Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Inferring Multi-Dimensional Ideal Points for US Supreme Court Justices

Mohammad Raihanul Islam*, **K.S.M. Tozammel Hossain***,
Siddharth Krishnan, and **Naren Ramakrishnan**
Discovery Analytics Center, Department of Computer Science, Virginia Tech
Email: {raihan8, tozammel, siddkris, naren}@cs.vt.edu

Abstract

In Supreme Court parlance and the political science literature, an ideal point positions a justice in a continuous space and can be interpreted as a quantification of the justice's policy preferences. We present an automated approach to infer such ideal points for justices of the US Supreme Court. This approach combines topic modeling over case opinions with the voting (and endorsing) behavior of justices. Furthermore, given a topic of interest, say the Fourth Amendment, the topic model can be optionally seeded with supervised information to steer the inference of ideal points. Application of this methodology over five years of cases provides interesting perspectives into the leaning of justices on crucial issues, coalitions underlying specific topics, and the role of swing justices in deciding the outcomes of cases.

1 Introduction

The Supreme Court of the United States (SCOTUS) is one of the most fascinating institutions of study for political scientists, constitutional scholars, and laymen alike. All members (justices) of this nine-member body are appointed by the President and have lifetime tenure. As a result, Supreme Court Justices can have a long-lasting influence on the country's affairs. Their opinions and decisions are scrutinized very carefully in order to fathom (and forecast) the overall direction of the country on key societal issues.

It has long been accepted that how a justice votes along with his/her rationale for the vote is a window into the justice's judicial temperament, personal philosophy, and policy ideology (Lauderdale and Clark 2014; Martin and Quinn 2002). Since cases build upon each other for rationale, opinions are typically well reasoned, carefully scoped, and set in the context of prior rulings. The classical approach to model ideological leanings of justices is using ideal point models, a well-established methodology in political science for analyzing legislative and judicial voting records (Sim, Routledge, and Smith 2014). An ideal point for a justice is a position of the justice in a continuous space, which can be either unidimensional (Martin and Quinn 2002) or multidimensional (Lauderdale and Clark 2014), and is interpreted as a quantification of policy preferences.

Existing research for modeling ideological preferences and voting behavior of justices uses metadata of cases as a primary data source and represents ideological preferences of justices with real values in a dimension (Martin and Quinn 2002; Segal and Cover 1989). Such simplistic models do not capture ideological preferences beyond a vanilla liberal-conservative ideological spectrum, and limit our ability to pose and answer more interesting research questions regarding the leanings of justices (Lauderdale and Clark 2014). Although there is an abundance of textual artifacts (e.g., opinions, briefs) associated with various stages of a case, few methods incorporate such textual data into their modeling.

A notable exception is the work of Lauderdale and Clark (2014) who propose the idea of higher-dimensional preferences for justices. They propose to solve the identification problem in multidimensional ideal point estimation by harvesting text underlying the opinions using topic models. We were motivated by this work and interested in generalizing it since, in modeling higher-dimensional preferences, this approach merges all opinions in a given case to estimate the proportion of issues discussed in the case and uses these issue proportions for estimating ideal points of justices. Merging all the opinions for a case together without considering their types and endorsement attributes is likely to lose many important aspects of judicial preferences. In particular, as we will show, this approach tends to dichotomize justices into partisan divides (which holds for only some issues). Incorporating opinions into the models by modeling their types and endorsement information promises to capture greater rationale behind judicial decisions. Note that there are methods that model legislative voting behavior and preferences of lawmakers using textual data (e.g., bills) (Gerrish and Blei 2012; Gu et al. 2014). As there are differences between bills (mostly single authored) and opinions (multi-authored and multi-endorsed) these models are not applicable to fine-grained modeling of the Court's opinions.

We propose SCIPM (Supreme Court Ideal Point Miner), a new approach to model justices' judicial preferences and voting behavior by mining the different types of opinions delivered in cases. SCIPM is a generative model that couples the modeling of opinions written and taken part in by justices with the modeling of their ideological preferences and voting behavior. SCIPM estimates multidimensional ideal points and situates them in either an inferred or imposed

*Both authors contributed equally to this work.

topic space, enabling the methodology to adapt to new issues and topics that arise over a court's term. In particular, because SCIPM is a true joint generative model of opinions and decisions, it can be used in a predictive strategy to forecast the outcomes of specific cases (by specific justices). This approach can thus be leveraged to answer questions about the dynamics of coalitions underlying voting behavior, e.g., how will the justices vote in a case about a particular topic? Which coalitions are likely to become relevant? Who is likely to be the swing justice in a given case or term? Thus SCIM can be used for both descriptive and predictive goals.

2 SCOTUS for the Data Miner

We begin with a brief introduction to the functioning of the US Supreme Court. The Supreme Court of the U.S. (SCOTUS) resolves various types of cases, which are appealed from the lower courts, or that have a State as a party, or those cases that affect ambassadors, public ministers, and consuls[1]. The Court consists of nine justices who hold lifetime appointments. New appointments are made typically when an existing justice retires, leaving a vacancy. Since appointments are made by the US President (and ratified by the Senate), the President has immense control in shaping the direction of the court (and thus this is one of the issues that Presidential candidates are typically quizzed on). It is hence normal (but technically incorrect) to think of justices as Republican or Democratic (depending on the party of the President who appointed the given justice). A case in SCOTUS goes through various stages, some of which are visible while others are typically out of the public view. The visible phases are the court's announcement of its decision to grant certiorari (i.e., reviewing the verdict of a lower court), the court's oral arguments, and the announcement of the court's decision. After the oral arguments of a case are completed, the justices meet in conference and determine the majority viewpoint of the court. A justice is assigned to write the opinion of the court and the other justices either support the opinion by "joining" the opinion or express reservations with her own opinion. Thus, when a verdict on a case is announced, there is typically a summary, the (majority) opinion of the court, along with optional concurring and dissenting opinion(s).

