
<td>52.74581</td>
<td>12.70875</td>
<td>0.9472</td>
<td>0.8115</td>
<td>0.7672</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young appealing</td>
<td>5.315642</td>
<td>1.525995</td>
<td>52.63966</td>
<td>12.75085</td>
<td>0.9477</td>
<td>0.7979</td>
<td>0.7184</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latest style</td>
<td>5.307262</td>
<td>1.526793</td>
<td>52.64804</td>
<td>12.76524</td>
<td>0.9481</td>
<td>0.7871</td>
<td>0.7214</td>
<td>0.952739</td>
<td>0.952875</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reputed brand</td>
<td>5.307262</td>
<td>1.51019</td>
<td>52.64804</td>
<td>12.75031</td>
<td>0.9474</td>
<td>0.8079</td>
<td>0.7478</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruff on roads</td>
<td>5.307262</td>
<td>1.523119</td>
<td>52.64804</td>
<td>12.78432</td>
<td>0.9485</td>
<td>0.7756</td>
<td>0.663</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long life</td>
<td>5.332402</td>
<td>1.574876</td>
<td>52.62291</td>
<td>12.78415</td>
<td>0.9496</td>
<td>0.7463</td>
<td>0.6446</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall performance</td>
<td>5.120112</td>
<td>1.639699</td>
<td>26.58101</td>
<td>6.860995</td>
<td>0.9134</td>
<td>0.619</td>
<td>0.4034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IIMS Journal of Management Science 133
Assessing the Stability of Market Segment – A fuzzy Clustering Approach

The discussion of the results

The results of the final cluster centers of both K-means and Fuzzy C-means clusters are portrayed in the below Table 2.

Table 2: Cluster centroid value for both Hard and Soft Segments.

<table>
<thead>
<tr>
<th>Segmentation descriptors</th>
<th>K Means Cluster centroids value</th>
<th>Fuzzy C means Cluster centroids value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster No 1</td>
<td>Cluster No 2</td>
</tr>
<tr>
<td>Price</td>
<td>6</td>
<td>2.17</td>
</tr>
<tr>
<td>Low maintenance</td>
<td>6</td>
<td>2.83</td>
</tr>
<tr>
<td>Rel. efficiency</td>
<td>5.33</td>
<td>2.67</td>
</tr>
<tr>
<td>Young appealing</td>
<td>2.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Latest style</td>
<td>3</td>
<td>5.67</td>
</tr>
<tr>
<td>Reputed brand</td>
<td>2.17</td>
<td>6.33</td>
</tr>
<tr>
<td>Ruff on roads</td>
<td>2.67</td>
<td>2.33</td>
</tr>
<tr>
<td>Longevity</td>
<td>3.33</td>
<td>2.17</td>
</tr>
<tr>
<td>Overall performance</td>
<td>2.5</td>
<td>2.33</td>
</tr>
</tbody>
</table>

The results of final cluster centers of both K-means and Fuzzy C-means resemble almost similar measure. This indicates that the clusters are well separated and reflects both the K-means and Fuzzy C-means algorithms effectively classified or clustered in the given data set. The label of each segment was made based on the cluster center values. Cluster 1 has relatively high value on the following variables: Price, Low Maintenance and Fuel efficiency. Hence, Cluster 1 could be labeled as “Economic Class”. Cluster 2 has relatively high value on the following variables: Young appealing, Latest Style and Preferences of reputed brands. Hence, Cluster 2 could be labeled as “Elation Class” and the Cluster 3 has relatively high value on the following variables: Ruff on roads, Longevity and the overall performance of bike. Hence, Cluster 3 could be labeled as “Robust Class”. From the results of FCM, The strongest possibility of the sample 9 belongs to “Robust class”, while there is a weaker possibility that it belong to other segments too. However, the hard segmentation shows that the sample only belongs to “Robust Class” Sample 9, 11 and 17 do not totally belong to one segment but they also involve in other segments. Based on the percentage of fuzziness, Economic class is considered to be the stable segment, which is followed by robust class. Among the three segments “Elation class” is said to be more volatile segment. To substantiate further and to have apparent depiction, a new segmentation index is created based on below Equation (6).
<table>
<thead>
<tr>
<th>Case/ sample Number</th>
<th>Segmentation Label</th>
<th>Hard Cluster Membership</th>
<th>Soft Cluster Membership Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cluster Numbers</td>
<td>Cluster No 1 Economic Class</td>
</tr>
<tr>
<td>1</td>
<td>1*</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2**</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3***</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>3</td>
<td>0.0001</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>3</td>
<td>0.8496</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>3</td>
<td>0.0001</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Market Size (Soft Segment)**
- of 220 samples: 64.339, 66.1441, 89.5169

**Market Size (Hard Segment)**
- of 220 samples: 64, 68, 88

**Differences**
- Compared to Hard Segments: 0.53%, 2.73%, 1.72%

* (Economic Class)
** (Elation Class)
*** (Robust Class)
Assessing the Stability of Market Segment – A fuzzy Clustering Approach

Where $SI_i$ is the Stability Index for the given segments $i$, and $diff$ is the differences between the sizes of hard segmentation and soft segmentation of particular market segment.

CONCLUSIONS AND MANAGERIAL IMPLICATIONS

This paper reviews various approaches to market segmentation. It is argued that the fuzzy segmentation method is potentially a useful technique to assess the stability of different segments. To substantiate the argument, it’s proved that the fuzzy clustering analytical method has the following advantages: (i) The FCM uses membership grade to do cluster numbering, so the membership value can help the marketers to understand the real market size in each segment. (ii) The FCM sketches membership value of every sample, which supports marketers to visualize individual’s level of multiple preferences. The concern to redolent here is that the extent to which market segment is reached depends on the extent to which they can be profiled with consumer descriptors. The grade value of each member will facilitate marketers to profile individuals who lie in different segments. This makes easy for marketers to design market offerings and to sketch out the strategies for business development actions. (iii) The results of soft segmentation solutions are attuned for comparing with hard cluster solutions. So it’s made easy for managers to assess the consistency of each segment.

However, this research is not without its limitations, which may be addressed in future research. (i) The FCM involves computation in the choices of fuzziness weight exponent and the maximum termination tolerance. Further research might investigate on developing valid thumb rule value for fuzziness weight exponent as well for the maximum termination tolerance. (ii) This study concentrates only on the respondent’s decision making descriptors on buying the motor bikes but, it has not pondered on the other usual segmentation variables like demographics, lifestyle, etc.

REFERENCES


D.M. Sezhiyan


