

Replicação de Vafa, Naidu e Blei 2020

Matheus Alves dos Santos
Departamento de Sistemas e Computação
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brasil
matheusalves@copin.ufcg.edu.br

I. ESTUDO ORIGINAL

Os modelos de pontos ideais são métodos quantitativos que caracterizam o posicionamento de atores políticos, associando-os a valores (pontos ideais) que se distribuem em uma única dimensão (eixo ou espectro político). Tipicamente, esses modelos são treinados a partir de dados das votações legislativas nas quais tais atores tomaram parte. Contudo, o processo político não é constituído somente por legisladores e, ainda que o fosse, as preferências políticas de um indivíduo não são expressas apenas através de seus votos.

Os autores do artigo original propõem o modelo *Text-Based Ideal Points* (TBIP) para caracterizar o posicionamento político a partir de dados textuais como, por exemplo, os discursos, os comunicados à imprensa e/ou os *tweets* de atores políticos. O TBIP é um algoritmo de modelagem de pontos ideais, probabilístico, baseado em estatística Bayesiana, não supervisionado e que adota a abordagem *bag-of-words*. Ainda, a ideia que dá origem ao seu funcionamento é o *political framing*, o mecanismo que descreve como o posicionamento político afeta o vocabulário adotado por um indivíduo quando este deseja transmitir mensagens políticas acerca de um tema. Por exemplo, em discussões sobre o porte de armas, os Republicanos estão mais propensos a utilizar expressões como “*constitutional rights*”, enquanto os Democratas estão mais propensos a expressões como “*gun violence*”.

Em uma análise comparativa, quando aplicado sobre os discursos e os *tweets* dos senadores estadunidenses, o TBIP apresentou resultados muitos semelhantes e fortemente correlacionados aos dos modelos mais tradicionais de pontos ideais, isto é, àqueles baseados em votações. Todavia, a principal contribuição desse trabalho diz respeito à capacidade do modelo proposto de determinar pontos ideais para atores políticos que não participam das votações legislativas. Os autores validaram experimentalmente tal capacidade, aplicando o TBIP sobre *tweets* dos pré-candidatos Democratas das eleições presidenciais de 2020. A avaliação empírica desses resultados corroborou a validade e a importância da contribuição desse artigo para os estudos relacionados a pontos ideais.

O artigo original está disponível [aqui](#).

II. OBJETO DA REPRODUÇÃO OU REANÁLISE

O objeto de reprodução deste trabalho é o cálculo de pontos ideais para os senadores estadunidenses da 114ª Legislatura (2015 a 2017) usando o modelo *Text-Based Ideal Points*. Mais especificamente, esse cálculo será feito apenas a partir dos discursos proferidos pelos senadores ao longo da legislatura supracitada, de modo a reproduzir os resultados apresentados no artigo original através da Figura 1.

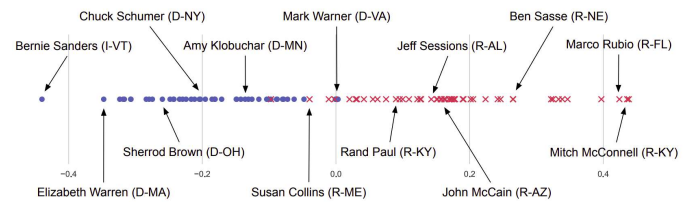


Fig. 1. Distribuição dos pontos ideais definidos pelo modelo *Text-Based Ideal Points* para senadores estadunidenses a partir de seus discursos.

Ainda que os valores dos pontos ideais não possuam significado intrínseco, observá-los em conjunto nos permite diferenciar os senadores mais progressistas dos mais conservadores – como Bernie Sanders e Marco Rubio, respectivamente – e visualizar uma separação ideológica bem evidente entre os Democratas e os Republicanos.

III. METODOLOGIA ORIGINAL

A. Materiais e métodos

Para esse experimento, os autores do artigo original utilizam quatro conjuntos de dados distintos para cada um dos períodos analisados. Os dados referentes à 114ª Legislatura estão disponíveis [aqui](#) e descrevem, respectivamente:

- O vocabulário do *corpus* de discursos dos senadores;
- Os nomes e partidos dos senadores;
- O índice associado a cada um dos senadores;
- A frequência dos termos do vocabulário nos discursos de cada um dos senadores.

Esses conjuntos de dados foram produzidos a partir das transcrições dos discursos proferidos no Senado estadunidense entre os anos de 2015 e 2017. Essas transcrições, por sua

vez, são mantidas e disponibilizadas pela *Stanford University* neste [repositório](#). Além disso, em virtude do volume de informações, os autores do artigo original disponibilizam os dois últimos conjuntos de dados listados apenas em formato de matriz esparsa.

