

# Data Mining in Market Segmentation: A Literature Review and Suggestions

Saibal Dutta, Sujoy Bhattacharya and Kalyan Kumar Guin

**Abstract** The importance of data mining techniques for market segmentation is becoming indispensable in the field of marketing research. This is the first identified academic literature review of the available data mining techniques related to market segmentation. This research paper provides surveys of the available literature on data mining techniques in market segmentation. A categorization has been provided based on the available data mining techniques used in market segmentation. Eight online journal databases were used for searching, and finally, 103 articles were selected and categorized into 13 groups based on data mining techniques. The utility of data mining techniques and suggestions are also discussed. The findings of this study show that neural networks is the most used method, and kernel-based method is the most promising data mining techniques. Our research work provides a comprehensive understanding of past, present as well as future research trend on data mining techniques in market segmentation. We hope this paper provides reasonable insight and clear understating to both industry as well as academic researchers.

**Keywords** Market segmentation · Data mining · Literature review · Classification

## 1 Introduction

Market segmentation is one of the important concepts both in marketing theory and in practice. In the last few decades, market segmentation has become a central issue in marketing theory. In marketing theory literature, many researchers have proposed

---

S. Dutta (✉) · S. Bhattacharya · K.K. Guin  
Indian Institute of Technology Kharagpur, Kharagpur, India  
e-mail: saibal.13@gmail.com

S. Bhattacharya  
e-mail: sujoy.research@gmail.com

K.K. Guin  
e-mail: kalyan@vgsom.iitkgp.ernet.in

different techniques but still proper market segmentation is a burning problem among market researchers. The concept of market segmentation was first proposed by American marketing researcher Smith [45] and then was further developed by many researchers. Market segmentation is the process of differentiating a large market into some groups or the clusters of customers from available market information or similar behavior customer [18, 23, 31, 45, 46, 51]. The groups or subset may be demographic, economic, and choice based.

This research paper surveys the development of data mining techniques in market segmentation. Data mining is the process of discovering pattern in data, and market researchers are frequently used for understanding the behavior pattern in marketing data. Nowadays, data mining is becoming one of the emerging fields of research due to explosion of data.

The aim of this paper is to present a comprehensive review of literature which published in academic journals related to the application of data mining techniques in market segmentation. The classification of framework is adopted here as per previous literature [42, 43]. The paper is organized as follows: First, the research methodology of the study is presented; second, introduced available list of data mining techniques which used for market segmentation and are also presented into several groups; third, articles about data mining in market segmentation are analyzed, and the results of the classification are showed through table; and finally, the conclusions, limitations of the study, and suggestion are discussed.

## 2 Research Methodology

Nowadays, the nature of research in market segmentation becomes multidisciplinary and many researchers across the different field worked on and have published relevant materials in various journals. The following online journal databases were used for searching comprehensive bibliography of academic literature on market segmentation.

1. Science Direct
2. ABI/INFORM Database
3. Emerald Full Text
4. IEEE Transaction
5. JSTOR
6. Springer
7. Google Scholar
8. Wiley Online Library.

The above academic literature was searched with key word “market segmentation” or “target marketing” or “data mining and market segmentation,” and originally around 750 articles were produced. Each full-text article was reviewed carefully and eliminated those articles which were not related to data mining techniques in market segmentation or those were not having main focus for improvement of market

segmentation. Below criteria followed for selection and the number reduced to 103 articles which are related to data mining techniques and development of methodologies in market segmentation. Those articles have been included which were published in above-mentioned academic journal. Doctoral thesis, master thesis, conferences, and unpublished work are avoided as per previous literature [43].

Each article has been carefully reviewed and separately classified into several categories of data mining techniques. Although this search tried to cover all the available used data mining techniques so far, this paper will serve as a comprehensive base of data mining research in market segmentation area and will give a broad view of available techniques in marker segmentation.

### 3 Classification Method

The reviewed research papers were classified into thirteen categories, and each category also consists of several single or hybrid data mining techniques. Each categorywise several data mining techniques presented into table with reference papers.

#### 3.1 Classification Framework for Data Mining Techniques

This paper surveys and classifies different market segmentation techniques into thirteen broad categories as follows:

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 set
13. Others.

