Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

The Utility of Text: The Case of Amicus Briefs and the Supreme Court

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

Abstract

We explore the idea that authoring a piece of text is an act of maximizing one's expected utility. To make this idea concrete, we consider the societally important decisions of the Supreme Court of the United States. Extensive past work in quantitative political science provides a framework for empirically modeling the decisions of justices and how they relate to text. We incorporate into such a model texts authored by amici curiae ("friends of the court" separate from the litigants) who seek to weigh in on the decision, then explicitly model their goals in a random utility model. We demonstrate the benefits of this approach in improved vote prediction and the ability to perform counterfactual analysis.

1 Introduction

Some pieces of text are written with clear goals in mind. Economists and game theorists use the word *utility* for the concept of satisfying a need or desire, and a huge array of theories and models are available for analyzing how utility-seeking agents behave. This paper takes the first steps in incorporating *text* into these models.

The Supreme Court of the United States (SCOTUS) is the highest court in the American judicial system; its decisions have far-reaching effects. While the ideological tendencies of SCOTUS' nine justices are widely discussed by press and public, there is a formal mechanism by which organized interest groups can lobby the court on a given case. These groups are known as *amici curiae* (Latin for "friends of the court," hereafter "amici," singular "amicus"), and the textual artifacts they author—known as amicus briefs—reveal explicit attempts to sway justices one way or the other. Taken alongside voting records and other textual artifacts that characterize a case, amicus briefs provide a fascinating setting for empirical study of influence through language.

We build on well-established methodology from political science known as *ideal points* for analyzing votes. Specifically, Lauderdale and Clark (2014) combined descriptive text and ideal points in a probabilistic topic model. Although the influence of amici has been studied extensively by legal scholars (Collins 2008), we are the first to incorporate them into ideal points analysis (§2). Drawing on decision theory, we then posit amici as rational agents seeking to maximize their expected utility by framing their arguments to influence justices toward a favorable outcome (§3). We derive appropriate inference and parameter estimation procedures (§4).

Our experiments (§5) show that the new approach offers substantial gains in vote prediction accuracy. More importantly, we show how the model can be used to answer questions such as: How effective were amici on each side of a case? What would have happened if some or all amicus briefs were not filed? How might an amicus have changed her brief to obtain a better outcome? Since our approach characterizes the amicus brief as a (probabilistic) function of the case parameters, our approach could also be used to ask how the amici would have altered their briefs given different merits facts or a different panel of justices. Although we focus on SCOTUS, our model is applicable to any setting where textual evidence for competing goals is available alongside behavioral response.

SCOTUS Terminology. SCOTUS reviews the decisions of lower courts and (less commonly) resolves disputes between states.[1] In a typical case, the **petitioner** writes a brief putting forward her legal argument; the **respondent** (the other party) then files a brief. These, together with a round of responses to each other's initial briefs, are collectively known as **merits briefs**. **Amicus briefs**—further arguments and recommendations on either side—may be filed by groups with an interest (but not a direct stake) in the outcome, with the Court's permission. After oral arguments (not necessarily allotted for every case) conclude, the justices vote and author one or more opinions. In this paper, we relate the votes of justices to merits and amicus briefs.

2 Ideal Point Models

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 her in this space.

Unidimensional Ideal Points The simplest model for judicial votes is a unidimensional IP model (Martin and Quinn

[1]Details about the procedures and rules of the SCOTUS can be found at http://www.uscourts.gov.

2002), which posits an IP $\psi_j \in \mathbb{R}$ for each justice j.[2] 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 with probability

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

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 the justice's ideology ψ_j: more 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). While they recover dimensions that maximize statistical fit, IP models conflate many substantive dimensions of opinion and policy, making it difficult to interpret additional dimensions.[4] Indeed, such embeddings are ignorant of the issues at stake, or any content of the case, and they cannot generalize to new cases.

Issues and Ideal Points Lauderdale and Clark (2014) incorporate text as evidence and infer dimensions of IP that are grounded in "topical" space. They build on latent Dirichlet allocation (LDA; see Blei, Ng, and Jordan 2003), a popular model of latent topics or themes in text corpora. In their model, each case i is embedded as $\boldsymbol{\theta}_i$ in a D-dimensional simplex; the dth dimension $\theta_{i,d}$ corresponds to the proportion of case i that is about issue (or, in LDA terminology, topic) d. The probability of justice j's vote is given by

$$p(v_{i,j} = \text{petitioner} \mid \boldsymbol{\psi}_j, \boldsymbol{\theta}_i, a_i, b_i) = \sigma\left(a_i + \boldsymbol{\psi}_j^\top \left(b_i \boldsymbol{\theta}_i\right)\right)$$

where $\psi_{j,d}$ is an *issue-specific* position for justice j. Therefore, the relative degree that each dimension predicts the vote outcome is determined by the text's mixture proportions, resulting in the issue-specific IP $\boldsymbol{\psi}_j^\top \boldsymbol{\theta}_i$. In their work, they inferred the mixture proportions from justices' opinions, although one can similarly use merits briefs, appeals court opinions, or any other texts that serve as evidence for inferring the issues of a case.

