

Friends with Motives: Using Text to Infer Influence on SCOTUS

Yanchuan Sim

Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
ysim@cs.cmu.edu

Bryan R. Routledge

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213, USA
routledge@cmu.edu

Noah A. Smith

Computer Science & Engineering
University of Washington
Seattle, WA 98195, USA
nasmith@cs.washington.edu

Abstract

We present a probabilistic model of the influence of language on the behavior of the U.S. Supreme Court, specifically influence of amicus briefs on Court decisions and opinions. The approach assumes that amici are rational, utility-maximizing agents who try to win votes or affect the language of court opinions. Our model leads to improved predictions of justices' votes and perplexity of opinion language. It is amenable to inspection, allowing us to explore inferences about the persuasiveness of different amici and influenceability of different justices; these are consistent with earlier findings.

“Language is the central tool of our trade.”

John G. Roberts, 2007 (Garner, 2010)

1 Introduction

The Supreme Court of the United States (SCOTUS), the highest court in the American judiciary, makes decisions with far-reaching effects. In a typical case, there are four participating parties: **petitioners** and **respondents** who file briefs arguing the merits of their sides of a case (“merits briefs”); third-party entities with an interest (but not a direct stake) in the case, who file **amicus curiae**¹ briefs to provide further arguments and recommendations on either side; and **justices** who, after oral arguments and discus-

¹*Amicus curiae* is Latin for “friends of the court.” Hereafter, we use *amicus* in singular and *amici* in plural to refer to these interested third parties. It is common for several amici to co-author a single brief, which we account for in our model.

sions, vote on the case and write “opinions” to explain the Court’s decisions.²

In recent years, amicus briefs are increasingly being employed as a lobbying tool to influence the Court’s decision-making process (Franze and Anderson, 2015; Kearney and Merrill, 2000). The content of these briefs reveals explicit attempts to persuade justices and provides a fascinating setting for empirical study of influence through language. As such, we take the perspective of an amicus, proposing a probabilistic model of the various parties to a case that accounts for the amicus’ goals.

Our model of SCOTUS is considerably more comprehensive than past work in political science, which has focused primarily on **ideal point** models that use votes as evidence. Text has been incorporated more recently as a way of making such models more interpretable, but without changing the fundamental assumptions (Lauderdale and Clark, 2014). Here, we draw on decision theory to posit amici as rational agents. We assume these amici-agents maximize their expected utility by framing their arguments to sway justices towards favorable outcomes.

We build directly on Sim et al. (2015), who used utility functions to explicitly model the goals of amici in a probabilistic setting. Their approach only considered amici in aggregate, inferring nothing about any specific amicus, such as experience or motivation for filing briefs. Here, we enrich their model to allow such analysis and also introduce Court opinions as evidence. By modeling the justices’ author-

²We use the term *opinions* to refer to these decision-explaining documents, not to abstract positions as the term is used in much NLP research.

ing process as well, we can capture an important aspect of amici’s goals: influencing the text of the opinions.

In §3, we demonstrate the effectiveness of our approach on vote prediction and perplexity. Furthermore, we present analyses that reveal the persuasiveness of amici and influenceability of justices that are consistent with past findings.

2 Generative Models of SCOTUS

Our approach builds on a series of probabilistic models only recently considered in NLP research. To keep the discussion self-contained, we begin with classical models of votes alone and build up toward our novel contributions.

2.1 Modeling Votes

Ideal point (IP) models are a mainstay in quantitative political science, often applied to voting records to place voters (lawmakers, justices, etc.) in a continuous space. A justice’s “ideal point” is a latent variable positioning him in this space. Martin and Quinn (2002) introduced the unidimensional IP model for judicial votes, which posits an IP $\psi_j \in \mathbb{R}$ for each justice j . Often the ψ_j values are interpreted as positions along a liberal-conservative ideological spectrum. Each case i is represented by popularity (a_i) and polarity (b_i) parameters.³ A probabilistic view of the unidimensional IP model is that justice j votes in favor of case i ’s petitioner (as opposed to the respondent) with probability

$$p(v_{i,j} = \text{petitioner} \mid \psi_j, a_i, b_i) = \sigma(a_i + b_i \psi_j)$$

where $\sigma(x) = \frac{\exp(x)}{1 + \exp(x)}$ is the logistic function. When the popularity parameter a_i is high enough, every justice is more likely to favor the petitioner. The polarity b_i captures the importance of a justice’s ideology: polarizing cases (i.e., $|b_i| \gg 0$) push justice j more strongly to the side of the petitioner (if b_i has the same sign as ψ_j) or the respondent (otherwise).

