Product Cannibalization and Synergy Estimation via MaxDiff Data

W. Michael Conklin and Stan S. Lipovetsky
GfK, 8401 Golden Valley Rd, Minneapolis, MN 55427

Abstract
Maximum Difference (MaxDiff) modeling is widely used for finding probabilities of choice among multiple items. We apply data and results of MaxDiff to solving another problem – of finding the products cannibalization and synergy. For each product we estimate its probability to be chosen as the best one within all the data, and also conditionally to each other product's presence or absence. For a given product, each other one behaves as a catalyzer or inhibitor of choice of the product in consideration. Constructing the entire matrix of such relations for all the products, we compare its symmetrical elements for each pair of products. This shows which of the pairs of products are mutually synergic, or complementary, so their chances to be chosen as the best ones are higher in the presence of each other. In other cases, the products can be of negative impact of one onto another, so one is a cannibalizer of another; or both products suppress each other. Various estimations on a real marketing data are considered, including the relative risk, *t*-statistics, and Shapley value for key driver analysis. Synergy relations among the products can also solve the bundle optimization problem for subsets of the products.

Key Words: MaxDiff, conditional choice probability, product cannibalization and synergy, game theory, Shapley value, key driver analysis, bundles optimization.

1. Motivation

Choice at the shelf can be complicated, and consumers likely use a number of simplification heuristics to determine which product they will choose. It is clear from experimental work that location on the shelf and the context of product adjacencies affects consumer choice.

But testing shelf layout and product adjacencies is expensive and difficult task. Is there a way to get some insight into which products are affecting each other (positively or negatively) from data we would normally collect for other purposes? The answer is yes, Max-Diff or Best Worst Scaling data.

Max-Diff or Best-Worst Scaling is a procedure commonly used in Marketing Research to identify a priority order of a set of items. It had been developed initially by Louviere (Best Worst Scaling: A model for the largest differences, 1991, *Working Paper, University of Alberta)* with roots in earlier work by Luce and Marley, and became extremely popular in Market Research with the introduction of software from Sawtooth Software for Hierarchical Bayesian (HB) estimation of the models. The procedure provides a convenient way to obtain priorities from respondents for a large number of items. Items are presented to respondents in blocks of 4 or 5 and respondents indicate which one is the best and which is the worst item in the set. Each respondent evaluates a number of blocks (typically 10 to 20). Block sizes and number of blocks are typically set so that each item is seen at least 3 times by each respondent. Data are evaluated commonly using a HB multinomial-logit (MNL) model.

For example, a respondent sees a number of screens like this one:

When buying an MRT system, which of the following benefits is most important and which one is least important for your purchase decision?

Most important	Attribute	Least important
X	Image Quality	
	Low Running Costs	X
	Low Purchase Price	
	Brand	

.
He/she selects the item that is “best” (most important) and “worst” (least important) in the set. A new set appears and the process repeats. Data is analyzed using HB-MNL model and the probability of choosing A as preferred to B is estimated.

Pros and Cons of this approach are as follows. It is an easy method to elicit overall preference structure for a large (10-20) number of items; ordinal judgments are more discriminating than traditional Likert type scales; probabilities of preferences for individual items are easily calculated; and HB technique provides estimates of preference structure at an individual level. On the other hand: ordinal judgments provide a relative measure of preference – most preferred item is not necessarily “good”; usually a fit is a main effects only model, there is no interaction effects; and HB modeling can be time consuming.

Recalling that we would like to understand which products should be adjacent to each other on the shelf, we could design an experiment in a virtual shopping exercise with different shelf configurations to explicitly answer this question, but it could be too expensive for implementation. We could try to fit a Mother Logit model to the data to account for all cross-effects, but the number of parameters makes the modeling difficult.

We are looking for a simple way to estimate these effects with the data we have, and we propose the following way to find Synergistic or Suppressor effects.