Lauderdale and Clark (2014) found that incorporating textual data in this manner[5] addresses the labeling problem for multidimensional models, and is especially useful for small voting bodies (e.g., SCOTUS), where estimating multidimensional models is difficult due to few observations and variation of preferences across issues.

Amici and Ideal Points The merits briefs describe the issues and facts of the case. It is our hypothesis that amicus briefs serve to "frame" the facts and, potentially, influence the case outcome. Collins (2008) argued that these organized interest groups play a significant role in shaping justices' choices. Public interest groups, such as the ACLU and Citizens United, frequently advocate their positions on any case that impinges on their goals. These briefs can provide valuable assistance to the Court in its deliberation; for example, they can present an argument not found in the merits.[6]

When filing amicus briefs, amici are required to identify the side they are supporting (or if neither). However, it is not trivial to automatically tell which side the amici are on as these intentions are not expressed consistently. We solve this by training a classifier on hand-labeled data (§4).

We propose that amici represent an attempt to shift the position of the case by emphasizing some issues more strongly or framing the case distinctly from the perspectives given in the merits briefs. The effective position of a case, previously $b_i\boldsymbol{\theta}_i$, is in our model $b_i\boldsymbol{\theta}_i + c_i^{\mathrm{p}}\boldsymbol{\Delta}_i^{\mathrm{p}} + c_i^{\mathrm{r}}\boldsymbol{\Delta}_i^{\mathrm{r}}$, where c_i^{p} and c_i^{r} are the *amicus polarities* for briefs on the side of the petitioner and respondent. $\boldsymbol{\Delta}_i^{\mathrm{p}}$ and $\boldsymbol{\Delta}_i^{\mathrm{r}}$ are the mean issue proportions of the amicus briefs on the side of the petitioner and respondent, respectively. Our amici-augmented IP model is:

$$\begin{aligned} &p(v_{i,j} = \text{petitioner} \mid \psi_j, \boldsymbol{\theta}_i, \boldsymbol{\Delta}_i, a_i, b_i, c_i) \\ &= \sigma\left(a_i + \boldsymbol{\psi}_j^\top \left(b_i\boldsymbol{\theta}_i + c_i^{\mathrm{p}}\boldsymbol{\Delta}_i^{\mathrm{p}} + c_i^{\mathrm{r}}\boldsymbol{\Delta}_i^{\mathrm{r}}\right)\right) \end{aligned} \quad (1)$$

In this model, the vote-specific IP is influenced by two forms of text: legal arguments put forth by the parties involved (merits briefs, embedded in $\boldsymbol{\theta}_i$), and by the amici curiae (amicus briefs, embedded in $\boldsymbol{\Delta}_i^{\{\mathrm{p,r}\}}$), both of which are rescaled independently by the case discrimination parameters to generate the vote probability. When either $|c_i^{\mathrm{p}}|$ or $|c_i^{\mathrm{r}}|$ is large (relative to a_i and b_i), the vote is determined by the contents of the amicus briefs. Hereafter, we let $\boldsymbol{\kappa}_i = \langle a_i, b_i, c_i^{\mathrm{p}}, c_i^{\mathrm{r}}\rangle$.

By letting $\boldsymbol{\Delta}_i^s$ be the average mixture proportions inferred from text of briefs supporting side s, we implicitly assume that briefs supporting the same side share a single parameter, and individual briefs on one side influence the vote-specific IP equally. While Lynch (2004) and others have argued that some amici are more effective (i.e., influence on justices' votes varies across amicus authors), our model captures the collective effect of amicus briefs and is simple.

3 Amici as Agents

In the previous section, the IP models focus on justices' positions embedded in a continuous space. However, we want to account for the fact that amici are purposeful decision makers who write briefs hoping to sway votes on a case. Suppose we have an amicus curiae supporting side s (e.g., petitioner), which is presided by a set of justices, $\mathcal{J}$. The amicus is interested in getting votes in favor of her side, that is, $v_j = s$. Thus, we assume that she has a simple evaluation

[2] Martin and Quinn (2002) describe a dynamic unidimensional IP model where justice IPs vary over time. In this work, we assume each justice's IP is fixed over time, for simplicity.

[3] This model is also known as a two parameter logistic model in item response theory literature (Fox 2010), where a_i is "difficulty" and b_i is "discrimination."

[4] A seminal finding of Poole and Rosenthal (1985) is that two dimensions, corresponding to left-right ideology and geographical latitude, explain most of the variance in U.S. Congressional votes.

[5] Of course, LDA is not the only way to "embed" a case in a simplex. One can take advantage of expert categorization of case issues. For example, Gerrish and Blei (2012) used bill labels as supervision to infer the proportions of issues.

