Maximum resolution dichotomy for customer relations management

J. K. Ho
Department of Information & Decision Sciences,
University of Illinois at Chicago, USA

Abstract

A topological model for a survey of customer attitude is a simultaneous graphical display of all the dimensions of its relevant data base, which provides a geometrical shape as a descriptive, visual statistics of the customer base at any particular instance of its development. In particular, various dimensions were identified from available data for constructing a multi-attribute dichotomy that can help discern customers' orientation toward costs and benefits. Such a model provides a visual cue for the preference structure of the market. Using a reference subset of prejudged cases, the configuration of the dimensions and the angles among them can be optimized for a topology that maximizes the resolution of such dichotomies. The approach is illustrated in a survey of fans of a major league sports team. It can be a useful tool for data mining and visualization in market research, product/service design, and strategic planning for customer relations management.
Keywords: customer relations management, market research, cost-benefit focus, multi-attribute dichotomy, data mining, visualization.

1   Introduction

In customer relations management (CRM), consumers are often surveyed for their attitude toward various attributes of products and services for insight into their perception of price, quality and value [1]. The results of analysis can be used in market segmentation for more effective design and promotion of future offerings. In this work, we focus on surveys that ask customers to rate the importance of the attributes directly. As an additional tool for exploratory data analysis in market research, a new methodology in data mining is presented to

classify customers according to whether they tend to be cost or benefit focused. A framework for surveying customer attitudes and a conventional approach for data visualization are summarized in Section 2. In Section 3, the star plot method to data presentation is discussed in the context of conventional Exploratory Data Analysis and the latest development in Data Mining. A new methodology that casts customer attitudes in a multi-attribute dichotomy is presented in Section 4. A constrained optimization model to derive a maximum resolution topology [2] is introduced in Section 5. By helping to rationally classify the price versus quality orientation of customers, this approach provides an additional analytic tool in data mining for customer relations management. It can be used in strategic planning of either price-reduction or quality improvement for the corresponding segments of the customer base. The approach is illustrated in a survey of fans of a major league sports team in Section 6. Finally, directions for future work are discussed in Section 7.

2 A framework for cost versus benefit focus

We consider customer attitude surveys that ask for the rating, on some given scale of relative importance, of various cost and benefit attributes of a product, service, or experience. Typically, a number of questions relate to different aspects of cost, including direct pricing, and indirect opportunity costs. The other questions, not necessarily in the same number, relate to benefits in terms of perceived quality, and customer satisfaction. A high importance rating of a cost attribute means the customer is cost conscious in that regard, and a high actual cost is likely to detract from overall satisfaction. Conversely, a high importance rating of a benefit attribute means the customer is particular about its value and is likely to pay for improved quality there.

As an example, we use the case of a major league sports team and a survey of its fans' attitude toward the following attributes of their entertainment experience at a home game.

On a scale of 1 to 7 (most important), how do you rate the importance of:

Question 1: the view of the game from your seat?
Question 2: amenities available in the stadium?
Question 3: the team's performance in the games?
Question 4: cost of tickets?
Question 5: cost of transportation and/or parking?
Question 6: cost of concessions?

The purpose of the survey is to gain insight into the cost versus benefit orientation among customers. Since there is obvious trade-offs involved, any serious response should not be a simplistic rating of 7 on all the questions. Although even such a case can be trivially classified as cost-benefit neutral based on the equal weights on all attributes. For the non-trivial cases, we take the approach of exploratory data analysis instead of seeking a formal consumer behavioral model to aggregate the attribute measures.