The opinions written or joined in by a justice reflects his or her judicial standpoints or preferences. An opinion can be joined and written by a number of justices. We say that a justice is voting for an opinion if he or she either writes or joins the opinion. Henceforth, for the purposes of this paper, we will use the word "vote" and "join" interchangeably. The majority opinion is called the Court Opinion (CO). However, justices who disagree on the decision of the majority may write their own opinion, which is known as a Dissenting Court Opinion (DCO). Justices sometimes agree with the outcome but write a separate opinion called a Concurring Court Opinion (CCO). Justices can write a Concurring in Judgment (CIJ) opinion when they agree with the opinion of the court, but not with the arguments presented in the CO. And finally, justices can write Concurring and Dissenting (CND) opinions wherein they simultaneously agree and disagree (partially) with the opinion of the court (CO). Note that voters of CCO, and CIJ agree with the court's decision but differ on the reasoning to reach this decision. In other words, by writing or joining CO, CCO and CIJ opinions, the justice's ideal points and the court decisions are aligned. It is typical for a justice to join in the CO, CCO and CIJ opinions for the same case. It is also normal to have multiple concurring and dissenting opinions (from different justices) for a case.

To instantiate the above concepts, consider the decision of SCOTUS on the Patient Protection and Affordable Care Act (PPACA), also known as *Obamacare* (Case No. 11-393)[2], a controversial case in recent times. This case was ruled 5-4, where many justices voted with reservations. Chief Justice Roberts delivered the opinion of the court which was joined in part by two sets of justices. These two sets of justices wrote their own opinions. In the minority side, there were two sets of justices who wrote their dissenting opinions. This case is a perfect example where we observe the multitude of CO, CCO, DCO, CIJ, and CND opinions. (Note that some justices participated in multiple opinions.)

3 Related Work

In this section we discuss pertinent research in three areas: supreme court and legislation modeling, topic modeling, and recommender systems.

There is an abundance of research conducted on Supreme Court decisions (Bailey 2012). Katz, Bommarito-II, and Blackman (2014) propose a decision tree classifier for predicting the court's decision using metadata about cases and the justices' backgrounds. Lauderdale and Clark (2014) combined the votes and opinions of the cases to model multidimensional ideal points of justices: each dimension is a topic which is identified using topic modeling. Along similar lines, Sim, Routledge, and Smith (2014) present a method for predicting the voting patterns of justices using amicus briefs. In general, thus, most analysis of SCOTUS cases focuses on prediction of votes on cases (Guimerà and Sales-Pardo 2011), estimation of judicial preferences (Martin and Quinn 2002; Lauderdale and Clark 2014), identification of polarities of justices (Clark 2008), and analysis of judicial activism (Bailey 2012) or restraint (Pickerill 2013). The most well-known measures of justices' ideology are the Martin–Quinn score (2002) and the Segal–Cover score (1989), which represent a justice's ideology with a real score on a liberal-conservative spectrum.

Research has also been conducted into predicting the voting patterns in legislative bills; here, the texts of bills are studied to help characterize how senators and congressmen will vote on bills. Gerrish and Blei (2012) propose an issue-adjusted ideal point model to compute the ideal points of legislators. A recent study focused on multi-dimensional ideal points instead of single-dimensional ideal points for legislators (Gu et al. 2014). The authors use both text and voting

[1] http://www.supremecourt.gov/about/briefoverview.aspx

[2] http://www.supremecourt.gov/opinions/11pdf/11-393c3a2.pdf

data for modeling such multidimensional ideal points.

Topic modeling algorithms such as LDA (Blei, Ng, and Jordan 2003) represent documents as a mixture of topics where each topic is a distribution over words in a vocabulary. Many variants of LDA have been proposed in the literature; most relevant to our purposes are variants aimed at modeling authorship and word distributions, e.g., the Author-Topic model (ATM) (Rosen-Zvi et al. 2004).

Prediction of votes on opinions is similar to a problem encountered in recommender systems research, viz. predicting ratings for products. A well-known approach is latent factor collaborative filtering (Koren, Bell, and Volinsky 2009; Mnih and Salakhutdinov 2007), where both users and items are mapped into a joint space, and the model aims to explain a given rating based on several latent aspects underlying how users rate items. Recent research combines such modeling of ratings with review text (McAuliffe and Blei 2008; McAuley and Leskovec 2013). Although there is a similarity between predicting ratings of a product and predicting votes of justcies on a case using texts (reviews, opinions), there is a key difference: each review is typically authored by a user, but each opinion is authored and/or endorsed by multiple justices. This makes the latter problem more challenging.

Algorithm 1 OpGen($\beta, D, U, K, \epsilon, N_{max}, \chi, \mu_{\mathbf{x}}, \sigma_{\mathbf{x}}, \mu_{\mathbf{a}}, \sigma_{\mathbf{a}}$)

Input: A set of parameters.
Output: A set of opinions written by the justices.

for each justice $u \in \{1, \dots, U\}$ **do**
  for each topic $k \in \{1, \dots, K\}$ **do**
    $x_{uk} \sim \mathcal{N}(\mu_{\mathbf{x}}, \sigma_{\mathbf{x}})$
  $\theta_u \leftarrow f(\mathbf{x}_u)$
for each opinion d **do**
  for each topic k **do**
    $a_{dk} \sim \mathcal{N}(\mu_{\mathbf{a}}, \sigma_{\mathbf{a}})$
for each topic k **do**
  $\phi_k \sim$ Dir(β)
for each opinion d **do**
  $\theta_d \leftarrow f(\theta_1, ..., \theta_U)$
  $A_d \leftarrow \{\}$
  for each user u **do**
    $r_{ud} \sim \mathcal{N}(f(x_u, a_{ad}, \theta_d), \epsilon)$
    if $r_{ud} \geq 0$ **then**
      $A_d \leftarrow A_d \cup \{u\}$
  $N_d \sim$ Bin(N_{max}, χ)
  for each word w_i, $1 \leq i \leq N_d$ **do**
    $z_i \sim$ Mult(θ_d)
    $w_i \sim$ Mult(ϕ_{z_i})

4 Methods

We present a generative model for SCOTUS decisions with opinions written and voted on by the justices. A graphical model describing the generative model is illustrated in Fig. 1. In the generative process (see Alg. 1) a justice u (total U justices) and dth opinion O_d have K-dimensional ideal points $\mathbf{x}_u$ and $\mathbf{a}_d$ respectively. Each dimension represents a topic. Each justice has a justice-topic distribution

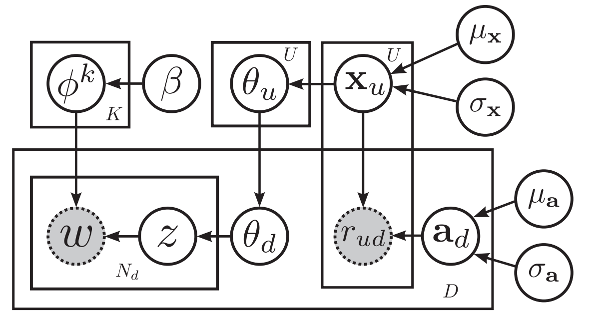

Figure 1: *Plate notation for SCIPM.* SCIPM is a graphical model for multidimensional ideal points of the justices of the Supreme Court.

