Replication: Measuring Political Sophistication through Textual Complexity

Zhiqiang Ji

Introduction

The original paper

Benoit, K., Munger, K. and Spirling, A. (2019), Measuring and Explaining Political Sophistication through Textual Complexity. American Journal of Political Science, 63: 491-508. https://doi.org/10.1111/ajps.12423

- Research Question: Political scientists lack domain-specific measures for the purpose of measuring the sophistication of political communication.

Methods

Data

- The text data used in the study primarily consists of paragraphs from the State of the Union (SOTU) addresses, parsed and prepared for analysis.

Develop statistically valid measures of textual complexity

1. **Obtain human judgments**
2. **Estimate latent easiness**: Apply an *unstructured Bradley-Terry model*
3. **Identify the best predictors**
4. **Fit structured Bradley-Terry Model**: Fit a *structured Bradley-Terry model*
5. **Prediction**: Use this model to predict the easiness of new texts

Crowdsourcing: Details

- Obtain human judgments on the relative easiness of political text snippets through crowdsourcing.
- Snippets were two-sentence segments from post-1950 State of the Union addresses
- Removed non-sentence text and disqualified snippets based on specific criteria, such as FRE score.
- Snippets were grouped into bands based on word count (345–60, 360–75, 375–90 words)
- A total of 2000+ snippet pairs were randomly chosen for crowdsourced comparison.
- “Gold pairs”(~15% of the tasks) were utilized as a quality control mechanism.

Develop the model

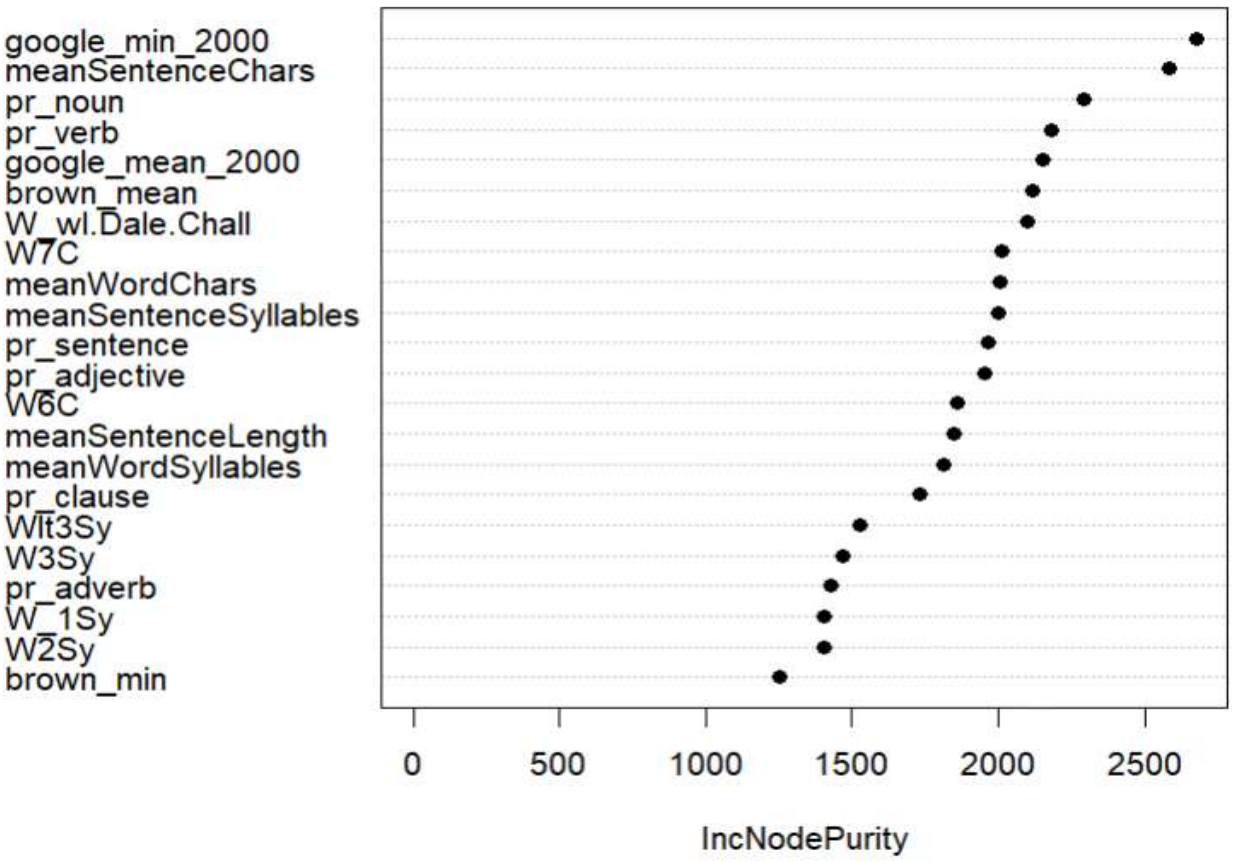
- Fit an *unstructured Bradley-Terry model* to the human judgments to estimate latent easiness.
- Add possible predictors/covariates to the model (22 predictors to test)