3   Star plots in exploratory data analysis

Data visualization, as with the use of Star Plots, is an essential aspect of Exploratory Data Analysis (EDA) [3]. In Star Plots (or Kiviat Diagrams [4]), the polygonal glyph obtained by connecting the plotted values gives "shape" to the data. Pattern and structure of such shapes can then provide useful visual cues for EDA [5]. While it is obvious and well known that the shape of the star glyphs depends on the configuration of the attributes along the radial dimensions of the plots, EDA to date simply encourages analysts to investigate the variations on an ad hoc basis [6]. An example of the Star Plot of a customer's attitude in the case of the major league sports team is given in Figure 1.
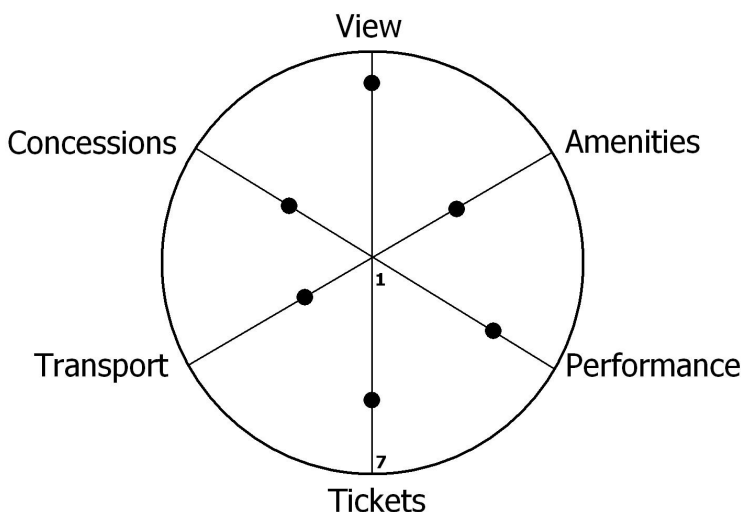

Figure 1:   Star plot of a cost-benefit profile.

A first step toward a more systematic approach is motivated by the topological analysis of online auction markets [7]. To address the question of "What is the 'shape' of an online auction market?" twelve attributes are identified. These attributes fall into two groups, giving rise to a buyer-seller dichotomy. By arranging the seller attributes on the left side and buyer attributes on the right side of a star plot, a glyph for the buyer-seller dichotomy can be drawn. Moreover, if the areas covered by the two parts can be used as a meaningful aggregate measure of their relative dominance, the result will be both visually and intuitively appealing. A larger area on the left (right) side of the glyph means dominance by the left (right) part. In the case of online auction markets, this asymmetry can be interpreted as market conditions being advantageous to either buyers or sellers. Before explaining how this can be done meaningfully, we first show why it may be applied to the cost-benefit framework.

4   Cost-benefit focus as a multi-attribute dichotomy

In general, a multi-attribute dichotomy is any multi-dimensional dataset [8–10] in which the dimensions can be partitioned into two groups, each considered to be contributing to one part of the dichotomy. In the cost-benefit framework, the six dimensions fall naturally into a price versus quality dichotomy. View of the game, amenities available in the stadium, and the team's performance are the three benefit dimensions. Cost of tickets, cost of transportation and/or parking, and cost of concessions are the cost dimensions.

To complete the model, two "dummy" dimensions named arbitrarily as Top and Bottom are introduced to serve as boundaries on the vertical axes. For reasons to be explained below, the average values for the six original dimensions are plotted on these two axes. Leaving the ordering of the dimensions within each group arbitrary for the time being, an example of the cost-benefit multi-attribute dichotomy is shown in Figure 2.
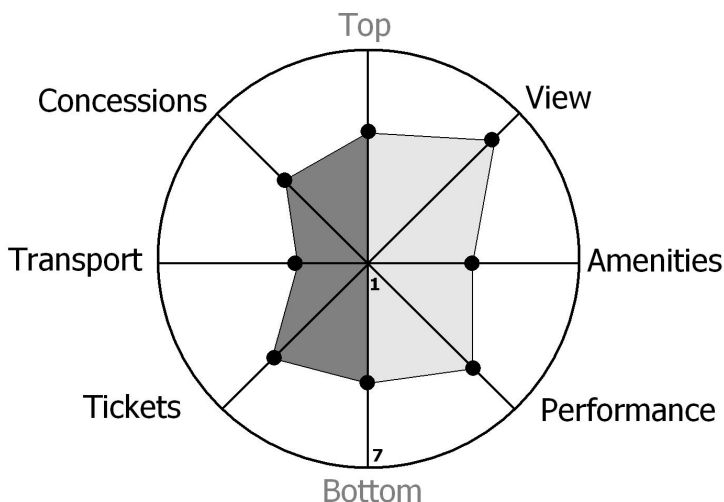