θ_u, which is derived from $\mathbf{x}_u$. Given $\mathbf{x}_u$, $\mathbf{a}_d$, and θ_u we estimate a topic mixture θ_d for O_d and the vote r_{ud} of u on O_d. For sampling O_d we draw the number of words N_d and for ith word w_{di} in O_d, a topic z_{di} is sampled from a multinomial distribution with parameter θ_d, and then w_{di} is sampled from a multinomial distribution with parameter $\phi_{z_{di}}$, a topic-word (Dirichlet) distribution with parameter β. At the end, we obtain a corpus of D opinions with words $\mathbf{w} = \{w_{11}, \dots, w_{di}, \dots, w_{DN_D}\}$ and topic assignments $\mathbf{z} = \{z_{11}, \dots, z_{di}, \dots, z_{DN_D}\}$. We infer the model's parameters using an algorithm that alternates between two steps: (a) estimating ideal points using gradient descent and (b) learning topics in opinions using Gibbs sampling.

Estimating Ideal Points using Gradient Descent. Given K topic-word distributions ϕ, we optimize ideal points for the justices ($\mathbf{x}_u$) and opinions ($\mathbf{a}_d$) using gradient descent (see Alg. 2 and equations in detail in Supplementary Material (Islam et al. 2016)). We estimate θ_{dk} using x_{uk} since justices take standpoints on a case after the case is argued and the ideal point $\mathbf{x}_u$ influences the justice to write and/or vote for the opinions. We derive the justice-topic distribution θ_u using $\mathbf{x}_u$. Since $\mathbf{x}_u \in \mathbb{R}^K$ and θ_u is a probability distribution, we use $\theta_{uk} = \frac{\exp^{x_{uk}}}{\sum_k \exp^{x_{uk}}}$, as a transformation function for calculating θ_{uk}. Now $\theta_{dk} = \frac{1}{|A_d|} \sum_{u' \in A_d} \theta_{u'k}$. Here A_d is the set of justices who either write or join opinion O_d. Note that we assume each justice who participates in an opinion has equal contribution toward writing the opinion. If u writes or joins opinion O_d, then $r_{ud} = 1$; otherwise, $r_{ud} = -1$. The decision r_{ud} is estimated as follows:

$$\hat{r}_{ud} = \begin{cases} 1, & \sum_k \theta_{dk} x_{uk} a_{dk} >= 0 \\ -1, & \sum_k \theta_{dk} x_{uk} a_{dk}) < 0 \end{cases} \quad (1)$$

We define the error function for the entire corpus as follows:

$$\mathcal{L} = \sum_{u,d} \left(r_{ud} - \sum_k (y_{ud} \theta_{dk} x_{uk} a_{dk}) \right)^2 + \lambda \sum_k x_{uk}^2 + \lambda \sum_k a_{dk}^2 - \gamma \prod_{d \in D} \prod_{j=1}^{N_d} \sum_{z_{dj}} \frac{1}{|A_d|} \sum_{u' \in A_d} \left(\frac{\exp^{x_{u'z_{dj}}}}{\sum_{k'} \exp^{x_{u'k'}}} \right) \phi_{z_{dj} w_{dj}}$$

Here y_{ud} is a binary variable which represents whether u joins O_d. For estimating the parameters x_{uk} and a_{dk} we differentiate $\mathcal{L}$ with respect to x_{uk} and a_{dk}.

$$\frac{\delta \mathcal{L}}{\delta x_{uk}} = \sum_{d}(r_{ud} - \sum_{k}\theta_{dk}x_{uk}a_{dk})\,(-\theta_{dk}a_{dk}) + \lambda x_{uk}$$
$$- \gamma \prod_{d \in D}\prod_{j=1}^{N_d}\sum_{z_{dj}}\frac{1}{|A_d|}\sum_{\substack{u' \in A_d \\ k=z_{dj}}}(\theta_{u'k} - x_{u'k}\theta_{u'k})\,\phi_{z_{dj}w_{dj}}$$
$$\frac{\delta \mathcal{L}}{\delta a_{dk}} = \sum_{u}(r_{ud} - \sum_{k}\theta_{dk}x_{uk}a_{dk})\,(-\theta_{dk}x_{uk}) + \lambda a_{dk}$$

For minimizing the error using the principles of gradient descent we get, $x_{uk} \leftarrow x_{uk} - \eta\frac{\delta \mathcal{L}}{\delta x_{uk}}$ and $a_{dk} \leftarrow a_{dk} - \eta\frac{\delta \mathcal{L}}{\delta a_{dk}}$. Here η is the learning rate. Once $\mathbf{x}$ and $\mathbf{a}$ are updated we derive θ_d and run a Gibbs Sampler (see next) to estimate ϕ.

