A Spectral Approach to Collaborative Ranking

Joachim Giesen
Max-Planck Institut für Informatik
Stuhlsatzenhausweg 85
D-66123 Saarbrücken
jgiesen@mpi-inf.mpg.de

Dieter Mitsche
Department of Computer Science
ETH Zürich
CH-8092 Zürich
dmitsche@inf.ethz.ch

Eva Schuberth
Department of Computer Science
ETH Zürich
CH-8092 Zürich
sceva@inf.ethz.ch

Abstract

Knowing your customers and their needs is mandatory when conducting business. In an e-commerce environment, where one often knows much less about individual customers than in a face-to-face business, questionnaires - besides historical transaction data and basic information like names and geographic location - are popular means to get to know customers. Here we report on an approach to analyze preference data that consumers provide in stated choice experiments which can be conducted on a company's web site. Naturally, the information one captures with web based questionnaires tends to be sparse and noisy. Nevertheless, our results show that if one collects data from enough consumers one can learn about different segments and their needs. The results are obtained with a spectral collaborative ranking algorithm that can be applied to stated choice data, especially, choice based conjoint analysis data.

Keywords. clustering, other algorithms, other applications

Introduction

In competitive markets it is essential to match consumer needs as well as possible. In some markets a single product fulfills all consumers' needs. In the other extreme there are markets where one strives to match each individual consumer's needs. But in most markets a compromise between these two extremes is the best strategy to pursue. Such a compromise is to segment the market and to match the needs of selected segments as closely as possible. So it is not surprising that market segmentation is one of the key activities in marketing. There are essentially two ways to segment consumers in a market, either indirectly by the consumer's socio-demographic profile, e.g., by age, gender, geographic location and so on, or directly by consumer's needs. Socio-demographic data are often readily available or at least less expensive to elicit than preference data. However, preference data more directly tell about consumers' needs. It is a non-trivial task to transform socio-demographic data together with a limited amount of preference data into reliable preference data, for first steps into this direction see (Hüllermeier & Fürnkranz 2004). On the other hand, although it is expensive, preference elicitation is daily practice in market research. Among the most popular preference elicitation methods is conjoint analysis[1], which assumes that products can be described in terms of attributes and levels, i.e., that the product set has a conjoint structure. In one of its many variants (choice based conjoint analysis) a respondent is faced with a few number of choice tasks. In each choice task the respondent has to pick one out of two to five product profiles presented to him. From all his choices the respondent's preferences over the whole product set are estimated. Remarkably, the product set typically is large compared to the number of choice tasks. Preferences are usually encoded in a value function for each respondent. A value function assigns to every product a value and induces a ranking of the products, i.e., the higher the value of a product, the higher is the product's position in the ranking. Often one even tries to extract metric information from a value function, e.g., information like the respondent likes product x five times better than product y. Since it is not straightforward to derive metric information from choice questions we restrict ourselves here to ranking information. In our approach we confront each respondent with very few choice tasks. The number of choice tasks should be independent of the size of the product set since even the most well meaning respondents get worn out after a certain fixed number of choice tasks. We make up for that by asking more respondents for larger product sets. This approach is suited for web based questionnaires where the respondent's answers are not always reliable and respondents get worn out easily.

Our approach also differs from the standard practice in market research (Sawtooth ; Chapelle & Harchaoui 2005) in that we first segment the set of respondents and then compute one ranking for each segment to approximate the individuals' rankings. Usually first individuals' value functions are estimated which are subsequently used to segment the market. That is, our approach builds directly on aggregated information collected from similar respondents—and as we will demonstrate, the similarity of respondents can often be decided even if the respondents answer only very few choice tasks each.

The approach was implemented using spectral techniques, i.e., building on the eigenvalues and -vectors of some matrix. We report on an evaluation of the approach on synthetic and

[1]See the section about conjoint analysis data or (Gustafsson, Herrmann, & Huber 2000) for more details.

real market research data and compare it with a more standard approach (Chapelle & Harchaoui 2005) on the real data in the section about conjoint analysis data.

