

ORIGINAL PAPER

Detecção de posicionamentos no nível do usuário em redes sociais online: uma revisão sistemática

User-Level stance detection in online social networks: a systematic review

Lucas Mendes Sales¹, Laís Carraro Leme Cavalheiro¹, Luciano Antonio Digiampietri ,¹

¹EACH-USP - Escola de Artes, Ciências e Humanidades da Universidade de São Paulo

*lucasm1sales@usp.br; laiscarraro@usp.br; digiampietri@usp.br

Recebido: 11/03/2025. Revisado: 30/03/2026. Aceito: 21/04/2026.

Resumo

As redes sociais online desempenham um papel central na comunicação contemporânea, funcionando como plataformas para a disseminação de informações, debates e expressão de opinião individual. Nesse contexto, a detecção de posicionamento surge como uma ferramenta promissora para a análise da opinião pública, permitindo a identificação de padrões e tendências em discussões sobre temas controversos. Essa detecção pode ser conduzida em dois níveis: em nível de postagem (*statement level*) ou em nível de usuário (*user level*). No entanto, a literatura tem se concentrado predominantemente na detecção de posicionamento em nível de postagem, resultando em lacunas quanto ao potencial da abordagem em nível de usuário. Diante desse cenário, este estudo apresenta uma revisão sistemática da literatura com o objetivo de mapear e analisar os métodos e estratégias empregados na detecção automática do posicionamento de usuários em redes sociais online. Os resultados fornecem uma visão estruturada das principais práticas adotadas na área, assim como desafios e oportunidades para pesquisas futuras.

Palavras-Chave: Detecção de Posicionamentos; Detecção de Posicionamentos de Usuários; PLN; Processamento de Língua Natural; Redes Sociais Online

Abstract

Online social networks play a central role in modern communication, serving as platforms for information dissemination, debates, and individual opinion expressions. In this context, stance detection emerges as a promising tool for public opinion analysis, enabling the identification of patterns and trends in controversial discussions. This task can be performed at two levels: statement level or user level. However, the literature has predominantly focused on stance detection at the statement level, leaving gaps regarding the potential of user-level approaches. In light of this, this study presents a systematic literature review aimed at mapping and analyzing the methods and strategies employed in the automatic detection of user stance in online social networks. The findings provide a structured overview of the key practices adopted in the field, as well as the challenges and opportunities for future research.

Keywords: Natural Language Processing; NLP; Online Social Networks; Stance Detection; User Stance Detection; User-level Stance Detection

1 Introdução

As redes sociais online são um componente central da comunicação e interação da sociedade atual, pois por meio delas as pessoas expressam suas ideias e opiniões diariamente, bem como se atualizam e se entretêm. Desta forma, demonstram ter um grande potencial de disseminação de informação (ALDayel and Magdy, 2021; Newman, 2011).

O posicionamento pode ser definido como a expressão do ponto de vista de um indivíduo em relação a um determinado assunto (Biber and Finegan, 1988). Detecção de posicionamentos é a tarefa de determinar a partir de um texto se seu autor é a favor, contra ou neutro em relação a um alvo (Zhang et al., 2023).

A detecção de posicionamentos pode ser categorizada em dois níveis: detecção de posicionamentos em nível de postagem (adaptação livre de *statement level*) e em nível de usuário (*user level*) (ALDayel and Magdy, 2021; Küçük, Dilek and Can, Fazli, 2020). Na literatura, a maioria dos trabalhos estuda o posicionamento em nível de postagem, como em Abeysinghe, Vulupala, Bourgeois and Sunderaman (2022), ALDayel and Magdy (2021) e Allaway and McKeown (2020).

Jia et al. (2022) definem o posicionamento de usuários como o ponto de vista do indivíduo em relação a um tópico em determinado recorte temporal. Referem-se, assim, à detecção de posicionamentos tradicional como a classificação do posicionamento do indivíduo em um dado momento no tempo (em uma postagem isolada) e à detecção de posicionamento de usuários como a classificação do posicionamento geral do indivíduo.

A detecção de posicionamento aplicada em redes sociais online se apresenta como uma das principais ferramentas para estudos analíticos da opinião pública, principalmente em assuntos de natureza controversa, como política e questões sociais, visto que a partir delas é possível analisar o ponto de vista de milhares de usuários em relação a inúmeros tópicos (ALDayel and Magdy, 2021; Han et al., 2020). Observa-se, por exemplo, que jornalistas recorrem às redes sociais online para analisar a opinião pública, especialmente na sequência de eventos midiáticos, como debates (McGregor, 2019). Esta opinião, combinada com informações espaço-temporais, pode servir como apoio para a tomada de decisão de governos em situações emergenciais (Han et al., 2020) e para medir a adoção de determinadas medidas pela população (Qazi et al., 2020).

A detecção de posicionamentos em nível de usuário pode apresentar vantagens em relação ao nível de postagem, principalmente por considerar o contexto do histórico de postagens (HP) e integrar as informações dos dados do perfil dos usuários. O nível de usuário tende a trazer, por exemplo, uma visão mais consistente e contextualizada das opiniões dos usuários ao longo do tempo e como elas podem ter evoluído, reduzindo possíveis ambiguidades em postagens individuais. Também possibilita a identificação de tendências e mudanças nas opiniões dos usuários, o que pode ser crucial para análises de longo prazo e estratégias de comunicação (Bechini et al., 2021). Além disso, demonstra maior eficácia na identificação de padrões de comportamento de *bots* e contas inautênticas, além de permitir a classificação de usuários com nenhuma ou poucas postagens (também conhecidos como “*silent*

users” ou “*lurkers*”) (Aldayel and Magdy, 2019; ALDayel and Magdy, 2021).

Embora a detecção de posicionamento em nível de usuário apresente um grande potencial para a área, esse campo ainda é pouco explorado na literatura. Essa lacuna também está presente nas revisões sistemáticas sobre o tema. Por exemplo, em Küçük and Can (2020), o foco está na definição do problema e na apresentação de áreas relacionadas à detecção de posicionamento, além da descrição das principais abordagens baseadas em aprendizado de máquina. De maneira semelhante, o estudo de ALDayel and Magdy (2021) aborda os diferentes níveis de detecção de posicionamento, porém mantém um foco alinhado ao do trabalho anteriormente citado, sem se aprofundar especificamente na detecção em nível de usuário.

A ausência de estudos de revisão sistemática voltados especificamente para a detecção de posicionamento em nível de usuário constitui a principal motivação deste trabalho. Mais especificamente, este estudo tem como objetivo identificar e analisar os métodos e técnicas empregados na detecção automática do posicionamento de usuários em redes sociais online, bem como os procedimentos utilizados para a obtenção e anotação dos dados empregados no treinamento e validação desses modelos. Dessa forma, buscou-se oferecer uma visão estruturada das práticas adotadas na área e das oportunidades existentes para avanços futuros.

2 Metodologia

Nesta seção, será apresentada a metodologia utilizada para a realização da revisão sistemática discutida neste trabalho. A seguir estão descritos o objetivo, questões de pesquisa, critérios de busca, critérios de seleção e condução da revisão.

2.1 Objetivo

Identificar e analisar os métodos e técnicas existentes para a tarefa de detecção automática de posicionamento de usuários em redes sociais.

2.2 Questões de Pesquisa

No contexto de detecção de posicionamentos de usuários em redes sociais, a presente revisão se propõe a responder às seguintes questões:

- i. Quais são os principais algoritmos utilizados para a criação de modelos de detecção de posicionamentos de usuários em redes sociais?
- ii. Quais foram as características utilizadas no treinamento dos modelos com os melhores resultados encontrados e que mais contribuíram para alcançá-los?
- iii. Como foram definidos os *baselines* utilizados?
- iv. Como foi realizada a avaliação dos resultados dos modelos?
- v. Como os dados utilizados foram obtidos e anotados?

2.3 Critérios de busca

A revisão bibliográfica foi conduzida com o objetivo de identificar trabalhos em inglês e português que abordem o uso de modelos de aprendizado de máquina para a tarefa de detecção de posicionamento de usuários em redes sociais. O foco da busca esteve em publicações revisadas por pares, incluindo artigos de periódicos e anais de conferências indexados em bases científicas da área.

