Text Based Ideal Point Model
(REVISITING THE ALGORITHM)

Daniela Pinto **Amelia Dai** **Marina Castano** **Simone Rittenhouse**

Abstract

This study explores the text-based ideal points model (TBIP), assessing its functionality and applicability in the context of the United Kingdom's political landscape. We aim to contribute to a deeper understanding of political discourse and ideological shifts. The study examines whether TBIP, a probabilistic topic model developed by Vafa et al. (2020), effectively positions politicians along an interpretable political spectrum in a multiparty context and explores its ability to capture the evolution of word choices and political preferences over time. Additionally, the paper delves into the model's design choices and its variational inference algorithm. Through experiments comparing TBIP with a traditional vote-based ideal point model, the study evaluates the model's performance and identifies potential areas for improvement. Notably, while TBIP demonstrates promise in classifying political stances, it falls short in comparison to the benchmark model, suggesting the need for further refinement and adaptation to diverse political contexts. Code is accessible here.

1 Introduction

Ideal point models offer valuable insights into lawmakers' political preferences by analyzing vote discrepancies on shared bills and determining a score that represents the lawmaker's ideology. However, political preferences extend beyond legislative votes, permeating all aspects of society, including the language used to discuss political issues. This paper assesses the functionality and applicability of the Text-Based Ideal Point Model (TBIP), a probabilistic topic model.

TBIP leverages word-choice differences on shared topics to infer estimates of ideal points. Designed to process unordered, unstructured, and unorganized speeches, TBIP reveals latent ideological topics and latent political positions of the authors behind these speeches (Vafa et al., 2020). By applying TBIP to broader political landscapes, this study aims to contribute to the understanding of political discourse and ideological shifts beyond traditional vote analyses.

Focusing on the complex political landscape of the United Kingdom, we aim to study TBIP's robustness and validity across diverse political systems. We examine whether TBIP adeptly positions politicians along an interpretable political spectrum in a multiparty context. The model reveals latent topics under discussion, providing invaluable insights into political discourse framing. Lastly, this study delves into the model's design choices and its variational inference algorithm.

2 Related Work

Existing literature predominantly emphasizes static ideal point estimates (Rosas et al., 2015; Heckman and Snyder, 1997; Carroll et al., 2013; Bafumi et al., 2005). Advancing this field requires integrating dynamic considerations, such as assessing trends and historical patterns in political preferences to anticipate future contexts effectively. Additionally, despite considerable progress in understanding political preferences, there is a noticeable gap in the external validity of these methodologies, largely due to their focus on the United States political landscape. Furthermore, although TBIP is extendable to any form of unordered, unstructured text as long as each text is clearly linked to its author, current applications are scarce, suggesting untapped potential for broader insights.

Our research builds upon the original TBIP model, which assumes a fixed set of topics influenced solely by ideological differences, neglecting shifts in term compositions due to changing importance. Moreover, we draw inspiration from other non-U.S. TBIP applications, such as Mendoza et al. (2024)'s work on uncovering social bots and their political preferences in Chile, and Feldkircher et al. (2022)'s analysis of central bankers' speeches in the Euro Area post-debt crisis. Our study aims to validate TBIP's utility, contributing to understanding evolving discourse and ideological shifts in diverse and broader political contexts.

3 Background

3.1 Overview of the United Kingdom Political System

The United Kingdom operates under a parliamentary democracy, where the government is elected by the people and must retain the confidence of Parliament. Its Parliament is the highest legislative body, comprising two chambers: the House of Commons and the House of Lords. The House of Commons, the lower house, consists of elected Members of Parliament. Meanwhile, the upper house, the House of Lords, includes appointed and hereditary members, such as life peers and bishops. The head of government in the UK is the Prime Minister, typically the leader of the party with the most seats in the House of Commons. The Prime Minister appoints Cabinet ministers to oversee specific government departments.