Segmentation and Ranking

Let X be a set of n products, typically substitution goods or services, and let m be the number of respondents. The basic data structure that underlies our algorithm is an $m \times \binom{n}{2}$-matrix A. The entries a_{ij} of A are in $\{\pm 1, 0\}$ and can be interpreted as follows: the index i refers to the i'th respondent and the index j refers to the j'th product pair $(x, y) \in \binom{X}{2}$. The entry a_{ij} is 1 if the i'th respondent prefers product x over product y, it is -1 if he prefers y over x and it is 0 if the respondent has not compared x and y. Since each respondent has to perform only very few choice tasks most entries of A will be 0.

Our algorithm has three phases. In the first phase we estimate the number of consumer types k. To this end we use the positive semi-definite $m \times m$ matrix $B = AA^T$. In the second phase we again use the matrix B to segment the m respondents into k classes that correspond to the types. In the third phase we use the segmentation obtained in the second phase to compute a ranking for each type.

Estimating the number of types. If the population can be segmented into a small number of types k, then this is simply a model order selection problem where we could systematically try out small values of k. But we also expect[2] that the k largest eigenvalues of B are significantly larger than the $m - k$ smallest eigenvalues. We estimate k as the number of eigenvalues larger than some threshold.

Respondent segmentation. Let us first give an interpretation of the entry b_{ij} of B. Let $X_{ij} \subset \binom{X}{2}$ be the set of pairs in $\binom{X}{2}$ that are compared by both the i'th and the j'th respondent. The entry b_{ij} is the number of pairs in X_{ij} on which the i'th and the j'th respondent agree in their comparison minus the number of pairs in X_{ij} on which they disagree. The intuition is that b_{ij} is large if both respondents belong to the same type and small otherwise. In other words the entries in the sub-matrices of the matrix B that have both indices in the same type are expected to be larger than the entries off these blocks. The goal now becomes identify these sub-matrices. We use the projector $P_k = \sum_{i=1}^{k} v_i^T v_i$ onto the space spanned by eigenvectors $v_i, i = 1, \ldots, k$ to the k largest eigenvalues for this purpose. For k we use the estimate from the first phase. Note that we can use the columns of P_k to associate each respondent with a vector in $\mathbb{R}^m$, namely, the i'th respondent corresponds to the i'th column of P_k. We use k-means clustering (MacQueen 1967) to segment the respondents, i.e., their corresponding vectors, into types.

[2] We will later experimentally test if this assumption is reasonable.

Computing the typical rankings. Once we have segmented the respondents we can compute for each type a m-dimensional characteristic vector $c_t,\ t = 1, \ldots, k$, whose i'th entry is 1 if the i'th respondent belongs to this type and 0 otherwise. The result of the matrix-vector product $A^T c_t$ is an $\binom{n}{2}$-dimensional vector each of whose entries corresponds to the comparison of two products. We interpret the entries of $A^T c_t$ as follows: if the entry that corresponds to the comparison of products x and y is positive then x is preferred over y by the t'th type. If the entry is negative then y is preferred over x by the t'th type, otherwise the type is indifferent between products x and y. Note that this procedure not necessarily provides us with a (partial) order on the products since it is not guaranteed that the resulting relation is transitive. When we refer in the following to the computed or reconstructed ranking for type t we think of the vector $A^T c_t$ which does not necessarily have to satisfy transitivity.

Synthetic data

We will first test our method on synthetic data that are generated according to a statistical model[3]. In (Giesen, Mitsche, & Schuberth 2007) we devised an algorithm for which we can prove theoretical performance guarantees using our statistical model. Unfortunately the algorithm is not practical, i.e., it only works if the number of respondents is very large.

Population model. We assume that the population can be partitioned into k types. Let $\alpha_i \in (0, 1)$ be the fraction of the i'th type in the whole population. For each type there is a ranking $\pi_i,\ i = 1, \ldots, k$, i.e., a permutation, of the n products.

Respondent model. We assume that the set of respondents faithfully represents the population, i.e., α_i is also (roughly) the fraction of respondents of type i among all respondents. Given a respondent let π be the ranking that corresponds to the type of this respondent. For any comparison of products x and y, with $x \prec y$ according to π, we assume that the respondent states his preference of y over x with probability $p > 1/2$. Note that this allows the respondents' answers to violate transitivity - something one also observes in practice.