Os termos “User” e “Stance” foram definidos como palavras-chave para a busca. As buscas foram realizadas verificando a presença das palavras-chave nos títulos e resumos das publicações. As seguintes fontes foram consultadas:

- Biblioteca Digital da ACM (<https://dl.acm.org/>)
- Biblioteca Digital do IEEE (<https://ieeexplore.ieee.org/Xplore/guesthome.jsp>)
- Portal de Periódicos da Capes (<http://novo.periodicos.capes.gov.br/>)
- ACL Anthology (<https://aclanthology.org/>)
- Springer (<https://link.springer.com/>)
- Scopus (<https://www.sciencedirect.com/search>)

2.4 Critérios de Seleção

A seleção dos trabalhos considerados nesta revisão seguiu critérios rigorosos de inclusão, garantindo a relevância e a qualidade das publicações analisadas. Foram considerados apenas estudos com foco em trazer contribuições diretas para a detecção de posicionamento de usuários em redes sociais, seja por meio da proposição de classificadores específicos ou da construção de um corpus aplicável à tarefa. Além disso, a disponibilidade integral dos textos em bases científicas foi um requisito essencial para assegurar a transparência e a reprodutibilidade das análises realizadas. Os critérios adotados estão detalhados a seguir, e, para ser considerado, o trabalho tem de estar de acordo com todos os critérios de inclusão e não satisfazer a nenhum critério de exclusão.

2.4.1 Critérios de inclusão

- i1. Trabalho publicado e disponível integralmente em bases de dados científicas.
- i2. O trabalho trata da detecção de posicionamento de usuários dentro do contexto de redes sociais.
- i3. O trabalho é um estudo primário que inclui propostas de classificadores específicos para o posicionamento de usuários ou criação de corpus utilizável para a tarefa de detecção de posicionamento de usuários.

2.4.2 Critérios de exclusão

- e1. Serão excluídos estudos que explorem somente a classificação de posicionamento em postagem (*statement level*).
- e2. Serão excluídos trabalhos que não especifiquem a origem dos dados utilizados.
- e3. Serão excluídos trabalhos que não tenham relação com Inteligência Artificial e Aprendizado de Máquina ou Análise de Redes Sociais.

2.5 Condução

A realização das buscas ocorreu em duas iterações. A [Tabela 1](#) contém os resultados da primeira iteração, com trinta e três artigos encontrados no total. Cinco destes foram considerados duplicados, pois foram encontrados em mais de uma fonte. Dessa forma, restaram vinte e oito artigos para as análises posteriores.

Dado o volume relativamente baixo de trabalhos encontrados na primeira iteração de buscas, optou-se pela realização da segunda iteração, expandindo os critérios de busca nas bases IEEE e ACM, principalmente por conta das funcionalidades de suas ferramentas de pesquisa, e pela inclusão de uma nova base, a base Scopus. A [Tabela 2](#) contém os resultados desta segunda iteração de buscas, incluindo apenas os números relacionados a artigos não inclusos na primeira iteração. Por fim, com os resultados das duas iterações, foi obtido um total de quarenta e quatro artigos encontrados, sendo vinte e oito da primeira iteração e dezesseis da segunda.

2.5.1 Análise de inclusão

A [Tabela 3](#) contém a relação dos critérios de inclusão descritos na [Seção 2.4.1](#) com os quarenta e quatro artigos únicos encontrados nas buscas. Quatro artigos não foram compatíveis com o critério “i1” pela impossibilidade de acessar seu conteúdo na íntegra. Pelos critérios “i2” e “i3”, identificou-se que cerca de metade dos artigos encontrados se encaixavam no tema de detecção de posicionamento de usuários, com o destaque de que somente um dos trabalhos descreve a construção de um conjunto de dados. Alguns dos trabalhos que não se encaixaram no critério “i3” discutiam o posicionamento de usuários, mas o foco dos modelos desenvolvidos era o posicionamento em nível de postagem.

2.5.2 Análise de exclusão

A [Tabela 4](#) contém a relação dos critérios de exclusão descritos na [Seção 2.4.2](#) com os quarenta e quatro artigos únicos encontrados nas buscas. O critério de exclusão com maior número de artigos relacionados foi o critério “e1”, ou seja, a maioria das exclusões se deu pelo fato de o trabalho tratar somente da detecção de posicionamentos em nível de postagem. Em contrapartida, poucos trabalhos se encaixaram nos critérios “e2” e “e3”, por não especificarem a origem dos dados ou estarem fora das áreas de Inteligência Artificial e Análise de Redes Sociais.

Com a aplicação dos critérios de exclusão, trinta e um estudos foram removidos da revisão, restando então treze artigos selecionados para a leitura integral.

3 Resultados

Nesta seção, apresentamos os resultados da revisão sistemática realizada conforme o protocolo estabelecido. Inicialmente, oferecemos uma visão geral dos 13 artigos selecionados e completamente analisados, destacando as principais ideias e contribuições dos autores para facilitar a compreensão e interpretação dos temas investigados. Posteriormente, organizamos os resultados conforme as questões de pesquisa definidas na [Seção 2.2](#). Por fim, são

Tabela 1: Descrição da primeira iteração de buscas.

Fonte	Data	Strings de busca	Filtros	Resultados
IEEE	23/10/2023	“User” and “Stance”	Título do documento	10
ACM	03/11/2023	“User” and “Stance”	Título do documento	5
ACL Anthology	03/11/2023	“User Stance”	Título do documento (Manual)	1
CAPES	03/11/2023	“User” and “Stance”	Título do documento + Periódicos revisados por pares	15
Springer Link	03/11/2023	“User Stance”	Título do documento (Manual) + Disciplina = Computer Science	2

Tabela 2: Descrição dos resultados inéditos da segunda iteração de buscas.

Fonte	Data	Strings de busca	Filtros	Resultados
IEEE	10/02/2023	“User Stance”	Título do documento + Resumo + Palavras-chave	6
ACM	11/02/2023	“User Stance”	Título do documento + Resumo + Palavras-chave	4
SCOPUS	12/02/2023	“User Stance”	Título do documento + Resumo + Palavras-chave	6

Tabela 3: Quantidade e porcentagem (não exclusivas) de atendimento aos critérios de inclusão.

Critério	Quantidade	Porcentagem
i1	40	90,9%
i2	21	47,7%
i3	16	36,3%

Tabela 4: Quantidade e porcentagem (não exclusivas) de atendimento aos critérios de exclusão

Critério	Quantidade	Porcentagem
e1	18	40,9%
e2	1	2,2%
e3	7	15,9%

apresentadas as considerações finais sobre os achados e suas implicações.

3.1 Visão Geral

Esta seção apresenta de forma resumida as abordagens, contribuições e resultados dos trabalhos analisados nesta revisão sistemática, organizados de forma sequencial por ano de publicação. A [Tabela 5](#) lista alguns dos elementos observados e discutidos ao longo desta seção.

O trabalho desenvolvido por [Dong et al. \(2017\)](#) utilizou dados em inglês do sistema de comentários do portal CNN News e do fórum 4Forums¹ com o objetivo de construir um conjunto fracamente guiado (*weakly-guided*), com baixa dependência de domínios específicos, e que atingisse resultados próximos aos apresentados na literatura sem a necessidade de processos manuais de anotação dos dados. A abordagem inicial escolhida foi a construção de um modelo generativo para descrever como usuários com um determinado posicionamento escolhem seus argumentos, assim utilizando somente o texto das postagens para a classificação do posicionamento de usuários. O desempe-

nho do modelo se mostrou insatisfatório para os autores ao tender a máximos locais, levando à criação de um novo modelo chamado STML (*Stance-based Text Generative Model with Link Regularization*). Este modelo combina informações textuais e um grafo de interações entre indivíduos para a classificação do posicionamento de usuários.

O grafo de interações foi criado a partir de conversas entre usuários identificadas em seções de comentários. Estes vínculos criados pelas conversas foram denominados “links”. Os links identificados foram anotados de forma automática a partir de um conjunto de regras definidas para indicar se os usuários envolvidos na conversa concordavam ou discordavam em relação ao tema alvo. Estas regras levam em consideração características como o número de turnos da discussão, citado como um indicativo útil de discordância entre usuários, e elementos do texto, como a presença de palavras normalmente relacionadas a expressões de desacordo.

Os resultados dos testes realizados com o STML sugerem que, com uma quantidade modesta de informações de interação dos usuários, o modelo é capaz de aprender distribuições de palavras utilizadas por usuários de diferentes posicionamentos e, desta forma, gerar previsões com desempenho comparável ao do estado da arte no momento para classificadores supervisionados ([Sridhar et al., 2015](#)). A avaliação final mostrou que o STML não superou o estado da arte. Os autores ressaltam que modelos supervisionados terão melhor resultado dependendo da qualidade dos dados e da anotação, e que, para um modelo não supervisionado, os resultados foram satisfatórios.