Amici IP models. Sim et al. (2015) introduced a multidimensional IP model that incorporated text

³This model is also known as a two parameter logistic model in item-response theory (Fox, 2010), where a_i is “difficulty” and b_i is “discrimination.”

from merits and amicus briefs as evidence. They inferred dimensions of IP that are grounded in “topical” space, where topics are learned using latent Dirichlet allocation (Blei et al., 2003). In their proposed model, the merits briefs describe the issues and facts of the case, while amicus briefs were hypothesized to “frame” the facts and potentially influence the outcome of the case. For case i and justice j , the vote probability is

$$p(v_{i,j} = \text{petitioner} \mid \psi_j, \theta_i, \Delta_i, a_i, b_i, \mathbf{c}_i) \quad (1) \\ = \sigma\left(a_i + b_i \psi_j^\top \left(\theta_i + \underbrace{\frac{1}{|\mathcal{A}_i|} \sum_{k \in \mathcal{A}_i} c_i^{s_{i,k}} \Delta_{i,k}}_{\text{case IP}}\right)\right)$$

where \mathcal{A}_i is the set of amicus briefs filed on this case, $s_{i,k}$ denotes the side ($\in \{\text{petitioner}, \text{respondent}\}$) supported by the k th brief, and c_i^p , and c_i^r are the amicus polarities for briefs on either side. The **case IP** is influenced by merits briefs (embedded in θ_i) and by the amicus briefs (embedded in $\Delta_{i,k}$), both of which are rescaled independently by the case discrimination parameters to generate the vote probability. The model assumes that briefs on the same side share a single embedding and that individual briefs on one side influence the vote-specific IP equally.

New IP model: Persuasive amici. Lynch (2004) and others have argued that some amici are more effective than others, with greater influence on justices. We therefore propose a new model which considers amici as individual actors. Starting from Eq. 1, we consider two additional variables: each amicus e ’s **persuasiveness** ($\pi_e > 0$) and each justice j ’s **influenceability** ($\chi_j > 0$).⁴

$$p(v_{i,j} = \text{petitioner} \mid \psi_j, \chi_j, \theta_i, \Delta_i, a_i, b_i, \boldsymbol{\pi}) \quad (2) \\ = \sigma\left(a_i + b_i \psi_j^\top \left(\theta_i + \frac{\chi_j}{|\mathcal{A}_i|} \sum_{k \in \mathcal{A}_i} \bar{\pi}_{i,k} \Delta_{i,k}\right)\right)$$

where $\bar{\pi}_{i,k} = \frac{\sum_{e \in \mathcal{E}_{i,k}} \pi_e}{|\mathcal{E}_{i,k}|}$ is the average of their π -values, with $\mathcal{E}_{i,k}$ denoting the set of entities who co-authored the k th amicus brief for case i .

Intuitively, a larger value of χ_j will shift the case IP more towards the contents of the amicus briefs,

⁴Note that the amici IP model of Sim et al. (2015), Eq. 1, is a special case of this model where $\chi_j = 1$ and each case has polarity parameters for each side; no information is shared across briefs written by the same amicus-entity for different cases.

thus making the justice seem more “influenced” by amicus. Likewise, briefs co-authored by groups of amici who are more effective (i.e., larger $\bar{\pi}_{i,k}$), will “frame” the case towards their biases. Unlike Sim et al. (2015), we eschew the amicus polarity parameters (c_i) and instead rely on the influenceability and persuasiveness parameters. Furthermore, we note that they performed a post-hoc analysis of amici influence on justices but we do so directly through χ_j .

With appropriate priors on the latent variables, the generative story for votes is:

1. For each topic $t \in \{1, \dots, T\}$, draw topic-word distributions $\phi_t \sim \text{Dirichlet}(\beta)$.
2. For each justice $j \in \mathcal{J}$, draw justice IP $\psi_j \sim \mathcal{N}(\mathbf{0}, \sigma_j^2 \mathbf{I} + \rho \mathbf{1})^5$ and influenceability $\chi_j \sim \log \mathcal{N}(\mathbf{0}, \sigma_j^2 \mathbf{I})$.
3. For each amicus-entity $e \in \mathcal{E}$, draw its persuasiveness $\pi_e \sim \log \mathcal{N}(\mathbf{0}, \sigma_p^2 \mathbf{I})$.
4. For each case $i \in \mathcal{C}$:
 - (a) Draw case parameters $a_i, b_i \sim \mathcal{N}(0, \sigma_C^2)$.
 - (b) Draw topic proportions for merits $\theta_i \sim \text{Dirichlet}(\alpha)$.
 - (c) For each word $w_{i,n}^{(m)}$ in the merits briefs, draw topic indicators $z_{i,n}^{(m)} \sim \text{Categorical}(\theta_i)$ and $w_{i,n}^{(m)} \sim \text{Categorical}(\phi_{z_{i,n}^{(m)}})$.
 - (d) For each amicus brief indexed by k :
 - i. Draw topic proportions $\Delta_{i,k}$ according to a distribution discussed in §2.3.
 - ii. For each word $w_{i,k,n}^{(a)}$ in the brief, draw topic indicators $z_{i,k,n}^{(a)} \sim \text{Categorical}(\Delta_{i,k})$ and $w_{i,k,n}^{(a)} \sim \text{Categorical}(\phi_{z_{i,k,n}^{(a)}})$.
 - (e) For each participating justice $j \in \mathcal{J}_i$, draw vote $v_{i,j}$ according to Eq. 2.