The political landscape centers on two main parties: the Conservative Party, advocating for free markets and a smaller state, and the Labour Party, championing social justice and workers' rights (Wright, 2013; Grayson, 2016). Alongside these parties are other significant political players. The Liberal Democrats, positioned in the center, prioritize social liberalism, civil liberties, and democratic reform. The Scottish National Party advocates for Scottish independence and social democracy, while Plaid Cymru, the nationalist party in Wales, supports Welsh independence and cultural preservation. The Green Party emphasizes environmentalism, social justice, and sustainability. In Northern Ireland, the Democratic Unionist Party backs Northern Ireland's place within the United Kingdom, while Sinn Féin, a nationalist party, seeks Irish reunification.

3.2 Data

Our work builds upon the ParlSpeech dataset (Rauh and Schwalbach, 2020), a collection of nearly 2 million speeches from the UK House of Commons, offering time series data ranging from 1988 to 2019. This dataset offers full-text vectors and metadata such as speaker, date, party, and agenda item for each speech. The pre-processing for the original TBIP model follows that outlined by Benjamin E. Lauderdale and Alexander Herzog for Wordshoal (Lauderdale and Herzog, 2016), a model that estimates ideal points from labeled text. We adapt this to the UK government where needed. This involves including speeches with over 50 words, Members of Parliament with more than 24 speeches, and bigrams spoken by 10 or more authors.

In the original U.S. Senate speech application of the TBIP model (Vafa et al., 2020), common U.S. state and city names were excluded, as were nonsubstantive speeches such as those centered around scheduling, prayer, tributes, recess, etc. as well as those delivered by the president pro tempore – who presides over the Senate. We apply this procedure to UK data by removing stopwords (including UK city names, dates, numbers, and titles), speeches with irrelevant agendas like business of the House and periods where the House is not in session (e.g. adjournments and prorogation), and speeches by the Speaker of the House. Our final dataset covers the years 2016 to 2019, containing 206,993 speeches from 742 Members of Parliament, and 10,576 words in the vocabulary.

Additionally, although we have a general idea of the ideology of each political party, it is important to note that there is no ground truth for the ideological score of each politician. Therefore, we are dealing with an unsupervised learning problem. To have a baseline for the performance of our model, we compare it to a traditional ideal point model that is based on the votes of the politicians. We should not expect a perfect match between the two models, but votes offer a much stronger and unambiguous signal in terms of ideology. For this reason, we obtain data on the votes of each Member of Parliament from 2016 to 2024 (Hub, 2024). After pre-processing the data, we are left with 360,137 votes from 972 Members of Parliament on 720 bills, from the year 2016 to 2019.

3.3 Preliminary Models

An ideal point is a single number summary used in politics to summarize the political preferences of lawmakers. For instance, more conservative justices have more positive ideal points and more liberal justices have more negative ideal points. Ideal points have been traditionally computed based on vote data. The text-based ideal points model builds on two preexisting models: the Bayesian Ideal Points Model and Poisson Factorization Topic Models. This section briefly introduces these models.

3.3.1 Bayesian ideal points

This probabilistic method is widely used to measure the political positions of legislators, based on voting records. Let i denote a lawmaker, j denote a bill, and v_{ij} denote the vote of lawmaker i on bill j, which can only take the values 0 for “no" and 1 for “aye". The likelihood of legislator i voting in favor of bill j is modeled as a Bernoulli-distributed variable:

$$v_{ij} \sim \text{Bernoulli}\Big(\sigma(\alpha_j + x_i\eta_j)\Big),$$

where the $\sigma(\cdot)$ represents the sigmoid function, x_i denotes the ideal point of lawmaker i, and α_j and η_j represent the popularity and the polarity of the bill j, respectively. The Bayesian model fixes the prior distributions:

$$x_i \sim \mathcal{N}(0,1), \quad \alpha_j \sim \mathcal{N}(0,1), \quad \text{and} \quad \eta_j \sim \mathcal{N}(0,1).$$

The model aims to estimate the latent variable x_i, that is, a scalar variable that shows the political ideology of that lawmaker. To approximate the value x_i for each lawmaker, two other latent variables are used, α_j and η_j to account for the popularity and the polarity of each bill. This allows us to understand how different factors influence the likelihood of a legislator voting for a particular bill:

