

Social media user stance detection without stance text

Pedro Semcovici

pedrosemcovici@usp.br

Universidade de São Paulo, EACH

São Paulo, SP, Brasil

Ivandré Paraboni

ivandre@usp.br

Universidade de São Paulo, EACH

São Paulo, SP, Brasil

Abstract

Context: User stance detection systems are intended to determine whether an individual holds a stance for or against a given target topic. For instance, ‘Vaccines save lives’ is a stance in favour of vaccination.

Problem: Most existing approaches to stance detection take as an input a piece of text assumed to convey a stance towards the intended target, and then predict its polarity. However, when stance text is unavailable (e.g., when the individual did not publish any opinion regarding that particular topic), models of this kind are unable to determine the user’s stance.

Solution: In the case of social media, we may consider as a substitute for stance text the use of other (non-stance) publications made by the user, or even by their close contacts.

Information Systems theory: The work was considering Social Network Theory and Social Information Processing Theory.

Method: We built a number of stance detection models from stance text and from non-stance data authored by the target users and their close contacts using supervised BERT, prompt-based LLaMA, and stacking.

Results: The non-stance publications made by an individual and, to a lesser extent, those made by his/her contacts, were found to be strong predictors of user stance and, in some cases, their combination was shown to outperform the actual use of text stance.

Contributions: Using network-related information may be indeed a substitute for stance text, and may enable the development of systems for scenarios in which stance text is unavailable.

CCS Concepts

- Computing methodologies → Natural language processing;
- Information systems → World Wide Web; Web mining;
- Applied computing → Document management and text processing.

Keywords

Natural language processing, user stance detection, social media

1 Introdução

A tarefa de detecção de posicionamentos (do inglês, *stance*) a partir de dados textuais como postagens em redes sociais e outros trata de decidir se um texto expressa uma atitude favorável ou desfavorável em relação a um determinado tópico-alvo de interesse [16]. Por exemplo, a frase ‘A situação das escolas públicas é lamentável e precisa ser resolvida com urgência’ pode ser vista como um posicionamento favorável ao alvo ‘ensino público’. Além disso, observa-se que este posicionamento foi expresso por meio de um sentimento negativo, do que decorre que a tarefa não se confunde com análise de sentimentos (AS) [1].

Posicionamentos são frequentemente expressos em redes sociais e discussões *online* em geral, e há grande interesse em aplicações computacionais capazes de reconhecê-los de forma automática. Aplicações deste tipo, que tipicamente fazem uso de métodos de aprendizado de máquina supervisionado, são assim amplamente estudadas na área de Processamento de Línguas Naturais (PLN) e afins.

Na sua forma mais tradicional, a detecção de posicionamentos é modelada como um problema de classificação de nível textual, ou seja, cujo objetivo é o de inferir um possível posicionamento expresso no texto fornecido como entrada [3, 11, 26]. Uma formulação alternativa do problema, entretanto, seria o objetivo de inferir o posicionamento de um indivíduo (e.g., um usuário de rede social) em situações em que este indivíduo não tenha manifestado nenhum posicionamento explícito sobre o alvo de interesse. Esta formulação do problema, conhecida por detecção de posicionamentos *em nível de usuário* (do inglês, *user stance detection*, cf. [10, 30]), é o foco do presente trabalho.

A detecção de posicionamentos em nível de usuário difere da detecção de nível textual tradicional em vários aspectos. Em especial lugar, cabe observar que, na detecção de posicionamento em nível de usuário, a ausência da informação de entrada (ou seja, o texto do posicionamento) precisa ser substituída por outras fontes de conhecimento. Estas fontes podem ser do tipo textual (e.g., outras postagens do mesmo indivíduo, tratando de outros assuntos, etc.) ou não textual (seu comportamento na rede social, suas conexões com outros usuários, etc.), e normalmente acrescentam complexidade à tarefa. Por exemplo, observa-se que a detecção de nível textual normalmente resulta em uma predição exata para cada texto de entrada (e.g., um rótulo favorável/desfavorável/neutro), enquanto que na detecção de posicionamentos em nível de usuários é possível encontrar evidências contraditórias (e.g., um indivíduo pode manifestar opiniões liberais e conservadoras ao longo de suas postagens na rede social, o que torna mais complexa a definição do seu posicionamento pessoal sobre o assunto).

De especial interesse para o presente trabalho, observa-se que, na ausência de informações que permitam detectar o posicionamento de um indivíduo sobre um alvo específico, podemos utilizar informações contextuais indicativas da sua atividade na rede social [4, 21]. Além disso, esta ideia pode ser expandida com base no princípio de homofilia [23] (i.e., o princípio segundo o qual indivíduos com afinidades tendem a formar grupos), o que sugere a possibilidade de uso de dados relativos aos contatos (e.g., amigos, seguidores, etc.) com quem o indivíduo se relaciona na rede social. Por exemplo, mesmo que uma pessoa não tenha manifestado seu posicionamento sobre um assunto específico (e.g., política) na rede social, é possível que seus contatos (e.g., amigos, seguidores, etc.) ofereçam pistas de qual seria o provável posicionamento deste indivíduo sobre o assunto.

Com base nestas observações, o presente trabalho apresenta um estudo sobre a tarefa computacional de detecção de posicionamentos em nível de usuários de redes sociais em português que, até onde temos conhecimento, é a primeira deste gênero. De forma mais específica, o estudo realizado considera como substituto ao texto do posicionamento outras postagens dos usuários (i.e., que não falam sobre o tópico-alvo de interesse), as postagens de seus contatos próximos, e a combinação destes, explorando métodos computacionais variados como classificadores do tipo BERT [12], métodos de engenharia de *prompts* para grandes modelos de linguagem (LLMs) do tipo LLama [32], e arquiteturas de *stack ensemble*.