2.2 Modeling Opinions

In most SCOTUS cases, a justice is assigned to author a majority opinion, and justices voting in the majority “join” in the opinion. Justices may author additional opinions concurring or dissenting with the majority, and they may choose to join concurring

⁵The positive off-diagonal elements of the covariance matrix for justice IPs (ψ_j) orient the issue-specific dimensions in the same direction (i.e., with conservatives at the same end) and provide shrinkage of IP in each dimension to their common mean across dimensions (Lauderdale and Clark, 2014).

and dissenting opinions written by others. Here, we extend the IP model of votes to generate the opinions of a case; this marks the second major extension beyond the IP model of Sim et al. (2015).

SCOTUS justices often incorporate language from merits (Feldman, 2016b; Feldman, 2016a) and amicus (Collins et al., 2015; Ditzler, 2011) briefs into their opinions. While amicus briefs are not usually used directly in legal analyses, the background and technical information they provide are often quoted in opinions. As such, we model opinions as a mixture of its justice-authors’ topic preferences, topic proportions of the merits briefs (θ), and topic proportions of the amicus briefs (Δ). This can also be viewed as an author-topic model (Rosenzvi et al., 2004) where justices, litigants, and groups of amici are all effective authors. To accomplish this, we introduce an explicit switching variable x for each word, which selects between the different sources of topics, to capture the mixture proportions.

Since any justice can author additional opinions explaining the rationale behind their votes, we concatenate all opinions supporting the same side of a case into a single document.⁶ However, we note that concurring opinions often contain perspectives that are different from the majority opinion and by concatenating them, we may lose some information about individual justices’ styles or preferences. Building on the generative model for votes, the generative story for each case i ’s two opinions-documents is:

5. For each justice $j \in \mathcal{J}$, draw topics $\Gamma_j \sim \text{Dirichlet}(\alpha)$.
6. For each case $i \in \mathcal{C}$:
 - (a) For each side $s \in \{\text{petitioner, respondent}\}$, draw “author”-mixing proportions:

$$\tau_i^s \sim \text{Dirichlet} \left(\begin{bmatrix} p(v_{i,1} = s) \\ \vdots \\ p(v_{i,|\mathcal{J}|} = s) \\ 1 \\ 1 \end{bmatrix} \right) \quad (3)$$

where the last two dimensions are for choosing topics from the merits and amicus briefs, re-

⁶Opinions where justices dissent from the majority are concatenated together, and those where justices concur with the majority are concatenated with the majority opinion.

spectively.⁷ Intuitively, our model assumes that opinions will incorporate more language from justices who agree with it.

- (b) For each side $s \in \{\text{petitioner, respondent}\}$ and each word $w_{i,s,n}^{(o)}$ in the opinion for side s ,
- i. Draw $x_{i,s,n} \sim \text{Categorical}(\tau_i^s)$.
 - ii. If $x_{i,s,n} \in \mathcal{J}_i$, draw $z_{i,s,n}^{(o)} \sim \text{Categorical}(\Gamma_{x_{i,s,n}})$, the justice’s topic distribution.
 - iii. If $x_{i,s,n} = \text{merits}$, draw $z_{i,s,n}^{(o)} \sim \text{Categorical}(\theta_i)$, the merits topic distribution.
 - iv. If $x_{i,s,n} = \text{amici}$, draw $z_{i,s,n}^{(o)} \sim \text{Categorical}(\Delta_i^s)$, side s ’s amicus briefs topic distribution.
 - v. Draw word $w_{i,s,n}^{(o)} \sim \text{Categorical}(\phi_{z_{i,s,n}^{(o)}})$.

Unlike in the Court, where an opinion is mainly authored by a single justice, all the participating justices contribute to an opinion in our generative story, with different proportions. This approach simplifies the computational model and reflects the closed-door nature of discussions held by justices prior to writing their opinions. Our model assumes that justices debate together, and that the arguments are reflected in the final opinions. In future work, we might extend the model to infer an authoring process that separates an initial author from “joiners.”

2.3 Amici Utility

Our approach assumes that amici are rational and purposeful decisionmakers who write briefs to influence the outcome of a case; this assumption leads to the design of the distribution over Δ (generative model step 4(d)i). When writing a brief Δ , an amicus seeks to increase the response to her brief (i.e., votes), while keeping her costs low. We encode her objectives as a utility function, which she aims to maximize with respect to the decision variable Δ :

$$U(\Delta) = R(\Delta) - C(\Delta) \quad (4)$$

where $R(\cdot)$ is the extrinsic response (reward) that an amicus gets from filing brief Δ and $C(\cdot)$ is the “cost” of filing the brief; dependency on other latent

⁷In cases where there are less than nine justices voting, the size of τ_i^p and τ_i^r may be smaller.

variables is notationally suppressed. When authoring her brief, we assume that the amicus writer has knowledge of the justices (IP and topic preferences), case parameters, and merits, but not the other amici participating in the case.⁸

Amicus curiae are motivated to position themselves (through their briefs) in such a way as to improve the likelihood that their arguments will persuade SCOTUS justices. This is reflected in the way a justice votes or through the language of the opinions. Hence, we investigate two response functions. First, an amicus supporting side s seeks to win votes for s ,

$$R^{\text{vote}}(\Delta) = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} p(v_j = s \mid \dots), \quad (5)$$

which is the expected number of votes for side s , under the model. This follows Sim et al. (2015).

