ISSN: 2278-3369

International Journal of Advances in Management and Economics
Available online at: www.managementjournal.info

RESEARCH ARTICLE

A Quantitative Approach in Market Research

Saibal Dutta*, Sujoy Bhattacharya

Indian Institute of Technology Kharagpur, West Bengal, India.

*Corresponding Author: Email: ***saibal.13@gmail.com***

Abstract

Data mining techniques successfully apply in market segmentation and considered one of the important fields of research in marketing. The aim of this paper is to summarize the academic literature review of the available data mining techniques and its application to the market segmentation. This research surveyed eight online journal databases and categorized into thirteen groups based on data mining techniques. The brief description of each of the data mining techniques and suggestions is also discussed. The results indicate that neural networks and kernel based method are the most emerging and well explored data mining techniques. This research paper provides as a direction to the industry as well as academic researchers.

Keywords: *Market segmentation, Data mining, Review, Neural network, Support vector machine, Research methodology.*

Introduction

Market segmentation is one of the major tools in the marketing research field. Last few decades, several theories proposed in this research area, but still considered one of the complex areas in marketing literature. The term "Market segmentation" first proposed by professor smith and it is defined as the process of dividing a large market into the some similar characteristic groups [1]. The aim of this work is to present a widespread review of literature which are reported in academic journals related to the application of data mining techniques in various business domains. This research paper followed previously published classification framework [2]. The paper is structured as follows: first, explained the research methodology of the study; second, performed a comprehensive literature review and result presented in tabular format. Discussed briefly each of data mining techniques with their strengths and weakness finally, the conclusions and suggestion are discussed.

Research Methodology

Market segmentation is now days becoming a complex and multidisciplinary research field. The following online journal databases were utilized for searching the academic literature on the market segmentation area.

- Science direct
- ABI/INFORM Database
- Emerald full text
- IEEE Transaction
- JSTOR
- Springer
- Google scholar
- Wiley online Library

The following academic literature was searched with several keyword like "Market Segmentation" or "Target marketing" or "Data mining and market segmentation" and got around 851 articles. Each of the article was reviewed carefully and filtered out the irrelevant research paper that were not related to market segmentation. The required steps exhibited in Fig 1.

Literature Review

The literature review identified thirteen categories and each category also consists of several single or hybrid data mining techniques. Each category-wise several data mining techniques discussed and shown in the table with reference papers. This study surveys and classifies various market segmentation techniques into thirteen broad categories:

1. Neural network
2. Evolutionary algorithm
3. Fuzzy theory
4. RFM analysis
5. Hierarchical clustering
6. K means
7. Bagged Clustering

