

Detecção de Posicionamento e Rotulação Automática de Usuários do Twitter: estudo sobre o embate científico-político no contexto da CPI da Covid-19

Patricia D. Santos¹, Denise H. Goya¹

¹Pós-Graduação em Ciência da Computação - Universidade Federal do ABC (UFABC)
Av. dos Estados, 5001 – Santo André – SP – Brazil

{patricia.santos,denise.goya}@ufabc.com.br

Abstract. *People's stance around social and political issues is often accomplished via messages posted on social media, however, predicting it without the help of manual labeling could be a challenging task. Using a specific case study, namely the Covid-19 Parliamentary Commission of Inquiry, this article proposes a method to detect and quantify the stance of Twitter users in relation to a politically controversial and polarizing topic. Through the use of computational approaches combined with social factors, such as homophily and network structure, it was possible to automatic label 98% of the users present on the studied databases, with very little human intervention, as well as categorize their positions using valence scores and two complementary metrics: degree of balance and engagement.*

Resumo. *O posicionamento das pessoas em torno de questões sociais e políticas é muitas vezes realizado via mensagens postadas nas mídias sociais. Prever esse posicionamento sem a ajuda de rotulações manuais pode ser uma tarefa desafiadora. Utilizando um estudo de caso específico, a saber, a CPI da Covid-19, este artigo propõe um método para detectar e quantificar o posicionamento de usuários do Twitter em relação a um tema politicamente controverso e polarizado. Por meio do uso de abordagens computacionais combinadas com fatores sociais, como homofilia e estrutura de rede, foi possível rotular automaticamente 98% dos usuários presentes nas bases de dados estudadas, com pouquíssima intervenção humana, bem como categorizar suas posições por meio de uma pontuação de valência de posicionamento e duas métricas complementares: grau de equilíbrio e engajamento.*

1. Introdução

A detecção de posicionamento tem uma longa tradição no domínio dos debates políticos e ideológicos online [Mohammad et al. 2016] e a sua aplicabilidade prática permite avaliar a opinião pública, identificar grupos que se cruzam e divergem, e compreender questões de interesse para diferentes comunidades de usuários [Magdy et al. 2016]. Estudos recentes sugerem que os usuários da web tendem a polarizar sua opinião [Conover et al. 2011] e a formar comunidades a partir de afinidades políticas [Figeac and Favre 2021], seguindo o princípio da homofilia [McPherson et al. 2001].

Apesar da natureza ruidosa, desestruturada e fragmentada dos tweets, eles têm sido analisados com sucesso por meio de técnicas de mineração de texto, processamento

de linguagem natural e aprendizado de máquina [D’Andrea et al. 2019]. Atualmente, uma das suas principais aplicações é o monitoramento de questões políticas, uma vez que as informações extraídas da web produzem um meio relevante de, juntamente com técnicas mais tradicionais como pesquisa de opinião [Jungherr et al. 2017], estimar o posicionamento médio de uma população em relação a um determinado tópico.

As conversas nas mídias sociais entre 2020 e 2021 no Brasil foram amplamente dominadas por discussões sobre as ações do governo federal no combate à pandemia [Maia et al. 2021] e a vacinação contra a Covid-19 [Malagoli et al. 2021], com discurso altamente politizado. Em maio de 2021 foram iniciados os trabalhos de uma Comissão Parlamentar de Inquérito, intitulada da Covid-19 CPI¹, criada para investigar possíveis omissões do governo no combate à pandemia. Uma base de dados composta por conversações sobre a CPI foi coletada e utilizada como estudo de caso nesse artigo por se tratar de um tema controverso, polarizado e atual ao ter a política e a ciência como temas centrais.

As contribuições desse trabalho são: (i) utilização de uma técnica não supervisionada de detecção de posicionamento para determinar o posicionamento de clusters de usuários do Twitter em relação à tópicos polarizados, (ii) rotulação automática individualizada de centenas de milhares de usuários utilizando uma combinação de atribuição de rótulos e cálculo de pontuação de valência, (iii) caracterização do comportamento dos usuários rotulados utilizando uma pontuação de engajamento e grau de equilíbrio, (iv) método de rotulação que pode ser generalizado para detecção automática de posicionamento em diferentes tópicos com indícios de polarização e contexto político.

O artigo está organizado da seguinte forma: a Seção 2 discute fundamentos e trabalhos relacionados, a Seção 3 detalha a base de dados estudada, a Seção 4 descreve a metodologia proposta e a Seção 5 analisa os resultados. Finalmente, a Seção 6 discute as ameaças à validade da abordagem proposta e a Seção 7 conclui o trabalho e oferece sugestões de trabalhos futuros.