An alternative is to maximize the (topical) similarity between her brief and the Court’s opinion(s) siding with s ,

$$R^{\text{opinion}}(\Delta) = 1 - H^2(\Delta, \Omega^s), \quad (6)$$

where $H^2(P, Q) = \frac{1}{2} \|\sqrt{P} - \sqrt{Q}\|_2^2$ is the squared Hellinger (1909) distance between two distributions, and Ω^s is the expected topic mixture under the model assumptions in §2.2 (which has a closed form). In short, the amicus gains utility by accurately predicting the expected opinion, thereby gaining publicity and demonstrating to members, donors, potential clients, and others that the language of the highly visible SCOTUS opinion was influenced. Both Eqs. 5 and 6 reward amici when justices “agree” with them, for different definitions of agreement.

We assume the cost $C(\Delta) = H^2(\Delta, \theta)$, the squared Hellinger distance between the mixture proportions of the amicus brief and merits briefs.⁹ The cost term defines the “budget” set of the amicus: briefs cannot be arbitrary text, as there is disutility or

⁸Capturing strategic amici agents (a petitioner amicus choosing brief topics considering a respondent amicus’ brief) would require a game-theoretic model and, we conjecture, would require a much richer representation of policy and goals. That idea is left for future research.

⁹Sim et al. (2015) used a Euclidean distance for cost rather than Hellinger distance, which we believe is a better fit for probability distributions without sacrificing symmetry (cf. KL divergence).

effort required to carefully frame a case, and monetary cost to hiring legal counsel. The key assumption is that framing is costly, while simply matching the merits is cheap (and presumably unnecessary).

Notationally, we use U^{vote} to refer to models where Eq. 5 is in the utility function (in Eq. 4) and U^{opinion} where it is Eq. 6.

Random utility models Recall our assumption that amici are purposeful writers whose briefs are optimized for their utility function. In an ideal setting, the Δ which we observe will be utility maximizing. We simplify computation by assuming that these amici agents’ preferences also contain an idiosyncratic random component that is unobserved to us. This is a common assumption in discrete choice models known as a “random utility model” (McFadden, 1974). We view the utility function as a prior on Δ ,

$$p_{\text{util}}(\Delta | \dots) \propto \exp \eta U(\Delta),$$

where our functional equations for utility imply $-1 \leq U(\cdot) \leq 1$. η is a hyperparameter tuned using cross validation. The behavior which we observe (i.e., the amicus’ topic mixture proportions) has a likelihood that is proportional to utility.

2.4 Parameter Estimation

The models we described above can be estimated within a Bayesian framework. We decoupled the estimation of the votes model from the opinions model; we first estimate the parameters for the votes model and hold them fixed while we estimate the new latent variables in the opinions model. In our preliminary experiments, we found that estimating parameters for both votes and opinions jointly led to slow mixing and poor predictive performance. Separating the estimation procedure into two stages allows the model to find better parameters for the votes model, which are then fed into the opinions model as priors through the vote probabilities.

We used Metropolis within Gibbs, a hybrid MCMC algorithm, to sample the latent parameters from their posterior distributions (Tierney, 1994).¹⁰ For the Metropolis-Hastings proposal distributions, we used a Gaussian for the case parameters a , b , and justice IPs ψ , log-normal distributions for χ and π ,

¹⁰The details of our sampler and hyperparameter settings can be found in §A and §B of the supplementary materials.

and logistic-normal distribution for the variables on the simplex θ , Δ , τ , and Γ . We tuned the hyperparameters of the proposal distributions at each iteration to achieve a target acceptance rate of 15–45%. We used $T = 128$ topics for model and initialized topic proportions (θ , Δ) and topic-word distributions (ϕ) using online LDA (Hoffman et al., 2010).

3 Experiments

Data. In our experiments, we use SCOTUS cases between 1985–2014; votes and metadata are from Spaeth et al. (2013) and brief texts come from Sim et al. (2015). We concatenate each of the 2,643 cases’ merits briefs from both parties to form a single document, where the text is used to infer the representation of the case in topical space (θ ; i.e., merits briefs are treated as “facts of the case”). Likewise, opinions supporting the same side of the case (i.e., majority and concurring vs. dissents) were concatenated to form a single document. In our dataset, the opinions are explicitly labeled with the justice who authored them (as well as other justices who decide to “join” it).