[3] Our model is such that if we would change our elicitation procedure slightly to enforce transitivity in each respondent's answers (simply by asking only for comparisons not in the transitive hull of the previous answers), then the rankings of the respondents (if we would ask them until we have elicited the complete ranking) that are of type i follow a Mallows distribution (Mallows 1957), i.e.,

$$P[\pi] = ce^{-\log(p/(1-p))\delta(\pi,\pi_i)}, \quad c^{-1} = \sum_{\pi} e^{-\log(p/(1-p))\delta(\pi,\pi_i)}$$

where $\delta(\pi, \pi_i)$ is the Kendall distance (Kendall 1970) between π and π_i. In (Feigin & Cohen 1978) and (Critchlow 1985) it was shown that Mallow's model often provides a good fit to ranking data.

We use the following measures to assess the quality of our method on data generated according to the model[4].

Misclassifications. We use the following intuitive notion of misclassified respondents, see also (Lange *et al.* 2004; Meila & Verma). From the model we have a partition function

$$\varphi : \{1, \ldots, m\} \to \{1, \ldots, k\}$$

that assigns every respondent to his type. The segmentation algorithm provides us with another partition function

$$\psi : \{1, \ldots, m\} \to \{1, \ldots, k'\},$$

where k' is the number of estimated types. The number of misclassified respondents is m minus the size of a maximum weight matching on the weighted, complete bipartite graph, whose vertices are the classes $\varphi^{-1}(i), i \in \{1, \ldots, k\}$ and the classes $\psi^{-1}(j), j \in \{1, \ldots, k'\}$ produced by the algorithm. The weight of the edge $\{\varphi^{-1}(i), \psi^{-1}(j)\}$ is $|\varphi^{-1}(i) \cap \psi^{-1}(j)|$, i.e., the size of the intersection of the classes. The matching gives a pairing of the classes defined by φ and ψ. Assume without loss of generality that always $\varphi^{-1}(i)$ and $\psi^{-1}(i)$ are paired in the maximum weight matching. Then the number of misclassifications is given as

$$m - \sum_{t=1}^{\min\{k,k'\}} |\varphi^{-1}(t) \cap \psi^{-1}(t)|.$$

In our analysis we only compute the number of misclassifications if we estimate the number of types k correctly.

Number of inverted pairs. We only compute this measure if we estimate the number of types correctly. In this case we get a one to one correspondence of actual and reconstructed types via the maximum weight matching that we compute in order to count the number of misclassified respondents. We compute the number of inverted pairs in the $\{\pm 1\}^{\binom{n}{2}}$ vector that encodes the product ranking for type t and its reconstruction $A^T c_t \in \mathbb{Z}^{\binom{n}{2}}$, where c_t is the computed characteristic vector for class t. That is, we count entries that have different signs in the two vectors (not counting zero entries). We always state the fraction of inverted pairs compared to all pairs, i.e., $\binom{n}{2}$.

Hit rate. For a respondent of type t that we classify correctly we compute the *hit rate* as the average success probability of correctly predicting the outcome of any pairwise product comparison from the vector $A^T c_t$. Note that if there is a zero entry in $A^T c_t$ for a product pair we just flip a fair coin to predict the outcome of a comparison performed by a respondent of type t.

[4]Intuitively speaking our method should work well on data generated according to the model since it can be shown that if the typical rankings π_i are sufficiently well separated, then the expected matrix B has k eigenvalues in $\Theta(m)$ and $m-k$ eigenvalues 0, i.e., it has a large spectral gap. Furthermore it has a nice block structure with blocks corresponding to the different types in the population.

Experimental results. In all our experiments we fix $n = 30$ (number of products), $l = 12$ (number of pairwise comparisons performed by each respondent) and $\alpha_i = m/k$ (fraction of the i'th consumer type of the population) for all types $i = 1, \ldots, k$. This leaves m (number of respondents), k (number of types), $1-p$ (deviation probability) and δ (separation of type rankings) as free parameters. Note that the information theoretic lower bound on the number of pairwise comparisons needed by a respondent to rank $n = 30$ products is $n \log_2 n \approx 147$ and a sorting algorithms like QUICKSORT needs ≈ 300 pairwise comparisons to rank all products. For any fixed parameter setting we ran ten independent trials of the experiment and report on our findings in the following tables.