Learning Topics in Opinions using Gibbs Sampling. Given θ_d in each gradient descent iteration we use collapsed Gibbs Sampling to estimate ϕ (see Alg. 3 in Supplementary Material (Islam et al. 2016)). We sample topic z_{di} for word w_{di} in O_d using the following distribution:

$$P(z_{di} = k|w_{di} = w, \mathbf{z}^{-di}, \mathbf{w}^{-di}, \beta) \propto \frac{C_{kw}^{-di}+\beta}{C_{k}^{-di}+W\beta}\frac{C_{dk}^{-di}}{C_{d}^{-di}} \quad (2)$$

where $\mathbf{w}^{-di}$, and $\mathbf{z}^{-di}$ represents all the words except the ith word of O_d and their topic assignments respectively. Moreover, C_{kw}^{-di} and C_{k}^{-di} represent the number of times word w is assigned to topic k and the total number of words assigned to topic k respectively, not including the current instance in consideration; C_{dk}^{-di} and C_{d}^{-di} denote the number of times topic k is assigned to the words in opinion d and the total number of words in the opinion d respectively, not including the current instance. An iteration of Gibbs sampling completes by drawing a sample for z_{di} according to Eq. 2. Given the sampled value for z_{di} the counts C_{kw}, C_{dk} are updated. Given $\mathbf{z}$ we estimate $\hat{\phi}_{kw} = \frac{C_{kw}+\beta}{C_k+W\beta}$ and $\hat{\theta}_{kd} = \frac{C_{dk}}{C_d}$.

Supplementary Material (Islam et al. 2016) shows how to choose the values for various parameters: number of iterations (N), number of topics (K), likelihood weight (γ), regularizer weight (λ), and learning rate (η). Besides the above unsupervised approach we also run our model using a supervised topic setting: we learn a set of topics in opinions using LDA (Blei, Ng, and Jordan 2003) and use the learned topic distribution ϕ in Alg. 2 (Supplementary Material) for learning $\mathbf{x}_u$ and $\mathbf{a}_d$. For both unsupervised and supervised settings, we use $\mu_{\mathbf{a}} = 0$ and $\sigma_{\mathbf{a}} = 0.1$ as priors for opinions' ideal points. We apply two settings of priors for justices' ideals points: an unbiased prior and a biased prior (Gerrish and Blei 2011; Clinton, Jackman, and Rivers 2004; Martin and Quinn 2002). With an unbiased prior for justices' ideal points we do not have any prior assumptions about justices' ideology, and we set $\mu_{\mathbf{x}} = 0$ and $\sigma_{\mathbf{x}} = 1$. With a biased prior we assume either 'Democratic' justices have positive ideal points while 'Republicans' are negatives, or that 'Republicans' are positives while 'Democrats' are negative. For a biased prior we set $\sigma_{\mathbf{x}} = 0.5$ with $\mu_{\mathbf{x}} = 1$ for 'Republicans' and $\mu_{\mathbf{x}} = -1$ for 'Democrats'.

Predicting Votes of Opinions. We predict justices who vote for an opinion using a 5-fold cross-validation over both real-world and synthetic datasets (see Sec. 5). Unlike the prediction of votes on a synthetic dataset, the true values for $\mathbf{x}_u$, $\mathbf{a}_{t_d}$, and r_{ut_d} are unknown for a real-world case given a justice u and a test opinion t_d (see Supplementary Material for prediction on synthetic datasets). For performing the prediction, we compare estimated value for true r_{ut_d} (denoted as r'_{ut_d}) with the predicted value for r_{ut_d} (denoted as r''_{ut_d}). The estimated true value r'_{ut_d} is calculated using known authors of t_d. We estimate $\mathbf{a}'_{t_d}$ using θ'_{t_d} (inferred using the known author list) and $\hat{\mathbf{x}}_u$ (inferred using the training set), and then estimate r'_{ut_d} using $\hat{\mathbf{x}}_u$, θ'_{t_d}, and $\mathbf{a}'_{t_d}$. For predicting votes of each justice on t_d, we estimate the predicted value r''_{ut_d} given that the authors of t_d are unknown. We assume each justice is equally likely to have authored the opinion. We calculate θ''_{t_d} given that each author is equally likely, and then estimate $\mathbf{a}''_{t_d}$ using $\hat{\mathbf{x}}_u$ and θ''_{t_d}. Finally, we estimate r''_{ut_d} using $\hat{\mathbf{x}}_u$, θ''_{t_d}, and $\mathbf{a}''_{t_d}$. If $|r'_{ut_d} - r''_{ut_d}| \leq (\max(r''_{ut_d}) - \min(r''_{ut_d})) \times \epsilon$, then justice u is predicted as a voter of t_d. Here ϵ is a user-defined error threshold . For nine judges, we calculate the accuracy, τ, of the prediction on t_d. If τ is greater than a user-defined accuracy threshold, ξ , we assume that the voters of t_d are predicted correctly. For the entire test set, we compute the recall value as the proportion of opinions for which authors are predicted correctly. We report the average recall value over five test-folds for various parameter settings.

Comparison with Existing Methods. We compare SCIPM with three methods: Lauderdale and Clark (2014), the Author-Topic model of (Rosen-Zvi et al. 2004), and a naïve Bayes classifier (Rish 2001). We use 5-fold cross validation for the analysis. Lauderdale and Clark (Lauderdale and Clark 2014) do not explicitly describe how to estimate model parameters and perform prediction for a test set. (Their method estimates topics separately from ideal point inference.) To circumvent this problem we learn ideal points and other model parameters using the entire dataset. We then perform prediction over the entire dataset using learned ideal points and other model parameters, and estimate the recall value of prediction. For comparing SCIPM with Lauderdale and Clark (2014), we aggregate the votes of CO, CCO, and CIJ of a case to predict the overall voting of the case: the justices who support any of CO, CCO, and CIJ are voting for the court's decision and the rest of the justices are voting against the court's decision. If the decisions of all the justices over a case are predicted correctly then we consider the method to be accurate in predicting the votes of a case. Finally we calculate the accuracy of predictions over all the cases. The Author-Topic Model (ATM) (Rosen-Zvi et al. 2004), a generative process, assumes that a text corpus has a set of authors and that each author has a distribution over topics, which are distributions over words of a vocabulary. Within this model a word is generated by sampling an author uniformly followed by the sampling of a topic from the sampled author's topic distribution. We predict justices using a maximum likelihood approach described in (Song 2009). If the probability of a justice u given t_d is greater than uniform probability, then we consider u be the author of t_d. We then calculate the accuracy of the prediction, and if the accuracy is greater than a user-defined accuracy threshold, ξ , we assume that the voters of t_d are predicted correctly. We then calculate the recall value of the test-fold and report the average recall value. We also use a naïve Bayes classifier (Rish 2001) for classifying the justices of a test opinion, where features are the words in the vocabulary and the value of a feature is the tf-idf (Manning, Raghavan, and Schïtze 2008) score of the word in an opinion, the class labels are the justices, who authored or supported this opinion. This formulation is a multi-label classification problem for which we employ a one-versus-rest classification methodology.