Early 1960s and 1970s researchers had started K-means and hierarchical cluster analysis for market segmentation [44]. Even today also, many researchers successfully used K-means clustering for market segmentation [8, 33]. But above traditional clustering techniques have some drawbacks. For example, K-means algorithm cannot handle noise and outliers data [5, 6, 47]. K-means algorithm also

failed to give any exact or initial number of cluster and the statistical validity of the cluster formed [37], and hence, clustering falls into local minima [15, 35]. To address the above issue, the researcher combined K-means with genetic algorithm to reach the global minima [3, 36, 40].

To classify the complex consumer pattern in market, researchers also introduced other different approaches such as evolutionary algorithm, kernel methods, rough set, Taguchi method, etc. Nowadays, the above approaches are quite popular among the market researchers, and also these approaches are able to perform better market segmentation than traditional one [24]. Other important market segmentation techniques such as multidimensional scaling, random forest, RFM analysis, bagged clustering, etc., are also found in academic literature [10, 48].

### ***3.2 Classification Process***

Each of the selected research papers were reviewed carefully and classified into one of the thirteen categories according to the proposed classification framework. The research papers were analyzed by year of publication and distribution of journal. The classification process is adopted here as per previous literature [42, 43]:

- Online academic literature search
- First-phase classification by the researcher
- Second-phase classification of initial classification result
- Final verification and classification of result.

## **4 Classification of the Articles**

The distributions of 103 articles have been done according to the proposed classification model which is shown in Table 1. Articles are categorized into thirteen broad methodologies, and then, each methodology is further divided into major data mining techniques. Most of the data mining techniques used in market segmentation have been included here.

### ***4.1 Distribution by Journals***

The distribution of articles by journal has been shown in Table 2. Articles related to application of data mining techniques in market segmentation are distributed across 45 journals. The top three journals are as follows: Expert Systems with Applications, Journal of Marketing Research, and Management Science.

**Table 1** Distribution of articles according to the proposed classification model

Method	Major methodology
Neural network	Self-organizing map [7, 29, 32, 39]
	Multilayer perceptron and K-means [26]
	FSCl algorithm [4]
	Neural network Hopfield dNN [9]
	Adaptive resonance theory (ART2) neural network [14]
	Neural network and Monte Carlo [41]
Evolutionary algorithm	Genetic algorithm [30, 33]
	Artificial immune system and simulated annealing algorithm [1, 13]
Fuzzy theory	Fuzzy clustering [55]
RFM analysis	RFM model [16]
	LRFM model [53]
Hierarchical clustering	Hierarchical clustering [38]
K-means	K-means [34]
Bagged clustering	Bagged clustering [11]
Kernel methods	Kernel-based clustering algorithm [24, 50]
Multidimensional scaling	Multidimensional scaling [48]
	Bilinear multidimensional scaling [22]
Rough set	Rough set [54]
Taguchi method	Taguchi method [27]
Model-based clustering	Model-based clustering [28]
Others	Chi-square automatic interaction (CHAID) [17]
	Choice-based conjoint analysis [20]
	Multicriterion clusterwise regression [13]
	Regression model [2]
	Metric conjoint segmentation [49]
	Clusterwise linear regression [52]
	Constrained market segmentation [21]
	CRISP: Customer response-based iterative segmentation [19]

The above journals covered more than 45 % of the total number of articles published. However, other important journals are as follows: Decision Support Systems, European Journal of Operational Research, and Tourism Management.

### 4.2 Distribution of Articles by Year of Publication

The distribution of articles by year of publication is shown in Fig. 1. Last 12 years has taken for consideration. The amount of publication increased significantly from 2001 to 2012. More research work would be expected in the future.

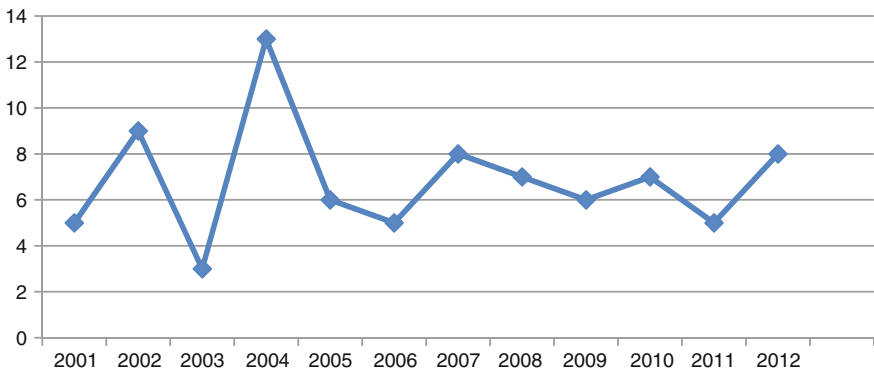
**Table 2** Distribution of articles by journal