O restante deste artigo está organizado da seguinte forma. A Seção 2 apresenta uma breve revisão da área de detecção de posicionamentos com uso de dados textuais e não textuais. A Seção 3 especifica o problema computacional a ser tratado nesta pesquisa. A Seção 4 descreve os métodos a serem investigados, cujos resultados são apresentados e discutidos na Seção 5. Finalmente, a Seção 6 apresenta um resumo das contribuições deste estudo e futuras direções de pesquisa.

2 Trabalhos relacionados

Foi conduzido um levantamento bibliográfico de estudos na área de detecção de posicionamentos com base em dados textuais e não textuais, incluindo tanto os que propõem novas abordagens computacionais para a tarefa como aqueles que descrevem a construção de um córpus textual ou multimodal para este fim. A Tabela 1 sumariza os principais estudos identificados. Estes estudos são categorizados por gênero (p=discurso político, r=Reddit, t=Twitter, j=jornalístico, o=opiniões), idioma (Ar=Árabe, Ca=Catalão, Da=Dinamarquês, En=Inglês, Fr=Francês, It=Italiano, Sp=Espanhol), representação textual (w=palavras, p=part-of-speech, c=caracteres, we=word embeddings, bp=BERT, n=ngrams), tipo de característica não textual (dem=informações demográficas, m=menções, r=respostas, rt=retweets, dom=informação específica do domínio ou tarefa, fr=amigos, fo=seguidores, h=hora da postagem, dist=distância para outros usuários, int=interações), e método computacional (SVM=máquinas de vetor suporte, LR=regressão logística, ADDA=adversarial discriminative domain adaptation, etc.).

Tabela 1: Detecção de posicionamentos usando dados textuais e não textuais.

Estudo	Gênero	Idioma	Textual	Não textual	Método
[24]	t	En	n	-	SVM
[22]	t	En	w	dem,m,r,rt	SVM
[31]	t	Ca,Sp		dem,m	córpus
[17]	t	Ca,Sp	w,p,c	m	SVM,LR
[9]	t	Ar	w	rt,m	similaridade
[20]	p	Da	we	dem,dom	LSTM,MLP
[18]	t	En,Fr,It,Sp,Ca	w,p,c	m,dom	LSTM,CNN
[7]	t	It		fr,fo,rt,m,h	córpus
[13]	t	It	bp	fr,fo,h,dist	voto maj.
[19]	t	En	w,c	fr,fo,h,dom	SVM
[15]	r	En	w,we	int,dem	SVM
[27]	t	Pt	bp		ADDA
[35]	t,j	En	bp		contrast. learning
[6]	t	En	graph,bp		ADDA
[25]	o	Pt	we		LSTM
[34]	t	En	ChatGPT		prompt
[33]	j	En	we		cond. generation
[29]	t	Pt	bp	fr,fo,m	LSTM
[28]	t	Pt	bp		LSTM, prompt

No que diz respeito aos conjuntos de dados disponíveis para detecção de posicionamentos, cabe destacar o córpus SemEval-2016 [24], uma coleção de 4.870 tweets rotulados manualmente, cobrindo seis alvos. Esse córpus, um dos primeiros do gênero a ser disponibilizado para pesquisa, tornou-se um conjunto de dados influente na área, sendo seguido por vários outros que abordaram outros tópicos-alvo e/ou outros idiomas [7, 9, 15, 18–20, 22, 25, 27, 29, 31]. Além disso, os estudos identificados são, em grande parte, baseados em postagens provenientes da plataforma Twitter/X, o que pode ser explicado pela maior facilidade de coleta de dados que era oferecida na época.

Quanto aos métodos computacionais empregados, o levantamento realizado mostra a evolução típica da pesquisa em PLN e áreas correlatas, com amplo uso de métodos de classificação ditos tradicionais (e.g., SVMs) em estudos mais antigos, gradativamente substituídos por métodos neurais (em especial, LSTMs). Mais recentemente, começam a surgir estudos baseados em engenharia de *prompts* para LLMs [28, 34].

Apesar da aparente afinidade de alguns estudos com o presente estudo – seja no foco no idioma português ou nos métodos computacionais, como será discutido nas próximas seções – cabe observar que, de modo geral, os estudos levantados abordam o problema de detecção de posicionamentos tradicional, ou seja, baseado em um texto de entrada que expressa um posicionamento cuja polaridade pretende-se estimar. Exceções dignas de nota são o estudo em [15], que aborda o problema de detecção de posicionamentos em nível de usuários de rede social, e o estudo em [29], que oferece suporte tanto para a detecção de posicionamentos em nível textual como de usuário.

3 Detecção de posicionamentos sem o posicionamento

O presente estudo contempla a condução de experimentos de aprendizado de máquina a partir de dados extraídos da rede social Twitter/X brasileira com o objetivo de comparar a abordagem de detecção de posicionamentos tradicional (i.e., baseada em um texto contendo um posicionamento a ser avaliado) com alternativas que são baseadas unicamente em informações externas ao texto do posicionamento propriamente dito. De forma mais específica, foram consideradas as três estratégias a seguir:

own-timeline: uso do conjunto de postagens do indivíduo cujo posicionamento deseja-se detectar excluindo-se as postagens que mencionam o tópico-alvo de interesse.

friend-timeline: idem, porém considerando as postagens do contato de amizade com quem o indivíduo apresenta maior volume de interação a respeito do tópico de interesse.

stance-tweet: a estratégia tradicional adotada em problemas de detecção de posicionamento a partir de texto, ou seja, o uso do texto (*tweet*) que expressa um posicionamento contrário ou favorável ao tópico de interesse, aqui empregada como *baseline* para comparação com as alternativas anteriores.

As três estratégias foram implementadas com uso de diferentes abordagens computacionais, incluindo modelos do tipo BERT [12] e Llama [32], e sua combinação em uma arquitetura de *stacking*.

