

Extracting Policy Positions from Political Texts Using Words as Data

Results of Wordscoring Reviews and Editors' letters

Kenneth Benoit

As an additional demonstration of the applicability of the word scoring technique to areas other than party manifestos, we decided to test it on the texts of the correspondence concerning our paper and its review by the *APSR*. The steps were as follows:

1. Using a scanner, convert the printed referee reports and editor's letters from both rounds of reviews into text files.
2. Generate a matrix of word counts from these texts, which were unusually long for reviews and editors letters. Our wordcount program reports the following:

Text	Ref Score	Total Words	Unique Words	Mean Freq.	Median Freq.
ed_1	3	1,660	624	2.66	1.00
R1_1	.	1,155	446	2.59	1.00
R2_1	.	3,081	882	3.49	1.00
R3_1	.	820	336	2.44	1.00
ed_2	2	1,157	468	2.47	1.00
R1_2	.	78	58	1.34	1.00
R2_2	.	1,704	563	3.03	1.00
R3_2	.	446	239	1.87	1.00

As can be seen from the total word column, some reports, notably R1's second report, were extremely brief.

3. We generated a vector of word scores based on the following notional metric. Considering a metric of 1 to 4, where 1 indicate accept, and 4 indicates rejection, and 2 and 3 are categories in between of revise and resubmit, with a 2 being more positive than 3. Because the style and approach of the reviewers' letters were quite different, we chose the two rounds of editor's letters as the reference texts. Based on our reading of the first and second editor's letters, we assigned a value of 3 to the first letter and a 2 to the second letter. We then used these two reference texts to generate the word scores needed by our procedure.
4. Our goal then was to score each referee's letter, testing to see how each changed between the first and second round of reviews. Before running the analysis, based on our reading of these letters, we expected R1 to remain the same (positive), R2 to have shifted from less to more positive, and R3 to remain the same (negative).

We focus here only on relative shifts in the same reviewer's position from round 1 to round 2, since the lexical differences in reviewer's letters and their clear differences in style (R2 for instance) should caution us against expecting to compare the different referees against one another.

Results:

Virgin Text		Raw Score	Raw SE	Unique Scored Words
R1_1		2.5179	0.0091	196
R2_1		2.4996	0.0057	319
R3_1		2.5043	0.0099	162
R1_2		2.4579	0.0343	42
R2_2		2.4718	0.0075	264
R3_2		2.5043	0.0140	134

Interpreting this, we see:

- a) **R1:** (with a very short and therefore relatively uncertain score for the second review) that the position of this reviewer cannot be said to have changed, based on the point estimate and 95% confidence interval of 2.52 [2.50, 2.54] from round 1, compared to 2.46 [2.39, 2.53], although the numeric value indicates a slight shift to an even more positive evaluation.
- b) **R2:** The shift here from 2.50 [2.49, 2.51] to 2.47 [2.46, 2.49] is a statistically significant shift to a more positive opinion, reflecting what is in fact clearly stated in R2's report!
- c) **R3:** The value from round 1 2.50 [2.48, 2.52] does not change in round 2 2.50 [2.48, 2.53], just as this reviewer's opinion remains negative.

Verdict: A qualified success!!