No estudo realizado por [Darwish et al. \(2017\)](#), explorou-se a possibilidade de utilizar a similaridade entre usuários, em vez do uso direto de texto ou interações, como características para a tarefa de detecção de posicionamento de usuários a partir de dados da rede social Twitter, com postagens nos idiomas inglês e árabe. A ideia central fundamenta-se em dois conceitos: a influência social, que se refere ao processo pelo qual os pensamentos, comportamentos e emoções de uma pessoa são influenciados ou moldados por outras pessoas ou grupos; e na homofilia, que descreve a propensão de indivíduos a interagir com

¹4Forums: <http://4forums.com>, acessado em 03/03/2025

Tabela 5: Síntese da revisão sistemática

Estudo	Rede social	Conjunto de dados	Anotação	Algoritmo(s)	Características	Avaliação	Resultado (max)
Dong et al. (2017)	CNN news commenting system; 4Forums.com	Original; Internet Argument Corpus (IAC); Idioma: inglês	Automática, Heurística	STML (original, generativo)	Texto (HP, conversas entre usuários)	Acurácia	75,6%
Darwish et al. (2017)	Twitter	Islands Dataset; Islam Dataset (HP coletado); Idiomas: árabe, inglês	Manual	SVM-light; reforço em grafo	Texto (HP); Perfis autores de retweets; Hashtags; Menções; URLs; Similaridade	Macro-F1	84,0%
Chen and Ku (2018)	Facebook	Original; Idioma: chinês	Manual	Joint-MRF (original, MRF)	Texto (HP); Likes	Macro-F1	67,2%
Aldayel and Magdy (2019)	Twitter	SemEval (HP, interações e amizades coletados); Idioma: inglês	N/A	SVM	Texto (HP); Favoritos; URLs; Amizades	Macro-F1	72,49%
Zhu et al. (2020)	Twitter	Original; Idioma: inglês	Automática; Manual	LSTM; GRU	Texto (HP, postagens de amigos); Amizades	Acurácia	85%
Samih and Darwish (2021)	Twitter	Fornecido por Stefanov et al. (2020); Idioma: inglês	Manual, clusterização	BERT; Mean Shift	Texto (HP); Retweets	Macro-F1	92,1%
Bechini et al. (2021)	Twitter	Original; Idioma: italiano	Manual	SVM	Texto (HP)	Acurácia	64,84%
Hisamitsu et al. (2022)	Twitter	Original; Idioma: Japonês	Manual; automática	DNN (original)	Texto (HP); Retweets; Menções; Respostas;	Macro-F1	66,5%
Abeysinghe, Vulupala, Bourgeois and Sunderraman (2022)	Twitter	Original; Idioma: inglês	N/A	Mean Shift	Texto (HP, artigos compartilhados)	Silhouette Score	0,543
Jia et al. (2022)	Twitter	Fornecido por Zhu et al. (2020); Idioma: inglês	N/A	CK-BiLSTM (original); BiLSTM, LDA	Texto (HP, vizinhos por tópico); Tópico; Bases de Conhecimento	Macro-F1	72,86%
Abeysinghe, Vulupala and Sunderraman (2022)	Twitter	Original; Idioma: inglês	N/A	N/A	N/A	N/A	N/A
Zhang et al. (2023)	Twitter	Original; Idioma: inglês	Automática, Supervisão Distante	XGBoost	Texto (HP); Perfil; Pos. em tópicos relacionados	F1	85%
Hashemian et al. (2023)	Twitter	SemEval; COVID-19 Cotfas et al. (2021) TW-BREXIT Lai et al. (2020); Idioma: inglês	N/A	CNN; LSTM; LSTM-GRU	Texto (HP, feed, favoritos); Características de linguagem	Macro-F1	78,18%

pessoas e conteúdos de crenças e interesses semelhantes aos seus levando à formação de grupos com mentalidade semelhante (Aldayel and Magdy, 2019; Darwish et al., 2017).

A similaridade entre os usuários foi calculada com base em suas interações, empregando uma combinação de grafo de passeio aleatório e reforço em grafo, visando minimizar o peso das interações muito comuns entre os usuários e, conseqüentemente, aumentar a relevância das interações mais discriminativas.

Tanto os modelos *baseline* quanto o modelo final proposto utilizaram o SVM-Light, uma implementação de código aberto do algoritmo SVM. A análise realizada revelou que o uso da similaridade melhorou os resultados em comparação com os *baselines*. A análise incluiu *baselines* treinados apenas com dados textuais e com os diferentes tipos de interação selecionados (menções, *retweets*, *hashtags* e URLs compartilhadas) separadamente. É discutido que alguns tipos de interação tiveram impacto positivo mais significativo nos resultados, sendo *retweets* o mais expressivo. Por outro lado, outros tipos se mostraram dispensáveis, como as URLs compartilhadas. É destacado ainda que a combinação da similaridade com *tweets* não relacionados ao tópico alvo auxiliou na classificação de usuários que não expressam seu posicionamento de forma direta e clara.

O trabalho de Chen and Ku (2018) apresenta um método que busca classificar simultaneamente postagens (*posts*) e usuários. A abordagem central é fundamentada na ideia de que os usuários tendem a curtir *posts* que expressam posições alinhadas com as suas próprias. O estudo empregou o Modelo de Campo Aleatório de Markov (*Markov Random Field - MRF*), no qual um grafo é construído a partir de usuários e *posts* extraídos do Facebook de grupos chineses contrários à energia nuclear, representando os nós, enquanto as interações entre eles correspondem às arestas. O modelo proposto é denominado Joint-MRF. O conjunto de dados foi anotado manualmente e o processo de classificação ocorre de maneira iterativa, com nós e arestas influenciando-se reciprocamente. O Joint-MRF alcançou o melhor desempenho, com uma vantagem estatisticamente significativa em comparação com diversos *baselines* estabelecidos. Entre esses *baselines*, as técnicas baseadas em grafos apresentaram os melhores resultados. Além disso, foi observado que as redes neurais enfrentaram dificuldades na classificação da classe “contra”, principalmente devido ao baixo volume de dados disponíveis para essa categoria.

Aldayel and Magdy (2019) abordam o efeito de diferentes grupos de características em nível de usuário na classificação do posicionamento de usuários. Esta abordagem é comparada à classificação de postagens individuais por modelos estado da arte no momento apresentados na competição SemEval (Mohammad et al., 2016). Além disso, uma análise detalhada foi conduzida para investigar a influência de cada característica nos resultados obtidos e como estas se comparam com a classificação baseada unicamente no conteúdo textual das postagens.

Foram explorados quatro grupos de características em nível de usuário:

- **Textuais:** conteúdo do histórico de postagens do usuá-

rio;

- **Interações:** usuários mencionados e usuários que mencionaram domínios compartilhados;
- **Preferências:** autores das postagens curtidas (ou favoritas) e domínios compartilhados das postagens curtidas;
- **Conexões:** seguidores e amizades.

Os resultados indicaram que a utilização das características de usuário, de maneira geral, melhorou significativamente o desempenho do modelo. Todos os conjuntos de características se mostraram relevantes, com ênfase nas interações. O desempenho mais promissor foi alcançado ao combinar texto e interações em um modelo de classificação binária, ou seja, com as classes “a favor” e “contra”, excluindo a classe “neutro”. Vale ressaltar que os resultados sugerem a possibilidade de classificar usuários com poucas postagens, frequentemente referidos como “*lurkers*” ou “*silent users*”, utilizando este conjunto de características, o que é explicado pela homofilia e pela utilidade de algumas das características, como as do grupo de preferências.

O tema principal do trabalho desenvolvido por Zhu et al. (2020) é a evolução do posicionamento de usuários do Twitter ao longo do tempo (*opinion dynamics*). Para abordar esse tema, foi construído um modelo baseado em Redes Neurais Recorrentes (RNN) e mecanismos de atenção, buscando considerar as postagens (texto) feitas pelo usuário em ordem cronológica, suas ligações de amizade com outros usuários e o contexto das postagens. Esse contexto é definido por postagens relacionadas ao tópico feitas por amigos, com pesos atribuídos por mecanismos de atenção. O modelo, denominado *Neural Opinion Dynamics* (NOD), recebe o histórico de postagens dos usuários dividido em épocas e prevê seu posicionamento em um dado tópico na época subsequente. Os resultados do NOD foram significativamente superiores aos dos *baselines* escolhidos, incluindo o DataStories (Baziotis et al., 2017), considerado o estado da arte para a classificação de posicionamento em nível de tweet na época.

O trabalho de Samih and Darwish (2021) visa fornecer uma forma de classificar o posicionamento de usuários do Twitter com poucos tweets relacionados ao tópico alvo. Para alcançar esse objetivo, são exploradas uma abordagem supervisionada e outra não supervisionada.