Figure 2:   Multi-attribute dichotomy (MAD) of a cost-benefit profile.

The concept of using the area of the parts of a dichotomy as an aggregate measure of their relative dominance is plausible, since increasing value of an attribute contributes positively to its designated part, and to the area in the glyph. However, the concept must be refined to realize meaningful interpretation. Refinement can be achieved by exploiting the degrees of freedom allowed by the topology of the glyph, namely, the configuration of the attributes and the angles between adjacent pairs of attributes. For any given arrangement of the attributes, the standard star plot produces a glyph along symmetrically spaced radial axes. Variations from this symmetry, together with permutations of the configuration, offer the choice of topologies that may make the enclosed areas a meaningful aggregate measure of the relative dominance of the two parts of the dichotomy.

To this end, we use the construct of a reference subset of the data instances to help define dichotic dominance. This is best explained in a medical scenario. Suppose a certain disease is monitored by a number of symptoms and tests, with a dichotic prognosis of "life" or "death". Judging from the combination of data for any particular case, it may be difficult to predict. A reference subset is a collection of non-trivial, non-obvious cases with known outcomes, namely life or death. Then an optimization model is used to derive a topology with maximum resolution in discerning dominance based on the reference subset [2].

5　A maximum resolution dichotomy for cost-benefit focus

In general, three options are open in choosing the reference subset of dichotomies. First, in the absence of significant insight into the classification of the dichotomies, and especially in the early stages of EDA, any initial dataset can be used on an ad hoc basis. An arbitrary configuration of the attributes within each part of the dichotomy is selected and evenly spaced. The classification of the dichotomies according to this topology is fed into the optimization model. The resulting optimal topology with respect to this reference set provides a working definition of dominance for the dichotomies. This is analogous to selecting a portfolio of stocks to provide an index for a stock market. The performance of any stock can be gauged relative to the index, which may be arbitrarily chosen initially. With better knowledge of the significance of individual stocks, more useful indices can be established. By the same token, the maximum resolution topology for multi-attribute dichotomies can be adaptively refined as the EDA for cost-benefit focus progresses.

Second, when there is expert opinion, the reference subset can be chosen subjectively. An expert (or a panel of experts) is asked to identify some cases that he or she judges with confidence to be cost-focused, and others (preferably more or less equal in number) that are benefit-focused. Obviously, a certain degree of consistency is expected of the subjective judgment for the optimization to be feasible.

Third, when there is empirical evidence, the reference subset can be established, e.g. with the use of a focus group. The subjects, suitably chosen for reliable results, may be asked or tested directly for their cost versus benefit orientation. Or in the event that tracking data is available, conclusions can be deduced from their actual pattern of behavior as customers.

Subject to the constraints of preserving the prejudged dominance in the reference subset of dichotomies, an optimal topology (configuration of attributes and angles between adjacent pairs) is derived from a Goal Programming Model [2] that maximizes the discriminating power (sum of absolute differences in left and right areas) for the reference subset. It should be remarked that maximum resolution in this context does not imply the classification of the most dichotomies. Any topology can classify all instances in a dataset (except in degenerate cases where the areas are numerically equal). The critical factor is that as aggregate measures, different topologies may classify the same dichotomy differently. The maximum resolution model rationalizes the choice of

one that best "enforces" the classifications in the reference subset. Also, the choice of setting the data values on the dummy dimensions at the average of the original six dimensions is necessary to preserve symmetry under optimization. This way, a case in which all dimensions are at equal levels will produce equal areas in both parts of the dichotomy.