Journal title	Amount	Percentage (%)
Expert Systems with Applications	38	36.89
Journal of Marketing Research	5	4.85
Management Science	4	3.88
European Journal of Operational Research	4	3.88
Tourism Management	3	2.91
Decision Support Systems	3	2.91
Marketing Letters	3	2.91
Computational Statistics & Data Analysis	2	1.94
Journal of Classification	2	1.94
Management Science	2	1.94
Marketing Science	2	1.94
International Journal of Research in Marketing	2	1.94
Computer-Aided Design,	1	0.97
Computers & Operations Research	1	0.97
Journal of Marketing	1	0.97
European Journal of Economic and Social Systems	1	0.97
International Marketing Review	1	0.97
Journal of Direct Marketing	1	0.97
Journal of Engineering Design	1	0.97
Journal of Organizational Computing and Electronic Commerce	1	0.97
Powder Technology	1	0.97
Psychometrika	1	0.97
Transportation Research	1	0.97
Applied Mathematical Modeling	1	0.97
Advanced Engineering Informatics	1	0.97
Annals of Tourism Research	1	0.97
Australian Marketing Journal	1	0.97
Journal of Econometrics	1	0.97
Omega	1	0.97
Computer-Aided Design	1	0.97
Computers & Operations Research	1	0.97
Decision Sciences	1	0.97
Fuzzy Sets and Systems	1	0.97
Housing Studies	1	0.97
International Journal of Applied Science and Engineering	1	0.97
International Journal of Intelligent Systems in Accounting, Finance and Management	1	0.97
Iranian Acc. Aud. Rev	1	0.97
Journal of International Consumer Marketing	1	0.97

(continued)

**Table 2** (continued)

Journal title	Amount	Percentage (%)
Journal of Management Information Systems	1	0.97
Journal of Strategic Marketing	1	0.97
Journal of the American Society for Information Science and Technology	1	0.97
Knowledge-Based Systems	1	0.97
Neurocomputing	1	0.97
Quality & Quantity	1	0.97
The International Journal of Management Science	1	0.97



**Fig. 1** Distribution of articles by year of publication

### 4.3 *Distribution of Articles According to the Proposed Classification Model*

The distribution of articles according to the major data mining techniques is shown in Table 3. Twenty-two major data mining techniques found in literature. The top six data mining techniques are as follows: neural network, evolutionary algorithm, fuzzy theory, RFM analysis, hierarchical clustering, and K-means.

More than 75 % of the total articles have published using above techniques. Researcher also used hybrid techniques for the better result and another reason was to overcome shortfall of another techniques. In last few years, many advance techniques are proposed for market segmentation.

**Table 3** Distribution of articles by data mining techniques

Data mining techniques	Amount
Neural network	30
Evolutionary algorithm	17
Fuzzy theory	12
RFM analysis.	10
Hierarchical clustering	5
K-means	5
Bagged clustering	3
Kernel methods	3
Multidimensional scaling	3
Taguchi method	3
Rough set.	2
Model-based clustering	1
Chi-square automatic interaction (CHAID)	1
Choice-based conjoint analysis	1
Simulated annealing algorithm	1
Multicriterion clusterwise regression	1
Regression model	1
Metric conjoint segmentation	1
Clusterwise linear regression	1
Combinatorial optimization approaches to constrained market segmentation	1
CRISP: customer response-based iterative segmentation	1
Total	103

## 5 Discussions, Limitations, and Suggestions

### 5.1 Discussions

This is the first identified research paper on data mining techniques in market segmentation. This paper covers available data mining techniques used in market segmentation so far. Eight online journal databases were used for searching. After careful analysis of several research papers on market segmentation, we found that various researchers adopted different data mining techniques in the last few decades. The development of methodologies on marketing research can be divided into two phases. In the first phase, early researchers who worked on market segmentation generally used K-means and hierarchical clustering. But after 2000, market researchers are motivated to use advanced data mining techniques that can handle complex consumer behavior pattern. Hybrid algorithms are found in literature to increase the performance of existing one. Some examples of hybrid algorithm include multilayer perceptron and K-means, genetic algorithm and K-means, particle swarm optimization and K-means, and self-organizing map and K-means. Neural network-based market segmentation is quite popular among the researcher.