A avaliação destas estratégias foi conduzida com o objetivo de investigar as seguintes questões de pesquisa:

Q1 - Desconsiderando-se as postagens contendo posicionamentos em relação ao alvo de interesse, o uso das postagens restantes da *timeline* de um indivíduo (como em *own-timeline*) permite detectar este posicionamento de forma análoga ao uso do texto propriamente dito (*stance-tweet*)?

Q2 - Qual a melhor abordagem computacional para detectar posicionamentos com base nas postagens da *timeline* de um indivíduo, excluídos seus posicionamentos sobre o alvo de interesse (*own-timeline*)?

Q3 - O uso das postagens do contato de amizade com quem o autor do posicionamento mantenha interação mais próxima sobre o tópico de interesse (como em *friend-timeline*) permite detectar este posicionamento de forma análoga ao uso do texto propriamente dito (*stance-tweet*)?

Q4 - Qual a melhor abordagem computacional para detectar posicionamentos com base na *timeline* do contato de amizade mais próximo (como em *friend-timeline*)?

Q5 - A combinação das estratégias (*own-timeline* e *friend-timeline*) permite detectar este posicionamento de forma mais eficiente que utilizando apenas as estratégias isoladas?

As questões Q1 e Q3 são motivadas pelo interesse em descobrir se outras fontes de conteúdo textual presentes na rede social, desconsiderando-se os posicionamentos conhecidos de um usuário, podem servir como substituto na detecção destes posicionamentos. As questões Q2 e Q4, de natureza mais computacional, são motivadas pelo interesse em comparar abordagens de classificação mais tradicionais, baseadas em BERT, com alternativas mais recentes baseadas em *prompts* para LLMs. Finalmente, partindo do pressuposto de que mesmo não superando a abordagem tradicional as abordagens que não usam o texto do posicionamento podem contribuir para a solução, a questão Q5 procura determinar se a combinação destas fontes de conhecimento externo pode servir como substituto na tarefa de detecção de posicionamentos em nível de usuário.

4 Abordagens computacionais consideradas

As estratégias *own-timeline*, *friend-timeline* e *stance-tweet* serão testadas com uso de abordagens computacionais aqui denominadas *BERT*, *LLama3* e *stacking* discutidas individualmente a seguir. O código utilizado nos experimentos, incluindo outras variações aqui omitidas, encontra-se disponível para reuso¹.

4.1 Abordagem *BERT*

A abordagem aqui denominada *BERT* foi desenvolvida utilizando-se *BERTabaporu*, um modelo BERT treinado especificamente em um córpus de postagens do Twitter/X brasileiro [8]. Em virtude do limite de 512 *tokens* de entrada deste modelo, entradas mais longas, como no caso das estratégias baseadas em *timelines*, foram reduzidas, como será detalhado a seguir.

¹<https://github.com/semcovici/stance-prediction-UstanceBR>

Por razões de custo computacional, nas abordagem baseada em dados de *timeline* (i.e., estratégias *own-timeline*, *friend-timeline*), somente uma pequena amostra das postagens é considerada para treino dos modelos. Essa amostra é selecionada a partir de uma lista não exaustiva de sugestões de termos de maior afinidade com cada tópico, obtida com uso da ferramenta ChatGPT, e truncada em 512 *tokens*. As sugestões de termos obtidas para cada tópico de interesse são relacionadas a seguir, onde os termos mais à esquerda são considerados menos importantes, e os mais à direita são considerados mais importantes.

Igreja: *vaticano, crisma, comunhão, batismo, culto, missa, hóstia, cálice, crucifixo, altar, sacerdote, papa, bispo, paróquia, templo, capela, catedral, pastor, padre, igreja*

Bolsonaro: *17, 22, ex-presidente, conservador, política, pl, partido liberal, governo bolsonaro, presidente, jair, bolsonaro*

Cloroquina: *droga, antimalárico, tratamento, medicamento, remédio, hidroxicloroquina, cloroquina*

Sinovac: *china, pandemia, covid-19, biontech, vacinação, imunização, vacina, vachina, coronavac, sinovac*

TV Globo: *jornalismo, mídia, emissora, televisão, tv, globo*

Lula: *13, política, governo, ex-presidente, luiz inácio lula da silva, partido dos trabalhadores, presidente, pt, lula*

Com base nessas listas de termos, as postagens da *timeline* são ranqueadas de acordo com a relevância para o tópico-alvo de interesse, e apenas as postagens consideradas mais relevantes são mantidas. Embora haja claramente muitos outros termos possíveis associados a cada um desses tópicos, é importante observar que, na prática, não se fez necessária a definição de uma lista mais completa dado que, por razões de custo computacional, os modelos desenvolvidos utilizam apenas uma fração das centenas de milhares de postagens que podem estar presentes em uma *timeline*. Assim, o uso de um conjunto reduzido de termos de filtragem já é o suficiente para a obtenção da quantidade de dados que serão efetivamente aproveitados, sem prejuízo ao resultado da tarefa de classificação.

4.2 Abordagem *LLama3*

A abordagem aqui denominada *LLama3* foi implementada com uso de *prompts* para o modelo Llama [32] em sua versão de 8 bilhões de parâmetros. Para este fim, foram submetidos ao LLM uma série de *prompts* que retornam respostas no formato '*x/10*', onde *x* representa o posicionamento do usuário, sendo 0 totalmente contrário e 10 totalmente favorável. Para a classificação binária final (contrário/favorável), um posicionamento é definido como contrário se *x* < 5, ou favorável nos demais casos.

