ICEB 2003 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-9-2003

The Study of Maximizing Customer Equity by Segmentation: A Modified K-Means Approach

Chun-Yao Huang

Russell K.H. Ching

Jashen Chen

The Study of Maximizing Customer Equity by Segmentation: A Modified K-Means Approach

Chun-Yao Huang
Dept. of Business Administration
Yuan Ze University, Taiwan
cyhuang@saturn.yzu.edu.tw

Russell K.H. Ching
Management Information Science Department
California State University, Sacramento, USA
chingr@csus.edu

Ja-Shen Chen
Dept. of Business Administration
Yuan Ze University, Taiwan
jchen@saturn.yzu.edu.tw

Abstract

As segmentation has been one of the central marketing tasks for decades and customer profitability valuation has seen wide study during the past few years, surprisingly, up to this date, there is a gap in marketing research that await a bridge to link up of these two important and closely related dimensions. In this paper, we introduce a decision support system with the goal of maximizing customer equity by segmentation. The decision support system introduced here is unique in that it accommodates the essence of customer profitability valuation into a segmentation scheme in a sensible and flexible manner, that it suggests the number of segments to be determined by the goal of profit maximization instead of some arbitrary numerical criterion, and that central to its technical core the outlier problem which is pervasive in cluster analysis has been addressed by a modified K-Means algorithm so that clustering can reflect the pattern of the majority of ordinary observations in a data set instead of being influenced by a handful of outliers. It followed by a number of test datasets from a public data source and a conclusion remark was made at the end.

1. Introduction

Central to the current Customer Relationship Management (CRM) thinking is the idea that a firm's customer base is the key to the business' profitability. Just as Blattberg and Deighton [1] put it, the essence of CRM emphasizes that "growing a business can be framed as a matter of getting customers and keeping them so as to grow the value of the customer base to its fullest potential". In this light, there are two stepping stones to successful CRM implementation: customer profitability valuation and customer segmentation.

What underlines the importance of customer profitability valuation is the recent emphasis on the concept of customer equity which, for a firm, can be defined as the total of the discounted lifetime values of its customer base [13]. At the individual level, therefore, an array of studies has aligned "lifetime value" with the revenue and cost associated with a customer over time [2], [6], [8], [12]. The essential assumption implied in this type of models is that managers can predict, at least with tolerable errors, *both* a customer's expected contribution to a firm's revenue over time *and* the costs that will incur for the acquisition, maintenance, and retention of this customer, again, over time. As a manager is likely to point out, this assumption may not be realistic for marketing decisions (in which costs are usually determined by marketing mix strategy and strategy in turn follows segment definition) in any marketplace that sees competition and undergoes structural change, which most industries do nowadays. That is, except for the role to play in long range planning, current customer profitability valuation models neither fit in nor provide practical guidance for manager's actual decision making.

The other stepping stone to CRM is customer segmentation. Having seen limelight in marketing for at least half a century, segmentation mainly "involves viewing a heterogeneous market as a number of smaller homogeneous markets, in response to differing preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants" [16]. Technically, segmentation issues that usually puzzle researchers and practitioners include the "optimal" number of segments and the bias of segmentation caused by outliers in a dataset. Yet more importantly, though various models have been proposed for segmentation in the marketing research literature (for a comprehensive review, see Wedel and Kamakura [17]), the relationships between most established quantitative segmentation models and marketing decision making, as implied managerial relevance in which should be the raison d'être of developing a model, are mostly neglected.

Meanwhile, to follow the CRM principles, a profit-maximizing firm will seek to maximize the profit to which its customer base can contribute. In terms of implementation, customer segmentation is likely to play a central role. Surprisingly, however, except for a few superficial discussions on the importance of market segmentation on customer profitability [11], [12] to the best of our knowledge, there seems to be a vacuum in marketing research that clarifies the relationship between customer profitability and segmentation, not to say attempts that formally links up these two important issues in a model.

This paper, therefore, aims to present a managerially relevant and easy-to-implement model in the form of a decision support system (DSS), to which we give the name of MCES (Maximizing Customer Equity by Segmentation), that addresses the several issues mentioned above. In the following sections, we will first discuss the relationship between customer profitability and segmentation, and then propose a segmentation decision framework for the maximization of customer profit—the backbone of our proposed DSS. Next, on the more technical front, we will introduce a modified K-means cluster algorithm that serves as the quantitative core of the DSS. Finally, we will discuss the implication of such a DSS as well as directions for future research.

