Sawtooth Software

RESEARCH PAPER SERIES

Political Landscape 2008: Segmentation Using MaxDiff and Cluster Ensemble Analysis

Bryan Orme & Chris King,
Sawtooth Software, Inc.

530 W. Fir St.
Sequim, WA 98382
(360) 681-2300
www.sawtoothsoftware.com

Political Landscape 2008:
Segmentation Using MaxDiff and Cluster Ensemble Analysis

September, 2008

Bryan Orme and Chris King, Sawtooth Software, Inc.[1]

This article is scheduled to be published in MRA's Alert Magazine, November 2008 as "Political Landscape 2008: Using Best-Worst Analysis for Measurement and Segmentation." The article as represented here has been renamed and slightly expanded for the Sawtooth Software audience.

Introduction

Regarding the Model T, Henry Ford famously wrote in his autobiography, "Any customer can have a car painted any colour that he wants so long as it is black." Well, Henry Ford's one-size-fits all approach wouldn't cut it today, either in business or politics.

If you've attended a marketing management course recently, you no doubt learned about targeting the market via segmentation and mass customization. Market segmentation involves the process of identifying and customizing a message or offering for key groups of people. Mass customization (often facilitated by the Internet) is the same approach applied to individuals.

Political consultants employ tools commonly associated with marketing. Identifying the attitudes, needs, and demographic profiles for target segments of constituents gives candidates the ability to motivate them with customized messages. And, beyond targeting broad segments of the population, microtargeting combines standard demographic and attitudinal information with voter records and even brand preference information to target very narrow segments or even individuals within the population. Once identified, these microsegments can be reached with customized mailing or phone calls, or even personal visits by an ever-growing corps of grass-roots volunteers.

The easiest approaches to segmentation involve categorizing people based on characteristics (variables) such as gender, income, education, and party registration. Thus, we could refer to a segment of the population as young, female, highly-educated, democrats. Such approaches are useful, but limiting, since they say nothing about a person's attitudes and preferences.

Attitudes, motivations, and preferences are not so easily classified as gender, age, or income—but they are extremely valuable for political segmentation and targeting strategies. To learn about these variables, researchers typically use opinion surveys and advanced statistical methods.

[1] The authors thank John Fiedler (Oreon, Inc.) and Rich Johnson (Chairman, Sawtooth Software) for their help in designing the items used in the questionnaire and for their helpful comments on the drafts of this article.

How to Ask about Preferences and Importances

For decades, researchers have struggled with how to ask respondents how much they like something or how important it is. A common approach is the 5-point scale.

The problem with asking the 5-point scale is that folks tend to say most everything is either very or extremely important, so there is little discrimination among the ratings. Even more tricky, people don't necessarily use scales in the same way, so a "3" for one person might mean the same as a "1" to another (this is called *scale use bias*). Finally, some respondents may find it very abstract or confusing when asked to map their internal feelings to an artificial rating scale.

These problems with the standard rating scale often lead to bar-chart displays showing relatively small differences among the items, and few meaningful differences across segments of the population. These outcomes don't bode well for developing insightful conclusions and correct strategy. And, if the data are submitted to advanced statistical methods, the resulting segments are not as meaningful and stable as they might be.

There are a few good approaches to asking people about preferences and importances that significantly reduce or even resolve the problems we've been discussing. One of the best new and emerging questioning approaches is called *best-worst measurement* (also known as MaxDiff).

In best-worst questions, we show a short list of brands or other items, and ask respondents which of these is the best and worst (or most and least important, etc.). For example:

Figure 1: Sample Best-Worst Question

Consider a candidate for President of the United States with the following positions.

Which position would make you **most likely** to vote for this candidate, and which would make you **least likely** to vote for this candidate?

Most Likely	Least Likely	**This candidate promises to...**
○	○	Enact policies to solve housing/mortgage crisis
○	○	Restrict gun ownership
○	○	Reduce our reliance on foreign oil imports
○	○	Increase worldwide humanitarian efforts
○	○	Increase defense/military spending

This type of question has the advantage that folks are not required to map their attitudes onto a rating scale; rather, they are just asked to choose. And, people are excellent at making choices—we do it all day long. Furthermore, this type of question leads to greater discrimination, since we don't allow people to tell us that everything is desirable or important (as is so easily done using a rating scale).