## 5.2 *Limitations*

This paper might have some limitations. A widespread comprehensive literature review of data mining techniques in market segmentation has been presented. This research paper tries to incorporate all the available data mining techniques without time bound. However, this work surveyed those articles which were extracted based on the keyword “market segmentation” or “target marketing” or “data mining and market segmentation.” In order to find out how data mining techniques for market segmentation developed, the categorization is done based on keyword index, abstract, and article methodology part. Articles without keyword index could not be extracted. This research work limited search for articles to 8 online databases. There might be the presence of other useful academic journals which can provide other important data mining techniques for market segmentation. We believe that many companies practice advanced techniques for market segmentation but failed to include here due to limited resource. In this study, we do not consider other non-English publications. However, other languages may have important techniques for market segmentation. Another major limitation is as the quality of market segmentation not only depends on good data mining techniques but also on the selection of proper segmentation variables [51]. Normally, researchers used general variables because general variables are easy to use [25]. The review of segmentation variable in market segmentation is out of scope of this paper.

## 5.3 *Suggestions*

1. Research on the application of data mining techniques in market segmentation is becoming an emerging field as shown in the Fig. 1 and Table 1, and the number of publication in term of research on the application of data mining in the market segmentation will gradually increase in the future.
2. Neural network-based market segmentation is the most used data mining technique. Neural network is used for classification, clustering, and prediction purpose. Many researchers preferred self-organizing map neural network model for visualization and determination number of cluster.
3. The result of market segmentation can easily improve with combination of good data mining techniques and proper selection of segmentation variable.
4. The K-means and hierarchical clustering are still preferred data mining techniques for market segmentation.
5. Kernel-based market segmentation is one of the promising techniques for robust market segmentation and also performed better than tradition techniques.
6. Many hybrid data mining techniques are used to increase the performance. For example, researchers used K-means with combination genetic algorithm for better performance.

7. Market segmentation is becoming more complex gradually and researchers are also working on development of more advanced data mining techniques that can handle outlier, noise, and big data-related problem.
8. Many data mining techniques are used from last few decades, but still there is always scope of improvement. Future researchers can improve the available algorithm for better performance in market segmentation. Some examples of such data mining techniques include kernel-based method, probabilistic fuzzy c-means, random forest, evolutionary algorithm.

## 6 Conclusion

Market segmentation is one of the primary and most critical parts of market research. This research work identified 103 articles which were published in 8 online databases. Our aim is to provide brief summary on available all data mining techniques which are most popular and successfully used so far. However, the research work cannot claim to be exhaustive. We believe that this work does provide reasonable insight of market segmentation and would help to give a clear picture for those who interested to work and research on this area.

## References

1. Abdi, K., Fathian, M., Safari, E.: A novel algorithm based on hybridization of artificial immune system and simulated annealing for clustering problem. *Int. J. Adv. Manuf. Technol.* **60**(5–8), 723–732 (2012)
2. Wildt, Albert R., McCann, John M.: Multicriterion market segmentation: a new model, implementation, and evaluation. *J. Mark. Res.* **17**(3), 335–340 (1980)
3. Babu, G.P., Murty, M.N.: A near-optimal initial seed value selection in K-means algorithm using a genetic algorithm. *Pattern Recogn. Lett.* **14**(10), 763–769 (1993)
4. Balakrishnan, P.V., Cooper, M.C., Jacob, V.S., Lewis, P.A.: Comparative performance of the FSCL neural net and K-means algorithm for market segmentation. *Eur. J. Oper. Res.* **93**, 346–357 (1996)
5. Bezdek, J.C., Pal, N.R.: Some new indexes of cluster validity. *IEEE Trans. Syst. Man Cybern. Part B: Cybern.* **28**(3), 301–315 (1998)
6. Blatt, M., Wiseman, S., Domany, E.: Clustering data through an analogy to the potts model. *Adv. Neural Inf. Process. Syst.* **8**, 416–422 (1996)
7. Bloom, J.Z.: Tourist market segmentation with linear and non-linear techniques. *Tour. Manage.* **25**, 723–733 (2004)
8. Boone, D.S., Roehm, M.: Retail segmentation using artificial neural networks. *Int. J. Res. Mark.* **19**(3), 287–301 (2002)
9. Boone, D.S., Roehm, M.: Evaluating the appropriateness of market segmentation solutions using artificial neural networks and the membership clustering criterion. *Market. Lett.* **13**(4), 317–333 (2002)