2. Assumptions of MCES

Current customer profit valuation techniques are developed upon the assumption that managers can reasonably predict the revenue a customer will contribute to the firm *as well as* the costs necessary for generating them. What is implied in models thus developed [2], [6] [8] then is that the revenue side and the cost side can be determined *simultaneously*. Arguably, though sound in terms of strategic thinking, of accounting, and perhaps of financial performance projection, current customer profitability valuation models thus developed seem distant from managers' daily marketing decision making. It needs not be so. The value of customer valuation and its accompanying CRM implications, we propose here, can be fully unleashed so that it can fit neatly in marketing decision making process if a more realistic view of customer profitability is taken. The key to managerial relevance for customer valuation in the context of marketing decision making, however, is the important but long neglected link (in terms of implementation) between segmentation and customer valuation. Just like Wedel and Kamakura [17] mentioned, traditionally, segmentation models put emphasis on goodness of fit and neglect the importance of relating marketing research with marketing strategy within a single model. The proposed MCES model in the form of a DSS, that will be introduced below, is a formal attempt to develop a managerially relevant framework which aims to combine inference and profit maximization and in which, in the words of Wedel and Kamakura ([17], p.340), "decisions on the proper number of segments would be based on managerial criteria (expected profits) rather than goodness of fit; inferences would be made directly in terms of the optimal marketing effort rather than mere descriptions of the segments."

The first task in actually bridging customer profitability and segmentation, given the essence of customer valuation and that profit is the difference between revenue generated and costs incurred, then, is to make more realistic assumptions than what current customer valuation models assume. In this light, the tenet of the decision support system proposed here is that for a manager with bounded rationality, certain experiences with the customer base to be managed, and perhaps some familiarity with established probabilistic models (not a necessary condition, though), the revenue a customer can generate in a reasonable period of time can be roughly predicted *a priori*, whereas the cost side can only be managerially estimated *after* marketing decision relating to segmentation is made. In the marketing research literature, the array of well-established models based on NBD has demonstrated that even without marketing mix information in the form of covariate, an NBD-based model performs well in a wide range of applications for the description of customer's repeat behaviors, which in our context can be directly translated into revenue generated by a customer base [10], [14]. A single customer's future contribution to a firm's revenue, then, can be easily derived by, say, empirical Bayesian method. For a nonstationary customer base, inferences can also be made by modified models with a nested NBD core [15]. Even without resorting to these formal probabilistic models, current practice has various heuristics for the prediction of customers' revenue generation potentials built in a wide range of CRM systems. Therefore, it can be reasonably expected that for a single customer, a firm can get a rough but useful estimate as to his/her future contribution to revenue stream, which can serve at least as a baseline in our model. On the other hand, the cost side in the profit equation is more subtle. In fact, in the segmentation context, it is only after the segments and the respective marketing strategy set are defined that the cost to maintain a specific customer can be reasonably calculated. Therefore, in our MCES DSS, we assume that:

1. Each customer's expected contribution to a firm's revenue, given the context, can be roughly calculated by certain heuristic or a formal probabilistic model.
2. Managers cannot figure out the cost to maintain any segment *a priori* (i.e., before a segment is defined), but are able to do it given the defined segments.
3.

3. MCES as a Decision Support System

Given the above-mentioned assumptions, here we propose a model to maximize customer equity by segmentation (MCES) in the form of a decision support system. Suppose that each customer's expected contribution to the firm's revenue stream in a period of time is given (either calculated by certain heuristic or estimated by a probabilistic model), the decision framework, which combines managerial judgment and a modified clustering method, can be summarized in the following steps:

1. Managerially define a desirable range of number of segments [S_i, S_a], S_i, $S_a \in$ N and $S_i < S_a$
2. Setting $m= S_i$
3. Assigning observations into m segments by modified K-means (to be introduced in the next section)
4. For j=1 to m, (managerially) judge the cost to maintain each segment thus derived, C_j. Meanwhile, calculate the sum of expected revenue to be generated by the segment members, R_j. Then calculate expected profit (loss) from each segment, $P_j=R_j-C_j$.
5. Summing up expected profits (losses) across the n_n segments, get gross expected profit E_m , $Em=\sum_{j=} P_j$.
6. If $m< S_a$ then $m=m+1$ and go to step 3; otherwise go to step 8
7. Compare the E_ms , $m= S_i, \ldots, S_a$, choose the solution among the S_a-S_i+1 choices which maximizes E_ms.
8. End

