**International Journal of Commerce and Management Research**
**ISSN: 2455-1627, Impact Factor: RJIF 5.22**
**www.managejournal.com**
**Volume 2; Issue 10; October 2016; Page No. 68-71**

# Cluster analysis technique and market segmentation

**[1] Anita Rani, [2] Dr. Balbir Singh**

[1] Research Scholar, Department of Commerce, Singhania University, Pacheri Bari, Jhunjhunu, Rajasthan, India
[2] Principal, B.A.R. Janta College, Kaul, Kaithal, Haryana, India

**Abstract**
Cluster analysis is a statistical technique that sorts observations into similar sets or groups. The use of cluster analysis presents a complex challenge because it requires several methodological choices that determine the quality of a cluster solution. This paper discusses a cluster analysis technique for market segmentation. Single linkage technique is used to conduct market segmentation. The solutions resulting from this approach result in market segments or groups each of which consists of people or organizations with similar demand of products and/or services in terms of price quality and functions. The similarity coefficient metric between any two customers is determined based on product type, price and functionality. The decision on number of clusters/market segments is based on stability of cluster structure. A sample numerical data set is constructed for a set of customer and the single linkage cluster analysis technique is tested on the sample data set. It is shown that the cluster analysis approach yields stable market segments for effective marketing strategies of the firm.

**Keywords:** Cluster Analysis, Segmentation, neighbor heuristic, Consumer Behavior

## Introduction
A market segment is a subset of market made up of people or organizations with one or more characteristics that cause them to demand similar products or services based on the qualities of the product such as price, function. A true market segment meets the following criteria it is district from other segments (different segments have different needs), it is homogenous within a segment, it is responds similarly to a market stimulus and it can be reached by a market intervention. The market segmentation term is also used when consumers with identical product/service needs are categorized into groups so that they can be charged different amounts.

Positive market segmentation means dividing the market into groups of individuals markets with similar wants or needs that company divides into district groups which have district needs, wants, behavior. Markets can be divided according to number of criteria such as industrial private, public. Industrial market segmentation is a scheme for categorizing industrial and business customers to guide strategic and tactical decision making. Goal of every industrial market segmentation is to identify most significant differences among current and potential customers that will influence purchase decisions or buying behavior. This allows industrial marketer to differentiate prices, programs or solutions for maximum competitive advantage. Consumer based market segmentation can be performed on a product specific basis to provide a close match between specific products and individuals. Improved market segmentation can lead to improved marketing effectiveness.

Cluster Analysis technique is a multi-variant statistical technique that seeks to organize information about variables so that relatively homogenous groups or clusters are formed. Clusters formed from this method should be highly internally homogenous and externally homogenous. The four basic steps in cluster analysis: Data collection and selection of variables for analysis, generation of similarity coefficient matrix, decision on the number of clusters and interpretation, validation of cluster solution. The output of cluster analysis technique is called cluster analysis diagram which is a tree diagram.

## Review of Literature
Kamakura, W.A (1999) [7] developed a probabilistic choice model for market segmentation elasticity structure. Haley, R. (1968) [5] proposes an approach where market segments are delineated first on the basis of factor with a causal relationship to future purchase behavior. The belief underlying this segmentation strategy is that benefits which people are seeking in consuming a product are basic reasons for existence of true market segment. Yankelovich D. (2014) [12] developed new criteria for market segmentation. The approach takes into account psychological and qualitative factors for segmentation of markets. Grover and Srinivasen benefit issues a simultaneous approach to market segmentation and market structuring. The cost segmentation plans. The paper offers an alternative view of normative segmentation and suggests some modification to standard segmentation practices. Beane, T.P. (2007) [2] proposes five different bases for segmentation markets: geographic, demographic, psychographic, behavioristic and image. This is followed by an overview of main techniques used to establish and verify segments including automatic interaction detector, conjoint analysis, multi-dimensional scaling and canonical analysis. Johnson, R (2011) [6] proposed a strategic management tool for doing market segmentation. Kotler (2003) [9] in his book has conducted an in-depth survey of market segmentation and different positioning strategies with respect to same.