10. Brida, J.G., Disegna, M., Osti, L.: Segmenting visitors of cultural events by motivation: A sequential non-linear clustering analysis of Italian Christmas market visitors. *Expert Syst. Appl.* **39**, 11349–11356 (2012)
11. Brida, J.G., Disegna, M., Scuderi, R.: Visitors of two types of museums: a segmentation study. *Expert Syst. Appl.* **40**(6), 2224–2232 (2012)
12. Bursco, M.J., Cradit, J.D., Stahl, S.: A simulated annealing heuristic for a bicriterion partitioning problem in marketing segmentation. *J. Mark. Res.* **39**(1), 99–109 (2002)
13. Bursco, M.J., Cradit, J.D., Stahl, S.: Multicriterion cluster wise regression for joint segmentation settings: an application to customer value. *J. Mark. Res.* **40**(2), 225–234 (2002)
14. Chen, C.H., Khoo, C.L., Yan, W.: A strategy for acquiring customer requirement patterns using laddering technique and ART2 neural network. *Adv. Eng. Inform.* **16**, 229–240 (2002)
15. Chen, M.C., Wu, H.P.: An association-based clustering approach to order batching considering customer demand patterns. *Omega* **33**(4), 333–343 (2005)
16. Cheng, C.H., Chen, Y.S.: Classifying the segmentation of customer value via RFM model and RS theory. *Expert Syst. Appl.* **36**, 4176–4184 (2009)
17. Chung, K.Y., Oh, S.Y., Kim, S.S., Han, S.Y.: Three representative market segmentation methodologies for hotel guest room customers. *Tour. Manag.* **25**(4), 429–441 (2004)
18. Croft, M.J.: *Market Segmentation: A Step-by-Step Guide to Profitable New Business*. Routledge, London, New York (1994)
19. Desarbo, W.S., Ramaswamy, V.: CRISP: customer response based iterative segmentation procedures for response modeling in direct marketing. *J. Direct Market.* **8**, 7–20 (1994)
20. Desarbo, W.S., Ramaswamy, V., Cohen, S.H.: market segmentation with choice-based conjoint analysis. *Market. Lett.* **6**(2), 137–147 (1995)
21. DeSarbo, W.S., Grisaffe, D.: Combinatorial optimization approaches to constrained market segmentation: an application to industrial market segmentation. *Market. Lett.* **9**, 115–134 (1998)
22. Desarbo, W.S., Grewal, R., Scott, C.J.: A cluster wise bilinear multidimensional scaling methodology for simultaneous segmentation and positioning analyses. *J. Mark. Res.* **45**, 280–292 (2008)
23. Frank, R.E., Massy, W.F., Wind, Y.: *Market segmentation*. Prentice Hall, Englewood Cliffs (1972)
24. Huang, J.H., Tzeng, G.H., Ong, C.S.: Marketing segmentation using support vector clustering. *Expert Syst. Appl.* **32**, 313–317 (2007)
25. Hammond, K., Ehrenberg, A.S.C., Goodhardt, G.J.: Market segmentation for competitive brands. *Eur. J. Mark.* **30**(12), 39–49 (1996)
26. Hruschka, H., Natter, M.: Comparing performance of feedforward neural nets and K-means for cluster-based market segmentation. *Eur. J. Oper. Res.* **114**, 346–353 (1999)
27. Hong, C.W.: Using the Taguchi method for effective market segmentation. *Expert Syst. Appl.* **39**, 5451–5459 (2012)
28. Kamakura, W.A., Russell, G.J.: A probabilistic choice model for market segmentation and elasticity structure. *J. Market. Res.* **26**, 379–390 (1989)
29. Kauko, T., Hooimeijer, P., Hakfoort, J.: Capturing housing market segmentation: an alternative approach based on neural network modeling. *Hous. Stud.* **17**(6), 875–894 (2002)
30. Kim, Y., Street, W.N.: An intelligent system for customer targeting: a data mining approach. *Decis. Support Syst.* **37**(2), 215–228 (2004)
31. Kotler, P., Gordon, M.: *Principles of Market*. Prentice Hall, Canada (1983)
32. Kuo, R.J., Ho, L.M., Hu, C.M.: Integration of self-organizing feature map and K-means algorithm for market segmentation. *Comput. Oper. Res.* **29**(11), 1475–1493 (2002)
33. Kuo, R.J., Chang, K., Chien, S.Y.: Integration of self-organizing feature maps and genetic algorithm- based clustering method for market segmentation. *J. Organ. Comput. Electron. Commer.* **14**(1), 43–60 (2004)
34. Li, Z., Wang, W., Yang, C., Ragland, D.R.: Bicycle commuting market analysis using attitudinal market segmentation approach. *Transport. Res. Part A: Policy Pract.* **47**, 56–68 (2013)