- The higher the popularity of the bill α_j, the more likely it is that any lawmaker will vote in favor of j regardless of their political affiliation, and inversely for a negative α_j. When α_j is close to 0, it means that the bill is highly politicized and the vote depends on the ideal point of the lawmaker.
- When the polarity of the bill η_j and x_i have the same sign, there is a higher probability that the lawmaker will vote “aye" than when they have different signs. If η_j is close to 0, then the bill is not politicized and the vote of the lawmaker depends on the popularity of the bill α_j.

3.3.2 Poisson factorization topic models

This is a bag of words model which factorizes a matrix of word counts per document into two positive matrices, a matrix $\boldsymbol{\theta}$ of topic intensities per document, and a matrix $\boldsymbol{\beta}$ of topics.

Let v denote each word and d each document. The count of word v in document d is assumed to follow a Poisson distribution:

$$y_{dv} \sim \text{Pois}\Big(\sum_k \theta_{dk}\beta_{kv}\Big),$$

where θ_{dk}, represents the intensity of topic k in document d, and β_{kv}, represents the appearance of word v in topic k. For some hyperparameters a, b we fix the prior distributions of our latent variables:

$$\theta_{dk} \sim \text{Gamma}(a,b), \quad \beta_{kv} \sim \text{Gamma}(a,b).$$

By estimating these latent variables from the observed word frequency matrix, the model seeks to provide insights into the topics present in the documents and how each word contributes to these topics.

4 Methodology

Our main focus of study, the text-based ideal points model (TBIP), is a mixture of the Bayesian ideal points model and the Poisson factorization topic models and introduced by Vafa et al. (2020).

4.1 Text-based ideal points model

Let v represent each word, d each document, and y_{dv} the count of word v in document d. The latent variables we consider are:

- $\boldsymbol{\theta_d}$, the topic intensities of document d. This is a non-negative vector of size K, for K the number of topics, that represents the topics that appear in each document.
- $\boldsymbol{\beta_k}$, the neutral topic k defined as a probability over words. This is a non-negative vector of size V, for V the number of words.
- $\boldsymbol{\eta_k}$, the ideology of each word in regards of the topic k. This is a vector of size V.
- x_s, the ideal point of author s. This is a real-valued scalar.

Let the prior distributions be fixed for some hyperparameters a, b:

$$\theta_{dk} \sim \text{Gamma}(a, b), \quad \beta_{kv} \sim \text{Gamma}(a, b), \quad \eta_{kv} \sim \mathcal{N}(0, 1), \quad \text{and} \quad x_s \sim \mathcal{N}(0, 1).$$

Considering a_d as the author of document d, the likelihood is then:

$$y_{dv} \sim \text{Poisson}\Big(\sum_k \theta_{dk}\beta_{kv}\exp(x_{a_d}\eta_{kv})\Big).$$

We note that this expression mirrors the Poisson factorization, with the addition of an ideological topic η_{kv}, which adjusts the neutral topic β_{kv} as a function of the ideal point of the author x_{a_d}. Then, the likelihood can be rewritten as:

$$y_{dv} \sim \text{Poisson}\Big(\sum_k \theta_{dk}\exp(\log\beta_{kv} + x_{a_d}\eta_{kv})\Big).$$

In this case, we can see its similarity to the Bayesian ideal points model with $\log\beta_{kv}$ as the popularity of a word, and η_{kv} as the polarity of that word. As it did in the ideal points model, the higher β_{kv} is, the more likely that word is to appear in the text, regardless of the political ideology of the author. That is why $\boldsymbol{\beta_k}$ represents the words that appear in a topic k neutrally, without considering ideology.

Each word v in a document d, is also assigned an ideology. If η_{kv} has the same sign as the ideal point of the author, x_{a_d}, then the word has a higher probability of appearing. If η_{kv} and x_{a_d} have different signs, then the word is less likely to appear. Therefore, the positive values in $\boldsymbol{\eta_k}$ represent the words in topic k with a certain ideology, and the negative values represent the words in the same topic but with opposite ideology.