TABLE 1 Determinants of Textual Complexity

Source of Complexity	Variable Name
Long Words	
Mean characters per word	meanWordChars
Words with at least 7 characters	W7C
Words with at least 6 characters	W6C
Mean syllables per word	meanWordSyllables
Words with at least 3 syllables	W3Sy
Words with fewer than 3 syllables	Wlt3Sy
Words with 2 syllables	W2Sy
Words with 1 syllable	W_1Sy
Rare Words	
Google Books baseline usage	google_min
	google_mean
Brown corpus baseline usage	brown_mean
	brown_min
Words in the Dale-Chall list	W_wl.Dale.Chall
Long Sentences	
Mean characters per sentence	meanSentenceChars
Mean sentence length in words	meanSentenceLength
Number of sentences per character	pr_sentence
Mean sentence length in syllables	meanSentenceSyllables
Complex Content	
Proportion of nouns	pr_noun
Proportion of verbs	pr_verb
Proportion of adjectives	pr_adjective
Proportion of adverbs	pr_adverb
Average subordinate clauses	pr_clause

Develop the model (cont'd)

- **Identify the best predictors**


Bias Reduced
google_min_2000
meanSentenceChars
pr_noun
pr_verb
google_mean_2000
brown_mean
W_wl.Dale.Chall
W7C
meanWordChars
meanSentenceSyllables
pr_sentence
pr_adjective
W6C
meanSentenceLength
meanWordSyllables
pr_clause
Wlt3Sy
W3Sy
pr_adverb
W_1Sy
W2Sy
brown_min
0
500
1000
1500
2000
2500
IncNodePurity

Develop the model (cont'd)

- Using the Selected Predictors to Fit a *Structured Bradley-Terry Model* to get a statistical model of textual complexity.
- Employing nonparametric *bootstrapping* for uncertainty in predictions.

Results

Model Performance

- The new model achieved a proportion correctly predicted (PCP) score of 0.585, which adjusted to the mean best possible performance is 74% (0.741).
- Flexibility of the model
- Interpretability
- Uncertainty in predictions

Coefficients and Performance of the Four Structured Models

	Coefficient	BT_basic_Flesch	BT_optimal_Flesch	BT_basic_RF	BT_best
1	Flesch[ID]	0.02 (0.001)			
2	google_min_2000[ID]			1318.889 (153.389)	1336.446 (155.916)
3	meanSentenceChars[ID]			-0.015 (0.001)	-0.014 (0.001)
4	meanSentenceLength[ID]		-0.062 (0.003)		
5	meanWordChars[ID]				-0.311 (0.024)
6	meanWordSyllables[ID]		-1.773 (0.07)		
7	pr_noun[ID]			0.489 (0.167)	0.348 (0.168)
8	AIC	26274.35	25923.47	25921.56	25748.68
9	PCP	0.718	0.737	0.736	0.742
10	[95% CI]	[0.709, 0.727]	[0.727, 0.745]	[0.727, 0.746]	[0.733, 0.752]

