

Computerized Text Analysis: Classwork 6

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 software to implement this example. Choices are Stata, using the Wordscores library (see <http://www.politics.tcd.ie/wordscores/>), or R using the Austin library (see <http://www.williamlowe.net/software/#austin>). Alternatively, you can program the algorithm yourself in a spreadsheet such as Microsoft Excel.

We will use two sets of files for this:

- The example from Table 1 of LBG (2003), available as [LBGexample.csv](#). This file is in .csv (comma separated value) format and can be loaded directly into Stata, R, or your spreadsheet.
- The Irish budget speeches from Exercise 5.

Instructions:

1. Load the `LBGexample.csv` into your numerical analysis program. Inspect the “word” frequency matrix.
2. Compute the wordscores using the procedure from LBG (2003). The reference scores should be set at -1.50, -0.75, 0.00, 0.75, and 1.50 for reference texts r_1 through r_5 respectively.
3. Score the virgin text and compare your results to LBG (2003) Table 1.
4. Now score the words with just two reference texts, the extremes, using -1.50 and 1.50 respectively. Compare the scores of the words from the previous scoring. Score the Virgin text and compare to the earlier result.
5. Run the wordscores scaling procedure on the Irish 2010 Budget speeches from Assignment 4. If you do this in the Stata library, then you can accomplish this analysis with the following commands — but first make sure your current working directory in Stata contains the files for the budget speeches.

```
sysdir set PLUS m:/
net install http://www.tcd.ie/Political_Science/wordscores/wordscores
wordfreqj *txt
setref t03_burton_lab -1 t05_cowen_ff 1 t01_lenihan_ff 1
describetext t*
wordscore proanti
textscore proanti t*
```

If you choose to do this exercise in R, then the equivalent commands will be the following:

```
install.packages("austin", repos="http://R-Forge.R-project.org",  
                dependencies="Depends", type="source")  
wf <- wordfreqj("*txt")  
ws <- classic.wordscores(wf[,c(9,2,14)], scores=c(-1,1,1))  
summary(ws)  
predict(ws, newdata=getdocs(wf))
```