Once we compute the posterior, we obtain the following results.

- The political position of authors, x_s.
- The latent topics in each document, $\boldsymbol{\theta_d}$.
- The words used for each topic neutrally, which are those with high β_{kv} and $\eta_{kv} = 0$.
- The words used for each topic with a certain ideology, which are those with high β_{kv} and $\eta_{kv} > 0$ or $\eta_{kv} < 0$ respectively.

Figure 1 presents the graphical view of this model through a directed acyclic graph. This graph illustrates the probabilistic dependencies between latent variables ($\boldsymbol{\theta_d}$, $\boldsymbol{\beta_k}$, $\boldsymbol{\eta_k}$, x_s) and observed word counts (y_{dv}), highlighting the flow of influence from document topic intensities, neutral and ideological topics, and author ideal points to the observed word counts.

4.1.1 Verbosity

To account for the difference in verbosity between authors, Vafa et al. (2020) add a parameter w_s, that is the average word count of author s with respect to the rest. With n_s as the average word count of author s across all their documents, and S as the total number of authors, the weight is calculated by

$$w_s = \frac{n_s}{\frac{1}{S}\sum_{s'} n_{s'}}.$$

For the author a_d of document d, the weight w_{a_d} is applied to the rate of the Poisson distribution as shown below:

$$y_{dv} \sim \text{Poisson}\Big(w_{a_d}\sum_k \theta_{dk}\beta_{kv}\exp(x_{a_d}\eta_{kv})\Big).$$

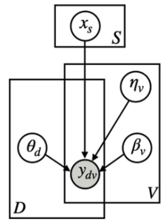


Figure 1: Graphical representation of the Text-based Ideal Points Model with latent variables. The directed acyclic graph illustrates the probabilistic dependencies between latent variables ($\boldsymbol{\theta_d}$, $\boldsymbol{\beta_k}$, $\boldsymbol{\eta_k}$, x_s) and observed word counts (y_{dv}).

4.2 Inference

The posterior distribution $p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x}|\boldsymbol{y})$ is intractable, so we use variational inference to approximate it. First, we denote $q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x})$ the variational family of approximate posterior distributions with parameters $\boldsymbol{\phi}$. Our goal is to find parameters $\boldsymbol{\phi}$ that minimize the KL-divergence between $q_{\boldsymbol{\phi}}$ and p. This is the same as maximizing the ELBO, the expectation of the sum of the log prior, the log likelihood, and the entropy of $q_{\boldsymbol{\phi}}$.

$$\mathbb{E}_{q_{\boldsymbol{\phi}}}\Big[\log p(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x}) + \log p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x}) - \log q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x})\Big].$$

We set the variational family to be the mean-field family. This means that we assume that the variational family factorizes, so each variable is independent.

$$q_{\boldsymbol{\phi}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{x}) = \prod_{d,k,s} q(\boldsymbol{\theta_d})q(\boldsymbol{\beta_k})q(\boldsymbol{\eta_k})q(x_s),$$

for d documents, k topics, and s author indices. For the non-negative latent variables, we set the log-normal distribution.

$$q(\boldsymbol{\theta_d}) = \text{LogNormal}_K(\mu_{\boldsymbol{\theta_d}}, I\sigma^2_{\boldsymbol{\theta_d}}) \quad \text{and} \quad q(\boldsymbol{\beta_k}) = \text{LogNormal}_V(\mu_{\boldsymbol{\beta_k}}, I\sigma^2_{\boldsymbol{\beta_k}}).$$

For the real latent variables, we set the normal distribution.

$$q(\boldsymbol{\eta_k}) = \mathcal{N}_V(\mu_{\boldsymbol{\eta_k}}, I\sigma^2_{\boldsymbol{\eta_k}}) \quad \text{and} \quad q(x_s) = \mathcal{N}(\mu_{x_s}, \sigma^2_{x_s}).$$

Finally, the ELBO is optimized with respect to $\boldsymbol{\phi} = \{\boldsymbol{\mu_\theta}, \boldsymbol{\sigma^2_\theta}, \boldsymbol{\mu_\beta}, \boldsymbol{\sigma^2_\beta}, \boldsymbol{\mu_\eta}, \boldsymbol{\sigma^2_\eta}, \boldsymbol{\mu_x}, \boldsymbol{\sigma^2_x}\}$ using stochastic gradient ascent.