Os *prompts* empregados na avaliação de *tweets* individuais (para modelos de *baseline stance-tweet*) e na avaliação de *timelines* do indivíduo (*own-timeline*) e *timelines* de seu contato de amizade mais próximo (modelos *friend-timeline*) foram adaptados do estudo em [28], que utiliza o mesmo córpus para detecção de posicionamentos

```

Below is an instruction that describes a task, paired with an
↳ input that provides further context. Write a response that
↳ appropriately completes the request

Stance is the attitude of the author of a text towards a given
↳ target.
Read the following text in portuguese and give a score between 0
↳ and 10 where 0 means that the text is totally against the
↳ target and 10 means the text is totally in favour of the
↳ target, in the format `x/10'.
Write only the score in the format 'x/10'. Do not write anything
↳ else.

### Input
Text: [comentário em Stance]
Target: [Alvo]

### Response:

```

Figura 1: Prompt utilizado nos modelos stance-tweet.

```

Below is an instruction that describes a task, paired with an
↳ input that provides further context. Write a response that
↳ appropriately completes the request

<instructions>
The following texts in <user_comments></user_comments> are
↳ comments made by a social media user.
Read the following user comments in portuguese and give a score
↳ between 0 and 10 where 0 means that the text is totally
↳ against the target and 10 means the text is totally in favour
↳ of the target, in the format `x/10'.
Write only the score in the format 'x/10'. Do not write anything
↳ else.
</instructions>

### Input
<user_comments>
[subconjunto de textos em Timeline]
</user_comments>

Target: [Alvo]

### Response:

```

Figura 2: Prompt utilizado nos modelos own-timeline.

```

Below is an instruction that describes a task, paired with an
↳ input that provides further context. Write a response that
↳ appropriately completes the request

<instructions>
The following texts are comments made by an user's friend on
↳ social media.
From these comments made by friends, try to guess the user's
↳ stance towards the target.
For example, if a large portion of friends' comments are against
↳ the target, it is very likely that the user is also against
↳ the target.
Give a score between 0 and 10 where 0 means it is unlikely that
↳ the user is against the target and 10 means it is very likely
↳ that the user is in favor of the target, in the format `x/10'.
Write only the score in the format 'x/10'. Do not write anything
↳ else.
</instructions>

### Input
<users_friends_comments>
[subconjunto de textos em User's Friends Timeline]
</users_friends_comments>

Target: [Alvo]

### Response:

```

Figura 3: Prompt utilizado nos modelos friend-timeline.

de nível textual (e não de nível de usuário como no presente caso), e são ilustrados nas Figuras 1, 2 e 3, respectivamente.

Por razões de custo computacional inerente ao uso deste tipo de LLM, as estratégias baseadas em *timelines* (*own-timeline* e *friend-timeline*) consideram apenas as N postagens mais pertinentes a cada tópico com base nos filtros baseados nas listas de termos discutidos na seção anterior, e também usados na abordagem *BERT*. Assim, a presente abordagem *Llama3* é avaliada em três versões com N representando 5, 10 ou 15 postagens, correspondendo às variações *Llama3:7b zero-shot-5*, *Llama3:7b zero-shot-10* e *Llama3:7b zero-shot-15* reportadas na Seção 5.

4.3 Abordagem stacking

Finalmente, a abordagem aqui denominada *stacking* combina as estratégias *own-timeline* e *friend-timeline* em uma arquitetura de *stacking* que utiliza como característica de aprendizado as probabilidades das previsões de cada modelo para compor um segundo nível de classificação responsável pela previsão final (contrário/favorável).

Para a classificação propriamente dita, foi utilizado o método de regressão logística e um modelo do tipo *bag-of-words* com contagens TF-IDF. A escolha dessa arquitetura, evidentemente mais simples em relação às abordagens anteriores, é motivada pelo custo computacional decorrente da classificação em dois estágios.

5 Avaliação

A seguir descrevemos o conjunto de dados e o procedimento de avaliação dos modelos discutidos na seção anterior, e seus resultados na investigação das questões de pesquisa enunciadas na Seção 3.

5.1 Conjunto de dados

No levantamento bibliográfico discutido na Seção 2, observa-se que a grande maioria dos estudos identificados trata da tarefa de detecção de posicionamento em nível textual, e não em nível de usuários de rede social como pretendido no presente trabalho. Além disso, dentre os estudos contemplando conjuntos de dados em português, observa-se que os trabalhos em [14] e [25] não fazem uso de dados de rede social. A exceção de maior interesse para o presente trabalho é o recentemente disponibilizado córpus *UstanceBR* [29], uma coleção de *tweets* anotados manualmente com informações de posicionamento. Este córpus, já utilizado em uma versão anterior para a tarefa de detecção de posicionamentos de nível textual em [27], será tomado por base para o presente estudo de detecção de posicionamentos em nível de usuário.

Foi utilizada a versão r3 do córpus *UstanceBR* [29], uma coleção de dados provenientes da plataforma Twitter/X brasileiro produzidos por usuários que manifestaram posicionamentos contrários ou favoráveis a seis tópicos-alvos de interesse: ex-presidentes brasileiros (Lula e Bolsonaro), medidas discutidas durante a pandemia Covid-19 (vacina e cloroquina), e instituições (rede Globo de televisão e a instituição igreja). Nos experimentos conduzidos, apenas as classes contrário/favorável foram consideradas, configurando assim um problema de classificação binária.

Os dados disponibilizados pelo córpus *UstanceBR* são divididos em três conjuntos: *tweets* individuais rotulados manualmente com um posicionamento contrário ou favorável ao tópico-alvo em questão, as *timelines* completas destes usuários (i.e., incluindo todas as suas postagens sobre quaisquer assuntos, e não apenas aquelas que

Tabela 3: Estatísticas descritivas das demais postagens do usuário (*own-timeline*) e de seus contatos próximos (*friend-timeline*) no córpus *UstanceBR* r3.