The first table contains data for varying m, fixed deviation probability $1 - p = 0.1$ and $k = 3$ types. The average minimum separation is $\delta = 45.2\%$.

m	miscl.	inv.	hit rate
150	52.8%	13.3%	70.7%
450	8.8%	2.9%	93.9%
750	6.1%	0.8%	98.4%
1050	4.3%	0.3%	99.5%

One can see that with increasing m also the quality of the segmentation and the reconstruction increases. The next table contains data for varying m, fixed deviation probability $1 - p = 0.2$ and $k = 3$ types. The average minimum separation is $\delta = 46\%$.

m	miscl.	inv.	hit rate
150	52.7%	20.9%	61.7%
450	41.2%	18.5%	74.8%
750	27.4%	9.6 %	86.7%
1050	6.7%	4.8%	93.2%

Finally we show a table containing data for varying m, fixed deviation probability $1 - p = 0.1$ and $k = 5$ types. The average minimum separation is $\delta = 40\%$.

m	miscl.	inv.	hit rate
150	61.1%	16.1%	59.6%
650	25.4%	7.3%	86.9%
1150	14.9%	2.2%	96.1%
1650	11.2%	0.63%	98.8%

The last table shows that also for an increasing number of types we have to ask more respondents to get comparably good results. This is plausible because for a fixed number of respondents a larger number of types means that a type contains less respondents and thus we have less ranking information about this type.

Conjoint analysis data

In this section we report on the performance of our algorithm on real data that were provided to us by SAWTOOTH SOFTWARE, INC.. Let us first give a short summary of conjoint analysis.

Conjoint analysis. Conjoint analysis deals with preference elicitation over a set of products that can be described by attributes and levels, i.e., the product set is a subset of a Cartesian product $A_1 \times \ldots \times A_n$, where each attribute A_i is a set of levels, e.g., an attribute *color* might have the levels $\{red, green, blue\}$. In conjoint analysis one is especially interested in the trade-offs customers make. For example if there are attributes *price* and *speed* to describe a car then the price attribute might range from \$20,000 to \$100,000 and the maximum speed might range from 100 mph to 160 mph, but it is reasonable that it is impossible to build a car that runs 160 mph for \$20,000. Thus when buying a car people are forced to deal with a trade-off decision between price and speed. It is reasonable to assume that the car market can be segmented into typical types regarding the price vs. speed trade-off, e.g., there are the extremes—sporting drivers and economic drivers—and probably a more diffuse compromise type[5]. Market research tries to find such segments and to estimate their size. Choice based conjoint analysis is a popular means to this end. In choice based conjoint analysis respondents are confronted with a series of choice tasks. In each choice task a respondent has to pick one out of a given number of (partial) product profiles. From a respondent's choices an individual value function or ranking is computed.

We got 21 disguised data sets from SAWTOOTH SOFTWARE, INC., i.e., we do not know in which context and for which application they were elicited. Here we report in more detail on the performance of our method on three representative ones of them. In the end of this section we will give a brief overview of the results of our algorithm applied to the remaining data sets. Before presenting the three representative studies we will introduce the measures that we use in the analysis. Note that in contrast to the synthetic data here we do not have a ground truth to compare against. Thus we have to use different quality measures.

Number of inverted pairs. Here we use a different definition than for the synthetic data. We compute for the respondents segmented into type i and the computed ranking for type j the average number of inverted pairs, i.e., the average number of product pairs on which a respondent of type i differs in his stated preference from the ranking computed for type j and denote it by inv_{ij}, i.e.,

$$\mathrm{inv}_{ij} = \frac{1}{\mid T_i \mid} \sum_{r \in T_i} \mathrm{inv}_j(r),$$

where T_i is the set of respondents segmented into type i and $\mathrm{inv}_j(r)$ is the number of inverted pairs that the respondent r has with respect to type j. We do not count the case when there is indifference in the type ranking for a product pair while the respondent states a preference for that product pair.