4.3 Experimental setup

To examine the performance of the TBIP model in a multi-party context, we utilize our clean UK data and compare the output with a benchmark vote-based ideal point model for the same politicians over the same time period. The evaluation metric is the correlation between the learned ideal points of the two models. Our TBIP model is trained for 30,000 steps with 512 batch sizes. Following the original setting, we assume 50 latent topics prior distributions θ_{dk}, $\beta_{kv} \sim$ Gamma(0.3, 0.3). Moreover, as the paper suggests, we include the average word count of each author $w_s = \frac{n_s}{\frac{1}{S}\sum_{s'} n_{s'}}$ as a weight to the Poisson likelihood. We find this crucial for better model performance because otherwise the author's verbosity rather than political ideology is reflected in the model's ideal points estimate.

For hyperparameter tuning, the best performance is achieved around 30,000 steps, with diminished performance noted when conducting fewer or greater numbers of steps. Other hyperparameters explored include parameters of Gamma prior, the number of topics, and log transformation, where the natural logarithm of the counts matrix is taken before the inference, aiming to mitigate the impact of longer speeches. However, as shown in Table 1, none of these have shown a significant improvement in our results. Note that when exploring different hyperparameters, we maintain the default settings for all parameters except the one under investigation. These default settings include 30,000 steps, a batch size of 512, 50 latent topics, Gamma priors configured as Gamma(0.3, 0.3), and the weight w_s being applied to mitigate the verbosity effect of the authors.

5 Evaluation

Correlation coefficients, such as Pearson and Spearman, are commonly used metrics for evaluating the performance of models in various fields, including political science. These coefficients provide a quantitative measure of the agreement between the ideal points generated by different models. A high correlation indicates a strong alignment between the text-based ideal points and the ground truth ideal points, derived from voting records. Our findings reveal notable disparities between these two sets of ideal points. Table 1 shows that the highest correlation observed between the TBIP-generated ideal points and the Bayesian vote-based ideal points model is 0.557 for Pearson correlation and 0.501 for Spearman correlation. These coefficients signify a moderate level of agreement between the two sets of ideal points, yet they fall significantly short from the correlation reported in the original paper by Vafa and colleagues (Vafa et al., 2020), where a correlation of 0.88 was attained for U.S. Senate speech data. This prompts critical reflections on the generalizability and applicability of text-based ideal point models across diverse political landscapes.

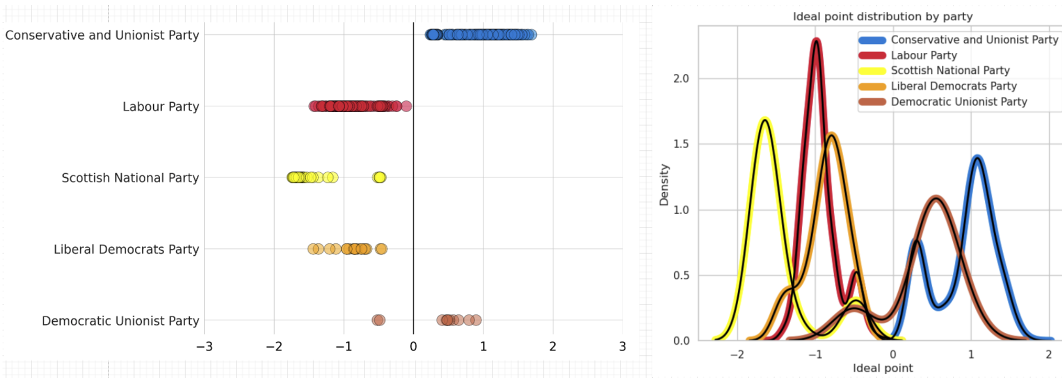
Table 1: The Pearson Correlation and Spearman Correlation between TBIP and vote-based ideal point model under different hyperparameter settings.