2. Fundamentos e Trabalhos Relacionados

A detecção de posicionamento visa determinar de forma automática se o autor de um texto é a favor, contra ou neutro em relação a um tópico ou proposição [Mohammad et al. 2017]. A sua aplicação é particularmente interessante no campo das mídias sociais, pois oferece a oportunidade de identificar o posicionamento de um elevado número de usuários sobre diferentes questões. Segundo [Darwish et al. 2020], apesar de pesquisas recentes se concentrarem principalmente em paradigmas de aprendizagem de máquina, essas abordagens têm vantagens e desvantagens: o paradigma supervisionado tem implementação simples, mas é dependente de dados de treinamento rotulados manualmente e sua acurácia varia de acordo com a técnica de classificação escolhida e o tamanho das bases de teste e treino. Já os paradigmas semi e não supervisionado requerem menor interferência humana, pois se baseiam em características da rede ou nas interações realizadas pelos usuários. Esses últimos são mais eficazes na classificação de usuários altamente vocais (que se expressam mais intensamente), entretanto produzem resultados abaixo do ideal para usuários que raramente expressam a sua opinião.

¹<http://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=464162&ori=1>

[Küçük and Can 2020] descrevem trabalhos recentes que utilizaram uma ampla gama de métodos de detecção de posicionamento, incluindo áreas de pesquisa como processamento de linguagem natural e ciências sociais computacionais. A detecção de posicionamento em redes sociais pode ser dividida em duas grandes abordagens: na primeira são utilizados algoritmos tradicionais de aprendizado de máquina [Wojatzki and Zesch 2016], com etapas de rotulação manual [Mohammad et al. 2017, Sobhani et al. 2017] enquanto que a segunda abordagem é baseada em técnicas de aprendizado profundo [Zarrella and Marsh 2016], em combinação com modelos BERT (*Bidirectional Encoder Representations from Transformers*) [Popat et al. 2019, Samih and Darwish 2021, Sirrianni et al. 2021]. A aplicação de técnicas de modelagem de tópicos também produz resultados interessantes: [Lin et al. 2017] utilizaram o modelo *Nonnegative Matrix Factorization* (NMF) para a identificação de instâncias múltiplas de posicionamento, com o uso de rotulação automatizada de textos.

Existem poucos trabalhos sobre detecção de posicionamento sobre temas políticos em idiomas distintos do inglês [Vamvas and Sennrich 2020]: [Rashed et al. 2021] coletaram tweets relacionados às eleições da Turquia em 2018 e aplicaram um método não supervisionado de detecção de posicionamento, enquanto [Ebeling et al. 2020] propõem um *framework* para analisar e caracterizar o comportamento de grupos com posições opostas e a relação com a polarização política, usando tweets sobre a Covid-19 no Brasil. Esses autores empregaram técnicas de modelagem de tópicos, detecção de comunidade e análise de características linguísticas para determinar aspectos psicológicos nos textos. [Christie et al. 2018] caracterizaram, de um amplo conjunto de tweets em português, a corrida presidencial brasileira de 2018.

Em estudo anterior [Santos and Goya 2021], nós combinamos a contagem de recorrência de palavras para rotular automaticamente o posicionamento de um usuário como contrário ou favorável em relação a um tema, com a classificação supervisionada de BERT para a modelagem de tópicos e a técnica de classificação não supervisionada proposta por [Darwish et al. 2020] para prever o posicionamento de usuários prolíficos do Twitter em relação a assuntos polêmicos em língua portuguesa. Nossos resultados mostraram que os usuários mais vocais eram responsáveis por cerca de 40 a 50% do volume de retweets com forte polarização em dois grupos majoritários, sendo que menos de 1,5% dos usuários apresentava posicionamento divergente ao longo do tempo em relação ao tópico estudado. Neste trabalho, nós expandimos significativamente o artigo anterior, a dimensão de usuários rotulados e introduzimos medidas que complementam a aferição do posicionamento para cada usuário rotulado e fornecem indicativos de engajamento e equilíbrio. Assim, enquanto que em nosso trabalho anterior e no método proposto por [Darwish et al. 2020] a rotulação de posicionamento ocorria no nível de agrupamentos de usuários mais vocais, neste trabalho consideramos o posicionamento no nível do usuário, ao inferir a valência de posicionamento para usuários únicos, em vez de agrupamentos de usuários.