Na abordagem supervisionada, o destaque principal é o uso do BERT para a geração de *embeddings* contextualizados, junto de um *fine-tuning* para a detecção de posicionamentos. Já na abordagem não supervisionada, adotou-se a clusterização por similaridade, cujos cálculos se baseiam em características semelhantes, como *hashtags* compartilhadas.

Foram realizados experimentos com e sem o enriquecimento dos dados pelo histórico de postagens dos usuários. Os resultados mais promissores foram alcançados na classificação supervisionada, especialmente quando o histórico não foi incluído. O artigo também inclui uma análise manual para investigar as razões das classificações incorretas. Observou-se que o BERT cometeu erros principalmente em situações de sarcasmo ou citações de perspectivas opostas. Por outro lado, o modelo não supervisionado enfrentou desafios ao classificar usuários com

poucos *retweets* e mostrou-se eficaz apenas com o conjunto de dados enriquecido pelos históricos de postagens.

No modelo descrito por D'Andrea et al. (2019) e aplicado na classificação do posicionamento de usuários em Bechini et al. (2021), a análise foi conduzida a partir da classificação de postagens individuais no Twitter sobre o tópico de vacinação na Itália. O modelo utiliza uma representação *bag-of-words* com índice TF-IDF para os textos dos tweets e emprega um SVM com *kernel* linear como algoritmo de classificação. O posicionamento de cada usuário é determinado pela proporção de tweets favoráveis e contrários em relação ao total de postagens.

Em Bechini et al. (2021), a dinâmica de engajamento e o posicionamento dos usuários são o foco principal. Foram investigadas a distribuição temporal da quantidade de tweets, com ênfase em datas de eventos relevantes (como a aprovação de uma lei de vacinação em escolas), comparações entre perfis verificados e não verificados, e análises do posicionamento de usuários em datas específicas. Também foi examinada a distribuição geográfica do posicionamento ao longo do tempo. Os resultados destacam que usuários verificados tendem a postar com maior frequência, o posicionamento pró-vacina aumentou progressivamente e algumas regiões apresentam um posicionamento majoritário mais pronunciado.

Hisamitsu et al. (2022) investigaram a evolução do posicionamento dos usuários do Twitter ao longo de cinco meses, analisando dados sobre vacinação no Japão. Os dados foram coletados em meses que apresentaram mudanças significativas na cobertura de cidadãos vacinados contra a COVID-19 na população japonesa.

A classificação dos usuários foi baseada no posicionamento de suas postagens individuais, determinada por voto majoritário em períodos de um terço de mês. As postagens foram processadas por meio de uma Rede Neural Profunda composta por três componentes: *text-encoder*, *reaction-encoder* e um classificador. Para o *text-encoder*, foi utilizado um modelo BERT específico para o japonês, com *fine-tuning* para abordar o problema em questão. O *reaction-encoder* gerou um vetor de interações entre usuários, e o classificador recebeu a concatenação dos vetores de texto e interações em uma camada totalmente conectada. A saída foi então encaminhada à função Softmax para realizar a previsão.

Com o posicionamento dos usuários identificado, foram conduzidas análises para investigar as mudanças de posicionamento e suas razões. O foco foi entender com quais usuários as pessoas que mudaram de opinião interagiram e quais fontes externas foram citadas. Os resultados indicaram que a maioria dos usuários era neutra, e as mudanças ocorreram principalmente de uma posição neutra para pró ou contra vacinação. Observou-se que raramente um usuário pró-vacina se tornava contra, e vice-versa. O número de usuários contra vacinação permaneceu estável ao longo do período, mas um grupo significativo mudou de uma posição neutra para pró-vacina. Estes usuários citaram perfis governamentais, médicos e compartilharam notícias de veículos de comunicação de massa, ressaltando a importância desses canais na formação de opinião.

Abeyasinghe, Vulupala, Bourgeois and Sunderraman (2022) propõem uma abordagem não supervisionada para classificar o posicionamento de usuários do Twitter com

base nos artigos de sites externos que eles citam, sem a prévia definição de tópicos. A arquitetura desenvolvida utiliza *embeddings* gerados por “Sentence Transformers” (incluindo BERT) e clusterização (*Mean Shift*) para agrupar os dados por tópico, denominando essa metodologia de “Article Based Stance”. Os dados foram obtidos via API do Twitter, e uma limitação evidente é que o usuário deve ter compartilhado ao menos um *link* de artigo externo para ser classificado por este método. A avaliação do modelo é realizada pelo *Silhouette Score*, que mede a pureza dos *clusters*, e o melhor modelo alcançou um *score* de 0,543. É ressaltada a dificuldade em avaliar os resultados devido à novidade da abordagem e à ausência de conjuntos de dados anotados para calcular outras métricas de avaliação. Também não são apresentados *baselines* para comparação.

A contribuição destacada em Jia et al. (2022) é o uso de bases de conhecimento, especificamente o SenticNet (Cambria et al., 2016), para enriquecer redes BiLSTM, resultando na versão CK-BiLSTM (CK - Common Knowledge). Esta abordagem incorpora informações de sentimento como filtros na rede durante cada iteração, com a entrada sendo o histórico de tweets de cada usuário e os tweets vizinhos. Informações sobre tópicos são extraídas usando LDA e mecanismos de atenção são aplicados para integrar e ponderar as informações dos tweets com base nos tópicos, favorecendo as postagens relevantes ao contexto.

O desempenho do modelo CK-BiLSTM foi comparado com diversos *baselines*, incluindo técnicas de aprendizado de máquina não neural, redes recorrentes, convolucionais e modelos que consideram sentimento ou contexto. Um estudo de ablação foi conduzido, removendo componentes-chave do modelo um por um. Os resultados indicaram que o modelo proposto superou os *baselines* e que os componentes incluídos, principalmente o mecanismo de atenção, tiveram impactos significativos nos resultados.

Abeyasinghe, Vulupala and Sunderraman (2022) apresentam a criação de um conjunto de dados para detecção de posicionamentos de usuário sobre vacinação durante a pandemia de COVID-19, segundo os autores, preenchendo uma lacuna de conjuntos de dados públicos específicos para essa tarefa. O conjunto de dados inclui informações de postagens do Twitter, como texto, likes, *retweets* e comentários, além de detalhes dos artigos compartilhados pelos usuários, como título, texto e URL. Os dados foram coletados via API do Twitter com apoio de outras bibliotecas, com destaque para newspaper3k², utilizada para extração de textos dos artigos externos. Foram aplicados filtros baseados em palavras-chave e *hashtags* no período de 01/01/2020 a 11/01/2020, resultando em 259.228 tweets que continham URLs compartilhadas. Não foi feito nenhum trabalho de anotação dos dados.

Zhang et al. (2023) emprega o uso do posicionamento de usuários a respeito de tópicos relacionados para a classificação do posicionamento em um tópico-alvo. Neste trabalho, o posicionamento a respeito do uso de máscaras e igualdade racial são utilizados como parte da entrada para classificar o posicionamento de usuários em relação ao ex-presidente dos EUA, Donald Trump. O modelo adotado foi o XGBoost, alimentado por três principais características:

²<https://newspaper.readthedocs.io/en/latest/>

conteúdo de tweets históricos, informações do perfil do usuário e seu posicionamento nos tópicos correlatos. Os resultados indicaram uma forte correlação entre o posicionamento em tópicos relacionados e o tópico alvo, validada pelo teste de significância de Spearman. Um estudo de ablação realizado revelou que a combinação das informações de perfil, conteúdo histórico e posicionamento em tópicos correlatos proporcionou o melhor desempenho do modelo.

O trabalho realizado por Hashemian et al. (2023) apresenta um modelo bifásico para rastrear o posicionamento dos usuários no Twitter ao longo do tempo, considerando o histórico de posicionamento do usuário e de outros usuários em seus feeds e favoritos como parte da entrada deste modelo. Na primeira fase, os tweets são classificados como não-subjetivos ou subjetivos (neutros ou com posicionamento), com base em um vetor de características linguísticas como subjetividade, polaridade e tipos de verbos. Este vetor alimenta um modelo de classificação de subjetividade, utilizando uma combinação de CNN e LSTM do PyTorch³. Na segunda fase, os tweets são categorizados como “a-favor” ou “contra”, usando os mesmos dados e uma LSTM-GRU adaptada para ajustar os intervalos temporais considerados.

O modelo destaca a importância da adaptação dos recortes temporais para o rastreamento eficaz do posicionamento. A influência do posicionamento de outros usuários foi mínima, enquanto o histórico de posicionamento do usuário e o mecanismo de adaptação temporal foram determinantes para o desempenho superior do modelo em relação aos *baselines* estabelecidos.