5 Experimental Results

We evaluate SCIPM using both synthetic and real-world datasets. Our experiments are designed to model how justices voted for specific opinions, to understand the quality of inferred ideal points for justices as well as for opinions and whether the inferred points can capture dynamics of coalitions. Due to space constraints, we present experimental results on real-world data for unsupervised and supervised settings with biased priors. Synthetic data with related experiments and experiments on real-world data for unsupervised and supervised settings with unbiased priors are described in the Supplementary Material (Islam et al. 2016). Code and Data for the experiments are available online (see Supplementary Material (Islam et al. 2016)).

Datasets. We collected opinion files[3] for the current Court (from 2010 to 2014) and curated them into a structured data format. These opinions are contributed by the current nine justices: Chief Justice John G. Roberts, Justice A. Scalia, Justice A. M. Kennedy, Justice C. Thomas, Justice R. B Ginsburg, Justice S. G. Breyer, Justice S. A. Alito, Justice S. Sotomayor, and Justice E. Kagan. Since we are interested in capturing differences between ideal points of justices on different issues, unanimous decisions are uninteresting (Supplementary Material (Islam et al. 2016) describes experiments on a dataset including both divided and unanimous decisions). We remove opinions that have partial participation of judges, but we aim to incorporate such opinions into SCIPM in the future. The resulting dataset after such preprocessing has 185 cases with 467 opinions. These opinions are of various types: court opinion (CO), concurring court opinion (CCO), concurring in judgment (CIJ), dissenting court opinion (DCO), and concurring and dissenting opinion (CND) (See Fig. S1 and Table S3).

Prediction of Votes on Real-World Data. We evaluate SCIPM's predictive capability in terms of votes on opinions within a case using real-world data. We observe that the model performs well with error threshold, $\epsilon \geq 0.2$, and accuracy threshold, $\xi = 80\%$, for a range of topic numbers and different settings (see Table 1 and Table S8). We compare SCIPM with three existing methods: Lauderdale and Clark (2014), ATM (Rosen-Zvi et al. 2004), and a naïve Bayes classifier (Rish 2001). We observe that SCIPM outperforms all of these three methods. The average recall value for SCIPM is 79.46%, which is greater than the average recall value (48.13%) for Lauderdale and Clark (2014). Table

Number of Topics (K)	Error Threshold (ϵ)	Accuracy Threshold (ξ)			
		90.00%	80.00%	70.00%	60.00%
5	0.15	37.42	59.14	71.18	80.86
	0.20	58.71	78.06	88.17	93.55
	0.25	71.83	87.10	94.41	97.20
	0.30	84.53	95.91	98.28	99.35
10	0.15	31.61	53.76	68.60	80.43
	0.20	55.27	72.90	86.24	92.90
	0.25	74.19	88.82	94.19	96.13
	0.30	84.95	94.41	96.77	98.06
15	0.15	26.88	49.46	63.23	75.48
	0.20	49.68	69.89	82.37	89.03
	0.25	68.82	84.95	93.33	96.56
	0.30	83.23	93.12	98.28	99.57
20	0.15	24.73	43.66	59.78	71.18
	0.20	48.17	65.59	77.85	87.31
	0.25	67.31	82.15	89.89	93.55
	0.30	81.51	89.46	93.76	97.63

Table 1: *Prediction of votes on opinions using SCIPM over real-world datasets for an unsupervised setting with a biased prior.* Using 5-fold cross validation average recall values for predictions are estimated for various settings of three parameters: number of topics (K), error threshold (ϵ), and accuracy threshold (ξ)

Num. of Topics (K)	ATM Model	SCIPM				
	Accuracy Threshold $\geq 50\%$	Error Threshold (ϵ)	Accuracy Threshold (ξ)			
			90%	80%	70%	60%
5	12.50	0.15	37.42	59.14	71.18	80.86
10	30.00	0.15	31.61	53.76	68.60	80.43
15	30.00	0.15	26.88	49.46	63.23	75.48
20	35.00	0.15	24.73	43.66	59.78	71.18

Table 2: *Comparison between the SCIPM for an unsupervised setting with a biased prior and ATM.* Average recall values for predictions are shown.

2 and Table S9 shows that SCIPM is superior to ATM in most of the constrained settings. Finally, the recall of the naïve Bayes classifier is 27.95%, which is lower than most of the recall values estimated by SCIPM (See Table 2 and Table S9).

Evaluation of Ideal Points of Justices. SCIPM estimates multidimensional ideal points for each of the justices. Such ideal points reflect judicial preferences of each justice over various issues (or dimensions). We use our model on the entire dataset for both unsupervised and supervised settings with unbiased and biased priors for learning ideal points. Fig. 2,3 illustrate the ideal points for each justice on ten issues using unsupervised and supervised settings with biased priors. Ideal points for unsupervised and supervised settings with unbiased priors are shown in Supplementary Material (Islam et al. 2016). In both settings, we infer meaningful topics including employment, legislation, the Fourth Amendment, and sentencing, to name a few (see Table 3,4). Using ideal points we can address interesting research questions. The following text discusses a few of the tidbits of