Alvo	own-timeline		friend-timeline	
	Tweets	Tokens	Tweets	Tokens
Bolsonaro	20.713.148	1.348.147	34.050.492	2.924.000
Church	65.389.115	4.728.279	115.720.349	9.595.698
Globo TV	55.040.219	3.683.586	76.812.192	6.567.638
Hydrox.	100.702.164	5.065.112	84.500.788	4.889.523
Lula	38.944.325	2.158.256	55.070.796	4.129.086
Sinovac	147.176.039	7.121.011	107.948.749	6.533.769
Total	427.965.010	24.104.391	474.103.366	34.639.714

tratam dos tópicos-alvo de interesse), e seus grafos de redes sociais, representando as relações de amizade, seguidores e menções a outros usuários na rede social.

Para construção dos modelos de *baseline stance-tweet*, foi utilizada a porção principal do córpus, ou seja, o conjunto de textos com posicionamentos produzidos pelos próprios usuários cujos posicionamentos se deseja estimar. A Tabela 2 apresenta estatísticas descritivas deste conjunto de dados.

Tabela 2: Estatísticas descritivas dos textos de posicionamentos (*stance-tweet*) do córpus *UstanceBR* r3.

Alvo	Tweets	Tokens/tweets	Contra	A favor	Total
Igreja	59.148	24,70	1.354	1041	2.395
Bolsonaro	15.901	21,17	649	102	751
Cloroquina	68.961	30,05	1.154	1141	2.295
Sinovac	92.079	29,77	1.416	1677	3.093
Globo TV	27.484	16,74	668	974	1.642
Lula	26.567	24,42	570	518	1.088
Total	290.140	24,47	5.811	5.453	11.264

Dentre os tópicos-alvo da Tabela 2, observa-se que apenas "Bolsonaro" possui um grande desbalanceamento entre as classes "Contra" e "A favor", o que pode afetar negativamente a assertividade dos modelos na predição da classe minoritária. É esperado assim que os modelos desenvolvidos para este tópico tenham desempenho inferior aos demais.

Para os modelos que não utilizam textos de posicionamento, foram utilizadas as *timelines* de cada indivíduo após a exclusão das postagens com posicionamentos sobre o tópico-alvo (modelos *own-timeline*), e as *timelines* dos contatos mais próximos com quem cada indivíduo conversa sobre o tópico-alvo (modelos *friend-timeline*). A Tabela 3 apresenta estatísticas descritivas destes dois conjuntos.

Dado que o custo de processamento de *timelines* completas para os modelos de usuários (*own-timeline*) e de contatos próximos (*friend-timeline*) é muito elevado para certos modelos (especialmente no caso de LLMs), somente uma porção dos dados sumarizados na Tabela 3 foi efetivamente utilizada na construção destes modelos, como será discutido a seguir.

5.2 Procedimento

Os textos das *timelines* empregadas nos modelos *own-timeline* e *friend-timeline* foram submetidos ao mesmo procedimento de pré-processamento do córpus *UstanceBR* descrito em [29], envolvendo a remoção de nomes (@username) de usuários. Além disso, no caso dos modelos *friend-timeline*, foi constatado que alguns poucos usuários não possuíam nenhum amigo, e portanto a *timeline* era vazia. Estes usuários foram removidos do conjunto de treino, porém mantidos no conjunto teste, o que acarreta uma penalidade aos classificadores deste tipo.

Para os modelos baseados em BERT, os textos foram truncados ao atingir o limite de 512 *tokens* da arquitetura. No caso dos modelos *stance-text*, que consideram apenas postagens (ou *tweets*) individuais, nenhum truncamento significativo foi realizado, dado que estas postagens são naturalmente breves. Já para os modelos *own-timeline* e *friend-timeline*, que levam em consideração várias postagens concatenadas, foram consideradas somente as primeiras postagens selecionadas (conforme discutido na Seção 4), até preencher o limite de 512 *tokens*. Isso resultou no descarte de uma quantidade significativa de postagens subsequentes.

Para avaliação dos modelos desenvolvidos, foi utilizada a divisão de treino e teste definida na versão r3 do córpus [29]. A métrica de avaliação utilizada foi a medida F1 macro. Por meio de *grid search*, diversas alternativas foram experimentadas (e.g., diferentes modelos BERT, classificadores e variações da seleção de N postagens nos modelos LLama, dentre outras). Por questões de simplicidade, entretanto, apenas as alternativas de melhor resultado são discutidas a seguir.

5.3 Resultados

A Tabela 4 apresenta os resultados de classificação binária obtidos utilizando-se as estratégias de *baseline stance-tweet*, e as estratégias propostas *own-timeline*, *friend-timeline* e *stacking*, juntamente com os melhores classificadores construídos para cada uma. Para fins ilustrativos, a abordagem de *baseline* é aqui apresentada nas versões BERT e LLama, embora a primeira tenha se demonstrado superior à segunda para todas as classes consideradas. Nas abordagens baseadas em dados de *timeline* usando *LLama3*, o número ao final do nome do modelo indica a quantidade (5, 10 ou 15) de postagens utilizada como entrada.

Como esperado, os resultados da Tabela 4 mostram que as estratégias baseadas em *timelines* (do próprio indivíduo em *own-timeline*, ou de seus contatos próximos, em *friend-timeline*) apresentam resultados inferiores ao uso do texto de posicionamento propriamente dito (em *stance-tweet*). No entanto, estas estratégias ainda apresentam um grau de acerto considerável e, mais importante, quando combinadas (em *stack*) são efetivamente superiores ao uso do texto de posicionamento. Detalhes desta avaliação são discutidos a seguir, retomando-se individualmente as questões de pesquisa Q1..Q5 enunciadas na Seção 3.