3. Coleta de Dados

A Tabela 1 mostra um resumo dos dados dos tweets coletados relacionados à CPI da Covid-19, utilizando-se da API v1.1 do Twitter². As coletas foram realizadas em tempo

²<https://developer.twitter.com>

Tabela 1. Dados dos tweets utilizados neste estudo

Dados	Descrição	Informação Utilizada	Tamanho
Tweets	Tweets coletados entre 4 de maio e 16 de julho de 2021	Id do tweet, Data de criação, Texto do retweet, Nome de usuário que fez o retweet, Nome de usuário retuitado	14,7 mi
Usuários	Dados dos usuários que fizeram retweets em cada base	Data de criação, Nome de usuário	662.793

real em intervalos de 24 horas para cada sessão de depoimentos. As *strings* de busca usadas na coleta foram compostas por termos e *hashtags* selecionados de acordo com os cronogramas de depoimentos divulgados na imprensa. As bases utilizadas neste artigo estão disponíveis no site do projeto Observa³. Como a primeira fase da CPI durou onze semanas, foi atribuída uma base de dados para cada semana de coleta.

4. Metodologia

O objetivo desse trabalho é desenvolver um método automatizado para estimar o posicionamento de um usuário do Twitter em um tema político e com indícios de polarização. A ideia geral é produzir um conjunto de dados rotulados de alta qualidade, mas generalizável e com menor risco de ocorrência de vieses e erros da rotulação manual, além de um *framework* que possa ser utilizado em diferentes corpus para rotulação automática de usuários do Twitter, considerando contextos políticos diversos e polarizados.

A abordagem proposta consiste em um método não supervisionado organizado em 6 etapas, tendo os resultados preliminares das primeiras quatro etapas sido discutidos no estudo [Santos and Goya 2021]:

1. Aplicação de um método de detecção de posicionamento não supervisionado que busca agrupar os usuários em dois clusters polarizados levando em conta os usuários que eles retuitam.
2. Recuperação do conteúdo retuitado em cada cluster e normalização dos dados textuais para aplicação de técnica de modelagem de tópicos.
3. Modelagem de tópicos em cada cluster utilizando *embeddings*.
4. Identificação do posicionamento dos clusters utilizando um dicionário de termos recorrentes utilizados por cada grupo polarizado, com confirmação humana.
5. Rotulação dos usuários, utilizando uma técnica de atribuição de rótulos e cálculo de pontuação de valência, e rotulação de tweets.
6. Caracterização do comportamento dos usuários rotulados.

4.1. Detecção de posicionamento baseada na iteração entre usuários

Nesta etapa foi aplicado um método de detecção de posicionamento não supervisionado que busca agrupar os usuários levando em conta os usuários que eles retuitam, com base no estudo desenvolvido por [Darwish et al. 2020]: (i) **Seleção de características**. Foi calculada a média de retweet dos usuários por base, bem como o intervalo de confiança dessas médias. Como esse intervalo variou entre 5,56 e 6,58, optou-se por selecionar

³Projeto CPI da Covid do Observatório de Conflitos da Internet, apoiado por Fapesp/MCTI/MC/CGI 2018/23022-3 <https://observa.pesquisa.ufabc.edu.br/list/cpi-da-covid/>

usuários que fizeram no mínimo 5 retweets únicos em cada base. Após isso, foi construído um vetor de categorias para cada usuário selecionado baseado nas contas que ele retuitou e calculada a similaridade do cosseno entre cada par de usuários, utilizando esse vetor. (ii) **Redução de dimensionalidade**. Após o cálculo da similaridade, os usuários foram projetados em um espaço bidimensional utilizando o algoritmo *Uniform Manifold Approximation and Projection* (UMAP) [McInnes et al. 2018]. (iii) **Clusterização**. Feita a projeção dos usuários em um espaço bidimensional, eles foram clusterizados utilizando o algoritmo *Mean Shift*. Isso permitiu a obtenção de dois grandes clusters polarizados.

É importante destacar que nesse estudo foram utilizados apenas retweets. Os retweets com comentários foram excluídos porque considerou-se que um usuário ao compartilhar um retweet com comentário pode expressar endosso ou desacordo, enquanto que o simples ato de retuitar normalmente sinaliza concordância ou apoio por parte dos usuários que compartilham o texto [Wischniewski et al. 2022].

4.2. Normalização de dados textuais para modelagem de tópicos

Essa etapa é necessária para a realização da modelagem de dados, dado que um tweet é composto de dados textuais altamente irregulares. Após a obtenção dos textos dos retweets compartilhados pelos usuários de cada cluster, estes foram normalizados utilizando as seguintes técnicas: (i) **Padronização**. Detecção e remoção de *links*, emoticons e *emojis*, *hashtags*, pontuações, espaços e linhas em branco, 'RT' e @ (informações de usuários), números, palavras reservadas do Twitter, aspas simples e duplas e cadeias de caracteres que possuam repetições, conversão de todas as palavras em minúsculas; (ii) **Tokenização**. Tokenização dos textos dos tweets separados por espaços em branco, removendo também as *stopwords* (verbos auxiliares, preposições, artigos, pronomes e as palavras mais frequentes) e palavras com menos de três caracteres; e (iii) **Remoção de textos duplicados**.