2. Conditional Probabilities for Product Synergy

Let us look at an example data set with 3062 respondents and Max-Diff exercise to prioritize 17 items. Respondents went through 10 tasks each where, in each task, they chose the best and worst item from a set of 4 of the 17 items. Design was balanced so that each item was seen an average of 2.35 times by each respondent. There were three different versions of the design so that for the overall sample the number of exposures of each item and each pair of items was balanced. In MaxDiff data pre-processing, we stack the data by task so that we have a final dataset with 3062*10 = 30,620 rows. Each row contains information on which items were shown in the task as well as which item was "best" and which was "worst".

We calculate Conditional Probabilities for each pair of items: the probability $P(A_i \mid B_j)$ of A being chosen as best when B is present, and the probability $P(A_i \mid \overline{B}_j)$ of A being chosen as best when B is not present; the overall probability of choosing A as best is always between these two quantities. Then the relative risk ratio can also be calculated as $RR_{ij} = P(A_i \mid B_j) / P(A_i \mid \overline{B}_j)$. If this quantity is greater than 1 then B is a catalyzer and if it is less than 1 then B a suppressor for A – see Table 1.

But, how to tell a big effect from a small effect: we can treat these two components of the RR, the conditional probabilities as estimates of choice from two non-overlapping samples. Therefore we can calculate a t-statistic for their difference:

$$t_{ij} = \frac{P(A_i \mid B_j) - P(A_i \mid \overline{B}_j)}{\sqrt{\left(\frac{P(A_i \mid B_j)(1 - P(A_i \mid B_j))N(B_j) + P(A_i \mid \overline{B}_j)(1 - P(A_i \mid \overline{B}_j))N(\overline{B}_j)}{N(B_j) + N(\overline{B}_j) - 2}\left(\frac{1}{N(B_j)} + \frac{1}{N(\overline{B}_j)}\right)\right)}}$$

The bigger positive t-values identify the catalysts for a product A from the other products B – see Table 2. For instance, for the 1st product, the products 2, 6, 8, 10, 11, and 14 are the catalysts.

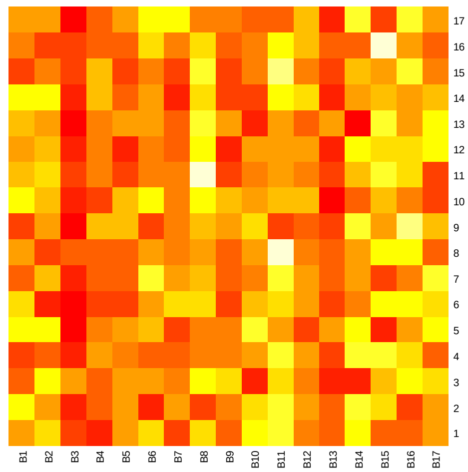
Table 1. Relative Risk values

	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}
RR_1	-	1.64	0.28	0.35	0.92	1.71	0.28	1.71	0.47	1.72	2.04	0.90	0.47	2.15	0.47	0.51	0.92
RR_2	2.11	-	0.27	0.58	1.09	0.41	1.01	0.58	0.91	1.63	2.62	1.01	0.58	2.62	1.63	0.28	1.09
RR_3	0.87	1.16	-	0.87	0.99	0.99	0.96	1.21	1.14	0.73	1.18	0.93	0.73	0.73	1.06	1.21	1.18
RR_4	0.68	0.73	0.34	-	0.83	0.73	0.83	0.83	0.83	1.08	2.04	1.15	0.68	2.04	2.04	1.38	0.71
RR_5	1.63	1.63	0.42	0.86	-	1.15	0.65	0.86	0.90	1.86	0.98	0.50	0.98	1.51	0.43	0.98	1.63
RR_6	1.67	0.28	0.21	0.31	0.49	-	1.86	1.67	0.33	1.39	1.86	1.10	0.31	0.97	2.32	2.32	1.86
RR_7	0.49	1.70	0.28	0.59	0.56	3.03	-	1.70	0.73	0.78	3.03	1.02	0.44	1.13	0.16	0.78	3.03
RR_8	1.32	0.40	0.36	0.22	0.22	1.32	0.54	-	0.22	1.32	5.17	0.73	0.25	0.99	2.54	2.71	0.56
RR_9	0.53	1.00	0.39	1.26	1.27	0.47	0.90	1.26	-	1.62	0.38	0.67	0.56	1.90	1.04	2.26	1.26
RR_{10}	1.95	1.29	0.16	0.56	1.28	1.88	0.89	2.04	1.29	-	1.36	1.36	0.22	0.63	1.29	0.89	0.34
RR_{11}	1.37	1.84	0.10	0.68	0.22	0.77	0.77	3.65	0.10	0.81	-	0.81	0.22	1.27	2.31	1.77	0.37
RR_{12}	1.01	1.17	0.31	0.94	0.31	0.87	0.73	1.62	0.39	1.08	1.08	-	0.39	1.79	1.59	1.32	1.79
RR_{13}	1.17	1.03	0.59	0.97	1.01	1.03	0.86	1.48	1.01	0.72	1.01	0.86	-	0.59	1.41	1.01	1.24
RR_{14}	1.72	1.72	0.35	1.17	0.70	1.10	0.41	1.49	0.64	0.67	1.58	1.49	0.35	-	1.17	0.97	1.28
RR_{15}	0.34	0.78	0.29	1.46	0.29	0.85	0.29	2.19	0.50	0.78	3.18	0.85	0.46	1.46	-	2.39	0.69
RR_{16}	0.80	0.47	0.47	0.75	0.64	1.61	0.77	1.52	0.53	0.77	1.69	1.23	0.64	0.53	2.69	-	0.53
RR_{17}	1.09	1.09	0.21	0.69	1.09	1.64	1.64	0.87	0.93	0.69	0.82	1.24	0.58	1.73	0.57	1.88	-