6   Case of a CRM survey for a major league sports team

To demonstrate the approach of maximum resolution topology, we present a case in customer relations management for a major league sports team in the US. Due to propriety of information, the identity of the organization and details of the survey instrument are not disclosed. The customer survey was a major CRM campaign conducted in the summer of 2004. The six questions concerning the three cost attributes and three benefit attributes in our MAD model were part of a total of over 40 items in a questionnaire. Apart from typical demographic and market-oriented queries [11, 12], respondents were also asked whether they were current season ticket holders, and if they were likely to renew their subscription in the following season. Fans of the team attending its home games were invited to respond to the survey before and after a game, and during intermission. They were enticed with inexpensive souvenirs and entry to a lottery drawing with more valuable prizes. Results were made available on the team's Web site as an effort to promote interest and interaction with supporters of the team.

For this work, a sample of 369 responses collected in June and July 2004 were used. For illustrative purposes, an ad hoc reference subset was chosen with 14 plausible cases of benefit focus, and 10 plausible cases of cost focus. To simply the presentation, the records in this reference subset were labeled from 1 through 24 and shown in Table 1. The shaded entries for Q1, Q2, and Q3 correspond to the benefit-focused records (1–14), while the shaded entries for Q4, Q5, and Q6 correspond to the cost-focused records (15–24).

The optimal angles (in units of π from the Top axis) are (0.125, 0.625, 0.125, 0.125) for the cost dimensions, and (0.125, 0.125, 0.625, 0.125) for the demand dimensions, respectively. When a MAD is plotted in a MRT, the resulting glyph is called a Maximum Resolution Dichotomy (MRD). See Figure 3 for an example of a MRD for a generic customer survey record. Since the star glyph topology accentuates pair-wise correlation among the attributes, we observe that *Performance* and *Amenities* have the highest correlated effect for the benefit dimensions, whereas *Tickets* and *Transport* do so for the cost dimensions, with assigned weights of 0.625 in both cases.

This input to the optimization model produces a maximum resolution topology (MRT) with an ordering (from the Top axis):

Cost Dimension	Benefit Dimension
Concession	Performance
Tickets	Amenities
Transport	View

Table 1:   Reference subset for the major league sports team MAD.

Record	Q1	Q2	Q3	Q4	Q5	Q6
1	5.00	5.00	5.00	4.00	2.00	3.00
2	7.00	5.00	7.00	5.00	5.00	5.00
3	4.00	4.00	5.00	1.00	3.00	4.00
4	6.00	6.00	6.00	5.00	5.00	5.00
5	7.00	1.00	5.00	4.00	3.00	2.00
6	6.00	5.00	7.00	6.00	5.00	5.00
7	6.00	4.00	5.00	4.00	5.00	4.00
8	7.00	7.00	6.00	7.00	6.00	6.00
9	6.00	6.00	2.00	4.00	5.00	3.00
10	4.00	4.00	3.00	4.00	3.00	3.00
11	6.00	4.00	2.00	4.00	3.00	4.00
12	6.00	6.00	5.00	7.00	5.00	5.00
13	3.00	3.00	4.00	4.00	2.00	4.00
14	5.00	5.00	4.00	4.00	5.00	5.00
15	7.00	1.00	4.00	7.00	5.00	1.00
16	4.00	3.00	2.00	4.00	3.00	3.00
17	5.00	5.00	3.00	5.00	5.00	4.00
18	6.00	5.00	5.00	6.00	7.00	4.00
19	7.00	6.00	7.00	7.00	7.00	7.00
20	5.00	4.00	6.00	4.00	6.00	7.00
21	4.00	3.00	7.00	6.00	6.00	4.00
22	6.00	1.00	1.00	5.00	4.00	3.00
23	4.00	5.00	5.00	6.00	6.00	5.00
24	6.00	6.00	5.00	6.00	7.00	7.00