4.3. Modelagem de tópicos

Partindo-se do princípio da homofilia [McPherson et al. 2001, Figeac and Favre 2021], considerou-se que os usuários de cada cluster teriam predominantemente a mesma posição (contrária ou a favorável) em relação à CPI. Assim, foi realizada a modelagem de tópicos separadamente nos textos dos retweets compartilhados pelos usuários de cada agrupamento majoritário e em cada base coletada. Para a modelagem de tópicos foi utilizada a técnica BERTopic desenvolvida por [Grootendorst 2022]. Esta técnica consiste na extração de *embeddings* de documentos, agrupamento para a criação de grupos de documentos semelhantes com UMAP e HDBSCAN, e na geração de tópicos através da obtenção das palavras mais importantes por meio de agrupamento com TF-IDF baseado em classe (c-TF-IDF). Na extração dos *embeddings* foi utilizado o modelo BERT pré-treinado em português conhecido como BERTimbau [Souza et al. 2020], o qual utiliza um corpus em português brasileiro (BrWaC [Wagner Filho et al. 2018]).

4.4. Identificação de posicionamento dos clusters

Um pequeno dicionário de termos recorrentes foi elaborado a partir da identificação de conjuntos de palavras que eram usadas predominantemente por um dos grupos: contrário [**circo**, **palhaç***, **ladr***, **pedof***] ou favorável [**bolsonar***, **genoc***, **negac***, **imunidade**, **desgoverno**]. Após isso, foi realizada uma busca dos termos do dicionário e feita uma contagem de ocorrências de palavras nos conjuntos de tópicos gerados por cada cluster

de cada base. Feito isso, foi realizada a rotulação automática de cada cluster baseada no peso relativo associado a cada palavra e suas variações.

4.5. Rotulação automática dos usuários

A detecção de posicionamento não supervisionada só funciona de forma eficaz na tarefa de classificar os usuários mais vocais (com quantidade de retweets acima da média), os quais constituem a minoria dos usuários [Darwish et al. 2020]. No conjunto de bases estudadas neste trabalho, os usuários mais vocais variaram entre 13 a 21% do total.

Um dos principais objetivos deste trabalho é rotular o posicionamento do maior número possível de usuários adicionais. Para tanto, foram utilizados os usuários rotulados automaticamente em cada cluster na etapa anterior para rotular os usuários menos vocais, através das contas que cada usuário clusterizado retuitou. Calculou-se então o número total de retweets que as contas de usuário receberam em cada cluster, bem como a sua porcentagem. Feito isso, foi atribuído um rótulo (contrário ou favorável) a partir da maior porcentagem de retweets recebidos obtida em cada cluster. As contas cuja diferença de porcentagem entre os clusters era inferior ao limiar de 20% não foram consideradas para a rotulação dos menos vocais. Essa escolha foi feita com o intuito de diminuir a incerteza (evitar casos de contas igualmente retuitadas em ambos os clusters) e realizar a atribuição, num primeiro momento, de rótulos dos usuários mais fortemente posicionados. O percentual de contas retuitadas por usuários de ambos os clusters variou entre 3,4 a 5,2% ao longo das bases e consistiu em muitos casos de contas de jornalistas e veículos de imprensa, enquanto que o percentual de contas desconsideradas variou entre 0,28 e 0,69%, o que é relativamente pouco expressivo. Após isso, as contas de usuários foram agrupadas em duas listas (contra ou a favor) e foram contados quantos retweets os usuários fizeram para as contas de cada lista.

4.6. Pontuação de valência para posicionamento

Para rotular o posicionamento dos usuários menos vocais, foi calculada uma pontuação de valência. Essa pontuação variou entre [-1, 1], onde um valor próximo a 1 indica que o usuário tem posicionamento similar a usuários favoráveis à CPI; -1 indica posicionamento contrário; e 0, uma posição de equilíbrio. Para calcular a pontuação de valência foi utilizada a equação (1), que é uma adaptação da proposta por [Conover et al. 2011].

$$p(u) = 2 * \left(\frac{r(u, F)}{N(F)} \middle/ \left(\frac{r(u, F)}{N(F)} + \frac{r(u, C)}{N(C)} \right) \right) - 1 \quad (1)$$

Onde $r(u, F)$ é o número de retweets feitos pelo usuário u para contas de usuários da lista com rótulo favorável e $N(F) = \sum_u r(u, F)$ é o número total de retweets recebidos pelas contas da lista com rótulo favorável em cada base. Analogamente, $r(u, C)$ e $N(C)$ são definidos para posicionamento contrário.