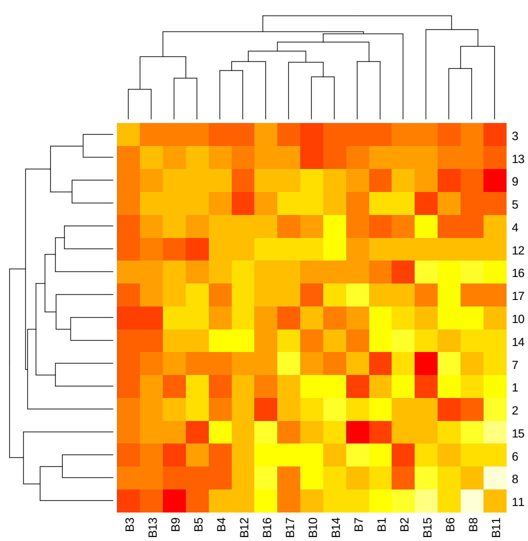
Table 2. t-statistics of two-sample proportions

	B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}	B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}
t_1	-	11.8	-14	-18	-1.4	10.9	-14	10.9	-9.7	13	17.7	-2.3	-9.7	16.9	-9.7	-8.8	-1.4
t_2	12.9	-	-14	-5.3	1.1	-11	0.1	-7.3	-1.4	6.8	15.7	0.1	-5.3	15.7	6.8	-9.5	1.1
t_3	-6.5	9.7	-	-6.5	-0.6	-0.9	-2.8	10.1	8.4	-14	8.4	-3.6	-14	-14	2.8	10.1	8.4
t_4	-11	-6.5	-17	-	-3.9	-6.5	-5.4	-3.9	-3.9	2.4	21.6	3.3	-10	21.6	21.6	10.4	-6.9
t_5	11.9	11.9	-19	-3	-	3.7	-10	-3	-2.7	16	-0.3	-11	-0.3	11.9	-13	-0.3	11.9
t_6	6.0	-11.	-13	-7.4	-7.4	-	7.5	6.0	-7.2	4.4	7.5	1.2	-7.4	-0.3	10.8	10.8	7.5
t_7	-9.5	11.1	-21	-10	-11	28.6	-	11.1	-6.4	-3.9	28.6	0.4	-11	2.1	-17	-3.9	28.6
t_8	2.9	-9.0	-6.8	-8.6	-8.6	2.9	-4.8	-	-8.6	2.9	26.5	-3.7	-8.1	-0.1	13.5	14.5	-6.3
t_9	-10	0.1	-19	5.0	6.5	-11	-2.6	5.0	-	11.5	-13	-6.9	-13	19.5	1.1	22.1	6.3
t_{10}	14.4	4.1	-14	-9.7	4.0	13.6	-1.6	13.5	4.1	-	5.0	5.0	-20	-7.9	4.1	-1.6	-11
t_{11}	2.9	5.1	-7.1	-2.3	-6	-1.7	-1.7	13.3	-7.1	-1.3	-	-1.3	-6	2.2	8.5	5.6	-6.8
t_{12}	0.2	3.8	-18	-1.4	-18	-3.9	-8.6	15.4	-15	1.7	1.7	-	-15	16	12.3	8.4	16
t_{13}	5.7	1.0	-16	-1.2	0.4	1.0	-5	16	0.5	-14	0.4	-5	-	-16	16.9	0.4	10.4
t_{14}	7.0	7.0	-7.6	1.8	-4.5	1.1	-6.9	4.9	-5.6	-5.1	6.9	4.9	-7.6	-	1.82	-0.3	3.6
t_{15}	-7.9	-2.5	-8.5	4.6	-8.5	-1.6	-8.5	12.4	-8	-2.5	19	-1.6	-8.9	4.6	-	13.9	-3.5
t_{16}	-2.3	-6.3	-6.3	-3.7	-4.1	6.2	-2.6	6.4	-5.5	-2.6	8.1	3.0	-4.1	-5.5	16.4	-	-5.5
t_{17}	1.1	1.1	-11	-3.9	1.1	6.8	6.8	-2.1	-1.1	-3.9	-2.8	2.7	-7.1	9.2	-5.4	9.2	-

In opposite cases, of RR<1 or t-statistics negative, we have B as the inhibitors for the choice of the product A. More analysis can be performed on this data. We can create a heat map to identify key pairs of items that have high synergy. Finding hotspots (white) where t-values are high is easy, however, note that the matrix is not symmetric, thus, the items may not be mutually synergistic.



For identifying the synergistic pairs we can multiply the relative risk matrix by its transpose and get a symmetric matrix. Sorting its rows with a hierarchical clustering algorithm gives us a better picture of items that can work together, and taking logs of the elements makes the picture even clearer. Because it is symmetric we only have to look in one half of the matrix.