For a fictitious example, let's assume that the expected contributions to revenue in the coming year of a small bank's 500,000 customers (by whatever heuristic/model on which its CRM system may rely) are given. The manager, upon making the annual strategic marketing plan, applies MCES for customer segmentation. Taking the resources available and past experiences into consideration, she first decides that the number of segments with managerial meaning as well as operational efficiency/effectiveness should fall between 2 and 5. Therefore, MCES guides her through scenarios with 2, 3, 4 and 5 segments, respectively. For each scenario, once the segments are objectively defined (to be elaborated in the next section), the manager judges the appropriate cost for the maintenance of each segment depending on segment characteristics that can be easily summarized in the form of descriptive statistics (e.g., segmental means). The system then calculates the expected profit of each segment. The total expected profit for the scenario is then derived by summing over the segments. Having done this for the four scenarios with different number of segments, the system list all the scenarios' expected profit in descending order. The "best" segmentation scheme is now obvious.

Simple as it seems, though, this proposed MCES framework has several features that positively distinguish it from established segmentation schemes:

1. It accommodates the essence of customer profitability valuation into a segmentation scheme in a sensible and flexible manner.
2. Unlike most segmentation approaches that make the defined segments *independent* of strategic considerations, MCES allows for, actually depends on, managerial judgment as an important input in finding the "best" definition of segments. In this way, a segment is no longer the end *per se* (as most segmentation models imply), but the means used by managers for the attainment of managerially desirable goal—which is what segmentation is managerially meant for.
3. The application of MCES addresses the issue that usually puzzles modelers as well as managers, namely: how many segments is the *most appropriate*? As Milligan and Cooper [9] concluded from their empirical comparisons among 30 stopping rules for the determination of number of clusters in cluster analysis, lacking a universal theoretical criterion and susceptible to data dependency in empirical settings, it is almost impossible for any numerical rule to claim as the "most appropriate" for the number of segments. Under MCES, however, the thorny problem of number of segments is solved, no longer by an arbitrary rule, but by the interplay of cluster analysis and managerial judgment (it should be highlighted here that it is not the number of segments a priori, but the costs associated with segments post-hoc to be judged by managers).

4. Modified K-Means as the Technical Core of MCES

Given a set of relevant individual level customer information, be it behavioral and/or demographic, cluster analysis is the dominant approach to segmenting a customer base. Among various cluster analysis methods, K-means is the most popular one. As Green et al. [5] and Huang [7] indicated, K-means is capable of handling large database that the business uses for their marketing efforts. Therefore, we have K-means as the technical core of the above-proposed MCES (Step 3). However, K-Means is highly susceptible to out-of-proportion influences of outliers [4]. Therefore, here we propose a modified K-Means algorithm by temporarily skipping outliers in the step of cluster formation. To implement this approach, the mean and standard deviation of the whole data set is first calculated. Data points whose distance to this mean going over some multiple z of the standard deviation are then temporarily skipped. Once the cluster analysis of the remaining, "ordinary", data points is done, the extreme-value points are then ascribed to the nearest group according to the distances calculated between the extreme-value point and group means.

Algorithm

Program Initiated
Step 1. Calculate the mean and standard deviation of the data set. Skip the points whose distance to this mean is beyond z standard deviation.
Step 2. Randomly generate seeds for non-extreme-value points.
Step 3. Following traditional K-Means, start grouping.
Step 4. Check the grouping by MSE. For example, in Figure , calculate the MSE from point C to group A and the MSE from point C to group B, respectively. If MSE(C, group B)<MSE(C, group A), then accept that C belongs to B. Otherwise, regroup point C into group A.
Step 5 After all non-extreme-valued points are segmented, assign extreme-valued points into groups whose group means are nearest to these points.
End of program

The Network Data Examples

We use four data sets that have been previously analyzed and published that are available in the public domain for an empirical test of this modified K-means. According to the requirement stated above, we first set the criterion of extreme value, z, equal 2. That is, those data points that are more than 2 standard deviations away from the data mean are treated as outliers. The model results, both from traditional K-Means and from the modified K-means are compared in Table 1.

