

Predicting recent Supreme Court outcomes

Jonas Heim, Matt Kaufmann, Michael Wagner, Núria Adell Raventós, Sergio Olalla Ubierna

Introduction

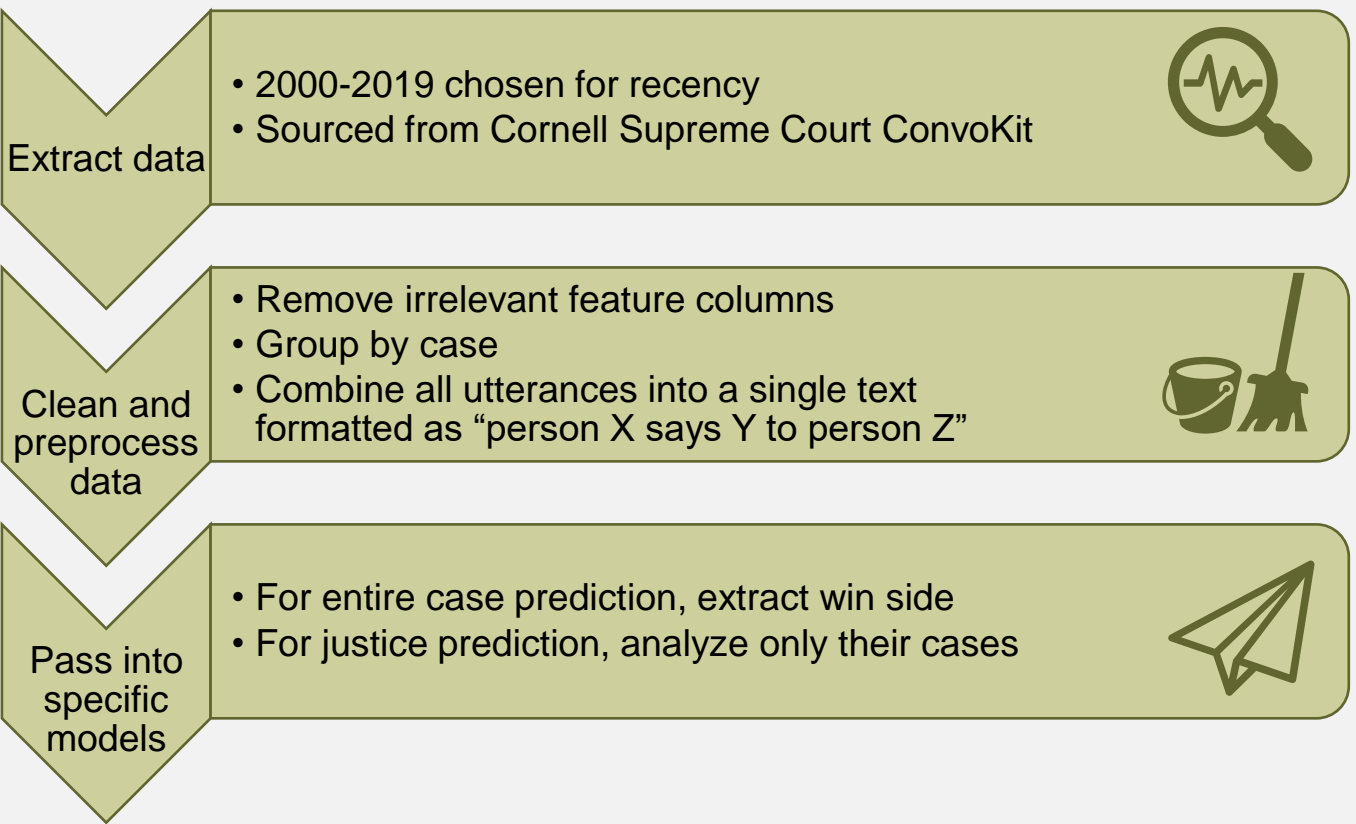
Context: The Supreme Court is one of the most important institutions in public policy, influencing the civil fabric of society. Recent advances in machine learning provide an innovative lens through which to predict the outcomes of Supreme Court cases and the voting behaviors of judges.

