

A DYNAMIC APPROACH TO FEATURE PRIORITIZATION

By Pankaj Kumar, Westley Ritz and Rajan Sambandam

White Paper Series



Have you had this problem? You need to measure preference for the features in your product/service but there are just so many of them it seems like an impossible task. Using conventional approaches you would have asked about the importance of each feature on a scale. But we all know how that story goes. With no other constraints, respondents don't have an incentive to say that anything is unimportant. You could use constraint-based methods like constant sum scales, but cannot realistically deal with more than a handful of features at a time. Over the last few years, the most popular technique to address this kind of feature prioritization problem has been Max-Diff (see white paper on Max-Diff). But using Max-Diff when there are more than a dozen attributes becomes a real chore. So what can you do when you have dozens of features that need to be efficiently culled? Let's first start with a look at a standard Max-Diff approach.

Standard Max-Diff

If there are ten features to prioritize, the Max-Diff algorithm can be set up such that respondents see grids of 3-5 features at a time, perhaps 8-10 grids in total. In each grid a respondent would indicate the feature that is most important (or some other relevant metric) and the one that is least important. At that point the respondent is done and the analysis of the data is conducted with Hierarchical Bayesian estimation to identify not only the rank ordering of the features but also the distances between them. The really neat outcome is that this information is available for each individual respondent, allowing further cutting and filtering of the results. But when the number of features to be tested increases, so does the number of grids. Since features have to be shown multiple times in this approach, the task quickly becomes monotonous and seemingly mindless for the respondent.

When you have 30 or more features, it is hard not to sympathize with the respondent's plight and at some point the quality of data obtained will come under question. So what can we do about that?

A New Approach

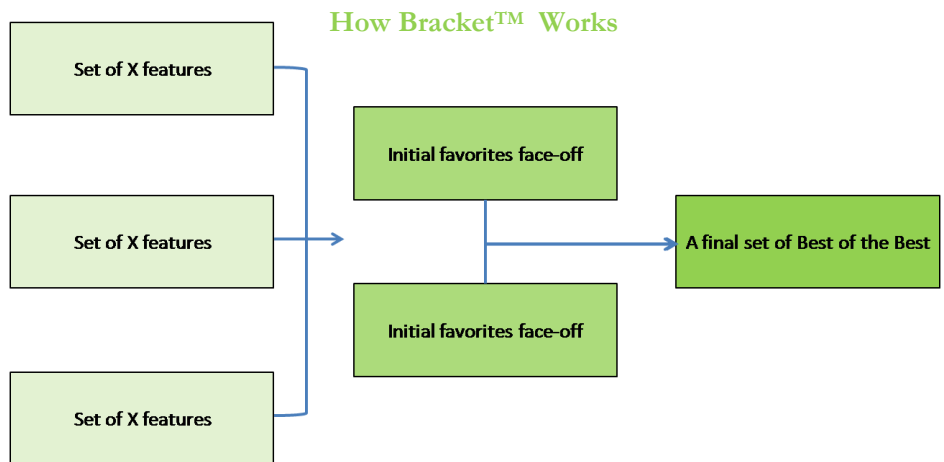
We use a dynamic approach (called Bracket™) that uses a tournament structure to successively eliminate the "losing" features, thus making the task more engaging and cognitively challenging. The first round is similar to a Max-Diff task in that respondents will see a series of grids with a few features in each one and indicate their preference. The losing features (in this case those that are not preferred) fall away and the winners live to compete in the next round and so on, till we narrow the features down to each respondent's final set.

There are several advantages to this approach. Most important is that it

regular Max-Diff task. Since only the preferred features move through the rounds of the tournament, the task becomes more engaging and difficult (in a good way) encouraging a more cognitive response than usual. Lastly, the most important features for each respondent are dynamically identified in real time allowing follow-up questions to be seamlessly asked in the same survey. Sounds like a winner, but how do we know that we are actually getting good quality data on the back end? We ran tests to confirm this and here is what we found.

A Bracket™ Example

The subject in this case was how movie-goers make decisions about which movie to see and where to see it. One can imagine many such factors: the stars, the director, the theater location, the show timing, etc. We imagined 18 of them and constructed



allows us to handle large numbers of features (upwards of 50) without running into respondent fatigue as the elimination process whittles down unimportant features very rapidly. What is also significant is that the process is more engaging to respondents as they don't have to keep choosing between features they have no interest in, as in a

study where one cell of respondents was provided a standard Max-Diff task, while another cell was provided a Bracket™ task. Why choose only 18 features? We could have gone with, say, 30 or 40 features. But to make the methods comparable we needed a number that both methods could handle comfortably. The higher we

went, the more the test would have been unfair to Max-Diff, so we chose to go with something in the neighborhood of 20.