4.7. Caracterização do comportamento dos usuários

O comportamento dos usuários rotulados a partir da pontuação de valência de posicionamento é caracterizado por meio das duas métricas complementares descritas a seguir, pois a valência sozinha nem sempre traduz as noções de engajamento, indefinição ou imparcialidade no debate, úteis para melhor compreender os eventos:

1. Grau de **engajamento** de um usuário u , representado pelo máximo valor entre (i) sua participação relativa em retweets emitidos e (ii) sua presença relativa ao longo das 11 semanas. O grau de engajamento é calculado pela equação (2), em que $r(u) = r(u, F) + r(u, C)$ e representa o número total de retweets emitidos por u e que puderam ser rotulados, $max(r)$ representa a maior quantidade de total de retweets realizados por um único usuário nas 11 semanas e $s(u)$ é o número de semanas em que o usuário u participou. Usou-se logaritmo para uniformizar as grandezas, pois a distribuição da quantidade de retweets seguiu uma função potência. Assim, os valores de engajamento variam entre 0 (baixa participação) e 1 (forte engajamento), capturando ambas as noções de comprometimento: volume de interações e presença no debate ao longo do tempo.

$$eng(u) = max\left(\frac{\log_{10}(r(u))}{\log_{10}(max(r))}, \frac{s(u) - 1}{10}\right) \quad (2)$$

2. Grau de **equilíbrio** de um usuário u , cujo cálculo é feito pela equação (3), representa a diferença entre as proporções de retweets a favor e contra realizados por u , considerando a totalidade de retweets $R(u)$ (incluindo os retweets realizados por u que não puderam ser rotulados). Assim, valores muito próximos dos extremos (-1 ou 1) indicam uma convergência de posicionamento para um dos polos, enquanto valores próximos de 0 podem representar indefinição, imparcialidade ou que o usuário retuitou mais postagens que não puderam ser rotuladas.

$$eq(u) = \frac{r(u, F) - r(u, C)}{R(u)} \quad (3)$$

A pontuação de valência de posicionamento e o grau de equilíbrio podem ser agrupados em 5 faixas com valores de -1 a 1 [Stefanov et al. 2020]:

$$pont = \begin{cases} --, & \text{se } p \in [-1, -0,6) \\ -, & \text{se } p \in [-0,6, -0,2) \\ 0, & \text{se } p \in [-0,2, 0,2) \\ +, & \text{se } p \in [0,2, 0,6) \\ ++, & \text{se } p \in [0,6, 1] \end{cases} \quad (4)$$

A escolha de um intervalo exato de 0,4 para cada faixa da equação (4) foi de certa forma arbitrária, mas considerada apropriada pois captura o posicionamento de um usuário mesmo em valores intermediários de pontuação. Além disso, o valor não poderia ser muito próximo de 0, pois a faixa de indefinição ou equilíbrio deve capturar usuários que não mostraram um posicionamento claro com relação aos seus retweets [Bechini et al. 2020].

5. Resultados e Discussão

As métricas gerais obtidas a partir da aplicação do método proposto são mostradas na Tabela 2. O percentual de retweets feitos para contas não rotuladas foi de pouco mais de um terço (35,7%). Apesar disso, é interessante observar que apenas 15.620 usuários (2,4% do total) não puderam ser rotulados pois não retuitaram nenhuma das contas pertencentes aos dois clusters majoritários polarizados ou retuitaram contas que foram descartadas por serem igualmente retuitadas por usuários de ambos os clusters. O volume de contas não rotuladas pode ser considerado baixo e parece indicar que a quase totalidade dos usuários da base retuitou as mesmas contas de usuário. Isso pode ser

um indício da ocorrência do efeito conhecido como câmara de eco (*echo chamber*), no qual os usuários têm suas crenças reforçadas por meio de interações com outros usuários que pensam da mesma forma, levando a padrões de interação homofílicos [Cinelli et al. 2020].