Objectives: (1) to predict case outcomes in favor of either the petitioner or the respondent, and (2) to predict the voting patterns of a subset of judges.

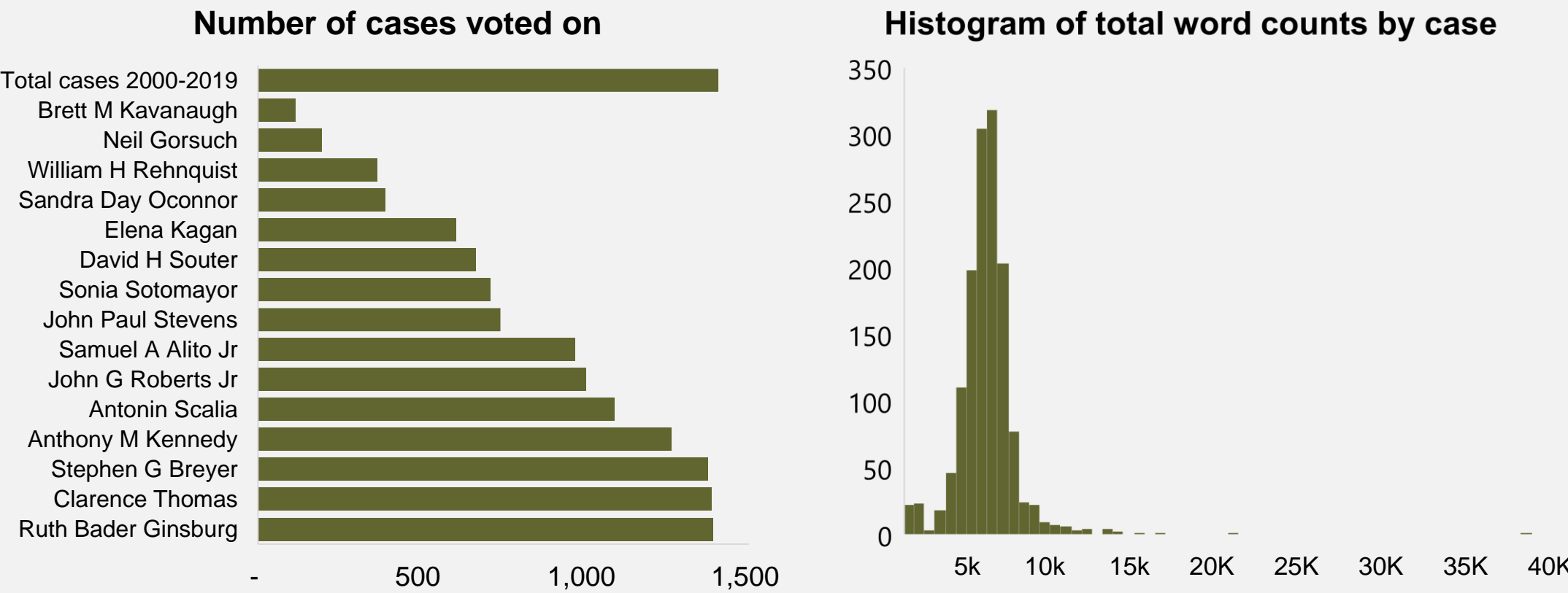
Significance: Based on how well our models predict a given judges vote, we can potentially understand how much predictive power utterances have on judges' decisions, or the impact of oral arguments on the judges' perspective of a case.