As the amicus briefs in the dataset were not explicitly labeled with the side that they support, Sim et al. (2015) built a binary classifier with bag-of- n -gram features that took advantage of cues in the brief content that *strongly* signal the side that the amici supports (e.g., “in support of petitioner”). We used their classifier to label the amici’s supporting side. Additionally, we created regular expression rules to identify and standardize amicus authors from the header of briefs. We filtered amici who have participated in fewer than 5 briefs¹¹ and merged regional chapters of amicus organizations together (i.e., “ACLU of Kansas” and “ACLU of Kentucky” are both labeled “ACLU”). On the other hand, we separated labeled amicus briefs by the U.S. Solicitor General according to the presidential administration when the brief is filed (i.e., an amicus brief filed during Obama’s administration will be labeled “USSG-Obama”). The top three amici by number of briefs filed are *American Civil Liberties Union* (463), *Utah* (376), and *National Asso-*

¹¹Briefs which have no authors as a result of the filtering process are removed from our dataset. This occurred in about 24% of amicus briefs.

| | |
|------------------------|-----------------|
| Cases / Votes | 2,643 / 23,465 |
| Merits / Amicus briefs | 16,416 / 16,303 |
| Opinions | 4,187 |
| Phrases | 18,207,326 |

Table 1: Corpus statistics.

ciation of Criminal Defense Lawyers (359).

We represent a document as a bag of n -grams with part of speech tags that follow the simple but effective pattern (Adjective|Cardinal|Noun)+ Noun (Justeson and Katz, 1995). We filter phrases appearing fewer than 100 times or in more than 8,500 documents, obtaining a final set of 48,589 phrase types. Table 1 summarizes the details of our corpus.

Predicting Votes. We quantify the performance of our vote model using 5-fold cross validation and on predicting future votes from past votes. The utility function in the vote model uses the response function in Eq. 5. Due to the specification of IP models, we need the case parameters of new cases to predict the direction of the votes. Gerrish and Blei (2011) accomplished this by using regression on legislative text to predict the case parameters (a, b). Here, we follow a similar approach, fitting ridge regression models on the merits brief topic mixtures θ to predict a and b for each case.¹² On the held-out test cases, we sampled the mixture proportions for the merits and amicus briefs directly using latent Dirichlet allocation with parameters learned while fitting our vote model. With the parameters from our fitted vote model and ridge regression, we can predict the votes of every justice for every case.

We compared the performance of our model with two strong baselines: (i) a random forest trained on case-centric metadata coded by Spaeth et al. (2013) to make predictions on how justices would vote (Katz et al., 2014) and (ii) Sim et al. (2015)’s amici IP model, which uses amicus briefs and their version of utility; it is a simpler version of our vote model that does not consider the persuasiveness of different amici or the influenceability of different justices. For prediction in Sim et al. (2015), we used the same approach described above to estimate the case parameters a, b , and regressing on amicus brief topics (Δ) instead for amicus polarities c^p and c^f . Table 2

¹²We tuned the parameters of the regression using 5-fold cross-validation on the training data.

| Model | 5-fold | 2013 | 2014 |
|--------------------------------------|--------------|--------------|--------------|
| Most frequent | 0.597 | 0.694 | 0.650 |
| Random forest | 0.651 | 0.648 | 0.633 |
| Vote model without U^{vote} | 0.661 | 0.655 | 0.660 |
| Sim et al. (2015) | 0.675 | 0.658 | 0.661 |
| Vote model with U^{vote} | 0.685 | 0.664 | 0.672 |

Table 2: Accuracy of vote prediction. There are 70 cases (625 votes) and 69 cases (619 votes) in the 2013 and 2014 test sets, respectively.

shows performance on vote prediction.

We evaluated the models using 5-fold cross validation, as well as on forecasting votes in 2013 and 2014 (trained using data from 1985 to the preceding year). Our model outperformed the baseline models. The improvement in accuracy over Sim et al. (2015) is small; most likely because both models are very similar, the main difference being the parametrization of amicus briefs. In the 2013 test set, the distribution of votes is significantly skewed towards the petitioner (compared to the training data), which resulted in the most frequent class classifier performing much better than everything else. Fig. 1 illustrates our model’s estimated ideal points for selected topics.

Predicting Opinions. We also estimated the opinion model using the utility function with response function in Eq. 6. We use perplexity as a proxy to measure the opinion content predictive ability of our model. Perplexity on a test set is commonly used to quantify the generalization ability of probabilistic models and make comparisons among models over the same observation space. For a case with opinion w supporting side s , the perplexity is defined as

$$\exp\left(-\frac{\log p(w | s, \dots)}{N}\right),$$

where N is the number of tokens in the opinion and a lower perplexity indicates better generalization performance. The likelihood term can be approximated using samples from the inference step.