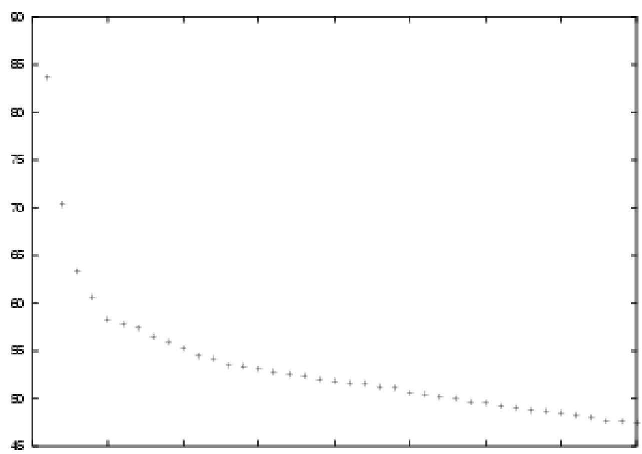
Maximum deviation. For each respondent we compute the deviation from his type as the number of inverted pairs for this respondent over the number of questions he had to answer. The maximum deviation for a given type is the maximum deviation of any respondent that belongs to this type. The maximum deviation for the segmentation is the maximum deviation of any respondent independent of his type.

Average deviation. For each type the average deviation is defined as the average number of inversions of respondents in that type over the number of questions they were asked, i.e., as $\frac{\mathrm{inv}_{ii}}{l}$. The average deviation for the whole segmentation is defined analogously. For the synthetic data the average deviation roughly corresponds to to $1 - p$.

Minimum separation. This is a measure for how different the computed type rankings are. It is a quality measure only in so far as types that are similar are more difficult to distinguish. For two computed type rankings the separation is the fraction of product pairs that are ranked differently in the two rankings. Note that 0 entries do not contribute to this measure. For the studies we compute the minimum over the separation of all pairs of type rankings.

In the following we present details for three typical out of the 21 studies.

Study 1. This study has 539 respondents and 4 attributes with 4, 3, 3 and 5 levels, respectively, i.e., the set of all product profiles has 180 elements. The number of questions per person is 30 (10 choice tasks with 4 products each). We first show a plot of the eigenvalues of the matrix B for this data set.



We can see a separation of the four largest eigenvalues from the other eigenvalues. If we compute the rankings for four types we get types (segments) with 81, 119, 130 and 209 respondents. That is, all four types contain a significant number of respondents.

In the following table we list the average number of inverted pairs, i.e., the average number of pairs on which a respondent of type $i = 1, \ldots, 4$ differs in his stated preference from the ranking computed for type $j = 1, \ldots, 4$.

[5]This observation motivates our statistical model of a segmented population.

Note that the off-diagonal entries are much larger than the diagonal entries.

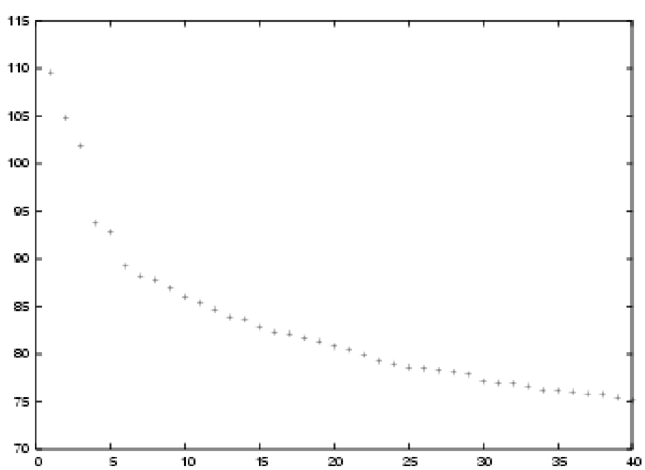
	ranking type 1	ranking type 2	ranking type 3	ranking type 4
type 1	0.44	6.77	5.11	6.53
type 2	5.58	0.92	6.92	7.98
type 3	3.56	6.1	0.84	5.67
type 4	3.56	5.08	4.25	1.16

In the following table we show the maximum deviation and average deviation for each type.

	max. dev.	avg. dev.
type1	10.0%	1.5%
type2	10.0%	3.1%
type3	13.3%	2.8%
type4	20.0%	3.9%

That the maximum deviation for all types is significantly larger than the average deviation means that the types contain outliers. This is expected since our method assigns every respondent to one of the four classes and not only respondents that fit nicely. This is not a problem since the "outliers" can be easily detected and removed from the types—doing this decreases the the average deviation even further. Also note that the minimum separation of the four types is only 1.4%, i.e., much smaller than the relative average deviation of $3.9.\%$ for the fourth type. Nevertheless the four types are clearly separated as can be seen from the average numbers of inverted pairs in the table above. This shows that our method is capable of detecting even small differences in types.