Configuration	Pearson Correlation	Spearman Correlation
No weights w_s	0.276	0.274
Number of steps		
10, 000 steps	0.156	0.140
30, 000 steps	**0.557**	**0.501**
50, 000 steps	0.351	0.306
Log transformation	0.557	0.499
Parameters of Gamma prior		
$\theta_{dk}, \beta_{kv} \sim$ Gamma(2, 1)	0.286	0.246
$\theta_{dk}, \beta_{kv} \sim$ Gamma(5, 1)	0.314	0.267
100 latent topics	0.292	0.249

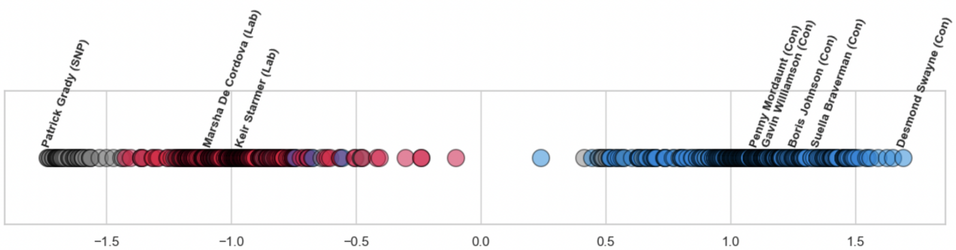
Figures 2a and 2b illustrate the performance of the Bayesian ideal points model, our "ground truth", showcasing the ideal points attributed to each political party. As anticipated, the ideal points are positioned to the right for parties with a more conservative ideology and to the left for those with a more progressive stance. Notably, there are local maxima observed in the neutral area for certain parties, a phenomenon we attribute to the influence of Brexit, a topic we will delve into further in our final discussion. On the other hand, Figures 3a and 3b present the results obtained from our TBIP model. Overall, the model effectively classifies political parties, aligning with our expectations based on their known ideologies. However, a notable overlap is observed around the neutral ideal point, particularly between the Conservative and Labour parties. This suggests a degree of ambiguity or similarity in the language used by these parties, potentially reflecting common policy positions or strategic communication tactics. Further analysis is warranted to explore the underlying factors contributing to this overlap and to assess the robustness of the TBIP model in capturing subtle distinctions in political discourse.

Figure 4 offers an in-depth exploration of topics identified by the model, revealing the associated words for each ideological category. To enhance clarity, we've categorized ideologies into "Progressive", "Neutral", and "Moderate", a progressive stance typically advocates for social and political change to address issues like inequality, social justice, and environmental sustainability, in contrast to a more conservative "Moderate" stance. Considering the topic of Brexit, regardless of the ideological category, key terms such as "brexit", "eu", "deal", and "leave" underscore discussions surrounding its implications for constituents, the economy, and job security. This topic underscores a pragmatic approach to Brexit, emphasizing the importance of clarity and consensus in navigating the complexities of withdrawal from the European Union. Expanding beyond Brexit, the analysis encompasses significant topics such as Social Programs. Here, a "Progressive" stance features keywords related to refugee rights and vulnerable populations, reflecting a compassionate approach to immigration policy. Conversely, a "Moderate" stance may focus more on workplace dynamics, systemic reforms, and financial considerations. Overall, the TBIP model offers a comprehensive overview of the political discourse surrounding Brexit and other critical issues.