Table 3 shows the perplexity of our model on opinions in the test set. As described in §2.4, we learn the vote model in the first stage before estimating the opinion model. Here, we compare our model against using vote models that do not include U^{vote} to evaluate the sensitivity of our opinion

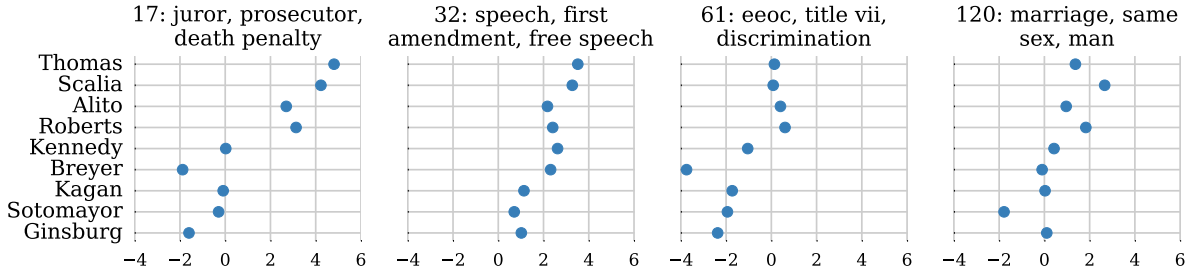


Figure 1: Justices’ ideal points for selected topics. Justices whose topic IPs are close to each other are more likely to vote in the same direction on cases involving those topics. The IP estimated by our model is consistent with publicly available knowledge regarding justices’ ideological stances on these issues.

model to the vote model parameters. Additionally, we compared against two baselines trained on just the opinions: one using LDA¹³ and another using the author-topic model (Rosen-Zvi et al., 2004). For the author-topic model, we treat each opinion as being “authored” by the participating justices, a pseudo-author representing the litigants which is shared between opinions in a case, and a unique amicus author for each side. Our model with U^{opinion} achieves better generalization performance than the simpler baselines, while we do not see significant differences in whether the first stage vote models use U^{vote} . This is not surprising since the vote model’s results are similar with or without U^{vote} and it influences the opinion model indirectly through priors and U^{opinion} .

In our model, the latent variable Γ_j captures the proportion of topics that justice j is likely to contribute to an opinion. When j has a high probability of voting for a particular side, our informed prior increases the likelihood that j ’s topics will be selected for words in the opinion. While Γ_j serves a similar purpose to ψ_j in characterizing j through her ideological positions, ψ_j relies on votes and gives us a “direction” of j ’s ideological standing, whereas Γ_j is estimated from text produced by the justices and only gives us the “magnitude” of her tendency to author on a particular issue. In Table 4, we identify the top topics in Γ_j by considering the deviation from the mean of all justice’s Γ , i.e., $\Gamma_{j,k} - \frac{1}{|\mathcal{J}|} \sum_j \Gamma_{j,k}$.

Amici Persuasiveness. The latent variable π_e captures the model’s belief about amicus e ’s brief’s ef-

| Model | 5-fold | 2013 | 2014 |
|--|----------------|-----------------------|----------------|
| LDA | 2.86 | 2.67 | 2.63 |
| Author-Topic | 2.62 | 2.36 | 2.25 |
| Opinion model without U^{opinion} | \dagger 2.43 | \dagger 2.26 | \dagger 2.13 |
| | 2.45 | 2.27 | 2.11 |
| Opinion model with U^{opinion} | \dagger 2.10 | \dagger 1.91 | \dagger 1.96 |
| | 2.07 | 1.98 | 1.94 |

Table 3: Perplexity of Court’s opinions ($\times 10^3$). There are 30,133 phrases (98 opinions) and 23,706 phrases (109 opinions) in the 2013 and 2014 test set, respectively. Results marked \dagger are initialized with a vote model U^{vote} .

fect on the case IP, which we call “persuasiveness.” A large π_e indicates that across the dataset, e exerts a larger effect on the case IPs, that is, according to our model, she has a larger impact on the Court’s decision than other amici. Fig. 2 is a swarm plot illustrating the distribution of π values for different types of amicus writers.

Our model infers that governmental offices tend to have larger π values than private organizations, especially the U.S. Solicitor General.¹⁴ In fact, Lynch (2004) found through interviews with SCOTUS law clerks that “amicus briefs from the solicitor general are ‘head and shoulders’ above the rest, and are often considered more carefully than party briefs.”

Another interesting observation from Fig. 2 is the low π value for ACLU and ABA, despite being prolific amicus brief filers. While it is tempting to say that amici with low π values are ineffective, we find that there is almost no correlation between π and the proportion of cases where they were on the winning side.¹⁵ Note that our model does not assume that a

¹³We used `scikit-learn`’s LDA module (Pedregosa et al., 2011) which implements the online variational Bayes algorithm (Hoffman et al., 2010).

¹⁴The average π for *Federal*, *State/Local* and *Others* are 2.35, 1.11, and 0.929 respectively.

¹⁵The Spearman’s ρ between π and the proportion of winning

| |
|---|
| John G. Roberts |
| 32: speech, first amendment, free speech, message, expression |
| 61: eeoc, title vii, discrimination, woman, civil rights act |
| 52: sec, fraud, security, investor, section ##b |
| Ruth B. Ginsburg |
| 61: eeoc, title vii, discrimination, woman, civil rights act |
| 80: class, settlement, rule ##, class action, r civ |
| 96: taxpayer, bank, corporation, fund, irs |
| Antonin Scalia |
| 94: 42 USC 1983, qualified immunity, immunity, official, section #### |
| 57: president, senate, executive, article, framer |
| 80: class, settlement, rule ##, class action, r civ |

Table 4: Top three topics contributed to Court opinions for selected justices (Γ). The full list can be found in supplementary §C.