The key to making this line of questioning work is that people answer a series of questions (usually 12 to 18) just like that shown in Figure 1. But, in each question, a different combination of items is shown. So, across the full set of questions, each item will have been shown to respondents multiple times.

The analysis is not as straightforward as simply reporting a mean rating for an attitudinal statement across a sample of respondents. But, one can use hierarchical Bayes analysis (HB) to estimate a score for each respondent indicating how preferred (or important) each item is. The scores may be rescaled to values that sum to 100, so they can be very easy to work with. Obviously, the more times an item is chosen as best in a set and the fewer times it is tagged as worst, the higher its resulting score will be.

Segmenting Respondents Based on Attitudes and Preferences

The most popular statistical tool for segmenting respondents based on attitudes and preferences has been *Cluster Analysis*. Cluster analysis finds groups of respondents such that the people within each group are quite similar and the people in different groups are quite different. This all sounds well and good, but there are a few well-known problems, including:

1. Cluster analysis always gives you an answer, even if people really don't naturally fall into distinct groups.
2. The segmentation solution can vary quite a bit depending on the settings the analyst uses (and the choice of clustering methods, of which there are many!).
3. It isn't always very easy to decide whether a 3-group solution is more compelling (or justified by the data), than, say, a 4-group solution.
4. Interpreting and choosing among candidate segmentations is as much art as science. There is no one agreed-upon fit statistic that indicates success or failure, or which one solution is the best.

Not surprisingly, all of these problems are reduced (but not eliminated) if one segments the market using better data (such as provided by best-worst questions).

One of the newest approaches we know about comes out of the machine learning discipline, and it is called *Cluster Ensemble Analysis*. In 2008, we released a software system that does both K-means cluster analysis and cluster ensemble analysis. That system is called CCEA (Convergent Cluster & Ensemble Analysis). The academic literature (as well as our experience) suggests that cluster ensemble analysis provides more accurate and stable classification of respondents to segments than standard cluster methods (Strehl and Ghosh 2002, Retzer and Shan 2007, Orme and Johnson 2008). As for what it does, it looks at a variety of cluster solutions (an *ensemble* of segmentations) and develops a representative *consensus* solution across the ensemble. It turns out that the consensus solution usually has stronger characteristics than any one of the separate cluster solutions within the ensemble.

Presidential Politics 2008: An Example

To illustrate the use of best-worst analysis and cluster ensemble segmentation, we'll show some results from a recent opinion poll we conducted in August, 2008. The data were collected just prior to John McCain and Barack Obama choosing their running mates, and just prior to the party conventions. Please note that we are not claiming that our example demonstrates the absolute *one* best way to conduct such a study. Rather, we collected the data to illustrate the points in this article and to stimulate thought.