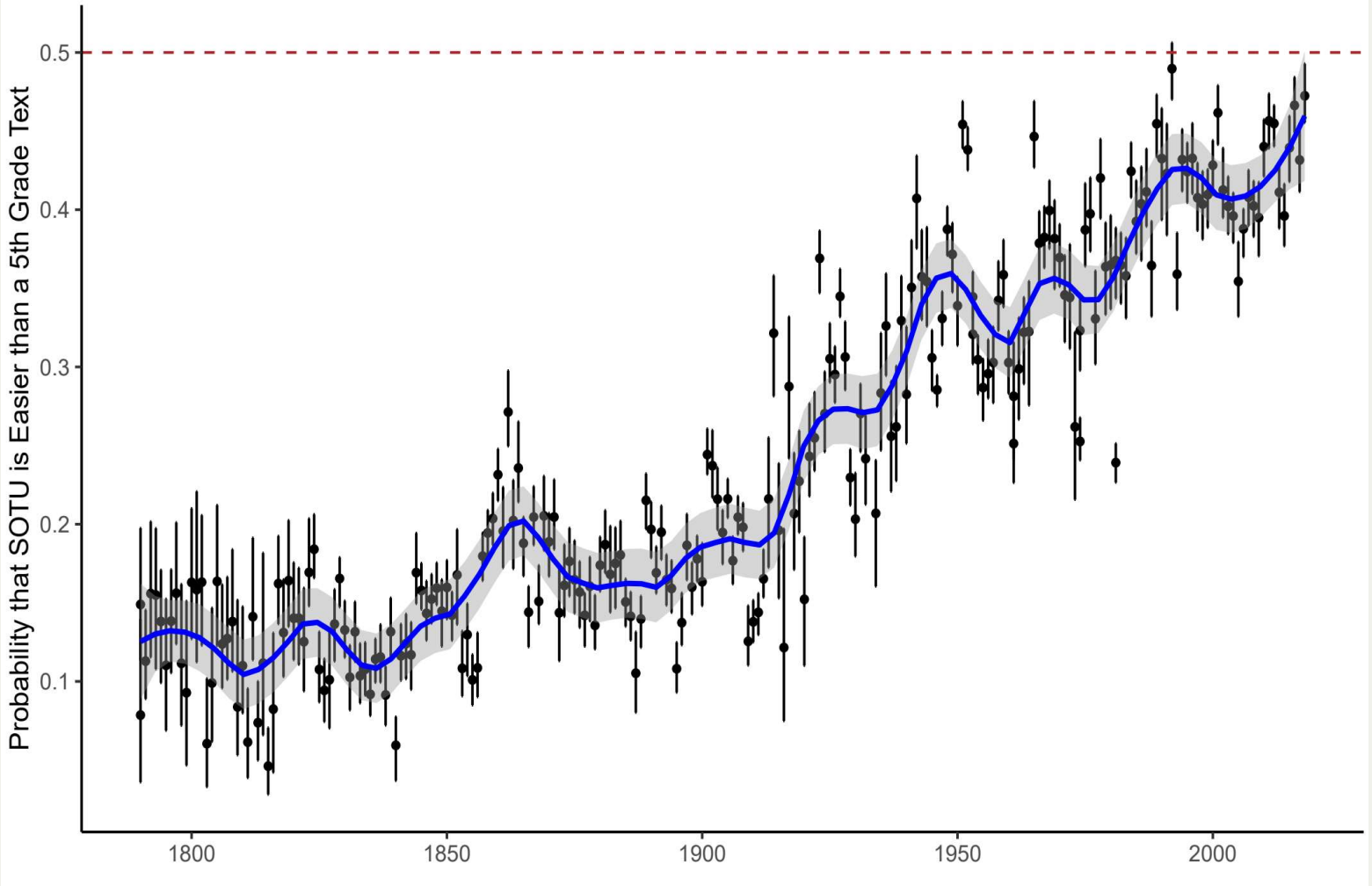
Example 1

- SOTU snippets in comparison with fifth-grade reading level (which is scaled to 100)
- Obama's snippet is a little bit easier for a fifth-grader 😂

	lambda	prob	scaled
Clinton_1999	-3.213821	0.2655429	39.04407
Bush_2005	-3.493522	0.2146608	22.28546
Obama_2012	-2.175249	0.5053040	101.27123
Trump_2019	-2.362582	0.4585663	90.04701

Example 2

- The readability of State of the Union addresses slight increase in simplicity over time.
- The baseline above is the fifth-grade reading level


Probability that SOTU is Easier than a 5th Grade Text
0.5
0.4
0.3
0.2
0.1
1800
1850
1900
1950
2000

Differences

- Did not replicate the crowdsourcing part of the study.
- Possible typo in Table 2
- Trivial differences such as the results of RF.

The PCP scores for "FRE Reweight" and "Basic RF" in the original code note is reversed in the paper 😟

FRE Reweight	Basic RF Model
−0.06	
(0.00)	
−1.79	
(0.07)	
	1298.14
	(153.07)
	−0.01
	(0.00)
	0.43
	(0.17)
19,430	19,430
25910.29	25915.01
→ 0.737	→ 0.738
[0.728, 0.747]	[0.729, 0.748]

```
## Original code note:
# Bottom row of TABLE 2
# Model                    PCP      AIC
# BT_basic_Flesch          0.719  26267.79
# BT_basic_RF          →   0.737  25915.01
# BT_optimal_Flesch    →   0.738  25910.29
# BT_best                  0.741  25740.25
```

Autopsy of the replication

- Compatibility Issues
- Discrepancies Between Theory and Practice

Extension

- Integration of Advanced Language Models like ChatGPT
- Enhanced the model
- Expanding the Scope of Text Sources