8. Kernel methods
9. Multidimensional scaling
10. Taguchi method
11. Model based Clustering
12. Rough sets
13. Others.

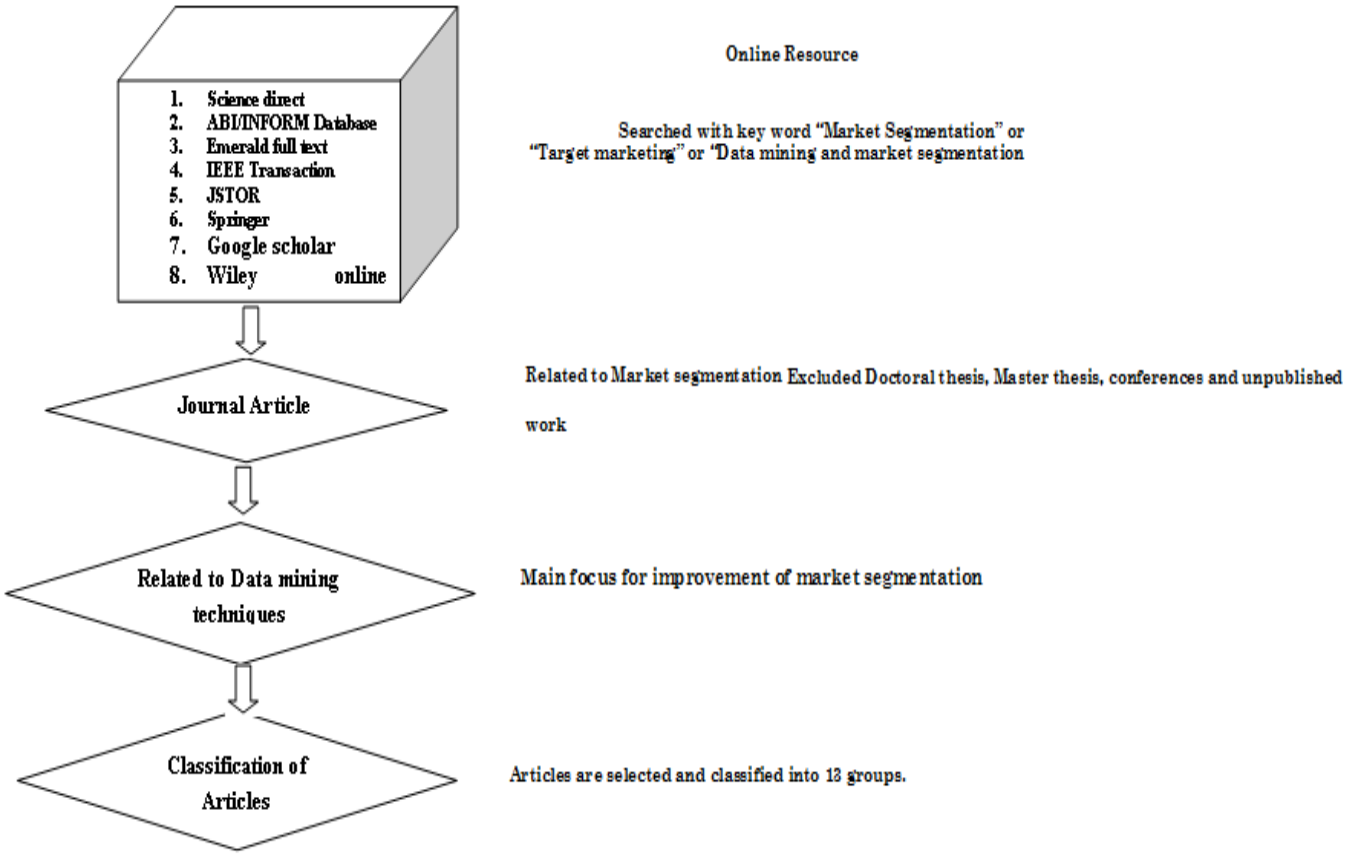

Figure1: Framework of research methodology

Literature review identified two basic approaches for market segmentation i.e. priori and posteriori. However, a priori segmentation bases are easy to implement using general segmentation variable, i.e. demographic, geographic and on the other side posteriori segmentation bases are done using data mining techniques called clustering. 60s and early 70s, researchers had used K-means and Hierarchical cluster analysis for market segmentation [3]. Even today, Researchers successfully applied the K means clustering for market segmentation [4,5]. But, above traditional clustering techniques have some drawbacks. For example, K-means algorithm cannot handle Noise and outliers of the data [6]. K-means algorithm also failed to give any exact or initial number of clusters and the statistical validity of the cluster formed [7] hence clustering fall into the local minima [8,9]. To overcome the above issue, the researcher proposed K means with Genetic algorithm to reach the global minima [10,11,12]. The steps of market segmentation showed in Fig 2.

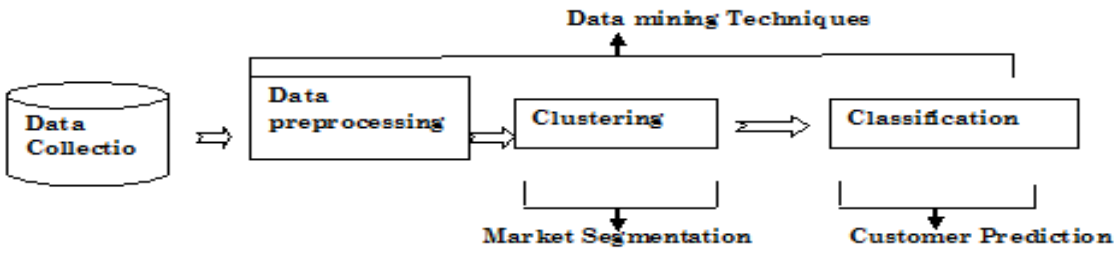

Figure 2: Steps for market segmentation

To classify the complex consumer pattern in the market, researchers also introduced other different approaches like Evolutionary algorithm, Kernel methods, rough set, Taguchi method, etc. Nowadays, the above approaches are quite popular among the market researcher and also these approaches are able to perform better market segmentation than traditional one [13]. Several other important market segmentation techniques like Multidimensional scaling, Random forest, RFM analysis, Bagged Clustering etc are also found in academic literature [14,15]. Major reviewed data mining techniques are as follows:

K means: K means algorithm is also known as square error based clustering and this algorithm can be implemented easily for solving practical marketing problem. Some of the advantages and disadvantages are as follows: efficient with a large number of variables and data. Computationally faster than hierarchical clustering. Manage to handle large data sets. Show good result if the data sets are well separated. Required number of cluster centers. Fall into the local optima of the squared error function. Failed to handle highly overlapping data and categorical data. Fail to handle noisy data and outliers. Fail to work when data sets are non-linear and with non-globular and very difficult to predict the number of clusters.

