Content Analysis and Word Scoring

Trinity Term 2016

Andrew Eggers
David Doyle

Course description:

Text is everywhere, and social scientists increasingly use statistics and algorithms to interpret it. We will examine two common approaches to analyzing "text as data" in political science: text scaling, which attempts to characterize the content of documents or speakers in terms of ideology, and text classification, which attempts to organize texts into categories. Our focus will be on how to design and interpret research using these methods, although we will point students toward resources that will help them acquire and process text for their own analysis.

Software:

We will be discussing techniques for text analysis that can be implemented in the R environment (http://www.r-project.org). We strongly suggest that students who are interested in applying these methods acquire a basic competency in R before the course.

R costs nothing to acquire and use. For those just getting started, we recommend RStudio (http://www.rstudio.com/products/rstudio/download/), which provides a congenial user interface for operating in R.

To learn R, we recommend these tutorials:
http://tryr.codeschool.com/ (Online: does not require you to install R or RStudio)
http://www.rstudio.com/resources/training/onlinelearning/#R
There are many more to be found.

Assessment:

Students taking the course for credit will submit an essay of not more than 2500 words, due at noon on Friday of 6th week (5 June) in the Courses of Office of DPIR.

We welcome two kinds of essays:
a) A critical review of 2-5 papers using text as data (possibly including replication)
b) An original analysis using methods from the course

There will also be worksheets designed to help you develop practical skills with text.

Course outline:

Week 1: Turning text into data

Core reading:

Justin Grimmer and Brandon Stewart, "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts" *Political Analysis* (2013) pp. 1-31.

Application readings: (In reading these, focus on understanding how the authors use text and how this fits into their research question.)

Andrew C. Eggers and Jens Hainmueller, "MPs for Sale? Returns to Office in Postwar British Politics" *American Political Science Review (2009).*

Scott R. Baker, Nicholas Bloom, and Steven J. Davis, "Measuring Economic Policy Uncertainty", unpublished manuscript (2013), available at http://www.policyuncertainty.com/media/BakerBloomDavis.pdf

Ricardo Puglisi and James M. Snyder, "Newspaper Coverage of Political Scandals", *Journal of Politics* (2011) 73(3): 931-950.

Matthew Gentzkow, Edward L. Glaeser, and Claudia Goldin, "The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered", in *Corruption and Reform: Lessons from America's Economic History* (2006) Edward L. Glaeser and Claudia Goldin, eds. Available at http://www.nber.org/chapters/c9984.pdf

Worksheet 1: Accessing text and counting words

Day 2: Classification, clustering, and topic modeling

Core reading:

Kevin Quinn, Burt L. Monroe, Michael Colaresi, Michael Crespin, and Dragomir Radev, "How to analyze political attention with minimal assumptions and costs", *American Journal of Political Science* (2010), 54:1 209-228.

Application readings:

Lucas et al, “Computer-Assisted Text Analysis for Comparative Politics“ *Political Analysis* (2015) pp. 1-24.

Amy Catalinac, “Pork to Policy: The Rise of National Security in Elections in Japan” (2014), unpubslished manuscript, available at http://scholar.harvard.edu/files/amycatalinac/files/catalinac_nov2014.pdf

Worksheet 2: Creating the term-document matrix and measuring similarity between documents

Day 3: Text scaling I - Introduction to Scaling Methods

Core reading:

Laver, Michael & John Garry. 2000. “Estimating Policy Positions from Political Texts.” *American Journal of Political Science*, 44 (3). pp. 619–634.

Application readings:

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge & Michael McDonald. 2006. *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990-2003*. Oxford: Oxford University Press. (Have a look at the introduction only)

Groseclose, Tim and Jeffrey Milo. 2005. “A Measure of Media Bias.” *Quarterly Journal of Economics*. 120 (4), pp. 1191-1237.

Mikhaylov, Slava, Laver, Michael and Benoit, Kenneth. 2012. “Coder reliability and misclassification in the human coding of party manifestos”. *Political Analysis*, 20 (1). pp. 78-91.

Dinas, Elias and Kostas Gemenis. 2010. “Measuring Parties’ Ideological Positions With Manifesto Data: A Critical Evaluation of the Competing Methods.” *Party Politics.* 16(4), pp. 427-450.

Software:

Yoshikoder at http://www.yoshikoder.org

Day 4: Text scaling II - Automated Document Scaling

Core readings:

Laver, Michael, Kenneth Benoit & John Garry. 2003. “Estimating the policy positions of political actors using words as data.” *American Political Science Review*. 97(2), pp. 311–331.

Slapin, Jonathan & Sven-Oliver Proksch. 2008. “A Scaling Model for Estimating Time Series Policy Positions from Texts.” *American Journal of Political Science*, 52(8), pp. 705-722.

Application readings:

Benoit, Kenneth, Conway, Drew, Lauderdale, Benjamin E., Laver, Michael and Mikhaylov, Slava (2015) *Crowd-sourced text analysis: reproducible and agile production of political data.* American Political Science Review.

Lowe, Will and Ken Benoit. 2013. “Validating Estimates of Latent Traits from Textual Data Using Human Judgment as a Benchmark.” *Political Analysis* 21(3), pp. 298-313.

Proksch, Sven-Oliver and Jonathan B. Slapin. 2010. "Position Taking in European Parliament Speeches", British Journal of Political Science 40(3), pp. 587-611.

Proksch, Sven-Oliver and Slapin, Jonathan. 2009. “How to Avoid Pitfalls in Statistical Analysis of Political Texts: the Case of Germany.” *German Politics* 18(3), pp. 323–344.

Arnold, Chris, David Doyle and Nina Wiesehomeier. “Presidents, Policy Compromise and Legislative Success.” 2015. unpublished manuscript.

Software:

Wordscores at http://wordscores.com

Wordfish at http://wordfish.org