Uso da timeline como substituto ao posicionamento. No que diz respeito à questão Q1 – referente ao uso das demais postagens do usuário em substituição ao texto do posicionamento – observa-se que, embora o uso do posicionamento (*stance-tweet*) ainda seja, em média, a melhor opção, a sua substituição pelo restante da *timeline*

Tabela 4: Resultados de *F1-Score macro*. O melhor resultado de cada classe, desconsiderando-se a estratégia de *baseline stance-tweet*, é destacado.

Estratégia	Abordagem	Igreja	Bolsonaro	Cloroquina	Sinovac	Globo TV	Lula	Média
stance-tweet (baseline)	bertabaporu-base	0,87	0,74	0,85	0,83	0,86	0,79	0,82
	Llama3:7b zero-shot	0,73	0,46	0,64	0,58	0,77	0,70	0,65
own-timeline	bertabaporu-base	0,70	0,79	0,88	0,86	0,51	0,71	0,74
	Llama3:7b zero-shot-5	0,61	0,59	0,64	0,53	0,56	0,51	0,57
	Llama3:7b zero-shot-10	0,56	0,61	0,61	0,53	0,51	0,52	0,56
	Llama3:7b zero-shot-15	0,60	0,59	0,61	0,52	0,52	0,53	0,56
friend-timeline	bertabaporu-base	0,58	0,46	0,57	0,64	0,59	0,60	0,57
	Llama3:7b zero-shot-5	0,50	0,30	0,47	0,47	0,49	0,51	0,46
	Llama3:7b zero-shot-10	0,56	0,33	0,44	0,46	0,49	0,48	0,46
	Llama3:7b zero-shot-15	0,50	0,29	0,42	0,46	0,48	0,54	0,45
stacking (own+friend)	regressão logística	0,70	0,73	0,90	0,87	0,59	0,75	0,75

do usuário (*own-timeline*) obtém resultados relativamente próximos e, para algumas classes, inclusive superiores a esse. Postagens do restante da *timeline* do indivíduo são, portanto, indicativos do posicionamento do usuário, e esta informação pode assim ser utilizada em auxílio à tarefa de detecção de posicionamentos, mesmo na ausência de informação mais detalhada a respeito.

Abordagens computacionais baseadas nas postagens do usuário. No que diz respeito à questão *Q2* – referente a qual abordagem computacional é mais efetiva dentre as alternativas do tipo *own-timeline* consideradas – observa-se uma ampla vantagem em todas as classes dos modelos baseados em BERT sobre as que utilizam Llama. Ainda assim, é importante destacar que são abordagens muito distintas e que, ao contrário da abordagem de classificação BERT tradicional, o uso de *prompts* para LLMs nas abordagens Llama não requer um córpus de dados rotulados para cada classe de interesse. Isso, em termos práticos, pode representar um fator importante na escolha do método computacional a ser utilizado.

Uso da timeline de contatos próximos. No que diz respeito à questão *Q3* – referente ao uso das postagens do contato mais próximo do usuário em substituição ao texto do posicionamento – observa-se um desempenho médio inferior ao uso da *timeline* do próprio usuário como na questão *Q1* anterior, embora ainda significativo se comparado, por exemplo, ao que seria obtido por uma medida simples de classe majoritária (não incluída na presente análise). Ou seja, o texto produzido pelo próprio usuário cujo posicionamento se deseja estimar é, conforme seria esperado, um indicativo de posicionamento mais forte do que os textos produzidos por seus contatos na rede social, mas os textos de terceiros também possuem um certo poder preditivo.

Abordagens computacionais baseadas nas postagens de contatos próximos. No que diz respeito à questão de pesquisa *Q4* – referente a qual abordagem computacional é mais efetiva dentre as alternativas do tipo *friend-timeline* consideradas – há o mesmo

efeito observado na questão *Q2* anterior. Novamente, modelos baseados em BERT superam as abordagens que utilizam Llama, embora neste caso a vantagem seja menor do que no caso das estratégias *own-timeline*.

Abordagens híbridas. No que diz respeito à questão *Q5* – sobre o uso combinado de *own-timeline* e *friend-timeline* em um modelo do tipo *stack ensemble* – os resultados apontam não só uma clara vantagem da estratégia *stacking* sobre seus componentes individuais, mas sobre o próprio *baseline stance-tweet*. Em outras palavras, o uso combinado destas duas fontes de conhecimento (ou seja, as postagens restantes da *timeline* do usuário e as postagens de seu contato mais próximo na rede social) pode servir como substituto efetivo ao próprio uso do texto que expressa o posicionamento do usuário, e pode inclusive apresentar um ganho positivo em relação à abordagem tradicional. Este ganho, conforme observado, decorre predominantemente do uso da *timeline* do usuário (*own-timeline*), e é impulsionado em menor grau pelo uso da *timeline* do contato próximo (*friend-timeline*).

Finalmente, embora omitido da presente discussão por razões de brevidade, constatou-se também que o desempenho da classificação para o tópico-alvo Bolsonaro foi de fato prejudicado pelo desbalanceamento de classes conforme discutido na Seção 5.1. Este desbalanceamento, entretanto, não é aparente no resultado reportado pelo fato de a métrica empregada combinar ambas as classes em um escore único.

6 Considerações finais

Este trabalho apresentou um estudo da tarefa computacional de detecção de posicionamentos em nível de usuários de redes sociais em português. Diferentemente das abordagens tradicionais de detecção de posicionamento em nível textual, nestes cenários o texto que expressa o posicionamento não está disponível, o que torna a tarefa possivelmente mais complexa.

Como substituto ao texto do posicionamento, o estudo investigou o uso de outras postagens dos usuários, postagens de seus contatos próximos, e a combinação destes. Os resultados obtidos com o uso

de classificadores do tipo BERT e *prompts* para o modelo LLama 3 sugerem que ambas as fontes de conhecimento – ou seja, postagens da *timeline* do indivíduo e de seus contatos – podem contribuir para a tarefa, e que a combinação de ambas em uma arquitetura de *stack ensemble* revelou-se superior até mesmo ao uso do texto do posicionamento. Assim, constatamos que é possível detectar o posicionamento de usuários de redes sociais mesmo que estes usuários não manifestem nenhum posicionamento explícito sobre o tópico de interesse.