[3] http://www.supremecourt.gov/opinions/opinions.aspx

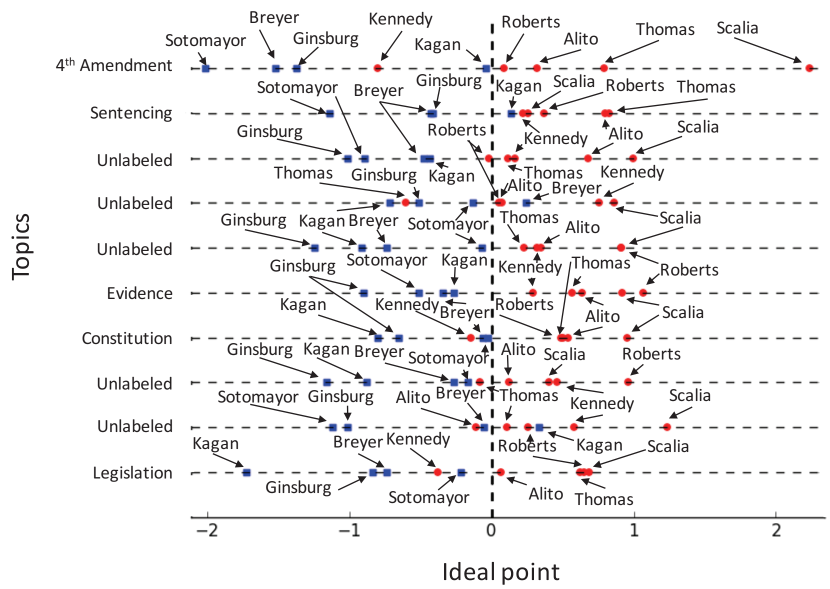

Figure 2: *Ideal points of justices in an unsupervised setting with a biased prior for ten topics.* The x-axis depicts the ideal points for justices and each horizontal line represents a topic or an issue (Table 3 shows the corresponding topics). Each justice is labeled with their last name. We color the ideal points with either red (the corresponding justice was appointed during a Republican presidency) or blue (the corresponding justice was appointed during a Democratic presidency).

inference captured in our analysis of ideal points which are corroborated with other studies.

Case Study 1. The Fourth Amendment to the U.S. Constitution guarantees freedom from unreasonable search and seizure. A well-discussed case about the Fourth Amendment in recent times is Heien v. North Carolina[4]. This case is about whether the Fourth Amendment is violated when a police officer pulls over a car based on a reasonable but mistaken belief (Kerr 2014b). The court ruled 8–1 that the Fourth Amendment was not violated; Justice Sotomayor was the lone justice who dissented (Root 2014). Our model learns a topic which can be labeled as the Fourth Amendment (see Table 3,4). The ideal points for this topic clearly show that Justice Sotomayor's ideal point is different compared to the other justices (see Fig. 2,3). Our inferred position of Justice Sotomayor on the Fourth Amendment is also supported by her stances on other cases regarding the Fourth Amendment (Root 2015; EPIC 2015; Bernick 2015; Kerr 2014a).

Case Study 2. The case Knox v. Service Employees International Union (SEIU) (case No 10-1121) deals with whether a public-sector union can impose a special assessment without the affirmative consent of a member or nonmember upon whom it is imposed (Oyez 2015). This case was ruled in 7–2 that SEIU went beyond the allowable rights and infringed upon nonmembers' First Amendment rights; Alito delivered the opinion of the court, in which Roberts and Scalia, Kennedy, and Thomas joined. Sotomayor filed a CIJ, in which Ginsburg joined; Breyer filed a dissenting opinion, in which Kagan joined. Fig. 4 illustrates the ideal points of the CO of the case. We observe relatively different ideal point values (dominant) on Topic 1 ("Petition"), Topic 2 ("Legal Action"), Topic 3 ("Employment"), and Topic 4 ("Legislation"). Roberts, Alito, Scalia, Kennedy, and Thomas have positive ideal ponts on Topics 1–4 (see Fig. 3). This suggests that these five justices should participate in an opinion together, which is supported by the case history (in this case, it is a court opinion). On the other hand Breyer and Kagan have negative ideal points for Topics 1–4, which justifies their vote for the dissenting opinion i.e. against the court opinion. Also, Ginsburg and Sotomayor is inclined toward the opinion on one topic but against the opinion on two and three topics, respectively, which resulted in a writing of a CIJ opinion by them.

We perform a qualitative comparison of ideal points learned by SCIPM versus that of Lauderdale and Clark (2014). Fig. S10 and Table S12 in Supplementary Material (Islam et al. 2016) respectively show ideal points of justices and topics learned by Lauderdale and Clark (2014). Although we observe similarities between the topics learned by Lauderdale and Clark (2014) and SCIPM (compare Table S12 with Table 3,4 and Table S10,S11), these two methods exhibit notable differences in ideal points (compare Fig. S10 with Fig. 2,3 and Fig. S7,S8): Fig. S10 shows a clear partisan (i.e., conservative and liberal) divide of justices on almost all of the ten issues, whereas ideal points learned by SCIPM (Fig. 2,3 and Fig. S7,S8) may not directly translate into a clear partisan divide. Ideal points learned by Lauderdale and Clark (2014) for five, fifteen, and twenty topics exhibit similar strict partisan divides on almost every issue (results not shown). Although it is widely believed that the

[4] http://www.supremecourt.gov/opinions/14pdf/13-604_ec8f.pdf

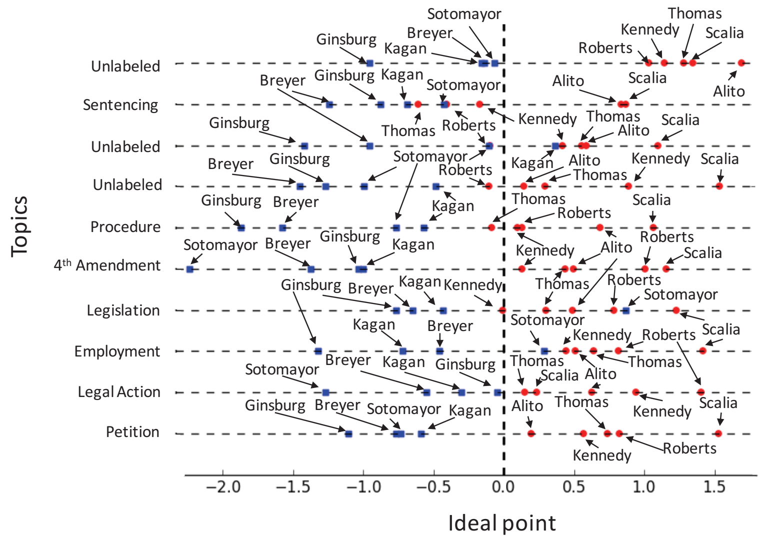

Figure 3: *Ideal points of justices in a supervised setting with a biased prior for ten topics.* The x-axis depicts the ideal points for justices and each horizontal line represents a topic or an issue (Table 4 shows the corresponding topics). Each justice is labeled with their last name. We color the ideal points with either red (the corresponding justice was appointed during a Republican presidency) or blue (the corresponding justice was appointed during a Democratic presidency).