Figure 2: Vote-based ideal points

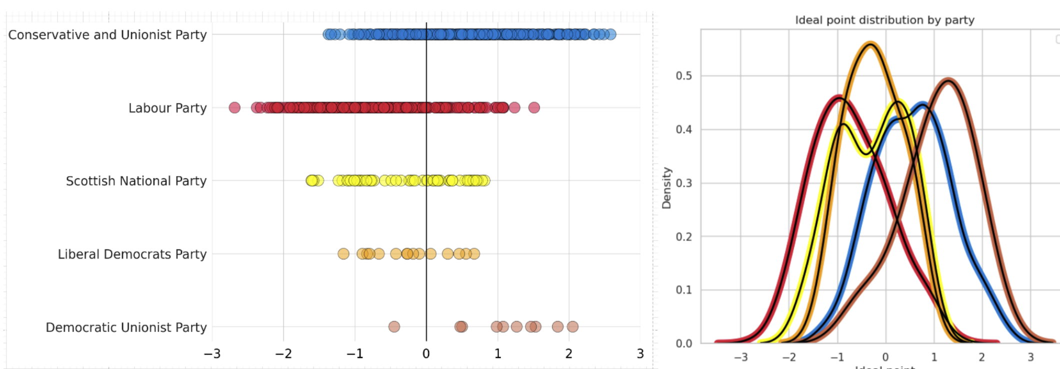


(a) Ideal points per party

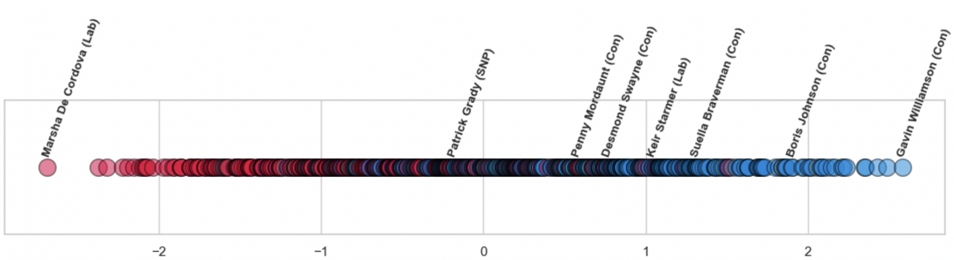


(b) Ideal points of main politicians, chosen based on popularity ranking.

Figure 3: Text-based ideal points



(a) Ideal points per party



(b) Ideal points of main politicians, chosen based on popularity ranking.

Topic	Ideology	Top Words
Brexit	**Progressive**	brexit, prime, deal, country, constituents, eu, jobs, economy, workers rights, deal brexit
	Neutral	deal, eu, prime, brexit, european union, leave, country, voted, referendum, parliament
	Moderate	ensure, european union, relation, deal, eu, united kingdom, negotiations, implementation, leave, clear
Social Programs	**Progressive**	disabled, social security, universal credit, poverty, cuts, food, million, families, un, sick
	Neutral	universal credit, poverty, benefits, payments, system, benefit, claimants, roll, payment, welfare
	Moderate	workplace, ensure, system, ensuring, important, benefits, payments, point, welfare, course
Foreign Affairs	**Progressive**	foreign, commonwealth, british, china, world, donald trump, countries, trump, abroad, nazanin
	Neutral	foreign, security, countries, world, international, commonwealth, country, russia, british, china
	Moderate	security, nato, allies, russia, intelligence, international, response, foreign, clear, sanctions

Figure 4: Topics identified by the Text-based Ideal Points Model along with the associated words for each ideological category, categorized as "Progressive," "Neutral," and "Moderate."

6 Discussion

Our research findings indicate that the TBIP model can effectively capture specific facets of political preferences evident within multi-party contexts, such as those observed in the UK Parliament. However, it falls short significantly when compared to the vote-based model utilized as a benchmark in the original paper. It's essential to recognize that adapting the TBIP model to different national contexts necessitates careful consideration and adjustments. The idiosyncrasies of each legislative body and political landscape demand tailored modifications to ensure the model's efficacy and relevance in diverse settings. Thus, the Bayesian ideal point model offers a simpler alternative, especially when the focus is solely on analyzing individuals within a voting assembly.