Study 2. This study has 1184 respondents and 3 attributes with 9, 6 and 5 levels, respectively, i.e., the set of all product profiles has 270 elements. The average number of questions per person is 48 (12 choice tasks with 5 products each). Again, we first show a plot of the eigenvalues of the matrix B for this data set.



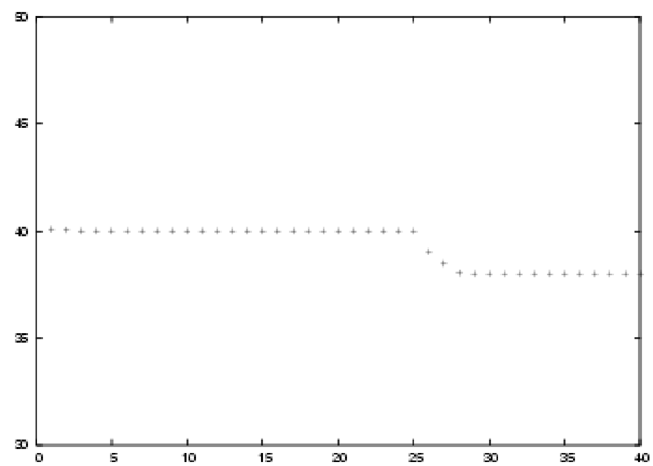
Like in the first study there is a gap in the eigenvalues. After computing five type rankings however, it turns out that four of the five types do not contain a significant number of respondents. The size of the types are 3, 3, 6, 8 and 1164, respectively. If we decrease the number of types to three the size of the types becomes 3, 6 and 1175, respectively. Only if we increase the number of types to seven, the large type gets split into two smaller types of size 942 and 201, respectively. In the following table we report on how the choice of k affects the maximum and average deviation for the corresponding segmentation.

k	max. dev.	avg. dev.
1	37.5%	12.1%
2	37.5%	5.7%
3	37.5%	5.8%
4	37.5%	5.7%

k	max. dev.	avg. dev.
5	37.5%	5.7%
6	37.5%	5.5%
7	31.3%	4.1%
8	33.3%	3.9%

Since the choice of k hardly affects the average deviation we conclude that there is only one large type in the population. The smaller types can be regarded as outliers.

Study 3. This study has 300 respondents and 6 attributes with 6, 4, 6, 3, 2 and 4 levels, respectively, the set of all product profiles therefore has 3456 elements. The average number of questions is 40 per person (20 choice tasks with 3 products each). There is no significant gap in the eigenvalue spectrum, see the following plot.



For this study our method did not provide us with good results because of the very small ratio between the number of respondents and the number of product profiles. The information gathered in this study was too sparse. Since for this study 300 respondents were asked 40 questions each we can only learn preferences for at most $300 \times 40 = 12,000$ product profile pairs. The number of all these pairs is $5,970,240$. That means that we have preference information for at most 0.2% of all pairs.

As mentioned earlier, altogether we received 21 data sets from SAWTOOTH, INC.. We presented results for three representative ones above. There are four more data sets which are similar to Study 1, i.e., there is a significant gap in the eigenvalues and all types contain a significant number of respondents. Five more data sets show the same characteristics as Study 2, i.e., there is a significant gap in the eigenvalues but after segmentation it turns out that basically there is only one big type while the number of respondents segmented into the other types is very small. Seven more data sets are similar to the one in Study 3 in the sense that the ratio between the number of respondents and the number of product profiles is so small that we cannot apply our method. Finally there are two data sets that we could not analyze due to data format problems.