In the Max-Diff case respondents saw 14 grids of four features and picked the features that were most (and least) important to them in choosing to go to a movie. In the Bracket™ case, respondents saw six grids of three features in the first round, two grids of three features in the second round and one grid in the third round. For validation purposes we asked four sets of choice questions set up like conjoint tasks. In each set respondents saw three profiles comprising the features tested in the study. They were asked which profile they would pick. So a person putting a lot of weight on a movie with their favorite stars would be expected to be more likely to pick the profile that offered that option. That was it for the data collection part.

Both datasets were analyzed using Hierarchical Bayesian estimation. The analysis is pretty standard in the Max-Diff case but is more nuanced in the case of Bracket™. Nevertheless, feature preferences for individual respondents can be obtained in both cases. The question then is how do they compare, and that is where the validation tasks come into play.

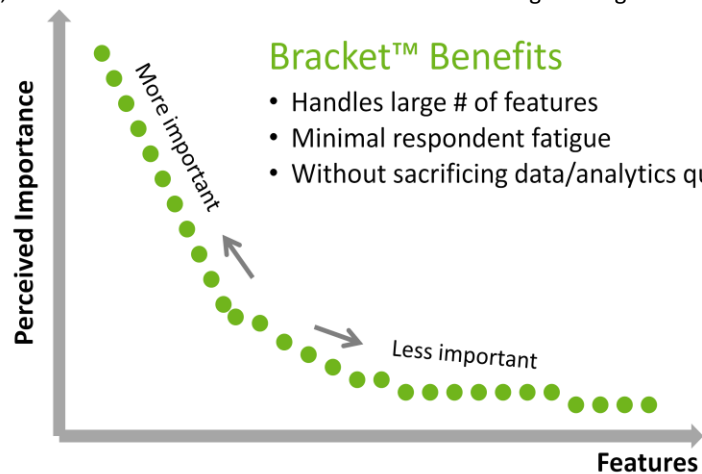
Validation

We know which profile options respondents picked in the holdout validation tasks in the study. Based on the individual level utilities (preferences for features) that we got from the analysis we can also make a prediction of which options they would pick. Comparing the two tells how well we did in estimating respondent preferences. So what did we find? Before we get there, it should be noted that in these kinds of comparisons Max-Diff should be expected to do better as it gathers more data. Each respondent in the Max-Diff cell is providing preference information on each feature approximately three times, whereas in the Bracket™ cell the winning features get a second and perhaps third look but a good number don't get more than one look. So, in

theory, it should be expected that Max-Diff will provide more accurate preference scores and thus validation. What we found surprised us. The validation hit rate for Max-Diff was 47% and for Bracket™ it was 45% (these numbers are in the expected range given the type of validation task and the domain). In other words, there was almost no difference between the more rigorous Max-Diff approach and the less rigorous, but more engaging, Bracket™ approach. How could this be? We further investigated the issue by examining carefully the information from each round of Bracket™ and found that information respondents provided in the very first round were highly useful. Think about that for a second. In the first round, respondents see each feature just once in random groupings of three. But the information they provide there, combined with the Hierarchical

test and we are hard pressed to argue against that.

We have applied this approach to several studies with feature sets varying from the teens to the fifties and have found it possible to elicit individual level preferences while keeping respondents engaged. We don't see why even larger feature sets cannot be practically prioritized. Another advantage we have learned by applying Bracket™ is that the tournament-based approach is particularly well suited in situations where the features are not well distinguished from each other. In such cases standard Max-Diff tasks become even more difficult for respondents, whereas Bracket™ tasks are well suited to help respondents focus on the most important ones. Consider message testing where many of the messages being tested may be only



Bayesian analysis is capable of producing a very robust foundation on which the next two rounds can sit. This tells us that Bracket™ is capable of producing very nearly Max-Diff type information content.

As mentioned before, we deliberately set the test up to be fair to Max-Diff and therefore chose to use only 18 features. Our contention is that when there are far more features Bracket™ will be the only useable method as Max-Diff would be too tedious. This validation study, well, validates that notion by establishing the robustness of Bracket™. However, it also raises the question of whether Bracket™ could be used in cases where there are fewer features to

subtly different from each other. As we found out in a test, regular Max-Diff can lead to a confusing mess, whereas Bracket™ can provide quite clear results.

In Conclusion

Feature prioritization is a very common marketing research problem. However, as the number of features increases into the teens and beyond it becomes difficult to use state-of-the-art methods like Max-Diff without substantially increasing the tedium quotient of the study. Bracket™ is a tournament-based approach that produces Max-Diff like results and can easily prioritize fifty or more features.