Topic 1	Topic 4	Topic 5	Topic 8	Topic 9	Topic 10
Legislation	*Constitution*	*Evidence*	*Property*	*Sentencing*	*Fourth Amendment*
government	senate	dna	tax	sentence	officers
amendment	president	testimony	indian	offense	search
political	constitution	victim	property	crime	warrant

Table 3: *Issues learned in a SCIPM unsupervised analysis with a biased prior for ten topics or issues*. The labels of issues are given manually based on the top words of each issues. Issues that do not have apparent labels are given a tag unlabeled and not shown here.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 9
Petition	*Legal action*	*Employment*	*Legislation*	*Fourth Amendment*	*Procedure*	*Sentencing*
claim	suit	employees	statute	police	trial	sentence
rule	jurisdiction	employee	congress	search	jury	crime
order	action	employment	provision	warrant	attorney	defendant

Table 4: *Issues learned in a SCIPM supervised analysis with a biased prior for ten topics or issues*. The labels of issues are given manually based on the top words of each issues. Issues that do not have apparent labels are given a tag unlabeled and not shown here.

(current) Roberts Court is more polarized than its predecessors, we observe, by analyzing the close votes (e.g., 5–4) and moderately close votes (e.g., 6–3 and 5–3), that a strict partisan divide occurs in 18.37% of the cases and a moderately partisan divide occurs in 11.35% of the cases (out of the 185 cases in our dataset, which contains only divided decisions). In many of the cases (70%) the justices from both ideological leanings (i.e. liberal and conservative) voted with each other. This result is supported by other studies (Devins and Baum 2014; Bartels In press). Bartels (In press) argues that polarization between justices occur more on the cases that are within the court's volitional agenda (politically salient issues) versus the cases that fall within the court's exigent agenda (institutional maintenance). In this scenario we would expect collaboration between justices from both ideological leanings on the issues, which is reflected in the ideal points learned by SCIPM.

Coalitions of Justices on Issues. We observe that 'Republican' judges and 'Democratic' judges seem to lie close to each other on many issues. For example, the pairs of Roberts and Alito, Scalia and Roberts, and Roberts and Thomas are close to each other on several issues for both unsupervised and supervised settings (see Fig. 2–3). We also observe Breyer and Ginsburg lying close to each other on nine issues for both unsupervised and supervised settings. To understand this better we infer frequent collaborations between justices (see Table S11 in Supplementary Material (Islam et al. 2016)) opinions' votes. We observe

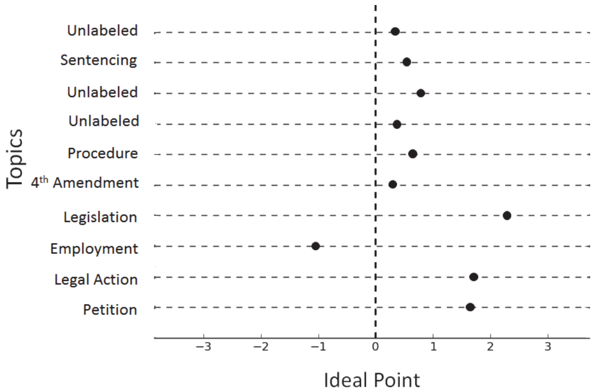

Figure 4: *Ideal point vector for the dissenting opinion of case 10-1121.* The figure shows that Topic 1 (“Petition”), Topic 2 (“Legal Action”), Topic 3 (“Employment”), and Topic 4 (“Legislation”) are dominant among the ten topics.

Justice	Score
Kennedy	1.03
Roberts	0.75
Kagan	0.75
Sotomayor	0.70
Scalia	0.66
Ginsburg	0.66
Breyer	0.65
Thomas	0.63
Alito	0.40

Table 5: *Swing scores ($\mathcal{S}_u$) for the current Court.*

pairs between Roberts and Alito, Scalia and Roberts, and Roberts and Thomas. These pairs vote together more frequently compared to other justices. Similarly Breyer and Ginsburg also frequently agree.

Identifying Swing Justices. We evaluate the quality of SCIPM in terms of whether we can identify the swing justices (Martin, Quinn, and Epstein 2004) from ideal points. A swing justice is the one whose vote generally decides the standing of the court on a case. A swing score can be defined in various ways. In this paper we hypothesize that a justice who has the largest average deviance of ideal points over the issues can be considered as a swing justice. We define a swing score, $\mathcal{S}_u = \frac{\sum_{k=1}^{K}(\mathrm{E}[x_{.k}]-x_{uk})^2}{|K|}$, for justice u. Table 5 shows the estimated swing measures for a supervised setting with an unbiased prior and highlights Kennedy as the one with the largest swing score, $\mathcal{S}_u$. The ranking of Chief Justice Roberts as the second swing justice is also explainable given his (surprising) votes to affirm the Patient Protection and Affordable Care Act (“Obamacare”).We also create a subset of cases which are ruled in 5–4. We identify frequently inferred collaborations between justices in this dataset (see Table 6). We observe that Justice Kennedy participates in two size-five patterns, where he is the common justice. Note that the exclusion of Justice Kennedy from these two size-five patterns yields two mutually exclusive sets that can be labeled as ‘Republican’

Pattern	Support(%)
Alito, **Kennedy**, Roberts, Scalia, Thomas	47.05
Breyer, Ginsburg, Kagan, **Kennedy**, Sotomayor	35.30
Alito, Breyer, Kennedy, Roberts, Thomas	08.82

Table 6: *Frequently inferred collaborations of justices in cases decided by a bare majority (5–4).*

and ‘Democrat’ based on the party of the president who appointed them. This finding can be verified by other quantitative scores for measuring ideological leanings such as the Segal–Cover score (Segal and Cover 1989) and the Martin–Quinn score (Martin and Quinn 2002). On both of these scores Justice Kennedy has the median position.