[6] On occasion, SCOTUS may adopt a position not advanced by either side, but instead urged solely by an amicus brief. Some notable cases are: *Mapp v. Ohio*, 367 U.S. 643, 646 (1961) and more recently, *Turner v. Rogers*, 131 S. Ct. 2507 (2011).

function over the outcome of votes $v_1, \ldots, v_9$,

$$u(v_1, v_2, \ldots, v_9) = \sum_{j \in \mathcal{J}} \mathbb{I}(v_j = s), \tag{2}$$

where $\mathbb{I}$ is the indicator function. This is her **utility**.

Cost of writing. In addition to the policy objectives of an amicus, we need to characterize her "technology" (or "budget") set. We do this by specifying a cost function, C, that is increasing in difference between $\boldsymbol{\Delta}$ and the "facts" in $\boldsymbol{\theta}$:

$$C(\boldsymbol{\Delta}, \boldsymbol{\theta}) = \tfrac{\xi}{2} \|\boldsymbol{\Delta} - \boldsymbol{\theta}\|_2^2$$

where $\xi > 0$ is a hyperparameter controlling the cost (relative to the vote evaluation). The function captures the notion that amicus briefs cannot be arbitrary text; there is disutility or effort required to carefully frame a case, or the monetary cost of hiring legal counsel. The key assumption here is that framing is costly, while simply matching the merits is easy (and presumably unnecessary). Note the role of the cost function is analogous to regularization in other contexts.

Expected utility. The outcome of the case is uncertain, so the amicus' objective will consider her *expected* utility:[7]

$$\max_{\boldsymbol{\Delta}} \ \mathbb{E}_{\boldsymbol{\Delta}}[u(v_1, \ldots, v_9)] - \tfrac{\xi}{2} \|\boldsymbol{\Delta} - \boldsymbol{\theta}\|_2^2$$

When an amicus writes her brief, we assume that she has knowledge of the justices' IPs, case parameters, and contents of the merits briefs, but ignores other amici.[8] As such, taking linearity of expectations, we can compute the expected utility for an amicus on side s using Eq. 1:[9]

$$\max_{\boldsymbol{\Delta}} \ \sum_{j \in \mathcal{J}} \sigma\left(a + \boldsymbol{\psi}_j^\top (b\boldsymbol{\theta} + c^s \boldsymbol{\Delta})\right) - \tfrac{\xi}{2} \|\boldsymbol{\Delta} - \boldsymbol{\theta}\|_2^2$$

Random utility models. There are several conceivable ways to incorporate amici's optimization into our estimation of justices' IP. We could maximize the likelihood and impose a constraint that $\boldsymbol{\Delta}$ solve our expected utility optimization (either directly or by checking the first order conditions).[10] Or, we can view such (soft) constraints as imposing a prior on $\boldsymbol{\Delta}$:

$$p_{\text{util}}(\boldsymbol{\Delta}) \propto \mathbb{E}_{\boldsymbol{\Delta}}[u(v_1, \ldots)] + \xi(1 - \tfrac{1}{2}\|\boldsymbol{\Delta} - \boldsymbol{\theta}\|_2^2) \tag{3}$$

[7]We use an evaluation function that is linear in votes for simplicity. The scale of the function is unimportant (expected utility is invariant to affine transformations). However, we leave for future work other specifications of the evaluation function; for example a function that places more emphasis on the majority vote outcome.

[8]Capturing strategic amici agents (a petitioner amicus choosing brief topics considering a respondent amicus' brief) would require a complicated game theoretical model and, we conjecture, would require a much richer representation of policy and goals.

[9]The first-order conditions for amicus' purposeful maximization with respect to $\boldsymbol{\Delta}$ lead to interesting brief writing trade offs, which can be found in supplementary §A.

[10]This is reminiscent of learning frameworks where constraints are placed on the posterior distributions (Chang, Ratinov, and Roth 2007; Ganchev et al. 2010; McCallum, Mann, and Druck 2007). However, the nonlinear nature of our expectations makes it difficult to optimize and characterize the constrained distribution.

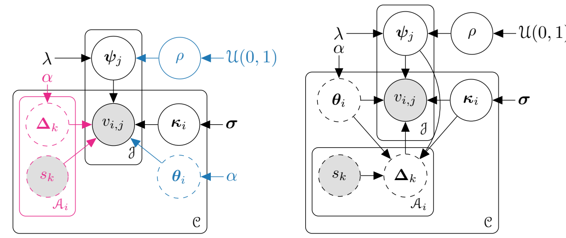

Figure 1: Plate diagrams for the IP models (left) and random utility model (right). $\mathcal{J}$, $\mathcal{C}$ and $\mathcal{A}_i$ are the sets of justices, cases, and amicus briefs (for case i), respectively. $\boldsymbol{\psi}_j$ is the IP for justice j; $\boldsymbol{\kappa}_i$ is the set of case parameters a_i, b_i, c_i^{p} and c_i^{r} for case i; and $\boldsymbol{\alpha}, \boldsymbol{\sigma}, \lambda$, and ρ are hyperparameters. The mixture proportion nodes (dashed) are fixed in our estimation procedure. On the left, black nodes comprise the basic IP model, blue nodes are found in both the issues IP and amici IP models, while magenta nodes are found only in the amici IP model.