O estudo realizado é de caráter preliminar, e deixa diversas oportunidades de melhorias. Dentre elas, destacamos o uso de outros LLMs recentes em substituição à abordagem adotada neste trabalho, e a investigação do uso de dados não-textuais como relacionamentos da rede social, possivelmente combinando os modelos atuais com a abordagem em [5]. Além disso, o córpus utilizado no presente trabalho será expandido de modo a ajustar o desbalanceamento de classes.

Cabe observar ainda que, no caso geral, modelos computacionais de detecção de posicionamentos precisam ser capazes de lidar com tópicos não vistos em tempo de treinamento [2]. Cenários de avaliação desse tipo, conhecidos por detecção *zero-shot*, representam o estado-da-arte para modelos deste tipo, e embora tenham avançado consideravelmente no caso da detecção de posicionamentos de nível textual [6, 28, 34], ainda não contemplam o uso de dados não textuais para detecção de posicionamentos em nível de usuário. Estas linhas de investigação são assim deixadas como sugestões de trabalhos futuros.

Acknowledgments

Este trabalho recebeu apoio parcial FAPESP # 2021/08213-0.

Referências

- [1] Abeer Aldayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Information Processing & Management* 58, 4 (2021), 102597.
- [2] Emily Allaway and Kathleen McKeown. 2020. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In *EMNLP-2020 proceedings*. Assoc. for Computational Linguistics, Online, 8913–8931. <https://doi.org/10.18653/v1/2020.emnlp-main.717>
- [3] Emily Allaway and Kathleen R. McKeown. 2020. Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations. In *EMNLP-2020 proceedings*. Assoc. for Computational Linguistics, Online, 8913–8931. <https://doi.org/10.18653/v1/2020.emnlp-main.717>
- [4] Adrian Benton and Mark Dredze. 2018. Using Author Embeddings to Improve Tweet Stance Classification. In *The 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, Wei Xu, Alan Ritter, Tim Baldwin, and Afshin Rahimi (Eds.). Association for Computational Linguistics, Brussels, Belgium, 184–194. <https://doi.org/10.18653/v1/W18-6124>
- [5] Lais Carrara Leme Cavalheiro, Matheus Camasmie Pavan, and Ivandré Paraboni. 2023. Stance prediction from multimodal social media data. In *Recent Advances in Natural Language Processing (RANLP-2023)*. INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 242–248. https://doi.org/10.26615/978-954-452-092-2_027
- [6] Wang Chunling, Zhang Yijia, Yu Xingyu, Liu Guantong, Chen Fei, and Lin Hongfei. 2023. Adversarial Network with External Knowledge for Zero-Shot Stance Detection. In *Proceedings of the 22nd Chinese National Conference on Computational Linguistics*. Chinese Information Processing Society of China, Harbin, China, 824–835. https://doi.org/10.1007/978-99-6207-5_26
- [7] Alessandra Cignarella, Mirko Lai, Cristina Bosco, Viviana Patti, and Paolo Rosso. 2020. SardiStance @ EVALITA2020: Overview of the Task on Stance Detection in Italian Tweets. In *CEUR Proceedings* vol. 2765. CEUR-WS.org, online, 177–186.
- [8] Pablo Botton da Costa, Matheus Camasmie Pavan, Wesley Ramos dos Santos, Samuel Caetano da Silva, and Ivandré Paraboni. 2023. BERTabaporu: assessing a genre-specific language model for Portuguese NLP. In *Recent Advances in Natural Language Processing (RANLP-2023)*. INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 217–223. https://doi.org/10.26615/978-954-452-092-2_024
- [9] Kareem Darwish, Walid Magdy, and Tahar Zanouda. 2017. Improved Stance Prediction in a User Similarity Feature Space. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. Assoc. for Computing Machinery, New York, USA, 145–148. <https://doi.org/10.1145/3110025.3110112>
- [10] Kareem Darwish, Peter Stefanov, Michael Aupetit, and Preslav Nakov. 2020. Unsupervised User Stance Detection on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 14, 1 (2020), 141–152. <https://doi.org/10.1609/icwsm.v14i1.7286>
- [11] André de Sousa and Karin Becker. 2023. SSSD: Leveraging Pre-trained Models and Semantic Search for Semi-supervised Stance Detection. In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*. INCOMA Ltd., Shoumen, Bulgaria, Varna, Bulgaria, 264–273. https://doi.org/10.26615/978-954-452-092-2_029
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-2019 proceedings*. Assoc. for Computational Linguistics, Minneapolis, USA, 4171–4186.
- [13] Maria S. Espinosa, Rodrigo Agerri, Alvaro Rodrigo, and Roberto Centeno. 2020. SardiStance: Combining Textual, Social and Emotional Features. In *7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2020)*. CEUR-WS.org, online.
- [14] Arthur Marçal Flores, Matheus Camasmie Pavan, and Ivandré Paraboni. 2022. User profiling and satisfaction inference in public information access services. *Journal of Intelligent Information Systems* 58, 1 (2022), 67–89. <https://doi.org/10.1007/s10844-021-00661-w>
- [15] Henri-Jacques Geiss, Flora Sakketou, and Lucie Flek. 2022. OK Boomer: Probing the socio-demographic Divide in Echo Chambers. In *10th International Workshop on Natural Language Processing for Social Media*. Assoc. for Computational Linguistics, Seattle, Washington USA, 83–105. <https://doi.org/10.18653/v1/2022.socialnlp-1.8>
- [16] Dilek Kucuk and Fazli Can. 2020. Stance Detection: A Survey. *ACM Computing Surveys* 53, 1 (2020), 1–37.
- [17] Mirko Lai, Alessandra Cignarella, and Delia Hernandez-Farias. 2017. itacos at ibereval2017: Detecting stance in Catalan and Spanish tweets. In *IberEval-2017 proceedings*. CEUR-WS.org, Murcia, Spain.
- [18] M. Lai, A. T. Cignarella, D. I. Hernandez Farias, C. Bosco, V. Patti, and P. Rosso. 2020. Multilingual stance detection in social media political debates. *Computer Speech and Language* 63 (2020), 101075.
- [19] Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. 2020. #Brexit: Leave or remain? the role of user's community and diachronic evolution on stance detection. *Journal of Intelligent & Fuzzy Systems* 39, 2 (2020), 2341–2352.
- [20] Rasmus Lehmann and Leon Derczynski. 2019. Political Stance in Danish. In *22nd Nordic Conference on Computational Linguistics*. Linköping University Electronic Press, Turku, Finland, 197–207.
- [21] Veronica Lynn, Salvatore Giorgi, Niranjan Balasubramanian, and H. Andrew Schwartz. 2019. Tweet Classification without the Tweet: An Empirical Examination of User versus Document Attributes. In *Third Workshop on Natural Language Processing and Computational Social Science*, Svitlana Volkova, David Jurgens, Dirk Hovy, David Bamman, and Oren Tsur (Eds.). Association for Computational Linguistics, Minneapolis, Minnesota, 18–28. <https://doi.org/10.18653/v1/W19-2103>
- [22] Walid Magdy, Kareem Darwish, Norah Abokhodair, Afshin Rahimi, and Timothy Baldwin. 2016. #ISISisNotIslam or #DeportAllMuslims? Predicting Unspoken Views. In *8th ACM Conference on Web Science*. Assoc. for Computing Machinery, New York, NY, USA, 95–106. <https://doi.org/10.1145/2908131.2908150>
- [23] Miller McPherson, Lynn Smith-Lovin, and James M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, 1 (2001), 415–444.
- [24] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 Task 6: Detecting Stance in Tweets. In *10th International Workshop on Semantic Evaluation (SemEval-2016)*. Assoc. for Computational Linguistics, San Diego, California, 31–41. <https://doi.org/10.18653/v1/S16-1003>
- [25] Matheus Camasmie Pavan, Vitor Garcia dos Santos, Alex Gwo Jen Lan, Joao Trevisan Martins, Wesley Ramos dos Santos, Caio Deutsch, Pablo Botton da Costa, Fernando Chin Hsieh, and Ivandré Paraboni. 2023. Morality Classification in Natural Language Text. *IEEE transactions on Affective Computing* 14, 1 (2023), 857–863. <https://doi.org/10.1109/TAFFC.2020.3034050>
- [26] Matheus Camasmie Pavan, Wesley Ramos dos Santos, and Ivandré Paraboni. 2020. Twitter Moral Stance Classification using Long Short-Term Memory Networks. In *BRACIS-2020 proceedings LNAI 12319*. Springer, online, 636–647. https://doi.org/10.1007/978-3-030-61377-8_45
- [27] Matheus Camasmie Pavan and Ivandré Paraboni. 2022. Cross-target Stance Classification as Domain Adaptation. In *Advances in Computational Intelligence - MICAI 2022 - Lecture Notes in Artificial Intelligence vol 13612*, Obdulia Richardson Lagunas, Juan Martínez-Miranda, and Bella Martínez Seis (Eds.). Springer Nature Switzerland, Cham, 15–25. https://doi.org/10.1007/978-3-031-19493-1_2
- [28] Matheus Camasmie Pavan and Ivandré Paraboni. 2024. A benchmark for Portuguese zero-shot stance detection. *Journal of the Brazilian Computer Society* 30, 1 (2024), 469–479. <https://doi.org/10.5753/jbcs.2024.3932>