Judging by Table 1, the modified K-means we propose here leads to more appropriate segmentation in terms of segment membership identification. The within-group SE criterion also indicates that our model outperforms the traditional one in most cases. The exception, however, has some interesting implication. An intuitive explanation of the exception is that by traditional K-Means, outliers are accounted for in the grouping process simultaneously with other points. Our two-stage modified K-means, instead, defines group boundaries solely on non-outlier points. As Figures 1 illustrated, in cases where outliers are really "extreme", traditional K-means are likely to be heavily biased by these extreme points. In that case, traditional K-Means may lead to better SE comparing with our method. However, managerially, the segmentation which is less susceptible to the influences of extreme-valued points are more relevant if the majority of "normal" data points are of central concern in a segmentation scheme.

5. An application of on line customer behavior survey

Since the MCES decision framework has not been introduced into practice, to demonstrate how it will actually work, here for mainly pedagogic purpose we simulate a project to exposit how managerial judgment would be fused with a decision support system under MCES. This simulated project aims to segment a bank's potential customers for the purpose of launching the bank's online transaction platform. That is, given a base of prospects, we, playing the role of the bank's marketing manager, would like to segment potential customers by their online behaviors so that various marketing mixes can be offered to different segments for the purpose of maximizing the transaction platform's expected profit. The data comes from the online survey from GVU's WWW user Survey Team. It includes 103 items and result in 9147 observations.

As the data consists of prospects instead of customers, expected size of customer base as well as expected contribution to online transaction revenue by future customers can only be predicted by heuristics concerning penetration, pattern of transaction amount and frequency, rate of retention, and so on. Therefore, in this simulation, the manager is able to assign an expected value to each observation with the aid of a heuristic based purely on demographic profile. Before the implementation of cluster analysis, the manager looks at the discriminating power of the 103 items in the survey by using a heuristic criterion of max-min difference. That is, for each attribute (item), the manager calculates the difference between the biggest and the smallest value among the 9147 observations. Only attributes whose max-min differences are larger than 0.2 are then included for the following cluster analysis. 21 attributes are therefore retained for analysis.

Initiating MCES, the manager first sets up the range of number of segments between 2 and 5. Here we only illustrate how the system works for the situation when it comes to the scenario of 4 segments. In this case, the technical core of MCES, the modified K-Means ascribed the 9147 data points into 4 segments. The result is summarized, in the form of descriptive statistics, which is presented to the manager for human input. The system then sums up expected contribution to revenue of each segment member to derive expected contribution to revenue by each segment.

Apparently, segments 1 and 3 are heavy Internet users. Between them, segment 1 users reflect a relative functional orientation that uses Internet mainly for information search, while segment 3 users in comparison reflect a relative leisure orientation that uses Internet mainly as the substitute for entertainment and interpersonal communications. On the other hand, segments 2 and 4 consist of light users. In terms of orientation, segment 4 is similar to segment 1, whereas segment 2 is close to segment 3.

Given the analysis (table 2), the four segments found are different in orientations and/or degree of usage. The manager therefore initiates a set of strategies for the four segments found.

1. Hard-core penetration

For segment 1, the heavy and functionally-oriented Internet users, direct benefits of online transaction with the bank should be intensely communicated through multiple online communication channels to reach rapid penetration.

2. Soft-core penetration

For the heavy and leisure-oriented Internet users of segment 3, another set of online marketing communication mixes should be initiated based on a tune with hedonic appeal. Rapid penetration is also the goal.

3. Functional communications

To segment 4, the light and functionally-oriented Internet users, well-designed and user-friendly online transaction platform should be explained via both online and offline channels. The short-term goal is not necessarily to attract these customers to directly apply for the service, but instead on lowering their psychological barrier to the Internet environment and on developing the trust of the bank's system from these customers.

4. Leisure communications

A soft-soft communication plan should be initiated for segment 2 which consists of light and leisure-oriented customers. Hedonic appeals should be stressed in design, and the goal is to attract these customers by, for example, word of mouth, in a longer run.