Based on the MRT derived from this particular reference subset, the entire sample of 369 records are classified as benefit-focused, cost-focused, or cost-benefit neutral according to whether the right area in the corresponding MRD is larger, smaller, or equal to the left, respectively. The results are summarized in Table 2, where importance is rated on a scale of 1 to 7, and all percentages are with respect to the group in the respective column. Of the 369 respondents, 58% are classified as benefit-focused, 33.3% as cost-focused, and 8.7% as cost-benefit neutral. The ratio of benefit-focused respondents to cost-focused respondents is 1.74. Management may use this indicator to, for example, devote more effort into quality improvement instead of price reduction.

Since we also have the statistics on season ticket holders, we observe further that of the 94 in this group for the current season, 62.8% are benefit-focused, which is above the overall level. Even more interestingly, the unlikely subscribers are more cost-focused while the undecided have essentially the same profile as the entire sample.
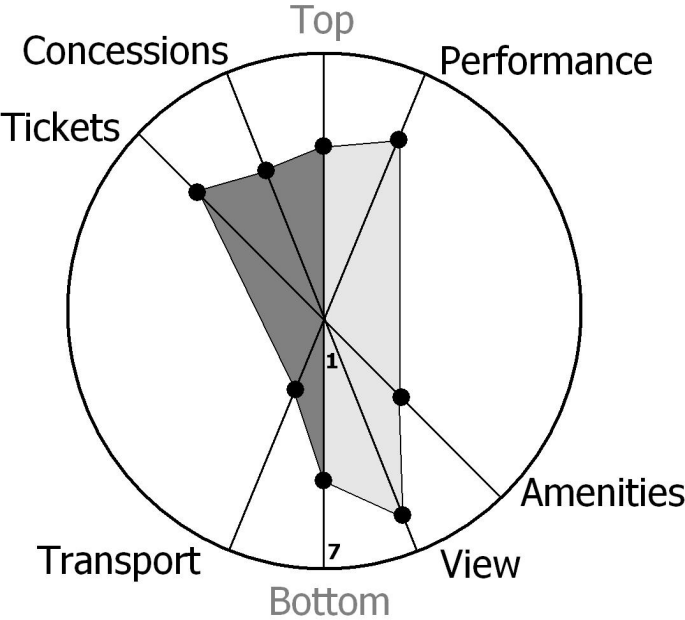
Figure 3:   Maximum resolution dichotomy (MRD) of a cost-benefit profile.

Table 2.

	CUSTOMERS	SEASON TICKET HOLDERS		
	ALL	CURRENT	NEXT SEASON		
			LIKELY	UNLIKELY	UNDECIDED
Number of complete surveys	369	94	136	100	133
Average importance of:					
Q1 View of Game	5.80	5.94	5.60	6.15	5.74
Q2 Amenities available	4.74	5.03	4.37	5.25	4.74
Q3 Performance of team	5.45	5.70	5.21	5.63	5.57
Q4 Cost of tickets	5.32	5.32	5.21	5.62	5.20
Q5 Cost of transport	4.99	5.22	4.76	5.45	4.89
Q6 Cost of concessions	4.53	4.79	4.13	5.02	4.57
Standard deviation of:					
Q1 View of Game	1.36	1.23	1.55	0.98	1.36
Q2 Amenities available	1.66	1.60	1.76	1.47	1.59
Q3 Performance of team	1.59	1.42	1.71	1.47	1.53
Q4 Cost of tickets	1.51	1.34	1.62	1.34	1.48
Q5 Cost of transport	1.66	1.56	1.72	1.51	1.64
Q6 Cost of concessions	1.76	1.67	1.84	1.54	1.73
% BENEFIT FOCUSED	57.99	62.77	59.56	56.00	57.89
% COST FOCUSED	33.33	26.60	33.09	34.00	33.08
% NEUTRAL	8.67	10.64	7.35	10.00	9.02
BF/CF	1.74	2.36	1.80	1.65	1.75