An alternative method for finding specific items of interest can be found in the key driver analysis developed in (Conklin et al., 2004). Using the relative risk calculations we can estimate Youden's J which can be thought of as the difference between the two conditional probabilities $J = P(B \mid A) - P(B \mid \bar{A})$. We can identify the best items to combine with a target item by using the Shapley Value, a tool from the cooperation game theory. Shapley Value in this case does provide a convenient way to identify a set of items that should be adjacent to an item of interest.

3. Summary

We look for a convenient way of analyzing data we already collect to elicit preference orderings of items. The goal is to find items that will work well together in the marketplace, or conversely items that will be antagonistic towards each other. Key requirements are ease and quickness of calculation and no need to collect different or new data. Our approach calculates simple conditional probabilities from standard Max-Diff data and easily identifies key sets of items that are synergistic or cannibalistic.

For a product of interest, each other one behaves as a catalyzer or inhibitor of choice. Constructing the entire matrix of such relations for all the products, we compare its symmetrical elements for each pair of products. It shows are the pairs of products mutually synergic, or complementary, so their chances to be chosen as the best ones are higher in presence of each other. In other cases, the products can be of negative impact of one onto another, so one is a cannibalizer of another; or both products suppress each other. More material on the considered problem can be seen in (Lipovetsky and Conklin, 2014 a, b).

References

Conklin M., Powaga K., and Lipovetsky S. (2004) Customer Satisfaction Analysis: Identification of Key Drivers, *European Journal of Operational Research*, 154/3, 819-827.

Lipovetsky S. and Conklin M. (2014a) Best-Worst Scaling in Analytical Closed-Form Solution, *The Journal of Choice Modelling* 10, 60–68.

Lipovetsky S. and Conklin M. (2014b) Finding Items Cannibalization and Synergy by BWS Data, *The Journal of Choice Modelling*, forthcoming; DOI: 10.1016/j.jocm.2014.08.001.