The same data sets have been analyzed in (Chapelle & Harchaoui 2005). The analysis follows the traditional way to estimate a respondent's value function and not a segmentation/aggregation strategy as we do. For the eleven data sets that they choose for their analysis which comprises three different methods they get an average deviation in the range of 13% to 58%. For the majority of studies and methods they report an average deviation above 30%, i.e., much larger than what we observe. This indicates that segmentation/aggregation can be a useful strategy for analyzing conjoint analysis data—note that in our approach so far we do not even exploit the conjoint structure of the data.

Discussion

The crucial parameter in our method is the number of respondents. This number is not easy to guess a priori[6] since it does not only depend on the number of products (which is known a priori) but also on the number of types, their sizes, the separation of the type rankings and the amount of noise in the data. The latter quantities are those that ideally we want to learn from the respondents' answers to the choice tasks.

If the number of respondents is small compared to the size of the product set, then our method is no longer directly applicable. Since the data in such scenarios are necessarily sparse there is the need for a (statistical) model that allows to extrapolate from the stated preference to products that have not been covered by the choice tasks. Work in this direction is done in discrete choice modeling, see for example (Train 2003).

In any case one should make sure to keep the set of product profiles as small as possible, which often can be achieved by careful modeling. For example one should not include products in the profile set that cannot be built or are overly expensive to build. Note that even for a conjoint structure with many attributes and levels the set of interesting products might be small. One should also try to keep the attribute sets as small as possible. That is, the different levels should differ significantly, e.g., it makes no sense to have the price attribute for a car as fine tuned as to resolve differences up to $100. One could start out with very few levels per attribute and then step by step refine the levels if one feels that more information is needed.

Finally, our analysis of the conjoint data also puts the statistical model from which we generated the synthetic data to the test. The results show that the model is reasonable and captures essential properties of markets and respondents. Furthermore, the analysis of the conjoint data gives us estimates for some of the model parameters, e.g., a deviation probability for a respondent from his type of more than 10% seems to be overly pessimistic. It would be interesting to theoretically analyze our approach for the statistical model but this seems not an easy task since the entries in the matrix $B = AA^T$ are not independent.

[6] For our query strategy a lower bound on the number of respondents can be obtained from the connectivity threshold for random intersection graphs if we consider respondents connected if they had to answer one common question.

Acknowledgement. Research was partly supported by the Swiss National Science Foundation in the project "Robust Algorithms for Conjoint Analysis".

References

Chapelle, O., and Harchaoui, Z. 2005. A machine learning approach to conjoint analysis. *Advances in Neural Information Processing Systems* 257–264.

Critchlow, D. 1985. *Metric Methods for Analyzing Partially Ranked Data*, volume 34. Springer Verlag Berlin.

Feigin, P., and Cohen, A. 1978. On a model for concordance between judges. *Jour. Royal Statist. Soc.* 40:203–213.

Giesen, J.; Mitsche, D.; and Schuberth, E. 2007. Collaborative ranking: an aggregation algorithm for individuals' preference estimation. *Proceedings of the 3d International Conference on Algorithmic Aspects in Information and Management (AAIM), Lecture Notes in Computer Science*.

Gustafsson, A.; Herrmann, A.; and Huber, F. 2000. Conjoint analysis as an instrument of market research practice. *Conjoint Measurement. Methods and Applications. A. Gustafsson, A. Herrmann and F. Huber (editors)* 5–45.

Hüllermeier, E., and Fürnkranz, J. 2004. Comparison of ranking procedures in pairwise preference learning. *Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*.

Kendall, M. G. 1970. *Rank Correlation Methods*. 4th ed. Hafner, New York.

Lange, T.; Roth, V.; Braun, M.; and Buhmann, J. 2004. Stability-based validation of clustering solutions. *Neural Computation* 1299–1323.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of 5'th Berkeley Symposium on Mathematical Statistics and Probability* 281–297.

Mallows, C. L. 1957. Non-null ranking models I. *Biometrica* 44:130–140.

Meila, M., and Verma, D. A comparison of spectral clustering algorithms. *UW CSE Technical report 03-05-01*.

Sawtooth. Sawtooth software. *Research paper series at* `www.sawtoothsoftware.com`.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press.