Hierarchical clustering: Hierarchical clustering evolves based on the proximity matrix and the result looks like a binary tree or dendrogram. Hierarchical clustering broadly classified as agglomerative methods and divisive methods. Many researchers used hierarchical clustering successfully in market segmentation. However, this algorithm is having serious drawback. Hierarchical clustering algorithm cannot handle large amount data and easily affected by the outlier. Termination criteria are not fixed. Failed to noise and outliers. Failed to handle large amount data and convex shapes.

Neural network: Due to increasing computer efficiency with less cost, Market segmentation using neural network has been quite popular in different domain of business research. Some of the popular areas are sales forecasting, bankruptcy prediction, direct marketing and target marketing. Self organizing map (SOM) is one of the popular neural network used for market segmentation. SOM helps to visualize input high dimension data into two dimensional map that allow us to find out important relationship among input data The main limitation of SOM is to find out exact clustering boundaries so researchers used a combination of SOM and K means to overcome the problem [16]. In programming point of view, SOM has serious drawbacks regarding determination of optimum cluster center, initial weights and stopping criteria.

Evolutionary algorithms: Several Evolutionary algorithms are available in the literature. Genetic algorithms (GA) and Particle swarm optimization (PSO) are the mostly used in market research. Both algorithms work in principle of evolutionary computation technique. For example PSO methodology was developed through the process of searching food by a group of birds. All the birds do not know the location of food, but if they follow the birds that is close to the food. PSO is also having pre-defined fitness value of fitness evaluation function similar to the GA and the objective is to optimize the fitness function.

RFM Analysis: RFM model based market segmentation is frequently found in marketing literature. The RFM model is used to find out the future consumer behavior pattern using a combination of past and present behavior pattern. Researchers generally used following RFM variables: Recency (R), Frequency (F), Monetary (M) [17]. RFM is cost-effective. RFM is very valuable in predicting the response of the customers. It is very effective to model with RFM variables as the purchase behavior can be summarized by using a very small number of variables. It is very easy to target particular customers. RFM is used to measure the strength of customer relationships. The RFM model generally identifies the best customer and failed to locate valuable customers RFM model can only use a limited number of selected variables. RFM focuses on a company's current customer and cannot be applied to the prospecting for new customers.

Fuzzy C means Algorithms: Fuzzy theory is based on the mathematics of fuzzy set. Fuzzy theory is used for the modeling of imprecise and the handling of uncertainty of qualitative knowledge various types of fuzzy clustering algorithms are available and fuzzy c means algorithm is one of the preferred techniques among the market researchers. Fuzzy c means also suffered similar problems of K means. Work on overlapped data set and comparatively better result than K-means algorithm. Required number of cluster centers, and membership value.

Bagged Clustering: Another useful market segmentation technique is Bagged clustering method which was first proposed by Leisch [15]. This algorithm is a combination of partitioning method and a hierarchical method [15] and result showed that this procedure performs better than available partitioning method.

Multidimensional scaling: Multidimensional scaling is a nonlinear mapping and generally used for 2-dimensional visualization. This technique is having poor generalization capability and highly noise sensitive [14].

Taguchi method: Taguchi method, a statistical method developed by Genichi Taguchi and used to select optimal or sub-optimal initial seeds for clustering. Taguchi method has been extensively used to improve the quality of manufactured goods through optimizing the design parameters and also other areas [18].

Kernel method: Kernel method worked based on cover's theorem and this method can handle complex non-linear data. This algorithm nonlinearly transforms data into higher dimensional feature space where we are able to separate data linearly [13]. The difficulty of dimensionality can be overcome by kernel trick. Many various kernel methods are available and kernel k means and support vector machine are few of them. The Kernel based method is successfully applied to obtain a linearly hyper plane in the high dimensional and easily handle, outlier even high noise [13]. The Algorithm is able to identify the non-linear structures. The Algorithm is best suited for real life data set. Number of cluster centers need to be predefined. The Algorithm is complex in nature and time complexity is large. The list of techniques shown in Table 1.