Data and preprocessing

Data pipeline



Data summary information



Model comparison

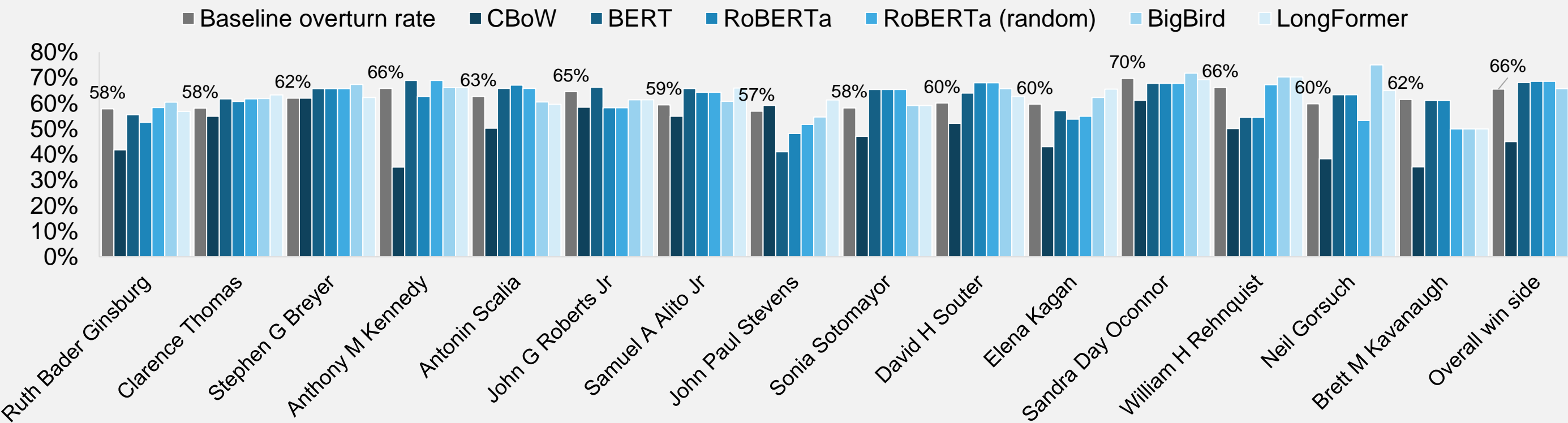
	Token limit	Depen-dencies	Benefits	Considerations
Continuous Bag of Words	N/A	None	<ul style="list-style-type: none">✓ Simplicity✓ Efficient representation	<ul style="list-style-type: none">✗ Limited contextual understanding
BERT		Short	<ul style="list-style-type: none">✓ Bidirectional context✓ Pre-trained	<ul style="list-style-type: none">✗ Complexity✗ Short attention window
RoBERTa		Medium	<ul style="list-style-type: none">✓ Generalizability✓ Enhanced relative to BERT	<ul style="list-style-type: none">✗ Relatively shorter attention window✗ Fine-tuning challenges
RoBERTa (randomized)		Medium	<ul style="list-style-type: none">✓ More robust training	<ul style="list-style-type: none">✗ Same as above
BigBird		Long	<ul style="list-style-type: none">✓ Longer depend.✓ Computationally efficient	<ul style="list-style-type: none">✗ Substantial memory
LongFormer		Short + Long	<ul style="list-style-type: none">✓ Long strings✓ Fine-grained attention	<ul style="list-style-type: none">✗ Training time✗ Limited context at boundaries of doc.



Note: Also ran sentiment analysis on justices to try and predict vote side, but results were inconclusive.

Results and conclusions

Accuracy by model type



Key conclusions

- RoBERTa and BigBird perform the best, underscoring the significance of structural information retained by more advanced architectures
 - RoBERTa (68.57%) predicts case outcomes at 3.04% higher than baseline
 - BigBird outperforms 12/15 justices on baseline

Next steps

- Improve data nuance:** Refine models to capture courtroom speaker dynamics
- Hyperparameter optimization:** Conduct grid search for performance improvement
- Train legal context embeddings:** Develop or use legal-specific embeddings for better language comprehension