We hypothesize that the model's underperformance may stem from the necessary pre-processing steps when dealing with natural language. The original model's cleaning procedures are tailored specifically to the Senate of the U.S., potentially limiting its applicability to other legislative bodies. For example, what text is considered nonsubstantive is open to interpretation and dependent upon having domain knowledge of a legislative body's procedures and structure. Moreover, the temporal context of our analysis, coinciding with the Brexit era, presents a significant consideration. The TBIP model relies on the concept of political framing – meaning that there must be a distinction in the terms used by opposing ideologies when discussing the same issue. Thus, it's conceivable that the intensity of discourse surrounding this historic issue led to a convergence of party stances.

Interestingly, our experimentation uncovered a noteworthy insight: while the choice of prior assumptions in the TBIP model does not notably impact its performance, the number of trials conducted does exert a discernible influence on the results. This finding underscores the robustness of the TBIP model to variations in prior assumptions, suggesting a certain level of insensitivity to the specific specifications of the prior distributions. However, it also highlights the importance of the computational aspect in model performance. The number of trials, which reflects the depth and breadth of the model's exploration of the parameter space, emerges as a critical factor affecting the model's ability to accurately capture political preferences.

Future research avenues could involve extending the application of the TBIP model to periods unaffected by significant political events like Brexit. By analyzing data from different temporal contexts, we can gain a more comprehensive understanding of how the model performs under varying circumstances. This approach would not only provide insights into the stability and generalizability of the model but also offer valuable insights into the dynamics of political preferences over time. Furthermore, investigating the efficacy of two-dimensional ideal points represents a promising direction for future exploration. By incorporating additional dimensions into the model, we could potentially uncover more nuanced patterns and relationships within the data which could prove useful for classification in a multiparty context.

References

Joseph Bafumi, Andrew Gelman, David K Park, and Noah Kaplan. Practical issues in implementing and understanding bayesian ideal point estimation. *Political Analysis*, 13(2):171–187, 2005.

Royce Carroll, Jeffrey B Lewis, James Lo, Keith T Poole, and Howard Rosenthal. The structure of utility in spatial models of voting. *American Journal of Political Science*, 57(4):1008–1028, 2013.

Martin Feldkircher, Paul Hofmarcher, and Pierre L. Siklos. Cacophony in central banking? evidence from euro area speeches on monetary policy. Available at SSRN: https://ssrn.com/abstract=4196226 or http://dx.doi.org/10.2139/ssrn.4196226, August 2022.

Richard S. Grayson. *British Politics: A Beginner's Guide*. Oneworld Publications, 2016. ISBN 9781780748825.

James J Heckman and James M Snyder. Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *RAND Journal of Economics*, pages S142–S189, 1997.

Developer Hub. Commons Votes API, 2024. URL https://commonsvotes-api.parliament.uk.

Benjamin E. Lauderdale and Alexander Herzog. Measuring political positions from legislative speech. *Political Analysis*, 24(3):374–394, 2016. URL https://benjaminlauderdale.net/files/papers/2016LauderdaleHerzogPA.pdf.

M. Mendoza, E. Providel, and M. et al. Santos. Detection and impact estimation of social bots in the chilean twitter network. *Sci Rep*, 14:6525, 2024. doi: 10.1038/s41598-024-57227-3. URL https://doi.org/10.1038/s41598-024-57227-3.

Christian Rauh and Jan Schwalbach. The ParlSpeech V2 data set: Full-text corpora of 6.3 million parliamentary speeches in the key legislative chambers of nine representative democracies, 2020. URL https://doi.org/10.7910/DVN/L4OAKN.

Guillermo Rosas, Yair Shomer, and Spencer R Haptonstahl. No news is news: Nonignorable nonresponse in roll-call data analysis. *American Journal of Political Science*, 59(2):511–528, 2015.

Keyon Vafa, Suresh Naidu, and David M. Blei. Text-Based Ideal Points. 2020.

Tony Wright. *British Politics: A Very Short Introduction*. Very Short Introductions. Oxford University Press, 3rd edition, 2013. ISBN 9780199661007.