B. Dados gerados no experimento

O objetivo desta replicação é produzir um conjunto de dados com a identificação dos senadores, os partidos a que estão filiados e os respectivos pontos ideais calculados por meio do *Text-Based Ideal Points*. A Tabela I exemplifica o formato desse conjunto de dados.

TABLE I
RESULTADOS HIPOTÉTICOS PARA A REPLICAÇÃO DO CÁLCULO DE PONTOS IDEAIS DOS SENADORES ESTADUNIDENSES.

Senador(a)	Filiação	Ponto Ideal
...
Bernie Sanders	Independente	-0.523
...
Chuck Schumer	Democratas	-0.211
...
Marco Rubio	Republicanos	0.437
...

C. Análise de dados

No artigo original, os pontos ideais calculados pelo *Text-Based Ideal Points* são quantitativamente avaliados com base em sua correlação com os pontos ideais definidos por um modelo mais tradicional, cujo cálculo se baseia em inferência variacional aplicada às votações dos senadores. Os resultados desse último podem ser obtidos a partir da execução desse [script](#). Nesse cenário, uma correlação positiva e forte indica bons resultados. Os autores também apontam a separação entre Democratas e Republicanos como indício dos bons resultados do TBIP, mas não apresentam evidências estatísticas (como intervalos de confiança ou testes de hipótese) acerca dessa distinção entre os partidos.

IV. DIFERENÇAS METODOLÓGICAS COM O ESTUDO ORIGINAL

Nesta replicação, a maioria das diferenças metodológicas são motivadas pelo custo computacional ou pela disponibilidade dos conjuntos de dados adequados. Por exemplo, os autores do artigo original calculam os pontos ideais de todos os senadores das Legislaturas 111 a 114, mas apenas essa última será considerada durante o presente experimento, visto que os dados referentes às legislaturas anteriores não estão integralmente disponíveis (ou acessíveis).

A execução do experimento será feita através de um [notebook do Google Colab](#) disponibilizado pelos autores do artigo original. Vale destacar, no entanto, que esse *notebook* adota uma versão mais recente do TBIP e que, visando reduzir o

custo computacional, também substitui a Fatoração de Poisson pela Fatoração de Matrizes Não-Negativas (implementada pela biblioteca *Scikit Learn*¹). Essas alterações podem ter algum impacto nos resultados do modelo.

Dada a indisponibilidade dos dados acerca das legislaturas anteriores, também não será possível reutilizar o código da análise dos resultados do experimento. Por fim, além da etapa de comparação a um modelo de pontos ideais baseado em votações, também será utilizada a inferência estatística para quantificar a diferença entre senadores Democratas e senadores Republicanos no que se refere aos seus pontos ideais.

V. RESULTADO DA REPLICAÇÃO OU REANÁLISE

Adotando a metodologia descrita anteriormente, o experimento do artigo original foi replicado a partir dos dados referentes à 114ª Legislatura do Senado estadunidense. Conforme apresentado na Figura 2, os pontos ideais foram calculados através do modelo *Text-Based Ideal Points* e do modelo baseado em votações. Vale ressaltar que, ainda que o Senado estadunidense seja composto por 100 membros, apenas 81 senadores dessa legislatura tomaram parte tanto nas votações quanto nos discursos desse período. Por isso, foram omitidos os pontos ideais dos demais.

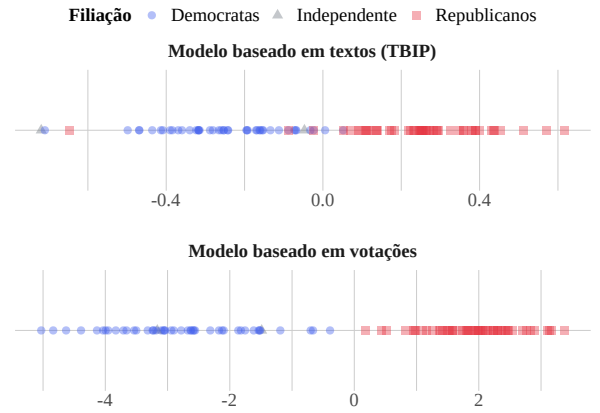


Fig. 2. Pontos ideais definidos pelos modelos *Text-Based Ideal Points* e de inferência variacional baseada em votações para os senadores estadunidenses com mandato entre 2015 e 2017.

Os pontos ideais definidos pelo TBIP estão fortemente relacionados aos do modelo baseado em votações, conforme apresentado na Figura 3. Nesta amostra, o coeficiente de correlação de Pearson alcança o valor de $\rho = 0.836$. Adotando reamostragem por *bootstrap* com 2000 repetições e nível de confiança de 95%, estimou-se que o valor desse coeficiente esteja entre 0.660 e 0.902, indicando que essa correlação varia

¹<https://scikit-learn.org>

de moderada a muito forte. Esses resultados corroboram a percepção de que o *Text-Based Ideal Points* é capaz de caracterizar posicionamentos políticos de maneira muito semelhante aos modelos de pontos ideais tradicionais.