where the constant is added so that p_{util} is non-negative. Note, were the utility negative, the amici would have chosen not to write a the brief. This approach is known as a **random utility model** in the econometrics discrete-choice literature (McFadden 1974). Random utility models relax the precision of the optimization by assuming that agent preferences also contain an idiosyncratic random component. Hence, the behavior we observe (i.e., the amicus' topic mixture proportions) has a likelihood that is proportional to expected utility. Considering all the amici, we estimate the full likelihood

$$\mathcal{L}(\boldsymbol{w}, \boldsymbol{v}, \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\Delta}, \boldsymbol{\kappa}) \times \left[\prod_{k \in \mathcal{A}} p_{\text{util}}(\boldsymbol{\Delta}_k)\right]^\eta \tag{4}$$

where $\mathcal{L}(\cdot)$ is the likelihood of our amici IP model (Eq. 1), and η is a hyperparameter that controls influence of utility on parameter estimation.

Eq. 4 resembles *product of experts* model (Hinton 2002). For the likelihood of votes in a case to be maximized, it is necessary that no individual component—generative story for votes, amicus briefs—assigns a low probability. Accordingly, this results in a principled manner for us to incorporate our assumptions about amici as rational decision makers, each of whom is an "expert" with the goal of nudging latent variables to maximize her own expected utility.

4 Learning and Inference

Model priors. The models we described above combine ideal points, topic models, and random utility; they can be estimated within a Bayesian framework. Following Lauderdale and Clark (2014), we place Gaussian priors on the justice and case parameters: $\rho \sim \mathcal{U}(0, 1)$, $\boldsymbol{\psi}_j \sim \mathcal{N}(\mathbf{0}, \lambda\mathbf{I} + \rho\mathbf{1})$ for each justice j, and $\boldsymbol{\kappa}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma})$ for each case i. The positive off-diagonal elements of the covariance matrix for justice IPs 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. Fig. 1 presents the plate diagram for the IP models (§2) on the left and the random utility model (§3) on the right.

For the non-utility IP models involving text (§2), LDA is used to infer the latent topic mixtures of text associated with the case and amicus briefs, thus $\boldsymbol{\theta}_i$ and $\boldsymbol{\Delta}_k$ are both drawn from a symmetric Dirichlet distribution with hyperparameter α and tokens in both sets of texts are drawn from shared topic-word distributions.[11]

Our random utility model can be described through a similar generative story. Instead of drawing amicus briefs ($\boldsymbol{\Delta}$) from a Dirichlet, they are drawn from the expected utility distribution (Eq. 3). The right side of Fig. 1 shows the corresponding plate diagram. Importantly, note that $\boldsymbol{\Delta}$ here serves as direct evidence for the justice and case parameters, rather than influencing them through v-structures.

Parameter estimation. We decoupled the estimation of the topic mixture parameters as a stage separate from the IP parameters. This approach follows Lauderdale and Clark (2014), who argued for its conceptual simplicity: the text data define the rotation of a multidimensional preference space, while the second stage estimates the locations in that space. We found in preliminary experiments that similar issue dimensions result from joint vs. stage-wise inference, but that the latter is much more computationally efficient.

Using LDA,[12] $\boldsymbol{\theta}$ (and, where relevant, $\boldsymbol{\Delta}$) are estimated, then fixed to their posterior means while solving for justice parameters $\boldsymbol{\psi}$ and case parameters $\boldsymbol{\kappa}$. For the second stage, we used Metropolis within Gibbs (Tierney 1994), a hybrid MCMC algorithm, to sample the latent parameters from their posterior distributions. We sampled $\boldsymbol{\kappa}_i$ for each case and $\boldsymbol{\psi}_j$ for each justice blockwise from a multivariate Gaussian proposal distribution, tuning the diagonal covariance matrix to a target acceptance rate of 15–45%. Likewise, ρ is sampled from a univariate Gaussian proposal, with its variance tuned similarly. For our random utility model, we used the same MCMC approach in sampling the latent IRT variables, but include the expected utility term for each brief in the likelihood function (eq. 4). Details of our sampler and hyperparameter settings can be found in the supplementary materials (§B), while topics and justices' IPs estimated by our model are found in §D and §E, respectively.