- [29] Camila Pereira, Matheus Pavan, Sungwon Yoon, Ricelli Ramos, Pablo Costa, Laís Cavalheiro, and Ivandré Paraboni. 2024. UstanceBR: a social media language resource for stance prediction. <https://doi.org/10.48550/arXiv.2312.06374>
- [30] Younes Sami and Kareem Darwish. 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*, Paola Merlo, Jörg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, Online, 2637–2646. <https://doi.org/10.18653/v1/2021.eacl-main.227>
- [31] Marionna Taulé, Maria Antònia Martí, Francisco Manuel Rangel Pardo, Paolo Rosso, Cristina Bosco, and Viviana Patti. 2017. Overview of the Task on Stance and Gender Detection in Tweets on Catalan Independence at IberEval 2017. In *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2017)*. CEUR-WS.org, Murcia, Spain, 157–177.
- [32] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. <http://arxiv.org/abs/2302.13971> cite arxiv:2302.13971.
- [33] Haoyang Wen and Alexander Hauptmann. 2023. Zero-Shot and Few-Shot Stance Detection on Varied Topics via Conditional Generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Toronto, Canada, 1491–1499. <https://doi.org/10.18653/v1/2023.acl-short.127>
- [34] Bowen Zhang, Daijun Ding, and Liwen Jing. 2023. How would Stance Detection Techniques Evolve after the Launch of ChatGPT? [arXiv:2212.14548](https://doi.org/10.48550/arXiv.2212.14548) [cs.CL]
- [35] Xuechen Zhao, Jiaying Zou, Zhong Zhang, Feng Xie, Bin Zhou, and Lei Tian. 2022. Zero-shot stance detection based on cross-domain feature enhancement by contrastive learning. <https://doi.org/10.48550/arXiv.2210.03380>

Received 1 November 2024; revised 10 February 2025; accepted 18 February 2025