Given the strategic outline and the information of expected segment contribution to revenue, the manager then allocates resources for the management of these four segments. The cost associated with each segment can then be estimated. Therefore, expected profit in the four-segment scenario can be derived. Repeating the same procedure over the scenarios of 2, 3, 4, 5 segments, respectively, the segmentation scheme that maximize the expected profit of the online transaction platform is then decided. The beauty of MCES is that upon the final definition of segments, the manager has already considered feasible alternatives and objectively chosen the best option among them. Meanwhile, once the segments are defined, the manager already has a set of strategic map as well as profitability projections.

6. Conclusions

As segmentation has been one of the central marketing tasks for decades and customer profitability valuation has seen wide study during the past few years, surprisingly, up to this date, there is a gap in marketing research that await a bridge to link up of these two important and closely related dimensions. By the introduction of a decision support system (MCES) with the goal of maximizing customer equity by segmentation, this paper attempts to fill the gap. We take the view that for a segmentation scheme to be managerially relevant, segments should not be the end of a one-shot model which is independent or antecedent of strategic decision making. Instead, managers should be provided with a set of possible segment alternatives as the means for them to initiate relevant strategies so that profit maximization, or customer equity maximization if one likes, can be objectively approached through the interplay of numerical classification rules and managerial judgment.

The MCES decision support system introduced here is unique in that it accommodates the essence of customer profitability valuation into a segmentation scheme in a sensible and flexible manner, that it suggests the number of segments to be determined by the goal of profit maximization instead of some arbitrary numerical criterion, and that central to its technical core the outlier problem which is pervasive in cluster analysis has been addressed by a modified K-Means algorithm so that clustering can reflect the pattern of the majority of ordinary observations in a data set instead of being influenced by a handful of outliers. More managerially relevant segmentation, therefore, is expected from the adoption of this decision support system. We also resort to fictitious case as well as real-world data to exposit and demonstrate, at least partially, the performance of the proposed framework. In a word, the system reflects our argument that it should be managers, rather than numerical methods, that should have a final say on the definition of customer segments.

Common sense MCES may seem if one buys in the popular CRM rhetoric. However, to the best of our knowledge, segmentation techniques, no matter how complicated and/or refined they are, have never been fused into the customer profitability valuation framework which we think to be of ultimate importance if a firm really aspires to leverage its expensive CRM system. And to actually achieve this goal, managerial judgment should play no lesser role than complicated numerical rules in segmentation tasks. Admittedly, being a pioneering research that attempts to fuse two important CRM dimensions together, this paper puts more emphasis on the principle side rather than details. Given the foundation we lay here, a few important aspects should be attacked in the future. For example, quoting various references, this paper takes each customer's expected contribution as exogenously given. One direction of future research is to handle this aspect within the system. Meanwhile, though a series of research based on probability models has shown that marketing covariates are not necessarily crucial for the prediction of customer's repeat behavior, a further refined MCES may allow for expected customer contributions to be changed given a set of marketing mix for a specific segmentation definition. Meanwhile, for some reasons that cannot be expected a priori, there may be cases that a firm decides to discard certain segment(s). It is therefore of interest for future research to take this possibility in to account.

To sum up, guided by the principle of profit maximization, this study adds to managers' decision making toolbox an objective and easy-to-implement framework for segmentation. Also, upon segmenting a customer database, the modified cluster algorithm proposed in this study will serve to lessen biases caused by outliers in a database as is frequent in segmentation practice. We believe that a decision support system thus developed will improve a business' efficiency and effectiveness in segmentation so that the potential of its CRM system can be tapped.

7. References

[1] Blattberg, R.C. and Deighton J. (1996), "Manage Marketing by the Customer Equity Test," *Harvard Business Review*, July/August, 136–44.

[2] Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12 (1) 17—30.

[3] Boyce, Gordon (2000), "Valuing Customers and loyalty: the Rhetoric of Customer Focus Versus the Reality of Alienation and Exclusion of (Devalued) Customers," *Critical Perspectives on Accounting*, 11, 649—689.

[4] Garcia-Escudero, L. A., and A. Gordaliza (1999), "Robustness Properties of K Means and Trimmed K Means," *Journal of the American Statistical Association*, 94 (447), pp. 956-969.

[5] Green, Paul E.; and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," *Journal of Marketing*, 54(4), pp. 3-19.