Table 1: List of data mining techniques applied in market segmentation

Method	Reference
Neural network	Mostafa [19], Hruschka and Natter [20], Balakrishnan [21]
	Boone and Roehm [22], Chen et al. [23], Bloom [24]
	Kuo et al. [16], Kauko et al. [25], Natter [26]
Evolutionary algorithm	Kim et al. [27], Ho et al. [28]
	Kuo et al. [5], Abdi et al. [29]
	Bursco et al. [30], Kuo et al. [31]
Fuzzy theory	Zhang [32]
RFM Analysis	Cheng and Chen [17]
	Wei et al. [33]
Hierarchical clustering	Miguéis et al. [34]
K- means	Li et al. [35]
Bagged Clustering	Brida et al. [15]
Kernel methods	Huang et al. [13], Wang [36]
Multidimensional scaling	Vishwanath and Chen [14] , Desarbo et al. [37]
Rough sets	Wu [38]
Taguchi method	Hong [18]
Model based Clustering	Kamakura and Russell [39]
Others	Chung et al. [40], Desarbo et al. [41]
	Bursco et al. [42], Albert et al. [43]
	Vriens et al. [44],Wedel and Kistemaker [45]
	DeSarbo and Grisaffe [46], Desarbo and Ramaswamy [47]

Conclusion

The purpose of paper is to provide a complete literature review to the academic and industrial people. This paper provides an extensive review of exiting data mining methodology in market segmentation. Thirteen different types of data mining techniques applied to the market segmentation process. This review process tries to incorporate all the relevant papers. The results showed that the neural network and support vector machine are quite favorable techniques. Researchers have also proposed hybrid algorithm for enhancing the performance of the market segmentation.

References

1. Smith WR (1956) Product differentiation and market segmentation as an alternative marketing strategy. Journal of Marketing, 21:3-8.
2. Ngai EW, Xiu L, Chau DCK (2009) Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2):2592-2602.
3. Punj G, Stewart DW (1983) Cluster analysis in marketing research: review and suggestions for application. Journal of marketing research, 134-148.
4. Boone DS, Roehm M (2002) Retail segmentation using artificial neural networks. International Journal of Research in Marketing, 19(3):287-301.
5. Kuo RJ, Chang K, Chien SY (2004) Integration of self-organizing feature maps and genetic algorithm- based clustering method for market segmentation. Journal of Organizational Computing and Electronic Commerce, 14(1):43-60.
6. Bezdek JC, Pal NR (1998) Some new indexes of cluster validity. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 28(3):301-315.
7. Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions of Pattern Analysis and Machine Intelligence, 24(12):1650-1654.
8. Chen MC, Wu HP (2005) An association-based clustering approach to order batching considering customer demand patterns. Omega, 33(4):333-343.
9. Maroosi A, Amiri B (2010) A new clustering algorithm based on hybrid global optimizationbased on a dynamical systems approach algorithm. Expert Systems with Applications, 37(8):5645-5652.
10. Babu GP, Murty MN (1993) A near-optimal initial seed value selection in K-means algorithm using a genetic algorithm. Pattern Recognition Letters, 14(10):763-769.
11. Murthy, C. A., & Chowdhury, N. (1996). In search of optimal clusters using genetic algorithms. Pattern Recognition Letters, 17(8), 825–832.
12. Maulik U, Bandyopadhyay S (2000) Genetic algorithm-based clustering technique. Pattern Recognition, 33(9):1455-1465.
13. Huang JH, Tzeng GH, Ong CS (2007) Marketing segmentation using support vector clustering. Expert Systems with Applications, 32:313-317.
14. Vishwanath A, Chen H (2006) Technology clusters: Using multidimensional scaling to evaluate and structure technology clusters. Journal of the American Society for Information Science and Technology, 57(11):1451-1460.
15. Brida JG, Disegna M, Scuderi R (2012) Visitors of two types of museums: A segmentation study. Expert Systems with Application, 39:11349-11356.
16. Kuo RJ, Ho LM, Hu CM (2002) Integration of self-organizing feature map and K-means algorithm for market segmentation. Computers & Operations Research, 29(11):1475-1493.