Data. We focused on 23 terms of the Court from 1990–2012 (Spaeth et al. 2013),[13] using texts from LexisNexis.[14] We concatenate each of the 2,074 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 ($\boldsymbol{\theta}$; i.e., merits briefs are treated as "facts of the case"). We did not make use of case opinions as did Lauderdale and Clark (2014) because opinions are written after votes are cast, tainting the data for predictive modeling. Each amicus brief is treated as a single document.

As the amicus briefs in our dataset were not explicitly labeled with the side that they support, and manually labeling each brief would be a tedious endeavor, we built a classifier to automatically label the briefs with the side the amici are supporting, taking advantage of cues in the brief content that *strongly* signal the side that the amici is supporting (e.g., "in support of petitioner" and "affirm the judgement").

Additionally, we find that using only phrases (instead of standard bag-of-words) gave us more interpretable topics (Sim et al. 2013). Details of our phrase extraction, data preprocessing steps, and brief "side" classification are in the supplementary materials (§C).

5 Experiments and Analysis

5.1 Vote Prediction

We evaluate each model's ability to predict how justices would vote on a case out of the training sample. To compute the probability of justices' votes, we first infer the topic mixture proportions for the case's merits briefs ($\boldsymbol{\theta}$), and amicus briefs ($\boldsymbol{\Delta}$). Given all the justice's IPs $\boldsymbol{\psi}_j$, we find the most likely vote outcome for the case by integrating over the case parameters $\boldsymbol{\kappa}$:

$$\arg\max_{\boldsymbol{v}} \int_{\boldsymbol{\kappa}} p(\boldsymbol{\kappa} \mid \boldsymbol{\sigma}) \prod_{j \in \mathcal{J}} p(v_j \mid \boldsymbol{\psi}_j, \boldsymbol{\theta}, \boldsymbol{\Delta}, \boldsymbol{\kappa}) \times \prod_{k \in \mathcal{A}} p_{\text{util}}(\boldsymbol{\Delta}_k \mid \boldsymbol{\psi}, \boldsymbol{\theta}, \boldsymbol{\Delta}, \boldsymbol{\kappa}, s_k)$$

where s_k is the side brief k supports, and the multiplier is the expected utility term (Eq. 3) which is ignored for the non-utility based models.

Due to the specification of IP models, the probability of a vote is a logistic function of the vote-specific IP, which is a symmetric function implying that justice j's probability of voting towards the petitioner will be the same as if she voted for the respondent when we negate the vote-specific IP. Thus, we would not be able to distinguish the actual side that the justice will favor, but we can identify the most likely partitioning of the justices into two groups.[15] We can then evaluate, for each case, an average pairwise accuracy score,

$$\binom{9}{2}^{-1} \sum_{j,j' \in \mathcal{J}: j \neq j'} \mathbb{I}[\mathbb{I}[\hat{v}_j = \hat{v}_{j'}] = \mathbb{I}[v^*_j = v^*_{j'}]]$$

where $\hat{\boldsymbol{v}}$ ($\boldsymbol{v}^*$) are predicted (actual) votes.

We performed 5-fold cross validation, and present the vote partition accuracy in Table 1. We have two naïve baselines, (i) where all justices vote unanimously, and (ii) where we trained an ℓ_1-regularized logistic regression classifier for each justice using the concatenated topic proportions of $\boldsymbol{\theta}$ and $\boldsymbol{\Delta}$ as features for each case. The baselines exhibit similar accuracies, and perform better than when adjusting for issues and/or amici. This suggests that justices' votes do not always align with their IPs, and that topic models alone may be inadequate for representing IPs. Furthermore, we believe

[11] We omit details of LDA, as it is widely known.

[12] We used the C++ implementation of LDA by Liu et al. (2011).

[13] The unit of analysis is the case citation, and we select cases where the type of decision equals 1 (orally argued cases with signed opinions), 5 (cases with equally divided vote), 6 (orally argued per curiam cases), or 7 (judgements of the Court). In addition, we dropped cases where the winning side was not clear (i.e., coded as "favorable disposition for petitioning party unclear").

[14] http://www.lexisnexis.com

[15] Given the partitioning of justices, domain experts should be able to identify the side each group of justices would favor.

Model	Accuracy
Logistic regression w/ topics	0.715 ± 0.008
Unanimous	0.714 ± 0.003
Unidimensional IP	0.583 ± 0.037
Issues IP	0.671 ± 0.008
Amici IP	0.690 ± 0.021
Random utility IP	$\mathbf{0.742} \pm 0.006$

Table 1: Average pairwise vote partition accuracy (five-fold cross-validation).

there may be insufficient information to learn the amicus polarity case parameters (c^s) in the amici IP model (which is slightly better than the issues IP model). However, in the random utility model, amici-agents-experts weigh in, providing additional signals for estimating these parameters, achieving significantly (paired samples t-test, $p < 0.001$) better predictive accuracy than the baseline.

As an additional qualitative validation of our approach, we compared the log-likelihoods between models that consider vs. ignore amicus briefs. We found correlations with anecdotal evidence of how justices view the influence of amicus briefs (see §F for details).

