Sensitivity Analysis in Automated Content Analysis Using a Political Party Platform Example

Stephen Purpura   John F. Kennedy School of Government, Harvard University

Three Critical Questions:

❖ Given training data and a test set with abortion references intact, what are three algorithms' estimates for attention to abortion in the 2004 Party Platforms?

❖ Removing the abortion sentences from the test set, what is each algorithm's estimate?

❖ Given answers of preceding, which automated method is closer to the "expert" estimate?

Acknowledgements

**Program on**
Networked Governance

Abstract

As documents become increasingly available in digital form, the use of computer-assisted content analysis is a natural fit. Humans typically review each unit of analysis in context within a document, but popular computer algorithms usually do not. *The resulting problem is a type of bias.* While a particular algorithm may seem to work well at the macro level of sorting concepts into categories, it may not work well for a specific concept of interest. Measuring the sensitivity of the algorithm is important for assessing validity, thus *concept sensitivity analysis* fills the gap.

The Algorithms

CMP Method
❖ Comparative Manifesto Project (CMP)
❖ Unit of Analysis: **Quasi-sentence**
❖ Method: **Human analysis**
❖ Abortion units are a subset of Codes 201, 603, and 604.
❖ Calculating attention to abortion: A human codes each sentence related to abortion. Sum the subset of sentences about abortion and divide by the total number of quasi-sentences in the platform.

$$Abortion\_Attention\_Score = \frac{\sum_{1 \to T} c}{c_T}$$

Sentences Mentioning Abortion by Platform
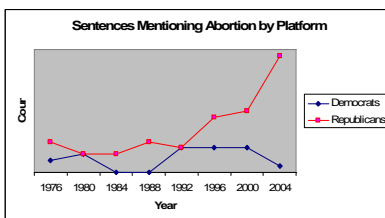
Wordscores Method
❖ Benoit, Laver, Garry (2003)
❖ Unit of Analysis: **Words-as-data**
❖ Method: Structured learning by example, Euclidian distance
❖ Expert scores training documents using CMP Method and feeds the documents and **Abortion Attention Scores into software** as "Training Data".

Purpura & Hillard (2006) Method
❖ Unit of Analysis: **Words-as-data but within Quasi-sentences**
❖ Method: Structured learning by example, SVMs
❖ Expert scores training documents using CMP Method. Feeds the documents, **the Abortion Attention Function, and each labeled quasi-sentence** into the software as "Training Data".

The Abortion Case

❖ Abortion is discussed in every U.S. election since 1976.

❖ Frequent Words: abortion, women/woman, choice, and right.

❖ In the 2004 Republican and Democratic platforms, expert analysis notes that 19 and 1 sentences, respectively, were about abortion, or 1% or less of the text.

❖ Using Wordscores, the estimate of the attention to abortion varies based upon the selected training texts. **Estimates are based on weights assigned to words.**

❖ Using Purpura & Hillard (2006), estimates also vary by training data. **Estimates are based on weights assigned to quasi-sentences.**

❖ **Can concept sensitivity analysis tell us which algorithm is doing better at capturing attention to abortion?**

Estimates using Full Testing Set

2004	Expert	Wordscores	P & H
Democratic			
Count	1	1	1
Abortion Attention Score	0.11%	0.11%	0.11%
Republican			
Count	19	25	18
Abortion Attention Score	1.04%	1.37%	0.99%
Algorithm - Expert			
Democratic			
Count	0	0	0
Abortion Attention Score	0.00%	0.00%	0.00%
Republican			
Count	0	6	-1
Abortion Attention Score	0.00%	0.33%	-0.05%

Estimates using Partial Testing Set (Abortion sentences removed)

2004	Expert	Wordscores	P & H
Democratic			
Count	0	1	0
Abortion Attention Score	0.00%	0.11%	0.00%
Republican			
Count	0	3	0
Abortion Attention Score	0.00%	0.33%	0.00%
Algorithm - Expert			
Democratic			
Count	0	0	0
Abortion Attention Score	0.00%	0.00%	0.00%
Republican			
Count	0	3	0
Abortion Attention Score	0.00%	0.33%	0.00%

The Sensitivity Control Study

❖ Remove Random Sentences and Calculate Scores.

**Effect of Deleting Random Sentences from the Test Set**
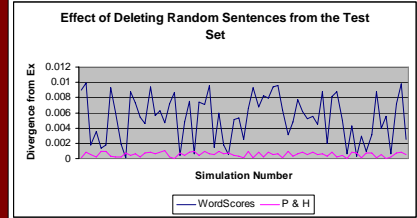
Which algorithm is better?

❖ **Purpura & Hillard (2006) is more accurate** at matching the CMP results. But more than 3 months and $1k in money were spent to transcribe expert results due to greater data requirements of analyzing by sentence.

❖ **Wordscores costs less to implement**, if you don't include the time spent searching for suitable reference documents and the initial coding is done without computers.

❖ The differences in the algorithms may appear slight, so a broader version is appropriate…

Wordscores vs. P & H on RILE

U.S. Democratic RILE, 1952 - 2000

U.S. Republican RILE, 1952 - 2000
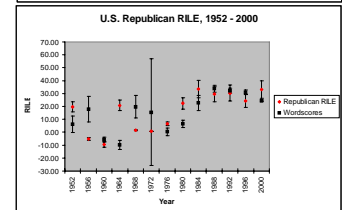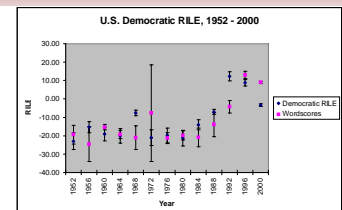
Future Research

❖ Future research should obtain more CMP data, and investigate validating/cleaning during the transcription process.