## Methodology for Solution
Traveling sales person problem (TSP) is a non-polynomial hard problem that seeks to find an optimal permutation of cities that a sales person should cover at the minimum cost. The constraint is that sales person should cover all cities and

come back to the home base. The order in which he covers the cities is important so as to minimize the total transportation cost tour. Nodes of the TSP tour represent cities and are weight between nodes represents transportation cost.

The analogy of TSP can be extended to market segmentation scenario where the nodes represent customers and are weight represented dissimilarity between a pair of customers. The TSP tour seeks to obtain an optimal permutation of customers visited so that the total dissimilarity of customers in minimized.

The nearest neighbor heuristic can be constructed the TSP tour. The city of origin is considered in first instance, the city at the minimum transportation cost from the first city is linked to the first city by an arc. The city at the minimum transportation cost from the second city is considered and linked to the second city by an arc. The city at the minimum transportation cost from the third city is considered and linked to third city by an arc. The procedure is continued till all the cities are covered. The procedure results in the formation of optimal permutation of cities that the sales person travels so that transportation cost is minimized.

Single linkage cluster analysis technique involves construction of a Traveling sales person (TSP) tour with minimum dissimilarity where the nodes is TSP tour represent customers and are weight the customers indicated the dissimilarity between customers. The nearest in first instance, the customer at the minimum dissimilarity from the first customer is linked to the first customer by an arc.

The customer at the minimum dissimilarity from the second customer is considered and linked to the second customer by an arc. The customer at the minimum dissimilarity from the third customer is considered and linked to third customer by an arc. The procedure is continued till the entire customers are covered. They are weight is the dissimilarity between customers. The procedure results in the formation of optimal permutation of customer that must be sequenced so that dissimilarity of customer is minimized.

$$S_{ij} = \frac{\text{Number of Products Commonly required by i and j}}{\text{Number of Products required by i and j}}$$

Dissimilarity $D_i = - S_{ij}$

Conceptually similarity coefficient indicates the degree of closeness or similarity between customers while the dissimilarity coefficient indicated degree of dissimilarity between customers in terms of product demand.

The next step in single linkage cluster analysis is to break the TSP tour first where the dissimilarity between the customers is maximum. The TSP tour is subsequently at intervals between the customers where dissimilarity between customers is $k^{th}$ maximum. The number of market segments formed is (K + 1). The value of k can be estimated from the staffed full time equivalents available with the company and budgetary constraints of the company.

**Numerical Example 1:** The Product Customer matrix is given in the below. The number 1 indicates that the customer buys the product and the number 0 indicates that the customer does not buy the product.

The RANDBETWEEN function is used to generate random number either 0 or 1. The matrix is given as:

**Table 1:** Product Customer Matrix

| | | Product Customer Matrix | | | | |
|---|---|---|---|---|---|---|
| | | Products | | | | |
| | | P1 | P2 | P3 | P4 | P5 |
| Customer | C1 | 1 | 1 | 1 | 1 | 0 |
| | C2 | 1 | 1 | 0 | 0 | 0 |
| | C3 | 0 | 1 | 1 | 1 | 0 |
| | C4 | 1 | 0 | 1 | 1 | 1 |

The table above is a product customer matrix which gives a clear idea of which product is required by which customer. From the above formulae the similarity coefficient values between customers can be calculated and tabulated as follows.

**Table 2:** Similarity Coefficient matrix

| | Similarity Coefficient Matrix | | | |
|---|---|---|---|---|
| | **C1** | **C2** | **C3** | **C4** |
| C1 | 1 | 0.5 | 0.75 | 0.6 |
| C2 | 0.5 | 1 | 0.25 | 0.2 |
| C3 | 0.75 | 0.25 | 1 | 0.4 |
| C4 | 0.6 | 0.2 | 0.4 | 1 |

On the basis of Similarity Coefficient Matrix, the Dissimilarity coefficient matrix also called Distance Matrix is obtained. The Dissimilarity value is the negative of Similarity value.