5.2 Post Hoc Analysis of Votes

On a case level, we can tease apart the relative contribution each textual component to a justice's decision by analyzing the case parameters learnt by our random utility model. By zeroing out various case parameters, and plotting them, we can visualize the different impact that each type of text has on a justice's vote-specific IP. For example, Fig. 2 shows the vote-specific IP estimates of justices for the 2011 term death penalty case *Maples v. Thomas* (132 S. Ct. 912). The issues-

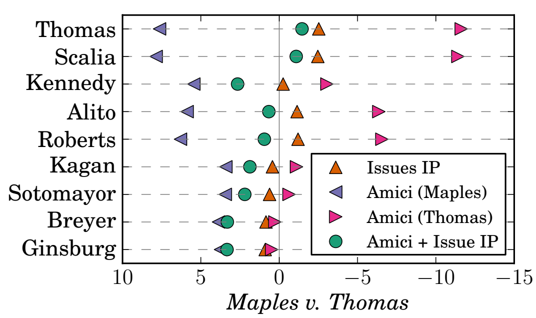

Figure 2: Vote-specific IP estimates decomposed into different influences on each justice's vote on *Maples v. Thomas*. An IP towards the left (right) indicates higher log-odds of vote that is favorable to Maples (Thomas).

only IPs are computed by zeroing out both the amicus polarity parameters (c^{p} and c^{r}). On the other hand, the IP due to amicus briefs supporting Maples (Thomas) is computed by zeroing out only c^{r} (c^{p}). We observe that the issues-only IPs are aligned with each justice's (widely known) ideological stance on the issue of capital punishment. For instance, the issues-only IPs of Thomas, Scalia, Alito, and Roberts, the strong conservative bloc, favor the respondents (that Maples should not be awarded relief); so did Kennedy, who is widely recognized as the swing justice. When effects of all amicus briefs are taken into account, the justices' IPs shift toward Maples with varying magnitudes, with the result reflecting the actual ruling (7–2 with Thomas and Scalia dissenting).

5.3 Counterfactual Analysis

Following Pearl (2000), we can query the model and perform counterfactual analyses using the vote prediction algorithm (§5.1). As an illustration, we consider *National Federation of Independent Business (NFIB) v. Sebelius (HHS)* (132 S. Ct. 2566), a landmark 2011 case in which the Court upheld Congress's power to enact most provisions of the Affordable Care Act (ACA; "Obamacare").[16]

In the merits briefs, the topics discussed revolve around *interstate commerce* and the *individual mandate*, while there is an interesting disparity in topics between briefs supporting NFIB and HHS.[17] Notably, amici supporting NFIB are found, on average, to use language concerning *individual mandate*, while amici supporting HHS tend to focus more on topics related to *interstate commerce*. This is commensurate with the main arguments put forth by the litigants, where NFIB was concerned about the overreach of the government in imposing an individual mandate, while HHS argued that healthcare regulation by Congress falls under the Commerce Clause. Our model was most uncertain about Roberts and Kennedy, and wrong about both (Fig. 3 top).

Choosing sides. The first type of counterfactual analysis that we introduce is, "What if no (or only one side's) amicus briefs were submitted in the ACA case?" To answer it, we hold the case out of the training set and attempt to predict the votes under the hypothetical circumstances with the random utility model. Fig. 3 (top) shows the resulting IP of hypothetical situations where no amicus briefs were filed, or when only briefs supporting one side are filed. If no amici filed briefs, the model expects that all but Kagan and Sotomayor would favor NFIB, but with uncertainty. With the inclusion of the amicus briefs supporting NFIB, the model becomes more confident that the conservative bloc of the court would vote in favor of NFIB (except for Alito). Interestingly, the model anticipates that the same briefs will turn the liberals *away*. In contrast, the briefs on HHS' side have more success in swaying the case in their favor, especially the crucial swing vote of Kennedy (although it turned out that Kennedy sided with the conservative bloc, and Roberts emerged as the deciding vote in HHS favor). Consequently, the model can provide insights about judicial decisions, while postulating different hypothetical situations.

[16] The case attracted much attention, including a record 136 amicus briefs, of which 76 of these briefs are used in our dataset. 58 (of the 76) were automatically classified as supporting NFIB.

[17] The merits briefs were estimated at 41% and 20% on the *interstate commerce* and the *individual mandate* topics, respectively. NFIB amicus briefs were 15% on *interstate commerce* and 41% on *individual mandate*; these figures switch to 36% and 22% for HHS amicus briefs.