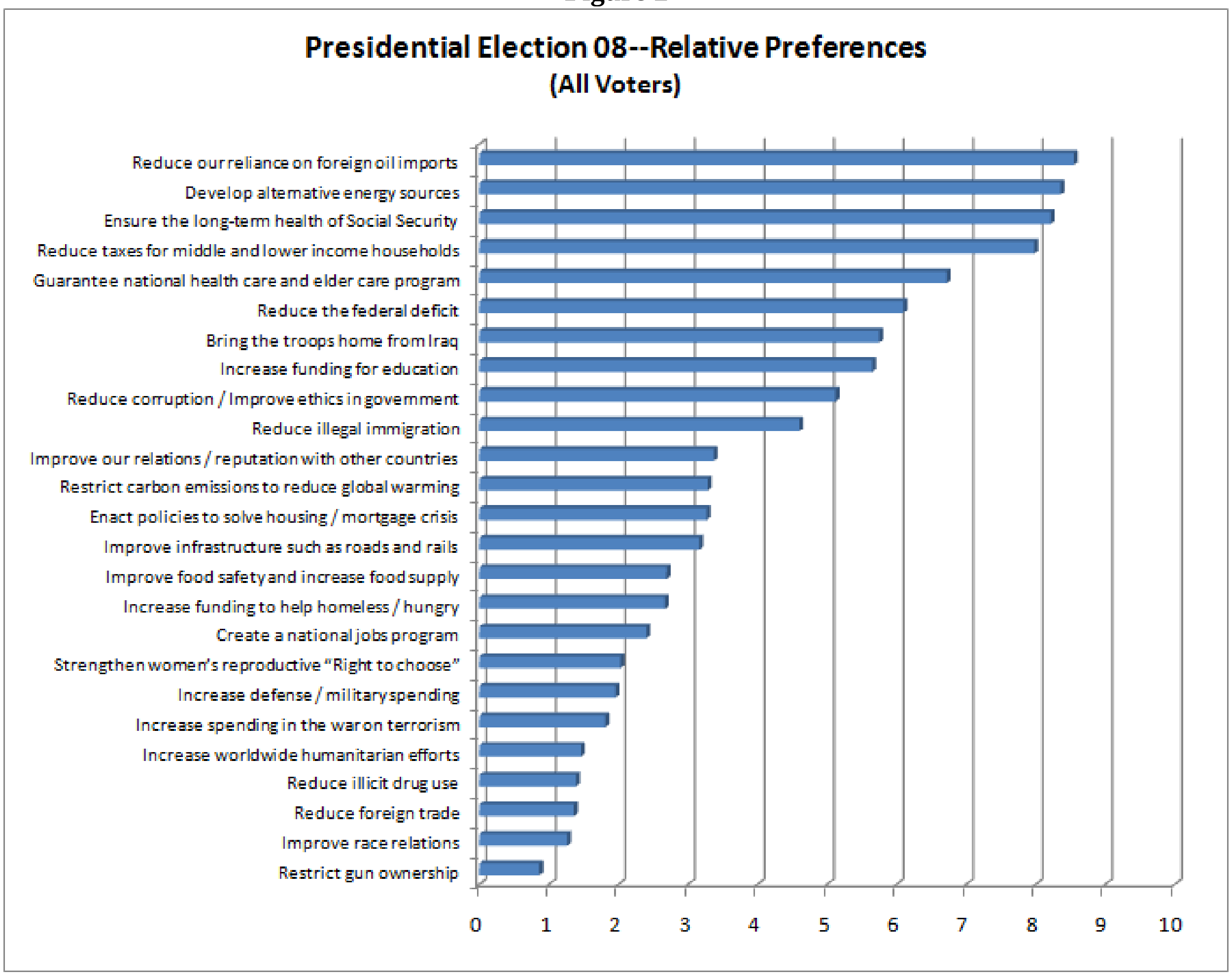
We asked people best-worst questions (see Figure 1) about the positions a presidential candidate might take on 25 issues (see Figure 2). Specifically, we wanted to know which positions would make people most want to vote for a candidate.

Respondents were recruited using the e-Rewards online consumer panel (www.e-rewards.com). (We chose e-Rewards because it was the most used panel provider among Sawtooth Software users based on our 2008 feedback survey. We are grateful to e-Rewards for providing the sample for this research and for their excellent project support.)

We asked respondents for which presidential candidate they would most likely vote. Among our online sample, the choices were: Obama 49%, McCain 40%, and Other 11%. (Note that this is quite close to a consensus of recent national opinion polls reported at www.realclearpolitics.com, which at the time of writing showed: Obama 47%, McCain 43%, and Other 10%.) We also asked respondents various demographics, including party affiliation.

The preference scores from best-worst analysis for the 25 items across the entire sample are shown in Figure 2, sorted from most to least preferred.