“persuasive” amicus tends to win. Instead, an amicus with large π will impact the case IP most, and thus explain a justice’s vote or opinion (even dissenting) more than the other components in a case.

Insofar as π explains a vote, we must exercise caution; it is possible that the amicus played no role in the decision-making process and the values of π_e simply reflect our modeling assumptions and/or artifacts of the data. Without entering the minds of SCOTUS justices, or at least observing their closed-door deliberations, it is difficult to measure the influence of amicus briefs on justices’ decisions.

Justice Influenceability. The latent variable χ_j measures the relative effect of amicus briefs on justice j ’s vote IP; when χ_j is large, justice j ’s vote probability is affected by amicus briefs more. Since χ_j is shared between all cases that a justice participates in, χ_j should correspond to how much they value amicus briefs. Some justices, such as the late Scalia, are known to be dubious of amicus briefs, preferring to leave the task of reading these briefs to their law clerks, who will pick out any notable briefs for them; we would expect Scalia to have a smaller χ than other justices. In Table 5, we compare the χ values of justices with how often they cite an amicus brief in any opinion they wrote (Franze and Anderson, 2015). The χ values estimated by our model are

sides is -0.0549 . On average, an amicus supports the winning side in 55% of cases. For the ACLU, ABA, CAC, and CWFA, the proportions are 44%, 50%, 47%, and 50% respectively.

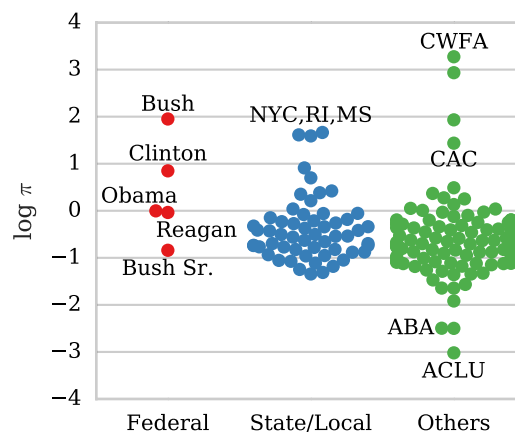


Figure 2: Amici “persuasiveness” by organization type. *Federal* refers to different presidential administration’s federal government (and represented by the U.S. Solicitor General) and *State/Local* refers to state and local governments. The abbreviated amici are New York City (NYC), Rhode Island (RI), Mississippi (MS), Concerned Women For America (CWFA), Constitution Accountability Center (CAC), American Bar Association (ABA), and American Civil Liberties Union (ACLU).

consistent with our expectations.¹⁶

We note that the χ values correlate considerably with the general ideological leanings of the justices. This might be a coincidence or an inability of the model’s specification to discern between ideological extremeness and influenceability.

4 Related Work

The ideal points model was first introduced by Poole and Rosenthal (1985) and has inspired a variety of IP models in SCOTUS (Lauderdale and Clark, 2014; Martin and Quinn, 2002) and Congressional bills (Clinton et al., 2004; Gerrish and Blei, 2011; Heckman and Snyder, 1996). IP has provided a useful framework to characterize voters using roll call information and textual evidence.

We view amicus briefs as “purposeful” texts, where authors are writing to maximize their utility function. This is related to work investigating news media for “slant” to maximize profit (Gentzkow and Shapiro, 2010) and economists choosing research topics maximize certain career outcomes (Jelveh et al., 2015). More generally, extensive literature in

¹⁶The Spearman’s ρ between χ_j and citation rates is 0.678.

| Justice | χ_j | Citation rate (%) |
|--------------------|----------|-------------------|
| Sonia Sotomayor | 1.590 | 45 |
| Elena Kagan | 0.714 | 40 |
| Stephen G. Breyer | 0.637 | 38 |
| Ruth B. Ginsburg | 0.515 | 41 |
| John G. Roberts | 0.495 | 42 |
| Anthony M. Kennedy | 0.468 | 42 |
| Samuel A. Alito | 0.286 | 27 |
| Antonin Scalia | 0.268 | 22 |
| Clarence Thomas | 0.162 | 25 |

Table 5: Justice χ values and their average amicus citation rates between 2010–2015, provided by Franze and Anderson (2015).

econometrics estimates structural utility-based decisions (Berry et al., 1995, *inter alia*).

Researchers have used SCOTUS texts to study authorship (Li et al., 2013), historical changes (Wang et al., 2012), power relationships (Danescu-Niculescu-Mizil et al., 2012; Prabhakaran et al., 2013), and pragmatics (Goldwasser and Daumé, 2014).

5 Conclusion

We presented a random utility model of the Supreme Court that is more comprehensive than earlier work. We considered an individual amicus’ persuasiveness and motivations through two different utility functions. On the vote prediction task, our results are consistent with earlier work, and we can infer and compare the relative effectiveness of an individual amicus. Moreover, our opinions model and opinion utility function achieved better generalization performance than simpler methods.