3.2 Questões de Pesquisa

A seguir, abordaremos os principais achados relacionados a cada questão de pesquisa, discutindo, portanto, os principais algoritmos, dados, características, métodos de avaliação e resultados observados nos estudos selecionados.

3.2.1 Algoritmos e modelos (QP1)

Inicialmente, nos trabalhos publicados entre 2017 e 2018, é notável que os autores buscaram utilizar conceitos de grafos como centro da proposta, em conjunto com outras técnicas. Dong et al. (2017) utilizam um grafo formado por usuários como nós e conversas como arestas. Estas arestas foram anotadas com base em um conjunto de regras (conversas com número elevado de turnos indicam discordância, por exemplo) e o modelo final foi uma Regressão Logística que classifica o posicionamento de novos usuários pela predição do sinal de novas arestas. O principal destaque do trabalho é a criação de um modelo pouco dependente do domínio dos dados. Destaca-se também que este trabalho foi o único analisado a dar o posicionamento de forma numérica, possibilitando a indicação da intensidade do posicionamento.

Em Darwish et al. (2017) é feita a classificação do posicionamento de usuários por um modelo SVM-Light treinado a partir de um grafo de similaridade entre usuários. Este

grafo foi criado pela combinação de grafo bipartido, reforço de grafo e a similaridade entre usuários foi computada a partir de interações entre eles. Esta abordagem facilita a classificação de usuários que publicam poucas postagens, e não expressam diretamente o seu posicionamento, mas podem fazê-lo por meio de favoritos ou compartilhamentos.

O trabalho de Chen and Ku (2018) foi o único a utilizar um método estocástico como centro da proposta. Os autores empregam um Campo Aleatório de Markov com um grafo composto de usuários e postagens como nós e interações como arestas para criar o modelo proposto, chamado de Joint-MRF. A classificação acontece de forma iterativa, com os nós e arestas influenciando seus vizinhos. A abordagem possibilita a classificação de usuários e postagens simultaneamente e, assim como a descrita em Dong et al. (2017), também se mostra pouco dependente do domínio dos dados.

Os trabalhos analisados demonstram que uma das principais vantagens do uso de grafos é a flexibilidade. Nota-se que a modelagem dos grafos empregada em cada um dos três trabalhos é bastante distinta. Dong et al. (2017) e Chen and Ku (2018) estão entre os poucos trabalhos analisados em que os resultados obtidos não estão diretamente vinculados a características específicas de uma rede (como *retweets*, por exemplo). Esta variabilidade na modelagem também indica que não foi definida uma estratégia ideal para uso de grafos, e a modelagem pode também ser um ponto de falha e complexidade a mais no desenvolvimento dos classificadores. Podemos notar também que poucos dos trabalhos recuperados nesta revisão e publicados nos anos seguintes utilizam grafos. Hisamitsu et al. (2022) fazem uso de um grafo de interações como parte da entrada, mas ele não é parte central do desenvolvimento do modelo como visto nos três primeiros, provavelmente porque os ganhos não compensam a complexidade adicionada.

O uso de aprendizado de máquina não neural também é bastante presente nos trabalhos analisados. Destacando-se um domínio das Máquinas de Vetores de Suporte e variações de implementação da mesma (como o SVM-Light), utilizadas em Darwish et al. (2017); Aldayel and Magdy (2019); Samih and Darwish (2021) e Bechini et al. (2021). Além do SVM, também foi utilizado o algoritmo XGBoost, empregado em Zhang et al. (2023).

A principal vantagem dos algoritmos tradicionais é a simplicidade da implementação, e suas principais desvantagens seriam a baixa interpretabilidade dos modelos e a ineficiência computacional com grandes volumes de dados, o que tende a ser o caso na área estudada. O motivo da escolha dos algoritmos tradicionais parece ser justamente sua simplicidade e a possibilidade de explorar diversos tipos de análise a partir deles. Nota-se que alguns trabalhos buscaram focar em outras contribuições além da assertividade dos modelos. Darwish et al. (2017) exploram mais o uso da similaridade para classificar usuários que não expressam diretamente o seu posicionamento. Aldayel and Magdy (2019) exploram uma grande variedade de características, com destaque para a utilização de modelos binários (sem a classe “neutra”, somente “a-favor” e “contra”), que mostrou bons resultados. Bechini et al. (2021) trouxeram análises detalhadas sobre as características dos usuários e a relação com o posicionamento.

³PyTorch <https://pytorch.org/>, acessado em 28/04/2024.

Zhang et al. (2023) focam em explorar a relação entre o posicionamento em diferentes tópicos. Dessa forma, é possível concluir que os algoritmos não neurais costumam gerar resultados sólidos e trazem simplicidade no desenvolvimento, mas tendem a não apresentar os melhores resultados.

Outra abordagem recorrente é a utilização de redes neurais. No trabalho desenvolvido em Zhu et al. (2020), o modelo é projetado para prever o posicionamento futuro de um usuário com base em seu histórico de postagens, dividido em épocas definidas por janelas de tempo fixas ou quantidade fixa de tweets. A rede é composta por diversas camadas, com destaque para a camada que gera a representação das postagens através de um LSTM, e a camada que concatena as características finais através de uma GRU. Os autores destacam a preferência pela rede GRU em todos os aspectos, exceto para a modelagem de linguagem.

Em Hisamitsu et al. (2022) utilizou-se uma Rede Neural Profunda composta por três componentes: um *text-encoder* que processa o texto nas postagens dos usuários, o chamado *reaction-encoder* representado por um vetor de interações entre usuários, e um classificador que recebe a concatenação dos vetores de texto (saída do primeiro componente) e interações (saída do segundo componente), processando-os em uma camada totalmente conectada e encaminhando a saída para a função softmax. Com todas as postagens anotadas quanto ao posicionamento, a classificação dos usuários foi realizada por voto majoritário dividida em períodos de um terço de mês, totalizando quinze períodos ao longo de cinco meses de dados.

Jia et al. (2022) têm como principal modelo uma BiLSTM alimentada pelo histórico de tweets dos usuários e de seus vizinhos. O modelo possui ainda informações de sentimento provenientes da base de conhecimento SenticNet (Cambria et al., 2016), que dá nome ao modelo final chamado de CK-BiLSTM (*Common Knowledge - BiLSTM*).

O trabalho desenvolvido em Hashemian et al. (2023) classifica os usuários em diferentes intervalos de tempo, considerando o posicionamento em períodos passados para o posicionamento futuro. Dos trabalhos que utilizam redes neurais, este é o único que não foca no desenvolvimento de uma rede original. É usado um modelo da biblioteca PyTorch composto por uma combinação de CNN e LSTM como classificador. No entanto, a novidade da abordagem é a inclusão de uma LSTM-GRU treinada para a adaptação dos intervalos de tempo usados no modelo final, e esta contribuiu significativamente para os resultados do modelo.

O uso de redes neurais para classificar o posicionamento de usuários é frequentemente adotado em cenários onde a evolução temporal do posicionamento é relevante (Zhu et al., 2020; Hisamitsu et al., 2022; Hashemian et al., 2023). Notavelmente, as redes LSTM são populares devido à sua eficácia em lidar com seqüências, como o histórico de postagens dos usuários, e em tarefas de Processamento de Linguagem Natural. A principal vantagem das redes neurais nesse contexto é sua aptidão para dados onde a ordem cronológica é crucial. Além disso, certos trabalhos, como Jia et al. (2022), apresentaram bons resultados apenas com dados textuais. No entanto, as desvantagens incluem a complexidade de implementação e a exigência de grandes conjuntos de dados. Por exemplo, Hisamitsu et al. (2022)

destacaram problemas de volume e desequilíbrio de classes que impactaram a assertividade do modelo.

Alguns dos trabalhos buscaram soluções não supervisionadas. A primeira aparição deste tipo de técnica foi vista em Dong et al. (2017). Neste trabalho, como mencionado na seção anterior, a abordagem principal foi o uso de um grafo formado por usuários como nós e conversas como arestas, e o posicionamento foi classificado a partir da atribuição de um sinal relativo à oposição ou ao acordo. Tal classificação foi feita de forma não supervisionada a partir da definição de um conjunto de regras. No entanto, também foi realizado um experimento com usuários selecionados aleatoriamente do conjunto de dados para a anotação deste sinal por humanos para fins de validação.

Em Samih and Darwish (2021), uma das estratégias investigadas envolveu a clusterização de usuários com base na similaridade de vetores de características, derivados das contas autoras de postagens que cada usuário repostou por via de *retweets*. Para essa tarefa, foram empregadas técnicas de projeção e clusterização utilizando UMAP e *Mean Shift*. Em seguida, os clusters resultantes foram manualmente anotados para representar o posicionamento dos usuários em relação ao tópico em foco.