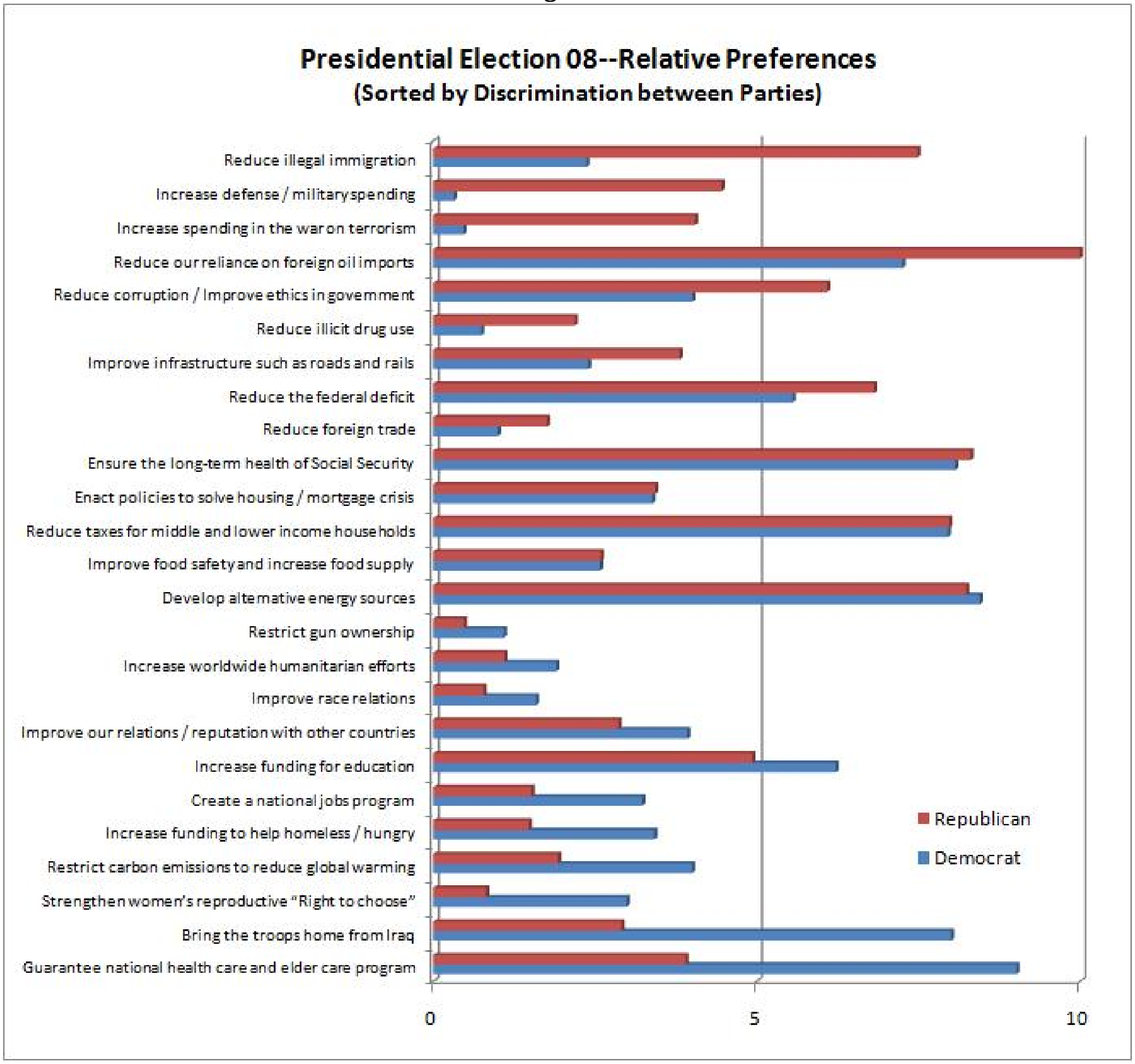
Figure 2



Viewing the top preferences, it is clear that the price of gas and the weakened US economy has had a huge impact on the national conscience. Candidates would do well to emphasize their energy policies.

The preferences for policies can vary strongly depending on the respondent's party affiliation. Figure 3 reports the preference scores by declared party affiliation, sorted by the difference between party preferences. Positions at the top of the list are those which most uniquely define Republicans' preferences (relative to Democrats). Those on the bottom most uniquely define Democrats' preferences.

Figure 3



What separates Republicans most from Democrats is preference for reducing illegal immigration, and increasing spending on both defense and the war on terrorism. What defines Democrats most (relative to Republicans) is preference for guaranteeing a national health care program, bringing the troops home from Iraq, and desire to strengthen women's reproductive "right to choose." Policies in the middle of the chart reflect those on which both parties agree. Overall, we see the best-worst questioning technique has done a fine job at eliciting discriminating preferences.

Segmentation Based on Preferences and Demographics

We used Cluster Ensemble Analysis to segment respondents based on the best-worst preference scores for our list of 25 position statements. For the purposes of this article, we decided to choose a number of segments (six) that would be both robust in terms of reproducibility and compact to display.