35. Maroosi, A., Amiri, B.: A new clustering algorithm based on hybrid global optimization based on a dynamical systems approach algorithm. *Expert Syst. Appl.* **37**(8), 5645–5652 (2010)
36. Maulik, U., Bandyopadhyay, S.: Genetic algorithm-based clustering technique. *Pattern Recogn.* **33**(9), 1455–1465 (2000)
37. Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans. Patt. Anal. Mach. Intell.* **24**(12), 1650–1654 (2002)
38. Miguéis, V.L., Camanho, A.S., Falcão e Cunha, J.: Customer data mining for lifestyle segmentation. *Expert Syst. Appl.* **39**, 9359–9366 (2012)
39. Mostafa, M.M.: Clustering the ecological footprint of nations using Kohonen's self-organizing maps. *Expert Syst. Appl.* **37**, 2747–2755 (2010)
40. Murthy, C.A., Chowdhury, N.: In search of optimal clusters using genetic algorithms. *Pattern Recogn. Lett.* **17**(8), 825–832 (1996)
41. Natter, M.: Conditional market segmentation by neural networks: a Monte-Carlo study. *J. Retail. Consum. Serv.* **6**(4), 237–248 (1999)
42. Ngai, E.W., Xiu, L., Chau, D.C.K.: Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst. Appl.* **36**(2), 2592–2602 (2009)
43. Park, D.H., Kim, H.K., Choi, I.Y., Kim, J.K.: A literature review and classification of recommender systems research. *Expert Syst. Appl.* **39**(11), 10059–10072 (2012)
44. Punj, G., Stewart, D.W.: Cluster analysis in marketing research: review and suggestions for application. *J. Market. Res.* **20**, 134–148 (1983)
45. Smith, W.R.: Product differentiation and market segmentation as an alternative marketing strategy. *J. Market.* **21**, 3–8 (1956)
46. Vellido, A., Lisboa, P.J.G., Vaughan, J.: Neural networks in business: a survey of applications (1992–1998). *Expert Syst. Appl.* **17**(1), 51–70 (1999)
47. Vesanto, J., Alhoniemi, E.: Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* **11**(3), 586–600 (2000)
48. Vishwanath, A., Chen, H.: Technology clusters: using multidimensional scaling to evaluate and structure technology clusters. *J. Am. Soc. Inform. Sci. Technol.* **57**(11), 1451–1460 (2006)
49. Vriens, M., Wedel, M., Wilms, T.: Metric conjoint segmentation methods: a Monte Carlo comparison. *J. Mark. Res.* **33**, 73–85 (1996)
50. Wang, C.H.: Apply robust segmentation to the service industry using kernel induced fuzzy clustering techniques. *Expert Syst. Appl.* **37**, 8395–8400 (2010)
51. Wedel, M., Kamakura, W.A.: *Market segmentation: conceptual and methodological foundations.* Kluwer Academic, Boston (2000)
52. Wedel, M., Kistemaker, C.: Consumer benefit segmentation using cluster wise linear regression. *Int. J. Res. Mark.* **6**, 241–258 (1989)
53. Wei, J.T., Lin, S.Y., Weng, C.C., Wu, H.H.: A case study of applying LRFM model in market segmentation of a children's dental clinic. *Expert Syst. Appl.* **39**(5), 5529–5533 (2012)
54. Wu, W.W.: Segmenting and mining the ERP users' perceived benefits using the rough set approach. *Expert Syst. Appl.* **38**, 6940–6948 (2011)
55. Zhang, Y., Jiao, J., Ma, Y.: Market segmentation for product family positioning based on fuzzy clustering. *J. Eng. Des.* **18**(3), 227–241 (2007)