No estudo de Abeysinghe, Vulupala, Bourgeois and Sunderraman (2022) foi proposto um método não supervisionado independente de tópicos para classificar o posicionamento de usuários, denominado *Article Based Stance*. Esta abordagem classifica o posicionamento dos usuários com base nos artigos que compartilham, utilizando a clusterização desses artigos pelo algoritmo *Mean Shift*. A classificação é realizada considerando a similaridade semântica entre as postagens dos usuários e os centros dos clusters. A escolha pelo *Mean Shift* foi motivada por este não exigir a definição prévia do número de clusters e sua eficiência em encontrar grupos em espaços com dimensionalidade elevada em comparação a outros métodos, como variações do KMeans. Devido à singularidade dessa estratégia, os autores encontraram desafios ao comparar seus resultados com outros estudos da área.

Em geral, a motivação predominante para o uso de técnicas não supervisionadas é a capacidade de lidar com dados não anotados, especialmente devido à escassez de conjuntos de dados disponíveis para a tarefa abordada nesta revisão. No entanto, essas abordagens enfrentam desafios ao lidar com volumes menores de dados (Samih and Darwish, 2021). Embora não tenham apresentado melhorias significativas em relação às abordagens supervisionadas, os resultados destas últimas estão intrinsecamente ligados à qualidade da anotação dos dados. Assim, sugere-se o uso de abordagens não supervisionadas onde há dificuldade na anotação e validação de sua qualidade nos dados explorados.

Em Zhu et al. (2020) e Jia et al. (2022), foi adotada uma abordagem combinando mecanismos de atenção e extração de tópicos para capturar a influência das postagens vizinhas na opinião do usuário. No trabalho desenvolvido por Zhu et al. (2020), as postagens vizinhas atuam como contexto para cada período, e sua influência é ponderada por meio de pesos estabelecidos por um mecanismo de atenção, complementado pela extração de tópicos por meio de uma HLDA. Já em Jia et al. (2022), as postagens do usuário, postagens vizinhas e informações de tópicos extraídas por

uma LDA são concatenadas e fornecidas a um mecanismo de atenção, gerando o vetor final para a classificação do posicionamento. Ambos os estudos empregaram variações de LDA para extração de tópicos e desenvolveram mecanismos de atenção que resultam em uma soma ponderada normalizada, refletindo o grau de influência das postagens vizinhas na opinião do usuário. Os autores ressaltam que essa combinação de técnicas impactou significativamente a qualidade dos resultados obtidos.

Por fim, vale destacar que muitos trabalhos incorporam o BERT como componente crucial do desenvolvimento e enfatizam sua relevância nos resultados alcançados (Samih and Darwish, 2021; Hisamitsu et al., 2022; Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022; Zhang et al., 2023). O uso mais frequente do BERT envolve seu *fine-tuning* para a tarefa específica em questão, sendo sua principal aplicação a geração de *embeddings*. Entre esses, destaca-se Zhang et al. (2023), que utiliza o *Sentence-Bert* na codificação de tweets.

3.2.2 Características mais relevantes (QP2)

Dentre os trabalhos analisados, nota-se que as características exploradas para o treinamento dos modelos podem ser agrupadas em algumas categorias principais: características relacionadas ao texto das postagens; interações entre usuários; conexões entre usuários; informações de perfil; similaridade entre usuários; e tópicos das postagens. O uso combinado de características de texto e interações é o mais comum, enquanto as outras características e estratégias variam conforme os objetivos específicos de cada estudo. A escolha das características também é frequentemente influenciada pela rede social de onde os dados foram coletados.

As características de texto podem ser derivadas de diversas fontes. A abordagem predominante nesta revisão é o uso do histórico de postagens dos usuários (Darwish et al., 2017; Aldayel and Magdy, 2019; Zhu et al., 2020; Samih and Darwish, 2021; Bechini et al., 2021; Hisamitsu et al., 2022; Jia et al., 2022; Zhang et al., 2023; Hashemian et al., 2023).

Observa-se que nove dos dez trabalhos que utilizaram dados do Twitter empregaram o histórico de postagens dos usuários. Essas características são descritas com maior importância em estudos que desenvolveram modelos não supervisionados (Samih and Darwish, 2021; Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022) e em classificações e análises da evolução do posicionamento dos usuários ao longo do tempo (Zhu et al., 2020; Hisamitsu et al., 2022; Hashemian et al., 2023).

Além do texto do histórico de postagens, outras fontes incluem: texto de postagens isoladas (Dong et al., 2017; Chen and Ku, 2018; Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022), texto de postagens vizinhas sobre o mesmo tópico (Jia et al., 2022), artigos compartilhados nas postagens (Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022; Abeysinghe, Vulupala and Sunderraman, 2022), comentários e conversas entre usuários (Dong et al., 2017), e texto de postagens presentes no *feed* dos usuários (Hashemian et al., 2023).

Quanto ao pré-processamento e extração de características dos textos, observa-se o uso frequente das seguintes técnicas:

- Tokenização;
- Categorias léxicas;
- Limitação do tamanho dos textos;
- Normalização de caracteres;
- *Stemming*;
- Remoção de links e URLs, menções, endereços de e-mail e *stop-words*;
- Verificação ortográfica.

Alguns trabalhos que analisaram o histórico de postagens dos usuários organizaram os dados em ordem cronológica (Zhu et al., 2020) e consolidaram todas as postagens de um usuário em um único documento ou vetor (Samih and Darwish, 2021). Além disso, técnicas de geração de dados a partir dos textos, como mineração de informações para coleta de frases de alta qualidade (Dong et al., 2017), análise de sentimento (Chen and Ku, 2018) e codificação de texto utilizando BERT (Hisamitsu et al., 2022) e *Sentence-BERT* (Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022) também foram empregadas, sendo a codificação consistentemente descrita com importância, ressaltando a utilidade dos mecanismos de atenção.

Em relação às interações, os principais elementos observados são os *retweets* e os favoritos (ou *likes*), que aparecem na maioria dos estudos que trabalham com dados do Twitter. Destacam-se os *retweets*, cuja influência na melhoria dos resultados se mostrou bastante consistente nesta revisão (Darwish et al., 2017; Samih and Darwish, 2021; Hisamitsu et al., 2022; Zhang et al., 2023).

Alguns estudos consideram como interações elementos presentes no texto das postagens que geram algum tipo de ligação entre os usuários, sendo os principais as *hashtags*, domínios compartilhados (quando dois ou mais usuários compartilham links da mesma fonte) e menções a outros usuários (Darwish et al., 2017; Aldayel and Magdy, 2019; Hisamitsu et al., 2022; Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022). Essas informações foram por vezes descritas como pouco influentes nos resultados, com destaque para os domínios compartilhados, que chegaram a mostrar uma piora nos resultados em Darwish et al. (2017).

Nos estudos de Dong et al. (2017); Hisamitsu et al. (2022) e Zhang et al. (2023), as interações são representadas como grafos. Em particular, Dong et al. (2017) trata como interações as conversas entre usuários nas seções de comentários das postagens e gera o grafo a partir dessas conexões.

Entre as características de treinamento observadas nesta revisão, as interações parecem ser as mais dependentes do domínio. A maioria dos trabalhos utiliza interações específicas e únicas da plataforma Twitter. Dong et al. (2017) e Chen and Ku (2018) são os únicos estudos selecionados que empregam dados de outras redes sociais online, sendo que Dong et al. (2017) utiliza conversas como interações e Chen and Ku (2018) faz uso somente de *likes*.

É comum entre os estudos que exploram características de texto e interações que os melhores, ou os únicos, resultados reportados sejam alcançados, não exclusivamente, pelo uso combinado desses grupos (Dong et al., 2017; Chen and Ku, 2018; Aldayel and Magdy, 2019; Hisamitsu et al., 2022). Destaca-se que as interações se mostraram relevantes, com cada tipo tendo uma influência diferente nos resultados. Estudos que buscaram classificar

usuários pouco ativos relataram grande importância das interações para os resultados obtidos (Darwish et al., 2017; Aldayel and Magdy, 2019; Samih and Darwish, 2021), com destaque para Samih and Darwish (2021), onde o melhor resultado do modelo não supervisionado desenvolvido foi obtido com base apenas nas contas autoras de postagens que cada usuário repostou por via de *retweets*, e Darwish et al. (2017), em que o uso de domínios compartilhados reduziu a precisão do modelo.