[6] Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart (2002), Valuing Customers, *Marketing Science Institute Working Paper Series*, No. 01-119.

[7] Huang, Samuel H. (1998), "Automated setup planning for lathe machining," *Journal of Manufacturing Systems*, 17(3), 196-208.

[8] Kim, Namwoon, Vijay Mahajan, and Rajendra K. Srivastava (1995), "Determining the Going Market Value of a Business in an Emerging Information Technology Industry: The Case of the Cellular Communications Industry," *Technological Forecasting and Social Change*, 49, 257—279.

[9] Milligan, Glenn W. and Martha C. Cooper (1985), "An Examination of Procedures for Determining the Number of Clusters in a Data Set," *Psychometrika*, 50 (2), 159—179.

[10] Morrison, D. G. and Schmittlein, D.C. (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, 6 (2), 145–66.

[11] Mulhern, Francis J. (1999), "Customer Profitability analysis: Measurement, concentration, and Research Directions," *Journal of Interactive Marketing*, 13 (1), 25—40.

[12] Peppers, Don and Martha Rogers, *Enterprise One-to-one,* Pitakus.

[13] Rust, Roland T., Katherine N. Lemon and Valarie A. Zeithaml (2001), "Driving Customer Equity: Linking Customer Lifetime Value to Strategic Marketing Decisions," *Marketing Science Institute Working Paper Series*, No. 01-108.

[14] Schmittlein, David C., Donald G. Morrison and Albert C. Bemmaor (1985), "Why Does the NBD Model Work? Robustness in Representing Product Purchases, Brand Purchases and Imperfectly Recorded Purchases," *Marketing Science*, 4 (Summer), 255—66.

[15] Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (1), 1—24.

[16] Smith, Wendell R. (1956), "Product Differentiation and Market Segmentation as Alternative Marketing Strategies," *Journal of Marketing*, 21, 3—8.

[17] Wedel, Michel and Wagner Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd eds., Kluwer Academic Publishers.

Table 1 Summary of cluster analysis results

		Glass	Iris Plants	Letter Image	Lung Cancer	Wine
No. of Data		214	150	20000	32	178
No. of Attribute		9	4	16	56	13
Extreme Value (Yes/No) – 2σ		Yes (No.106、107、108、111、112、113、132、164、172、173、185)	No	Yes (No.10、553、9518、11843、14741、15334、16750、17599、18469)	No	Yes (No.4、6、11、15、19、32)
No. of Cluser		2	3	26	3	3
MSE (Inter-Cluster)		27.01899	14.16062	53.32213	47.37889	304578.9
MSE (Intra-Cluster)		6.729922	0.527575	33.67757	16.82197	15482.57
Intra-*SE*	Modified K-Means	859.4714	78.85563	633859.7	551.2781	2461905
	K-Means	821.3103	78.87183	617196.2	567.6376	2634232

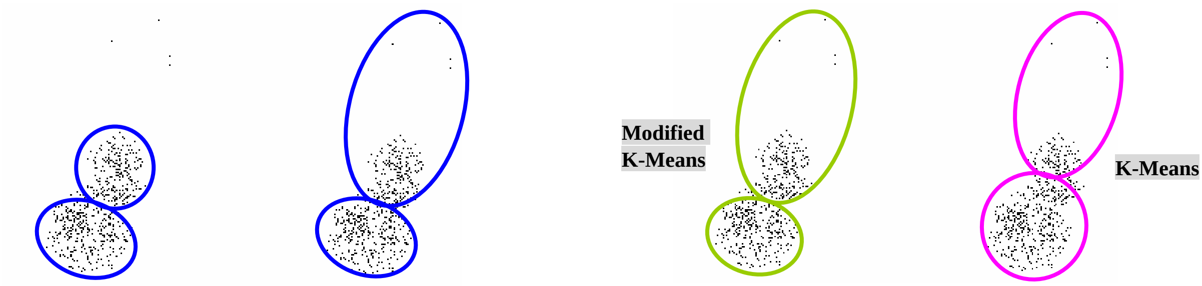

Figure 1 The comparison of dealing extreme value points by K-Means and modified K-Means

Table 2 The four segments found

		Orientation	
		Functional	Leisure
Usage	Heavy	Seg. 1	Seg. 3
	Light	Seg. 4	Seg. 2