

Computerized Text Analysis: Assignment 8

Wordscores

Kenneth Benoit

The objective of this class exercise is to better understand the Wordscores text scaling algorithm, using examples from Laver, Benoit and Garry (2003).

You will need some additional software to implement this example. We recommend that you use R for this, along with the the Austin library (see <http://www.williamlowe.net/software/#austin>). It is also possible to use Stata, with the Wordscores library (see <http://www.politics.tcd.ie/wordscores/>), although this code is badly needs updating. Alternatively, you can program the algorithm yourself in a spreadsheet such as Microsoft Excel.

We will use two sets of files:

- The example from Table 1 of LBG (2003). This data is built in to the `austin` library in R so if you are using R to do this, then you will not need to load this file in at all. If you are using some other software, then you can download the file `LBGexample.csv`. This file is in .csv (comma separated value) format and can be loaded directly into Stata or your spreadsheet.
- The Irish 2010 budget speeches data, available as the file `budget2010.csv`. The R file `exercise8.R` has all of the commands you will need to implement the steps outlined below, including the installation of the `austin` library.

Instructions:

1. Start the R program.
2. Open the file `exercise8.R` in the R text file editor.
3. Install the `austin` library as per the first command, and make sure this succeeds.
4. Estimate the wordscores model for the LBG (2003) example, and inspect the results. Follow the code for precise instructions. Here you will be using the reference scores set at -1.50, -0.75, 0.00, 0.75, and 1.50 for reference texts r_1 through r_5 respectively. Score the virgin text and compare your results to LBG (2003) Table 1.
5. Run the wordscores scaling procedure on the Irish 2010 Budget speeches. Here we will use the 5th text (Cowen, the FF Prime Minister) as one reference text, and the 6th text (Kenny, the FG opposition leader). We will score all words in the Cowen and Kenny texts, and then score all texts as if they were purely “virgin” documents.
6. Inspect the word scores as shown in the .R file. Compare these to your dictionary words from Assignment 7.