**Table 3:** Dissimilarity Coefficient Matrix

| | Dissimilarity Coefficient Matrix | | | |
|---|---|---|---|---|
| | C1 | C2 | C3 | C4 |
| C1 | -1 | -0.5 | -0.75 | -0.6 |
| C2 | -0.5 | -1 | -0.25 | -0.2 |
| C3 | -0.75 | -0.25 | -1 | -0.4 |
| C4 | -0.6 | -0.2 | -0.4 | -1 |

The nearest neighbor heuristic is used for computation of Traveling Sales Person tour. The customer 1 is considered first for assignment. The minimum dissimilarity coefficient value is considered. The customer 3 is appended to the customer 1 in a chain. The minimum dissimilarity coefficient value with respect to customer 3 is considered from the matrix row. The customer 4 is appended to the customer 3 in the chain. The customer 2 is thereafter appended to the customer 4 in the chain.

**Table 4:** The TSP Tour can be observed as

| | | | |
|---|---|---|---|
| 1 | 3 | 4 | 2 |

Theoretically a cut can be made between customer 1-3 and/or customer 3-4 and/or customer 4-2. Out of these three possibilities it can be observed that the maximum dissimilarity coefficient value is between customer 2-4 and equal to 0.2. The cut is made where the dissimilarity is maximum or similarity coefficient value is minimum. We decide to make two cuts.

Thus three clusters/segments can be formed

Cluster 1 (Market Segment 1): Customer C1 and C3

Cluster 2 (market Segment 2): Customer C4
Cluster 3 (market Segment 3): Customer C2

**Numerical Example 2:** The Product Customer matrix is given in the table below. The number 1 indicates that the customer buys the product and the number 0 indicates that the customer does not buy the product.

**Table 5:** The Rand Between function is used to generate random number either 0 or 1. The matrix is given as:

| | **P1** | **P2** | **P3** | **P4** | **P5** | **P6** | **P7** | **P8** | **P9** | **P10** |
|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| C2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| C3 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| C4 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| C5 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| C6 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| C7 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |
| C8 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| C9 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| C10 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 |

**Table 6:** The similarity coefficient matrix is given as follows from the formulae given in the paper.

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|
| C1 | 1 | .25 | .33 | .57 | .3 | .5 | .42 | .88 | .33 | .22 |
| C2 | .25 | 1 | .6 | .4 | .17 | .14 | .2 | .22 | .33 | 0 |
| C3 | .33 | .6 | 1 | .5 | .38 | .25 | .14 | .3 | .42 | .13 |
| C4 | .57 | .4 | .5 | 1 | .25 | .5 | .4 | .5 | .13 | .42 |
| C5 | .3 | .17 | .38 | .25 | 1 | .5 | .13 | .4 | .38 | .43 |
| C6 | .5 | .14 | .25 | .5 | .5 | 1 | .33 | .5 | .11 | .29 |
| C7 | .42 | .2 | .14 | .4 | .13 | .33 | 1 | .13 | .14 | .4 |
| C8 | .88 | .22 | .3 | .5 | .4 | .5 | .13 | 1 | .3 | .33 |
| C9 | .33 | .33 | .42 | .13 | .38 | .11 | .14 | .3 | 1 | .29 |
| C10 | .22 | 0 | .13 | .42 | .43 | .29 | .4 | .33 | .29 | 1 |

**Table 7:** The Dissimilarity Coefficient Matrix can be computed as follows:

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|---|---|---|---|---|---|---|---|---|---|---|
| C1 | -1 | -.25 | -.33 | -.57 | -.3 | -.5 | -.42 | -.88 | -.33 | -.22 |
| C2 | -.25 | -1 | -.6 | -.4 | -.17 | -.14 | -.2 | -.22 | -.33 | 0 |
| C3 | -.33 | -.6 | -1 | -.5 | -.38 | -.25 | -.14 | -.3 | -.42 | -.13 |
| C4 | -.57 | -.4 | -.5 | -1 | -.25 | -.5 | -.4 | -.5 | -.13 | -.42 |
| C5 | -.3 | -.17 | -.38 | -.25 | -1 | -.5 | -.13 | -.4 | -.38 | -.43 |
| C6 | -.5 | -.14 | -.25 | -.5 | -.5 | -1 | -.33 | -.5 | -.11 | -.29 |
| C7 | -.42 | -.2 | -.14 | -.4 | -.13 | -.33 | -1 | -.13 | -.14 | -.4 |
| C8 | -.88 | -.22 | -.3 | -.5 | -.4 | -.5 | -.13 | -1 | -.3 | -.33 |
| C9 | -.33 | -.33 | -.42 | -.13 | -.38 | -.11 | -.14 | -.3 | -1 | -.29 |
| C10 | -.22 | 0 | -.13 | -.42 | -.43 | -.29 | -.4 | -.33 | -.29 | -1 |