Em Darwish et al. (2017), foram realizados experimentos com diferentes *baselines*, variando o uso de apenas texto e apenas interações como características de treinamento. Os resultados indicaram que pode ocorrer, mas não é possível generalizar, das interações se mostrarem superiores ao texto na contribuição para a performance dos modelos de detecção de posicionamento.

Outra característica consistentemente relevante nos estudos analisados são as conexões entre usuários. No entanto, a forma como essas conexões são utilizadas varia significativamente entre os trabalhos. Em Aldayel and Magdy (2019), informações sobre seguidores e seguidos, extraídas dos perfis dos usuários, são usadas para melhorar a classificação de usuários pouco ativos. Por outro lado, em Chen and Ku (2018), devido às restrições da plataforma de origem dos dados (Facebook), as conexões são criadas a partir das postagens com as quais os usuários interagiram. Esse processo resulta na criação de um grafo, onde usuários e postagens são representados como nós, e a autoria ou interação dos usuários com postagens, como arestas. Esse grafo constitui a base para o modelo desenvolvido. Finalmente, Zhu et al. (2020) utilizam informações de conexão dos perfis dos usuários para coletar postagens vizinhas sobre o mesmo tópico e no mesmo período, com o objetivo de gerar um contexto para a classificação das postagens do usuário em questão.

Desconsiderando informações sobre amizade e conexões, apenas um trabalho relatou o uso de informações de perfil dos usuários. Em Zhang et al. (2023), foi demonstrado em um estudo de ablação que o melhor modelo observado recebeu como entrada informações de perfil, textos do histórico de postagens e o posicionamento do usuário em tópicos relacionados. Ressalta-se também que este trabalho foi o único a utilizar o posicionamento dos usuários em tópicos previamente classificados. Embora os resultados tenham mostrado uma influência positiva significativa dessa característica, a aplicação dessa abordagem se mostra complexa devido à necessidade de selecionar cuidadosamente os tópicos a serem considerados e de classificar os dados obtidos para eles.

Nos trabalhos de Darwish et al. (2017) e Samih and Darwish (2021), a similaridade entre usuários foi incluída no conjunto de características utilizadas. Ambos os estudos justificam o uso da similaridade com base no conceito de homofilia e têm como um de seus principais objetivos a classificação de usuários pouco ativos. Em Darwish et al. (2017), a similaridade é calculada com base nas interações e os resultados computados são utilizados como parte da entrada do modelo final. Essa abordagem demonstrou uma melhora nos resultados em comparação com a utilização direta das interações. Já em Samih and Darwish (2021), a similaridade serve como base para a estratégia de agrupamento não supervisionado para a classificação dos

usuários. O cálculo é feito a partir de um vetor de características formado por contas de *retweets* do usuário. Os resultados dessa abordagem foram semelhantes aos de um modelo supervisionado focado na geração de *embeddings* através do BERT, desenvolvido no mesmo estudo.

Por fim, em Zhu et al. (2020) e Jia et al. (2022), observa-se o uso de tópicos extraídos de postagens como parte do conjunto de características utilizadas. Ambos os trabalhos seguem abordagens semelhantes quanto à forma e ao objetivo do uso dos tópicos. Eles coletam postagens vizinhas e utilizam um algoritmo com base em LDA para extrair tópicos dessas postagens. Esses tópicos são então integrados com mecanismos de atenção para incluir a influência do contexto do usuário na classificação final de seu posicionamento. Em particular, Jia et al. (2022) destaca que essa combinação teve o maior impacto nos resultados dentre as características selecionadas.

3.2.3 *Baselines e Avaliação (QP3 e QP4)*

Os trabalhos analisados apresentam uma relativa variedade de *baselines*, com alguns focando em testar o conjunto de dados criado em outros trabalhos (Dong et al., 2017; Chen and Ku, 2018; Zhu et al., 2020; Samih and Darwish, 2021; Jia et al., 2022; Zhang et al., 2023; Hashemian et al., 2023), enquanto outros dão bastante atenção à variação das características fornecidas aos *baselines* em relação ao modelo final (Darwish et al., 2017; Aldayel and Magdy, 2019; Samih and Darwish, 2021).

Dentre os algoritmos utilizados, o mais presente foi o SVM, empregado em sete dos doze trabalhos revisados que desenvolveram modelos de classificação (Darwish et al., 2017; Chen and Ku, 2018; Aldayel and Magdy, 2019; Samih and Darwish, 2021; Jia et al., 2022; Zhang et al., 2023; Hashemian et al., 2023). A principal justificativa para esta escolha parece ser o uso e os resultados observados do SVM na competição SemEval 2016 (Mohammad et al., 2016).

Apenas dois estudos mencionaram o uso direto de modelos desenvolvidos em trabalhos anteriores. Zhu et al. (2020) utiliza o modelo DataStories (Baziotis et al., 2017), considerado o estado da arte em detecção de posicionamentos até então. Jia et al. (2022) selecionam trabalhos anteriores de forma a variar o tipo de algoritmo empregado, incluindo modelos baseados em RNN, CNN, SVM, além de modelos que consideram avaliações de sentimento do usuário e baseados em contexto.

Observamos também alguns casos específicos que merecem destaque. Em Dong et al. (2017), o modelo final proposto é feito a partir de duas classificações distintas, exigindo o desenvolvimento de um *baseline* para cada etapa. Outro ponto de destaque do trabalho foi a realização de testes variando a forma de anotação dos dados; os *baselines* e o modelo final foram avaliados com dados anotados de forma automática e manual separadamente. Um dos pontos discutidos nos resultados foi que a Regressão Logística utilizada como *baseline* aprendeu melhor os dados anotados automaticamente em comparação aos anotados manualmente.

Aldayel and Magdy (2019) não descreveram a criação de um *baseline* específico, mas realizaram variações na configuração do modelo em uma espécie de ablação, variando as classes (versões com e sem a classe “neutra”) e as características utilizadas. O modelo final foi dado pela

combinação que apresentou os melhores resultados.

Samih and Darwish (2021) foi o único estudo que indicou o uso de modelos públicos pré-treinados como parte dos *baselines*, utilizando o modelo FastText⁴, baseado em aprendizado profundo.

A respeito da avaliação, a métrica mais presente foi a medida Macro-F1, utilizada em seis dos doze trabalhos revisados que desenvolveram modelos de classificação (Darwish et al., 2017; Chen and Ku, 2018; Aldayel and Magdy, 2019; Hisamitsu et al., 2022; Hashemian et al., 2023; Jia et al., 2022). Não foi encontrada uma justificativa específica para essa escolha.

Alguns trabalhos avaliam os resultados observando os valores obtidos para diferentes classes e tópicos separadamente (Zhu et al., 2020; Samih and Darwish, 2021). Esta abordagem pode tornar mais claras as dificuldades relacionadas à classificação do posicionamento de usuários em vista dos dados e dos modelos escolhidos. Por exemplo, fica evidente nos resultados que a precisão para a classe “contra” foi menor. Destaca-se também que as métricas de avaliação finais podem ser apresentadas como a média desses resultados obtidos separadamente, como visto em Jia et al. (2022) e Hashemian et al. (2023), que empregam a Macro-F1 média.

Jia et al. (2022) e Zhang et al. (2023) descrevem um estudo de ablação detalhado junto dos resultados, variando características e componentes presentes no treinamento de seus modelos. Em particular, Zhang et al. (2023) também variaram técnicas de pré-processamento durante o treinamento dos *baselines*.

3.2.4 Conjuntos de Dados (QP5)

Os conjuntos de dados utilizados nos trabalhos desta revisão apresentam pouca variação em relação a características e forma de coleta dos dados. Alguns trabalhos fazem uso de conjuntos criados em pesquisas anteriores (Darwish et al., 2017; Aldayel and Magdy, 2019; Samih and Darwish, 2021; Hashemian et al., 2023), porém a maioria fez a coleta de dados especificamente para o trabalho (Dong et al., 2017; Chen and Ku, 2018; Zhu et al., 2020; Bechini et al., 2021; Hisamitsu et al., 2022; Abeysinghe, Vulupala, Bourgeois and Sunderraman, 2022; Jia et al., 2022; Abeysinghe, Vulupala and Sunderraman, 2022; Zhang et al., 2023).

É notável a falta de disponibilidade de conjuntos de dados públicos para a tarefa de detecção de posicionamentos de usuários. Os trabalhos que aproveitam dados de pesquisas passadas comumente descrevem o enriquecimento de um conjunto de dados originalmente criado para a tarefa de detecção de posicionamento em nível de postagem com o *feed* ou o histórico de postagens dos usuários ativos.