Acknowledgments

The authors thank the anonymous reviewers for their thoughtful feedback and Tom Clark, Philip Resnik, and members of the ARK group for their valuable comments. This research was supported in part by an A*STAR fellowship to Y. Sim, by a Google research award, and by computing resources from the Pittsburgh Supercomputing Center.

References

Steven Berry, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Economet-*

rica: Journal of the Econometric Society, pages 841–890.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Joshua Clinton, Simon Jackman, and Douglas Rivers. 2004. The statistical analysis of roll call data. *American Political Science Review*, 98:355–370.

Paul M. Collins, Pamela C. Corley, and Jesse Hamner. 2015. The influence of amicus curiae briefs on U.S. Supreme Court opinion content. *Law & Society Review*, 49(4):917–944.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proc. of WWW*.

Megan Ann Ditzler. 2011. Language overlap between solicitor general amicus curiae and Supreme Court majority opinions: An analysis. Master’s thesis, Southern Illinois University Carbondale.

Adam Feldman. 2016a. All copying is not created equal: Examining Supreme Court opinions’ borrowed language. *Journal of Appellate Practice and Process*, 17.

Adam Feldman. 2016b. A brief assessment of Supreme Court opinion language, 1946–2013. *Mississippi Law Journal*, 85.

J. P. Fox. 2010. *Bayesian Item Response Modeling: Theory and Applications*. Statistics for Social and Behavioral Sciences. Springer-Verlag New York.

Anthony J. Franze and R. Reeves Anderson. 2015. Record breaking term for amicus curiae in Supreme Court reflects new norm. *National Law Journal*, Supreme Court Brief. <http://www.nationallawjournal.com/supremecourtbrief/id=1202735095655/>, August 19, 2015.

Bryan A. Garner. 2010. Interviews with United States Supreme Court justices. In Joseph Kimble, editor, *The Scribes Journal of Legal Writing*, volume 13. American Society of Legal Writers.

Matthew Gentzkow and Jesse M Shapiro. 2010. What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71.

Sean Gerrish and David Blei. 2011. Predicting legislative roll calls from text. In *Proc. of ICML*.

Dan Goldwasser and Hal Daumé. 2014. “I object!” modeling latent pragmatic effects in courtroom dialogues. In *Proc. of EACL*.

James J. Heckman and James M. Snyder. 1996. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. Working Paper 5785, National Bureau of Economic Research.

- Ernst D. Hellinger. 1909. Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 1909(136):210–271.
- Matthew Hoffman, Francis R. Bach, and David M. Blei. 2010. Online learning for latent Dirichlet allocation. In *Advances in Neural Information Processing Systems 23*.
- Zubin Jelveh, Bruce Kogut, and Suresh Naidu. 2015. Political language in economics. *Columbia Business School Research Paper Series*, 14(57).
- John S. Justeson and Slava M. Katz. 1995. Technical terminology: Some linguistic properties and an algorithm for identification in text. *Natural Language Engineering*, 1:9–27.
- Daniel Martin Katz, Michael James Bommarito, and Josh Blackman. 2014. Predicting the behavior of the Supreme Court of the United States: A general approach. <http://ssrn.com/abstract=2463244>.
- Joseph D. Kearney and Thomas W. Merrill. 2000. The influence of amicus curiae briefs on the Supreme Court. *University of Pennsylvania Law Review*, pages 743–855.
- Benjamin E. Lauderdale and Tom S. Clark. 2014. Scaling politically meaningful dimensions using texts and votes. *American Journal of Political Science*, 58(3):754–771.
- William Li, Pablo Azar, David Larochelle, Phil Hill, James Cox, Robert C. Berwick, and Andrew W. Lo. 2013. Using algorithmic attribution techniques to determine authorship in unsigned judicial opinions. *Stanford Technology Law Review*, pages 503–534.
- Kelly J. Lynch. 2004. Best friends – Supreme Court law clerks on effective amicus curiae briefs. *Journal of Law & Politics*, 20.
- Andrew D. Martin and Kevin M. Quinn. 2002. Dynamic ideal point estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis*, 10(2):134–153.
- Daniel McFadden. 1974. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830. Available at <http://scikit-learn.org/>.
- Keith T. Poole and Howard Rosenthal. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):357–384.
- Vinodkumar Prabhakaran, Ajita John, and Dorée D. Seligmann. 2013. Who had the upper hand? ranking participants of interactions based on their relative power. In *Proc. of IJCNLP*.
- Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. The author-topic model for authors and documents. In *Proc. of UAI*.
- Yanchuan Sim, Bryan Routledge, and Noah A. Smith. 2015. The utility of text: The case of amicus briefs and the Supreme Court. In *Proc. of AAAI*.
- Harold J. Spaeth, Sara Benesh, Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, and Theodore J. Ruger. 2013. Supreme Court Database, Version 2013 Release 01. Database at <http://supremecourtdatabase.org>.
- Luke Tierney. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728.
- William Yang Wang, Elijah Mayfield, Suresh Naidu, and Jeremiah Dittmar. 2012. Historical analysis of legal opinions with a sparse mixed-effects latent variable model. In *Proc. of ACL*.