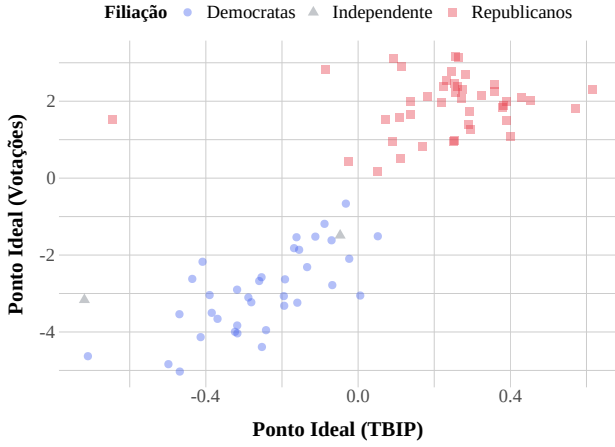


Fig. 3. Distribuição conjunta dos pontos ideais definidos pelos modelos *Text-Based Ideal Points* e de inferência variacional baseada em votações para os senadores estadunidenses com mandato entre 2015 e 2017.

Nessa mesma amostra, é possível notar que os pontos ideais definidos pelo TBIP para os senadores Democratas e Republicanos são quase disjuntos, com os respectivos valores médios de -0.250 e 0.243 . Novamente recorrendo à reamostragem por *bootstrap* com 2000 repetições e nível de confiança de 95%, estimou-se que o ponto ideal médio dos Democratas está no intervalo $[-0.297, -0.207]$, enquanto o dos Republicanos está no intervalo $[0.180, 0.287]$. Ademais, a diferença absoluta dessas médias está entre 0.425 e 0.556. Esses intervalos de confiança estão apresentados na Figura 4.

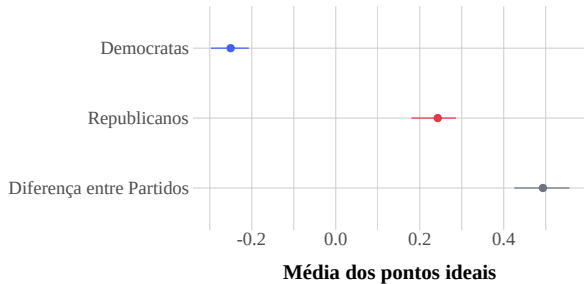


Fig. 4. Intervalos de confiança para os pontos ideais médios dos partidos (Democratas e Republicanos) e para a diferença desses pontos médios.

Visto que os pontos ideais definidos pelo modelo *Text-Based Ideal Points* assumiram valores entre -0.719 e 0.618 , a diferença absoluta entre os pontos ideais médios dos dois partidos pode ser considerada bastante alta. Essa distinção é

condizente com o contexto político estadunidense, marcado por um bipartidarismo fortemente polarizado.

VI. REPLICAÇÃO VS. ESTUDO ORIGINAL

No artigo original, os autores indicam uma correlação forte ($\rho = 0.88$) entre os pontos ideais definidos pelo modelo baseado em votações e aqueles definidos pelo *Text-Based Ideal Points* com base nos discursos dos senadores. Esse resultado está em concordância com os resultados obtidos nesta replicação – e incluso no intervalo de confiança estimado. Quanto à distinção entre os pontos ideais por partido, os autores do estudo recorreram apenas à observação dos valores na amostra, mas os seus resultados também são condizentes com os desta replicação.

Todavia, foram identificadas diferenças relevantes quanto aos valores dos pontos ideais apresentados no artigo original. O exemplo mais evidente é o senador Rand Paul, um republicano bastante conservador. Considerando seus votos ao longo da 114ª legislatura do Senado estadunidense, ele possui o 12º menor ponto ideal entre os republicanos. Já para o TBIP, ele não só possui o menor ponto ideal de seu partido, como também está posicionado muito próximo a Bernie Sanders, um senador democrata altamente progressista. As causas dessas diferenças não são claras, mas provavelmente estão relacionadas à quantidade de dados textuais utilizados no treinamento do modelo e/ou a filtros aplicados nesses conjuntos de dados que não foram mencionados no estudo original. Ainda, apesar de utilizar a mesma *random seed*, esta replicação adotou uma versão mais recente do TBIP e um método de fatoração menos custoso computacionalmente. Essas alterações também podem ter impactado, mesmo que discretamente, nos resultados obtidos.

VII. MATERIAL PARA REPLICAÇÃO

Todo o material produzido durante a presente replicação estará disponível no seguinte repositório do GitHub: <https://github.com/alvesmatheus/replicacao-tbip>.