7   Discussion

In customer relations management and marketing research, it is common interest to gain insight into consumers' attitude toward the costs and benefits associated with a product, service, or experience. Surveys on the multitude of attributes lead to high-dimensional data, for which meaningful aggregate measures remain a major challenge in data mining [13, 14]. As a contribution to the modeling of market segments, we presented a methodology to rationally classify customers as to whether they are benefit-focused or cost-focused. The cost-benefit survey framework is cast as a multi-attribute dichotomy, with a cost side and a benefit side. As an extension of the star plot to visualize multi-dimensional data, the areas spanned by the cost-side and benefit-side attributes of a data instance in a survey suggest an aggregate measure of the relative dominance of the corresponding parts. By optimizing over the ordering of the attributes as well as the angles among them, a configuration with maximum discerning power with respect to a reference subset of pre-classified cases is obtained. A case study with a major league sports team was introduced. As illustration, an ad hoc reference subset of 24 fans was used to derive a maximum resolution topology, based on which 58% of the 369 fans in the sample were classified as benefit-focused. Future work includes expounding the application of this dichotomy as a significant output of data mining for CRM, as well as refining the selection of the reference set by, for example, incorporating results from focus group studies, as well as expert judgment from researchers of specific markets.

References

[1]   Zeithaml, V. A., Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, *Journal of Marketing,* **52(3)**, pp. 2-22, 1988.

[2]   Ho, J. K & Chu, S. C. K., Maximum resolution topology for multi-attribute dichotomies, *Informatica* **16(4)** pp. 557-570, 2005

[3]   Tukey, J. W., *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, 1977.

[4]   Kolence, K. & Kiviat, P., Software unit profiles and Kiviat figures. *ACM SIGMETRICS, Performance Evaluation Review*, **2(3)** September, pp. 2-12, 1973.

[5]   Hoffman, P. & Grinstein, G., A survey of visualizations for high-dimensional data mining, in U. Fayyad et al (eds.) *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann, pp. 47-82, 2001.

[6]   Chambers, J., Cleveland, W., Kleiner, B. & Tukey, P., *Graphical Methods for Data Analysis,* Wadsworth Press, Belmont, CA, 1983.

[7]   Ho, J. K., Topological analysis of online auction markets, *International Journal of Electronic Markets* **14(3)** pp. 202-213, 2004.

[8]   Chen, M. S., Han, J. & Yu, P. S., Data mining: An overview from a database perspective, *IEEE Transactions on Knowledge and Data Engineering* **8(6)** pp. 866 – 883, 1996.

[9]   Ankerst M., Ester, M. & Kriegel, H. P., Towards an effective cooperation of the computer and the user for classification, *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining (KDD 2000)*, Boston, MA, pp. 179-188, 2000.

[10]   Keim, D. A., Information visualization and visual data mining, *IEEE Transactions on Visualization and Computer Graphics*, **7(1)** pp. 100-107, 2002.

[11]   Luna-Arocas, R. & Tang, T., The use of cluster analysis to segment clients of a sport center in Spain, *European Sport Management Quarterly* **5(4)** pp. 381-413, 2005.

[12]   Dale, B., van Iwaarden, J., van der Wiele, T. & Williams, R., Service improvement in a sports environment: a study of spectator attendance *Managing Service Quality* **15(5)** pp. 470-484, 2005.

[13]   Green, P. E. & Srinivasan, V., Conjoint analysis in consumer research: issues and outlook, *Journal of Consumer Research,* **5(2)** pp. 103-123, 1978.

[14]   Green, P. E. & Srinivasan, V., Conjoint analysis in marketing: new developments with implications for research and practice, *Journal of Marketing*, **54(4)** pp. 3-19, 1990.