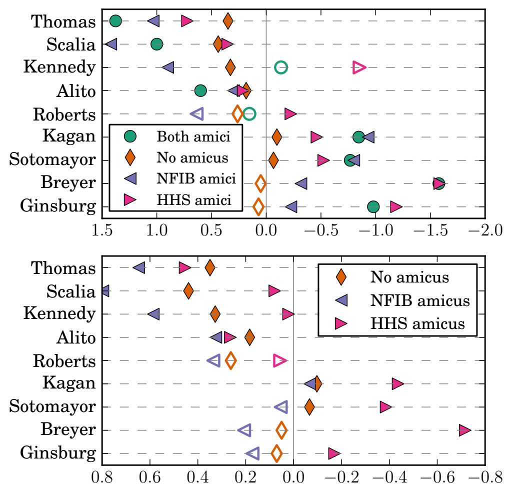

Figure 3: Counterfactual analyses for *NFIB v. Sebelius (HHS)*. *Top*: What if amicus briefs for one side were not filed? *Bottom*: What if a single amicus files an "optimally" written brief? An IP towards the left (right) indicates higher log-odds of vote favorable to NFIB (HHS). Hollow markers denote that prediction differed from the actual outcome.

Choosing what to write. Another counterfactual analysis we can perform, more useful from the viewpoint of the amicus, is, "how should an amicus frame arguments to best achieve her goals?" In the context of our model, such an amicus would like to choose the topic mixture $\boldsymbol{\Delta}$ to maximize her expected utility (Eq. 3). Ideally, one would compute such a topic mixture by maximizing over both $\boldsymbol{\Delta}$ and vote outcome $\boldsymbol{v}$, while integrating over the case parameters. We resort to a cheaper approximation: analyzing the filer's expected utility curve over two particular topic dimensions: the *individual mandate* and *interstate commerce* topics. That is, we compute the expected utility curve (see supplementary §G) faced by a single amicus as we vary the topic proportions of *individual mandate* and *interstate commerce* topics over multiples of 0.1.

Consequently, the amicus who supports NFIB can expect to maximize their expected utility (5.2 votes at a cost of 0.21) by "spending" about 70% of their text on *individual mandate*. On the other hand, the best that an amicus supporting HHS can do is to write a brief that is 80% about *interstate commerce*, and garner 4.7 votes at a cost of 0.31. We plot the justices' predicted IPs in Fig. 3 (bottom) using these "best" proportions. The "best" proportions IPs are different (sometimes worse) from that in Fig. 3 (top) because in the latter, there are multiple amici influencing the case parameters (through their utility functions) and other topics are present which will sway the justices. From the perspective of an amicus supporting HHS, the two closest swing votes in the case are Roberts and Kennedy; we know *a posteriori* that Roberts sided with HHS.

5.4 Discussion

Our model makes several simplifying assumptions: (i) it ignores the effects of other amici on a single amicus' writing; (ii) amici are treated modularly, with a multiplicative effect and no consideration of diminishing returns, temporal ordering, or reputation; (iii) the cost function does not capture the intricacies of legal writing style (i.e., choice of citations, artful language, etc.); (iv) the utility function does not fully capture the agenda of each individual amicus; (v) each amicus brief is treated independently (i.e., no sharing across briefs with the same author), as we do not have access to clean author metadata. Despite these simplifications, the model is a useful tool for quantitative analysis and hypothesis generation in support of substantive research on the judiciary.

6 Related Work

Poole and Rosenthal (1985) introduced the IP model, using roll call data to infer latent positions of lawmakers. Since then, many varieties of IP models have been proposed for different voting scenarios: IP models for SCOTUS (Martin and Quinn 2002), multidimensional IP models for Congressional voting (Heckman and Snyder 1996; Clinton, Jackman, and Rivers 2004), grounding multidimensional IP models using topics learned from text of Congressional bills (Gerrish and Blei 2012) and SCOTUS opinions (Lauderdale and Clark 2014).

Amici have been studied extensively, especially their influence on SCOTUS (Caldeira and Wright 1988; Kearney and Merrill 2000; Collins 2008; Corley, Collins, and Hamner 2013). Collins (2007) found that justices can be influenced by persuasive argumentation presented by organized interests. These studies focus on ideology metadata (liberal/conservative slant of amici, justices, decisions, etc.), disregarding the rich signals encoded in the text of these briefs, whereas we use text as evidence of utility maximizing behavior to study the influence of amicus curiae.

Our model is also related to Gentzkow and Shapiro (2010) who model the purposeful "slant" of profit-maximizing newspapers looking to gain circulation from consumers with a preference for such slant. More generally, extensive literature in econometrics estimates structural utility-based decisions (Berry, Levinsohn, and Pakes 1995, *inter alia*).

In addition to work on IP models, researchers have used court opinions for authorship (Li et al. 2013) and historical analysis, while oral argument transcripts have been used to study power relationships (Danescu-Niculescu-Mizil et al. 2012; Prabhakaran, John, and Seligmann 2013) and pragmatics (Goldwasser and Daumé III 2014).