Tabela 2. Métricas Gerais

Métrica	Quantidade	Porcentagem
Quantidade de usuários únicos	664.793	-
Quantidade de usuários rotulados	649.173	97,6%
Quantidade de usuários mais vocais	470.398	72,5%
Quantidade de usuários menos vocais	178.775	27,5%
Total de retweets	10.153.438	-
Total de retweets rotulados	6.524.578	64,3%
Total de retweets rotulados a favor	3.495.628	53,6%
Total de retweets rotulados contra	3.028.950	46,4%

O gráfico de tweets rotulados da Figura 1 mostra a distribuição dos tweets compartilhados pelos usuários ao longo das semanas. É interessante notar a ocorrência de uma espécie de “empate percentual” na maioria das bases, com alternância de uma maioria simples tendendo para um dos posicionamentos ao longo das semanas, com exceção das bases das semanas 2 e 3, em que a concentração de tweets favoráveis foi consideravelmente maior. No entanto, quando se observa o gráfico da pontuação da valência de posicionamento dos usuários ao longo das semanas (mesma figura, à direita), percebe-se que os usuários rotulados como muito contrários à CPI estiveram em menor número (18,5%), mas fizeram proporcionalmente mais retweets (46,4%) que os usuários muito favoráveis, podendo ser considerados uma espécie de “minoria barulhenta”.

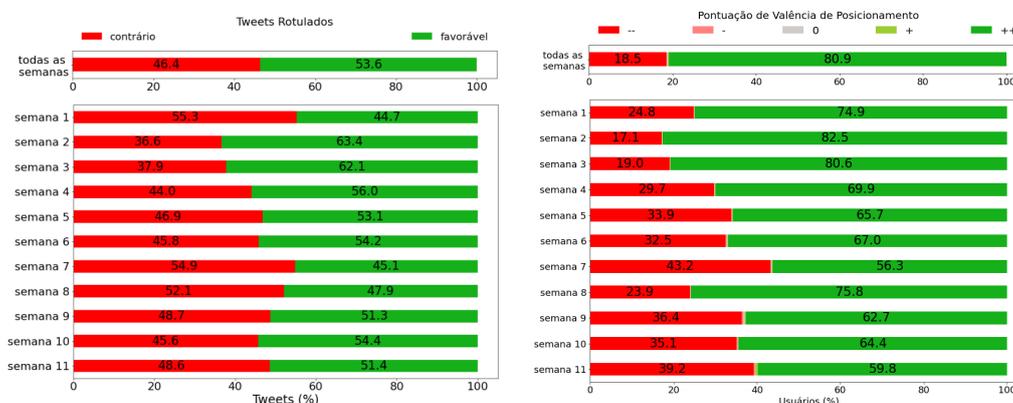


Figura 1. Distribuições dos tweets (esq.) e dos usuários rotulados por pontuação de valência de posicionamento por faixa (dir.)

5.1. Análise do comportamento dos usuários

Nesta subseção são analisadas as diferenças de inclinação e interesse entre os usuários mais e menos vocais utilizando duas medidas: o grau de equilíbrio e o grau de engajamento. A Figura 2 mostra o grau de equilíbrio dos usuários rotulados. Um ponto importante diz respeito à frequência de usuários na faixa de posicionamento de grau 0: a distribuição de usuários rotulados nessa faixa é de apenas 0,82%, quando se consideram todos os usuários. Isso quer dizer que os usuários convergiram para um dos extremos, com 78% com grau muito favorável e 17,3% com grau muito contrário quando se consideram todos os usuários. Ao analisar os usuários divididos entre mais

e menos vocais essas proporções mudam: a diferença na distribuição de usuários nos extremos se reduz para 59,9% com grau muito favorável e 28,1% com grau muito contrário dos usuários mais vocais, enquanto aumenta para 84,9% com grau muito favorável e 13,2% com grau muito contrário no caso dos usuários menos vocais. Esses percentuais reforçam a suposição de forte polarização no posicionamento dos usuários rotulados.

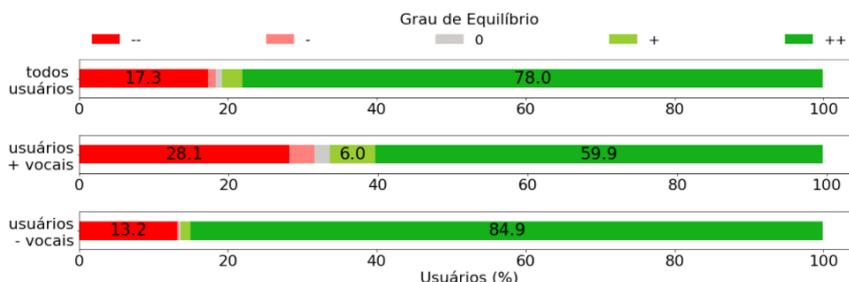


Figura 2. Distribuição do grau de equilíbrio por tipo de usuário

Em relação ao grau de engajamento (Figura 3), quando se observa a totalidade dos usuários, 76,6% se concentram na faixa com grau mais baixo, proporção que aumenta para 98,9% quando se consideram os usuários menos vocais. Em relação aos usuários mais vocais, a proporção de usuários é mais bem distribuída com 18,2% na faixa de grau mais baixo, 15,8% na faixa de grau mais alto e 46,8% com grau mediano. Esses valores indicam que a maioria absoluta dos usuários menos vocais participou da discussão em poucas semanas e/ou fez poucos retweets, enquanto que os usuários mais vocais tiveram participação mais ativa e foram significativamente mais engajados.