6 Discussion

We have presented SCIPM, a modeling approach that unifies textual analysis with voting records to identify multidimensional ideal points for judges on the US Supreme Court. In addition to supporting a probabilistic model of the underlying data, we have demonstrated how this approach yields meaningful interpretations about judging ideology, coalitions, and swing votes. Most importantly, SCIPM be used as not just an explanatory model but as a predictive model to forecast outcomes. The results here suggest that ideal point methods can be fruitfully extended with the latest developments from computer science in modeling of text. One of the areas of future research is to model the (often slow) evolution of court over the years, i.e., some justices are known to become more (or less) conservative as the court composition changes. Secondly, we are interested in determining the effect of public sentiment on issues that face the court and whether accounting for public sentiment predisposes the court to act in certain ways. Finally, we would like to combine our models with formal interpretations of the US constitution, (e.g. originalism, constructionism, and textualism).

References

Bailey, M. 2012. Measuring court preferences, 1950–2011: Agendas, polarity and heterogeneity. *Georgetown University*. Available at: http://faculty.georgetown.edu/baileyma/CourtPref_July2012.pdf.

Bartels, B. In press. The sources and consequences of polarization in the U.S. Supreme Court. In Thurber, J., and Yoshinaka, A., eds., *American Gridlock*. Cambridge University Press.

Bernick, E. 2015. Justice Sotomayor: Stop bending the constitution to favor cops. *The Huffington Post*. At: http://www.huffingtonpost.com/evan-bernick/justice-sotomayor_b_6534970.html.

Blei, D.; Ng, A.; and Jordan, M. 2003. Latent Dirichlet Allocation. *JMLR* 3:993–1022.

Clark, T. 2008. Measuring ideological polarization on the United States Supreme Court. *Political Research Quarterly*.

Clinton, J.; Jackman, S.; and Rivers, D. 2004. The statistical analysis of roll call data. *American Political Science Review* null(02):355–370.

Devins, N., and Baum, L. 2014. Split definitive: How party polarization turned the Supreme Court into a partisan court. *Social Science Research Network*. DOI: http://dx.doi.org/10.2139/ssrn.2432111.

EPIC. 2015. Sotomayor and privacy. Available at: https://epic.org/privacy/sotomayor/.

Gerrish, S. M., and Blei, D. M. 2011. Predicting legislative roll calls from text. In *In Proc. of ICML*.

Gerrish, S., and Blei, D. 2012. How they vote: Issue-adjusted models of legislative behavior. In *NIPS*, 2753–2761.

Gu, Y.; Sun, Y.; Jiang, N.; Wang, B.; and Chen, T. 2014. Topic-factorized ideal point estimation model for legislative voting network. In *the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 183–192.

Guimerà, R., and Sales-Pardo, M. 2011. Justice blocks and predictability of U.S. Supreme Court votes. *PloS One* 6(11):e27188.

Islam, M. R.; Hossain, K. T.; Krishnan, S.; and Ramakrishnan, N. 2016. Inferring multi-dimensional ideal points for US Supreme Court Justices. Available at: https://dl.dropboxusercontent.com/u/8921131/scotus/scipm-suppl.pdf.

Katz, D.; Bommarito-II, M.; and Blackman, J. 2014. Predicting the behavior of the Supreme Court of the United States: A general approach. *Social Science Research Network*. DOI: http://dx.doi.org/10.2139/ssrn.2463244.

Kerr, O. 2014a. Justice Sotomayor on her Privacy Opinion: "Remember, I was only one of nine.". *The Washington Post*, 2014. Available at: https://www.washingtonpost.com/news/volokh-conspiracy/wp/2014/01/27/justice-sotomayor-on-her-privacy-opinion-remember-i-was-only-one-of-nine.

Kerr, O. 2014b. Reasonable mistake of law can generate reasonable suspicion, Supreme Court holds. *The Washington Post*. Available at: http://www.washingtonpost.com/news/volokh-conspiracy/wp/2014/12/15/reasonable-mistake-of-law-can-generate-reasonable-suspicion-supreme-court-holds/.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.

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

Manning, C.; Raghavan, P.; and Schïtze, H. 2008. *Introduction to Information Retrieval*. Cambridge University Press.

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

Martin, A.; Quinn, K.; and Epstein, L. 2004. The median justice on the United States Supreme Court. *N.C.L.Rev.* 83:1275. Available at: http://scholarship.law.berkeley.edu/facpubs/1604.

McAuley, J., and Leskovec, J. 2013. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *the 7th ACM Conference on RecSys*, 165–172.

McAuliffe, J., and Blei, D. 2008. Supervised topic models. In *NIPS*, 121–128.

Mnih, A., and Salakhutdinov, R. 2007. Probabilistic matrix factorization. In *NIPS*, 1257–1264.

Oyez. 2015. Knox v. service employees international union. *Chicago-Kent College of Law at Illinois Tech*. Available at: https://www.oyez.org/cases/2011/10-1121.

Pickerill, J. M. 2013. Measuring judicial minimalism on the Roberts court. *Social Science Research Network*. DOI: http://dx.doi.org/10.2139/ssrn.2314135.

Rish, I. 2001. An empirical study of the naïve Bayes classifier. In *IJCAI Workshop on Empirical Methods in Artificial Intelligence*, 41–46.

Root, D. 2014. Supreme Court sides with police in 4th Amendment case arising from officer's 'mistake of law'. Available at: http://reason.com/blog/2014/12/15/supreme-court-sides-with-police-in-4th-a.

Root, D. 2015. Sotomayor to Justice Department lawyer: 'we can't keep bending the Fourth Amendment to the resources of law enforcement'. Available at: https://reason.com/blog/2015/01/22/sotomayor-to-justice-department-lawyer-w.

Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; and Smyth, P. 2004. The Author-Topic model for authors and documents. In *the 20th Conference on Uncertainty in artificial intelligence*, 487–494.

Segal, J., and Cover, A. 1989. Ideological values and the votes of U.S. Supreme Court Justices. *American Political Science Review* 83(02):557–565.

Sim, Y.; Routledge, B.; and Smith, N. 2014. The utility of text: The case of amicus briefs and the Supreme Court. *CoRR*. At: http://arxiv.org/abs/1409.7985.

Song, J. 2009. The Author-Topic model and the author prediction. Available at: http://pages.cs.wisc.edu/~jerryzhu/cs769/project/2009/JIASI_SONG.pdf.