17. Cheng CH, Chen YS (2009) Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications, 36:4176-4184.
18. Hong CW (2012) Using the Taguchi method for effective market segmentation. Expert Systems with Applications, 39:5451-5459.
19. Mostafa MM (2010) Clustering the ecological footprint of nations using Kohonen's self-organizing maps. Expert Systems with Applications, 37:2747-2755.
20. Hruschka H, Natter M (1999) Comparing performance of feed forward neural nets and K- means for cluster-based market segmentation. European Journal of Operation Research, 114:346-353.
21. Balakrishnan PV, Cooper MC, Jacob VS, Lewis PA (1996) Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. European Journal of Operational Research, 93:346-357.
22. Boone DS, Roehm M (2002a) Evaluating the appropriateness of market segmentation solutions using artificial neural networks and the membership clustering criterion. Marketing Letters, 13(4):317-333.
23. Chen CH, Khoo, CL, Yan W (2002) A strategy for acquiring customer requirement patterns using laddering technique and ART2 neural network. Advanced Engineering Informatics, 16:229–240.
24. Bloom JZ (2004) Tourist market segmentation with linear and non-linear techniques. Tourism Management 25:723-733.
25. Kauko T, Hooimeijer P, Hakfoort J (2002) Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modeling, Housing Studies, 17(6):875-894.
26. Natter M (1999) Conditional market segmentation by neural networks: a Monte-Carlo study. Journal of Retailing and Consumer Services, 6(4):237-248.
27. Kim Y, Street WN (2004) An intelligent system for customer targeting: a data mining approach. Decision Support Systems, 37(2):215-228.
28. Ho GTS, Ip WH, Lee CKM, Mou WL (2012) Customer grouping for better resources allocation using GA based clustering technique. Expert Systems with Applications, 39(2):1979-1987.
29. Abdi K, Fathian M, Safari E (2012) A novel algorithm based on hybridization of artificial immune system and simulated annealing for clustering problem. The International Journal of Advanced Manufacturing Technology, 60(5-8):723-732.
30. Bursco MJ, Cradit JD, Stahl S (2002a) A simulated annealing heuristic for a bicriterion partitioning problem in marketing segmentation. Journal of Marketing Research, 39(1):99-109
31. Kuo RJ, Akbaria K, Subroto B (2012) Application of particle swarm optimization and perceptual map to tourist market segmentation. Expert Systems with Applications, 39(10):8726-8735.
32. Zhang Y, Jiao J, Ma Y (2007) Market segmentation for product family positioning based on fuzzy clustering. Journal of Engineering Design, 18(3):227-241.
33. Wei JT, Lin SY, Weng CC, Wu HH (2012) A case study of applying LRFM model in market segmentation of a children's dental

clinic. Expert Systems with Applications, 39(5):5529-5533.

34. Miguéis VL, Camanho AS, Falcãoe Cunha J (2012) Customer data mining for lifestyle segmentation. Expert Systems with Applications, 39:9359-9366.
35. Li Z, Wang W, Yang C, Ragland DR (2013) Bicycle commuting market analysis using attitudinal market segmentation approach. Transportation Research Part A: Policy and Practice, 47:56-68.
36. Wang CH (2010) Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques. Expert Systems with Applications, 37:8395-8400.
37. Desarbo WS, Grewal R, Scott CJ (2008) A clusterwise bilinear multidimensional scaling methodology for simultaneous segmentation and positioning analyses. Journal of Marketing Research, 45:280-292.
38. Wu WW (2011) Segmenting and mining the ERP users' perceived benefits using the rough set approach. Expert Systems with Applications, 38:6940-6948.
39. Kamakura WA, Russell GJ (1989) A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research, 379-390.
40. Chung KY, Oh SY, Kim SS,nHan SY (2004) Three representative market segmentation methodologies for hotel guest room customers. Tourism Management, 25(4):429-441.
41. Desarbo WS, Ramaswamy V, Cohen SH (1995) Market Segmentation with Choice-Based Conjoint Analysis. Marketing Letters, 6(2):137-147.
42. Bursco MJ, Cradit JD, Stahl S (2002b). Multicriterion clusterwise regression for joint segmentation settings: an application to customer value. Journal of Marketing Research, 40(2):225-234.
43. Albert R. Wildt, John M. McCann (1980) Multicriterion Market Segmentation: A New Model, Implementation, and Evaluation. Journal of Marketing Research, 17(3):335-340.
44. Vriens M, Wedel M, Wilms T (1996) Metric conjoint segmentation methods: A Monte Carlo comparison. Journal of Marketing Research, 33:73-85.
45. Wedel M, Kistemaker C (1989) Consumer benefit segmentation using cluster wise linear regression. International Journal of Research in Marketing, 6:241-258.
46. DeSarbo WS, Grisaffe D (1998) Combinatorial optimization approaches to constrained market segmentation: An application to industrial market segmentation. Marketing Letters, 9:115-134.
47. Desarbo WS, Ramaswamy V (1994) CRISP: Customer response based iterative segmentation procedures for response modeling in direct marketing. Journal of Direct Marketing, 8:7-20.