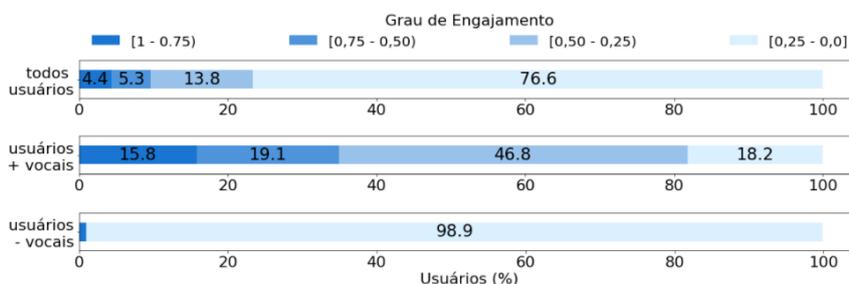


Figura 3. Distribuição do grau de engajamento dos usuários, por faixa

6. Ameaças à Validade

Esta seção discute alguns pontos a serem observados quanto à validade do estudo apresentado. O primeiro é a forma como os grupos foram definidos: o uso de redes de retweets é uma abordagem muito usada para rotulação de grupos em redes sociais, porém pode introduzir vieses. Para mitigar esse problema, excluíram-se retweets com comentários das bases estudadas. Outro ponto é a existência de grupos (minoritários) de usuários que não estejam discutindo temas relacionados ao tópico estudado. Tais grupos foram eliminados quando identificados durante a etapa de detecção não supervisionada de posicionamento. Um terceiro ponto é a separação de contas de usuário que são retuitadas por ambos os lados e, portanto, percebidas como neutras. Dada a significativa polarização verificada no conjunto de bases e a baixa proporção percentual de usuários nessas condições, conclui-se que não considerar os retweets para tais contas ao rotular os usuários menos vocais não deve afetar os padrões gerais identificados.

Um último ponto diz respeito à existência de *bots*. Foi utilizada a API Botometer⁴ para detectar a existência de *bots* em um recorte de 1.747 usuários que tinham maior probabilidade de ter rótulos divergentes (pontuação de valência entre [-0,25, +0,25]). Como 380 dessas contas foram suspensas ou deletadas pelo próprio usuário, foram analisados um total de 1.367 usuários, dos quais 457 (33%) possuía em torno de 70% de probabilidade de serem *bots*. Como esse recorte engloba apenas 0,2% da base, um estudo em um conjunto maior de usuários se faz necessário.

Por fim, a rotulação de usuários é uma tarefa complexa e subjetiva mesmo quando realizada de forma manual e caso a caso. Para analisar se o método proposto estava atribuindo rótulos adequados, foi selecionada uma amostra aleatória de mais de 100 usuários de todas as faixas de equilíbrio. Para cada usuário sorteado foram analisados em conjunto: valência de posicionamento, texto do tweet original e contas retuitadas. Verificou-se que o modelo teve um excelente resultado, tendo atribuído rótulos adequados a todos os usuários amostrados e analisados.

7. Conclusão e Trabalhos Futuros

Esse trabalho propôs um método de seis etapas para rotulação de posicionamento de usuários de forma a caracterizar a opinião pública sobre um tema político e controverso: a CPI da Covid-19. A análise mostrou que o método é bastante efetivo, tendo sido possível atribuir um rótulo para a quase totalidade dos usuários analisados. É importante destacar a diferença de comportamento dos usuários em relação ao posicionamento aferido: os usuários muito contrários à CPI, apesar de estarem em menor número, compartilharam em média uma quantidade maior de retweets do que os usuários favoráveis, cuja proporção é significativamente maior. Além disso, ao analisar a dimensão temporal dos dados coletados e o posicionamento dos usuários ao longo das semanas, é possível inferir prováveis mudanças da opinião pública, possivelmente associadas a eventos relacionados à CPI.

Como trabalho futuro, deseja-se ampliar o leque de características analisadas pelo método, como análise das principais narrativas de cada cluster. Além disso, considera-se interessante explorar novas métricas que possam apoiar análises qualitativas sobre os eventos, introduzindo conceitos como características de fluxo de informações entre os clusters e entendimento da dinâmica de rede de cada grupo de interesse.