We have limited space in this article to discuss the details of these segments. But, they differed substantially on the position statements, preference for presidential candidate, as well as the demographics, as shown in Figure 4. We've reported some demographic characteristics of each group as well as a heat map showing the relative preferences for the twelve candidate positions that discriminated most among these six groups. The depth of the color reflects absolute preference for each position.

Figure 4: Segmentation Summary

Segment#	I	II	III	IV	V	VI
Percent of voters	18%	15%	24%	8%	22%	14%
Voter preference: Obama	81%	73%	68%	52%	14%	2%
Voter preference: McCain	11%	19%	17%	38%	73%	90%
Commitment to candidate	++	–		+		++
Age				–	+	++
Affluence			–	+	+	++
Education	++		–	+		+
Religious service attendance	–	–			+	++
Increase defense/military spending						
Increase spending in the war on terrorism						
Reduce illegal immigration						
Guarantee national health care and elder care program						
Restrict carbon emissions to reduce global warming						
Bring the troops home from Iraq						
Improve our relations / reputation with other countries						
Reduce our reliance on foreign oil imports						
Create a national jobs program						
Reduce taxes for middle and lower income households						
Increase funding to help homeless / hungry						
Increase worldwide humanitarian efforts						

It's common practice to assign labels to the segments that are somewhat descriptive of the group as a whole. There were other demographic variables not shown in Figure 4 that we can also draw upon to develop appropriate labels. One possible set of descriptors is:

Segment I: Progressive Intellectuals
Segment II: Suburban Progressives
Segment III: New "New Dealers"
Segment IV: Young Independents
Segment V: Moral Middle Class
Segment VI: Conservative Right

Admittedly, dividing the country into just six segments and affixing labels involves a healthy amount of generalization and fiction. Many respondents would likely dispute that they fit the descriptor assigned to their group.

Latent Class vs. Cluster Ensemble Analysis

Those familiar with Sawtooth Software's white papers know that we typically advocate the use of latent class for discovering segments within choice data (such as CBC or MaxDiff). In this article, we have employed cluster ensemble analysis operating on the scores developed in a first-stage HB analysis. There are weaknesses to this two-stage approach. That said, we were interested in how well cluster ensemble analysis (a new tool to us) could perform with such a dataset. In addition to the cluster ensemble solution reported above, we examined solutions that leveraged both demographic segmentations and preference segmentations within the ensemble. We expect that the strength and flexibility of cluster ensemble analysis may provide tempting opportunities to apply it instead of latent class analysis for some problems involving choice data. In those instances, we recommend that there be a relatively large amount of information in the questionnaire relative to the number of parameters to be estimated. In our case, we displayed each item 3x for each respondent within the MaxDiff exercise. We would hesitate to shorten the questionnaire, and in hindsight it might be have been even better to include each item 4x within the questionnaire. With more information at the individual level, there would have been even less smoothing to population parameters via HB and probably even greater discrimination among the item scores.

Next Steps and Conclusion

In the real world, we would have collected a larger sample size so we could drill down deeper (more segments, as we saw strong cluster stability out to about 10 segments) and draw even stronger conclusions regarding each group. This would reduce the problems of generalization mentioned earlier. It would be natural to develop an entire cluster segmentation focused on those voters who are undecided or not very committed. Fence-sitter segments in battleground states could be targeted with the messages the candidate can espouse and demonstrate high credibility.

In summary, target strategies are a must for both politics and marketing. And, the questioning approach and statistical methods chosen can make a big difference on the quality of the results.

References:

Orme, Bryan and Rich Johnson (2008), "Improving K-Means Cluster Analysis: Ensemble Analysis instead of Highest Reproducibility Replicates," Sawtooth Software Research Paper, available at www.sawtoothsoftware.com.

Retzer, J. and M. Shan (2007), "Cluster Ensemble Analysis and Graphical Depiction of Cluster Partitions," Proceedings of the 2007 Sawtooth Software Conference, Sequim WA.

Strehl, A. and J. Ghosh (2002), "Cluster Ensembles — A Knowledge Reuse Framework for Combining Multiple Partitions," Journal on Machine Learning Research (JMLR), 3:583-617, December 2002.