7 Conclusion

We have introduced a random utility model for persuasive text; it is similar to a classical generative model and can be estimated using familiar algorithms. The key distinction is that persuasive text is modeled as a function of the addressee and the particulars of the matter about which she is being convinced; authors are agents seeking to maximize their expected utility in a given scenario. In the domain of SCOTUS, this leads to improved vote prediction performance, as the model captures the structure of amicus briefs better than simpler treatments of the text. Secondly, and more importantly, our model is able to address interesting counterfactual questions. Were some amicus briefs not filed, or had they been written differently, or had the facts of the case been presented differently, or had different justices presided, our approach can estimate the resulting outcomes.

Acknowledgements

This research was supported in part by an A*STAR fellowship to Y. Sim, NSF grant IIS-1211277, a Google research award, and by computing resources from the Pittsburgh Supercomputing Center.

References

Berry, S.; Levinsohn, J.; and Pakes, A. 1995. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.

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

Caldeira, G. A., and Wright, J. R. 1988. Organized interests and agenda setting in the U.S. Supreme Court. *American Political Science Review* 82:1109–1127.

Chang, M.-W.; Ratinov, L.; and Roth, D. 2007. Guiding semi-supervision with constraint-driven learning. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 280–287.

Clinton, J.; Jackman, S.; and Rivers, D. 2004. The statistical analysis of roll call data. *American Political Science Review* 98:355–370.

Collins, P. M. 2007. Lobbyists before the U.S. Supreme Court: Investigating the influence of amicus curiae briefs. *Political Research Quarterly* 60(1):55–70.

Collins, P. M. 2008. *Friends of the Supreme Court: Interest Groups and Judicial Decision Making*. Oxford University Press.

Corley, P.; Collins, P. M.; and Hamner, J. 2013. The influence of amicus curiae briefs on U.S. Supreme Court opinion content. In *APSA 2013 Annual Meeting Paper*.

Danescu-Niculescu-Mizil, C.; Lee, L.; Pang, B.; and Kleinberg, J. 2012. Echoes of power: Language effects and power differences in social interaction. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, 699–708.

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

Ganchev, K.; Graça, J. a.; Gillenwater, J.; and Taskar, B. 2010. Posterior regularization for structured latent variable models. *Journal of Machine Learning Research* 11:2001–2049.

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

Gerrish, S., and Blei, D. M. 2012. How they vote: Issue-adjusted models of legislative behavior. In Pereira, F.; Burges, C.; Bottou, L.; and Weinberger, K., eds., *Advances in Neural Information Processing Systems 25*, 2753–2761. Curran Associates, Inc.

Goldwasser, D., and Daumé III, H. 2014. "I object!" modeling latent pragmatic effects in courtroom dialogues. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 655–663.

Heckman, J. J., and Snyder, Jr., J. M. 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.

Hinton, G. E. 2002. Training products of experts by minimizing contrastive divergence. *Neural Computation* 14(8):1771–1800.

Kearney, J. D., and Merrill, T. W. 2000. The influence of amicus curiae briefs on the Supreme Court. *University of Pennsylvania Law Review* 743–855.

Lauderdale, B. E., and Clark, T. S. 2014. Scaling politically meaningful dimensions using texts and votes. *American Journal of Political Science* 58(3):754–771.

Li, W.; Azar, P.; Larochelle, D.; Hill, P.; Cox, J.; Berwick, R. C.; and Lo, A. W. 2013. Using algorithmic attribution techniques to determine authorship in unsigned judicial opinions. *Stanford Technology Law Review* 503–534.

Liu, Z.; Zhang, Y.; Chang, E. Y.; and Sun, M. 2011. PLDA+: Parallel latent Dirichlet allocation with data placement and pipeline processing. *ACM Transactions on Intelligent Systems and Technology*. Software available at http://code.google.com/p/plda.

Lynch, K. J. 2004. Best friends – Supreme Court law clerks on effective amicus curiae briefs. *Journal of Law & Politics* 20:33.

Martin, A. D., and Quinn, K. M. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953–1999. *Political Analysis* 10(2):134–153.

McCallum, A.; Mann, G.; and Druck, G. 2007. Generalized expectation criteria. Technical Report UM-CS-2007-60, University of Massachusetts, Amherst, MA 01003, USA.

McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In Zarembka, P., ed., *Frontiers in Econometrics*. New York: Academic Press. 105–142.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Poole, K. T., and Rosenthal, H. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science* 29(2):357–384.

Prabhakaran, V.; John, A.; and Seligmann, D. D. 2013. Who had the upper hand? ranking participants of interactions based on their relative power. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, 365–373.

Sim, Y.; Acree, B. D. L.; Gross, J. H.; and Smith, N. A. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 91–101.

Spaeth, H. J.; Benesh, S.; Epstein, L.; Martin, A. D.; Segal, J. A.; and Ruger, T. J. 2013. Supreme Court database, version 2013 release 01. Database at http://supremecourtdatabase.org.

Tierney, L. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics* 22(4):1701–1728.