The nearest neighbor heuristic is considered for computing the Traveling Sales Person Tour Solution. From the nearest neighbor heuristic, the TSP Tour obtained is:

**Table 8:** They are weights from the Dissimilarity matrix are given as:

| C1 | C8 | C4 | C3 | C2 | C9 | C5 | C6 | C7 | C10 |
|---|---|---|---|---|---|---|---|---|---|

C1 – C8 Dissimilarity = -0.88 Similarity = 0.88
C8 – C4 Dissimilarity = -0.5 Similarity = 0.5
C4 – C3 Dissimilarity = -0.5 Similarity = 0.5
C3 – C2 Dissimilarity = -0.6 Similarity = 0.6
C2 – C9 Dissimilarity = -0.33 Similarity = 0.33
C9 – C5 Dissimilarity = -0.38 Similarity =0.38
C5 – C6 Dissimilarity = -0.5 Similarity = 0.5
C6 – C7 Dissimilarity = -0.33 Similarity = 0.33
C7 – C10 Dissimilarity = -0.4 Similarity = 0.4
C10 – C1 Dissimilarity = -0.22 Similarity = 0.22

The cut is made between customer C8 and C4, Customer C2 and C9, Customer C6 and C7 to achieve homogeneity and similarity within a cluster/segment.

Cluster 1 (Market Segment1)
Cluster C1, Customer C8
Cluster 1 (Market Segment2)
Cluster C4, Customer C3, Customer C2
Cluster 1 (Market Segment3)
Cluster C9, Customer C5, Customer C6
Cluster 1 (Market Segment4)
Cluster C7, Customer C10

**Conclusions**

We conclude from our research studies in this cluster analysis is an effective method for doing market segmentation. Market Segments are obtained from single linked cluster analysis and nearest neighbor methods. The customers in a segment exhibit a high degree of similarity in terms of product purchases. Consumer behavior in terms of product selected by customers is taken into account in market segmentation. A market segment is considered to be dissimilar from other market segment. The similarity and dissimilarity between customers is quantified by a coefficient value. Single linkage cluster analysis methods are related to the Traveling Sales person problem and efficient solutions are computed with the nearest neighbor method. The number of segments / clusters is based on the financial budgetary policies of the company.

**References**


1. Arabie P, Hubert LJ. Cluster Analysis in Marketing Research. In Advanced methods in marketing research. Ed. R.P. Bagozzi. Blackwell: Oxford. 1994, 160-189.
2. Beane TP, Ennis DM. Market Segmentation a Review, European Journal of Marketing. 2007; 21(5):20-42.
3. Everitt BS. Cluster Analysis. New York: Halsted Press. 1993.
4. Grover R, Srinivasan V. An Approach for Tracking within Segment Shifts in market Shares, Journal of Marketing Research. 1989; 26:230-236.
5. Haley R. Benefit Segmentation a Decision Oriented Research Tool, Journal of Marketing. 1968; 32(3):30-35.
6. Johnson, Ravi. Market Segmentation, at Strategic Management Tool, Journal of Marketing Research. 2011; 8:13-18.

7. Kamakura WA, Russel GJ. A probabilistic choice model for market segmentation and elasticity structure, Journal of Marketing Research. 1999; 26:379-390.
8. Kohonen T. Self-Organizing Maps, 2nd edition. Berlin: Springer. 1997.
9. Kotler P. Marketing Management. 2003.
10. Myers JH, Tauber E. Market structure analysis. Chicago: American Marketing Association. 1977.
11. Yoram W, Thoman R. Conceptua & Methodical Issues in Origanizational Buying Behavior, European Journal of Marketing. 1980; 5:239-263.
12. Yankelovich D. New Criteria for market Segmentation, Harvard Business Review. 2014; 3:1-12.