Referências

- Bechini, A., Ducange, P., Marcelloni, F., and Renda, A. (2020). Stance analysis of twitter users: the case of the vaccination topic in italy. *IEEE Intelligent Systems*.
- Christie, W., Reis, J. C., Moro, F. B. M. M., and Almeida, V. (2018). Detecção de posicionamento em tweets sobre política no contexto brasileiro. In *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*. SBC.
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2020). Echo chambers on social media: A comparative analysis. *arXiv preprint arXiv:2004.09603*.
- Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 5, pages 89–96.
- D’Andrea, E., Ducange, P., Bechini, A., Renda, A., and Marcelloni, F. (2019). Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications*, 116:209–226.

⁴<https://botometer.osome.iu.edu/>

- Darwish, K., Stefanov, P., Aupetit, M., and Nakov, P. (2020). Unsupervised user stance detection on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 141–152.
- Ebeling, R., Sáenz, C. A. C., Nobre, J., and Becker, K. (2020). Quarenteners vs. chloroquiners: A framework to analyze how political polarization affects the behavior of groups. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 203–210. IEEE.
- Figeac, J. and Favre, G. (2021). How behavioral homophily on social media influences the perception of tie-strengthening within young adults’ personal networks. *New Media & Society*, page 14614448211020691.
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv e-prints*, pages arXiv–2203.
- Jungherr, A., Schoen, H., Posegga, O., and Jürgens, P. (2017). Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. *Social Science Computer Review*, 35(3):336–356.
- Küçük, D. and Can, F. (2020). Stance detection: A survey. *ACM Computing Surveys (CSUR)*, 53(1):1–37.
- Lin, J., Mao, W., and Zhang, Y. (2017). An enhanced topic modeling approach to multiple stance identification. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2167–2170.
- Magdy, W., Darwish, K., and Weber, I. (2016). # failedrevolutions: Using twitter to study the antecedents of isis support. In *2016 AAAI Spring Symposium Series*.
- Maia, M., Oliveira, E., and Gallegos, L. (2021). Covid-19 e tweets no brasil: coleta, tratamento e análise de textos com evidências de estados afetivos alterados em momentos impactantes. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 79–90. SBC.
- Malagoli, L., Stancioli, J., Ferreira, C., Vasconcelos, M., Silva, A. P., and Almeida, J. (2021). Caracterização do debate no twitter sobre a vacinação contra a covid-19 no brasil. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 55–66, Porto Alegre, RS, Brasil. SBC.
- McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29).
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.
- Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 31–41.
- Mohammad, S. M., Sobhani, P., and Kiritchenko, S. (2017). Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):1–23.

- Popat, K., Mukherjee, S., Yates, A., and Weikum, G. (2019). Stancy: Stance classification based on consistency cues. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6413–6418.
- Rashed, A., Kutlu, M., Darwish, K., Elsayed, T., and Bayrak, C. (2021). Embeddings-based clustering for target specific stances: The case of a polarized turkey. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 537–548.
- Samih, Y. and Darwish, K. (2021). A few topical tweets are enough for effective user stance detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2637–2646.
- Santos, P. D. and Goya, D. H. (2021). Automatic twitter stance detection on politically controversial issues: A study on covid-19’s cpi. In *Anais do XVIII Encontro Nacional de Inteligência Artificial e Computacional*, pages 524–535. SBC.
- Sirrianni, J. W., Liu, X., and Adams, D. (2021). Predicting stance polarity and intensity in cyber argumentation with deep bidirectional transformers. *IEEE Transactions on Computational Social Systems*, 8(3):655–667.
- Sobhani, P., Inkpen, D., and Zhu, X. (2017). A dataset for multi-target stance detection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 551–557.
- Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23 (to appear)*.
- Stefanov, P., Darwish, K., Atanasov, A., and Nakov, P. (2020). Predicting the topical stance and political leaning of media using tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 527–537.
- Vamvas, J. and Sennrich, R. (2020). X-stance: A multilingual multi-target dataset for stance detection. In *5th SwissText & 16th KONVENS Joint Conference 2020*, page 9. CEUR-WS. org.
- Wagner Filho, J. A., Wilkens, R., Idiart, M., and Villavicencio, A. (2018). The brwac corpus: A new open resource for brazilian portuguese. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.
- Wischniewski, M., Ngo, T., Bernemann, R., Jansen, M., and Krämer, N. (2022). “i agree with you, bot!” how users (dis) engage with social bots on twitter. *New Media & Society*, page 14614448211072307.
- Wojatzki, M. and Zesch, T. (2016). Itl. uni-due at semeval-2016 task 6: Stance detection in social media using stacked classifiers. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 428–433.
- Zarrella, G. and Marsh, A. (2016). Mitre at semeval-2016 task 6: Transfer learning for stance detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 458–463.