A respeito dos tópicos relacionados aos dados utilizados, foi comum a presença de tópicos controversos e, principalmente, políticos. Como o Brexit (Zhu et al., 2020) e a vacinação contra a COVID-19 (Bechini et al., 2021). Também há trabalhos que trataram mais de um tópico, como visto em Samih and Darwish (2021), cujos dados empregados continham postagens sobre oito tópicos polarizados nos Estados Unidos.

A forma de coleta de dados, em maioria, se deu pelo uso

de APIs por meio de buscas realizadas com filtros de tempo e palavras-chave. Em Zhang et al. (2023), por exemplo, o conjunto de dados utilizado foi coletado através da API do Twitter a partir de palavras-chave, tendo estas sido definidas após a análise de uma seleção de postagens de interesse feita manualmente. Foram inicialmente coletados dados relacionados aos três tópicos tratados, resultando em mais de 100 milhões de tweets. Após a remoção de *retweets*, tweets de bots e tweets de usuários pouco ativos, restaram 29 milhões. Foram então coletadas informações de perfil e histórico de postagens dos usuários e, para os 10 mais ativos, também foram coletados *retweets* usados na formação de uma rede.

O trabalho de Abeysinghe, Vulupala and Sunderraman (2022) foi o único desta revisão dedicado à criação de conjunto de dados. O trabalho descreve a criação de um conjunto de dados para detecção de posicionamentos no contexto de vacinação na época da pandemia de COVID-19. Segundo os autores, existia uma lacuna dada pela falta de conjuntos de dados publicamente disponíveis direcionados para esta tarefa.

O conjunto criado contém informações de tweets (texto, *likes*, *retweets*, número de comentários) e informações de artigos compartilhados por meio de links externos pelos usuários (título, texto, url). Os dados foram coletados utilizando a API do Twitter em conjunto com outras bibliotecas, destaque para a *newspaper3k*⁵, utilizada para extrair textos dos artigos compartilhados. Os filtros de busca foram a presença de um conjunto de palavras-chave, *hashtags* e texto das postagens dentro de um intervalo de dez meses.

O trabalho de Dong et al. (2017) foi o único analisado que não fez uso de APIs. Foram utilizados dados do “CNN news commenting system”, coletados a partir de um *crawling* de informações de artigos de notícias. Também foram utilizados dados provenientes do site “4Forums.com” a partir do conjunto de dados “Internet Argument Corpus (IAC)”. Vale ressaltar que o sistema de comentários citado foi desativado, e o website “4Forums.com” não está disponível para acesso no momento de escrita deste texto.

Sobre a anotação dos dados, a abordagem mais comum foi a combinação de anotação manual e automática, com a anotação manual frequentemente usada para validar a automática. Zhu et al. (2020) realizaram a anotação de forma automática por um modelo chamado DataStories (Baziotis et al., 2017), e uma fração dos dados foi anotada manualmente para a validação. Outro exemplo pode ser visto em Chen and Ku (2018), onde foram coletados dados de postagens em grupos do Facebook em um intervalo de um ano. Um terço dos dados (13.973) foi anotado manualmente por dois anotadores e a concordância entre eles foi avaliada com Cohen’s Kappa.

No trabalho de Dong et al. (2017), conversas entre usuários foram anotadas automaticamente com base em um conjunto de regras (por exemplo, um alto número de turnos na discussão indica discordância). Um experimento adicional foi realizado com quatrocentos usuários selecionados aleatoriamente para anotação manual.

No trabalho de Zhang et al. (2023), os dados foram anotados utilizando uma estratégia de Supervisão Distante,

⁴<https://fasttext.cc/>

⁵<https://newspaper.readthedocs.io/en/latest/>

classificando tweets com base nas hashtags presentes no conteúdo. Os usuários foram classificados de acordo com a porcentagem de tweets de sua autoria em cada classe. Se mais de 55% dos tweets de um usuário fossem classificados como a favor do tópico, o usuário também seria classificado como a favor. E para a classificação dos tweets sem hashtags, foi utilizado um classificador baseado no BERT com *fine-tuning* para a detecção de posicionamentos.

4 Discussão

Esta revisão teve foco em classificação de posicionamento de usuários. Dada a flexibilidade dos filtros utilizados para as buscas realizadas, as bases consideradas e o número de estudos incluídos ao final da análise de inclusão, podemos considerar que esta é uma subárea da detecção de posicionamentos pouco explorada de forma direta na literatura atual. Nesta seção, indicaremos nossas principais conclusões a respeito das lacunas e tendências identificadas.

Observamos que a maioria dos trabalhos indicou uma escassez de conjuntos de dados disponíveis para a tarefa de detecção de posicionamento de usuários. Por isso, optaram por fazer a coleta dos dados durante o desenvolvimento de seus estudos, seja construindo o conjunto do zero ou enriquecendo conjuntos criados para a detecção em nível de postagens com informações do histórico de postagens ou perfil dos usuários.

Foi clara a dificuldade de balanceamento dos dados em relação às quantidades para cada classe. Alguns trabalhos explicitaram o baixo volume para a classe “contra”, por exemplo. Apesar de existirem técnicas para mitigar este problema, é importante que trabalhos futuros tenham este ponto em foco no momento da coleta dos dados a serem utilizados.

Foram encontrados poucos esforços para o desenvolvimento de modelos fracamente guiados e não supervisionados, ou mesmo fora da rede social Twitter (atualmente chamada de X). Nota-se uma maioria considerável em relação à rede social de origem dos dados, em que apenas dois dos treze trabalhos analisados não tiveram base na plataforma Twitter, como é possível visualizar na Tabela 6. Essa prevalência gera certa dependência dos modelos criados em relação à plataforma, visto que boa parte dos melhores resultados obtidos estão ligados a características específicas do Twitter, como os *retweets*, por exemplo, conforme apresentado na Seção 3.1. Pode-se concluir que esta preferência pelo Twitter teve fundamento na adequação dos dados fornecidos por ela ao problema de detecção de posicionamentos, bem como na facilidade que existia para a obtenção destes dados.

No entanto, esta dependência de uma determinada plataforma controlada por empresas privadas se mostra um problema para o futuro da área, e se faz necessário encontrar formas alternativas para a construção de modelos e conjuntos de dados para a detecção de posicionamentos de usuários, resgatando a ideia do desenvolvimento de trabalhos que priorizem a baixa dependência de plataformas específicas.

Dentre as características utilizadas, o texto do histórico de postagens se mostra a mais consistente, mas interações parecem se destacar com maior relevância para os

resultados, principalmente para a classificação de usuários pouco ativos ou com poucas postagens dentro do tópico escolhido. O uso da similaridade nos trabalhos analisados se mostrou promissor como alternativa para estes problemas específicos, e parece ter sido pouco explorado.

Alguns trabalhos indicam os *retweets* como uma das principais características para os resultados de seus modelos finais. Este, que funciona como uma espécie de junção do compartilhamento direto e do comentário, é bastante dependente e muito particular da rede social Twitter. Isso mostra a necessidade de buscar um substituto para essa característica em outras redes sociais. Já para outras interações presentes no Twitter, como, por exemplo, os favoritos, é comum que haja interações análogas. Tal como as curtidas do Facebook.

Sobre os *baselines*, pode-se perceber o foco na variação de algoritmos e características. O uso de modelos desenvolvidos em trabalhos anteriores foi pouco utilizado. Dentre os motivos para tal parecem estar a falta de disponibilidade e a dificuldade de replicação, visto que seria necessário ter acesso aos mesmos dados ou ao modelo previamente treinado.

Dentre os modelos adotados, inicialmente, os grafos foram uma escolha comum, com um dos motivos sendo a flexibilidade dada por sua modelagem. No entanto, a complexidade de implementação e os resultados obtidos implicaram na mudança de estratégia para trabalhos mais recentes. O SVM se fez bastante presente, tanto na definição dos *baselines* quanto nos modelos finais. Já para os modelos que levam em consideração características temporais, as RNNs foram a escolha comum, com foco principalmente em derivados de redes LSTM.

Nota-se uma crescente adoção de mecanismos de atenção aliados à extração de tópicos e ao contexto no qual o usuário está inserido, seja por seu *feed* ou por postagens de amigos. O mesmo crescimento também é observado na utilização do BERT para a geração de *embeddings* contextualizados.

Observa-se também uma maioria de modelos supervisionados, utilizando dados já anotados de trabalhos anteriores, ou seguindo uma estratégia de anotação mista, com parte manual, comumente empregada para validação, e parte automática. No entanto, os poucos modelos não supervisionados desenvolvidos mostraram resultados relativamente próximos aos supervisionados, e a praticidade da não necessidade de anotação se mostra relevante para a detecção de posicionamentos, dada a velocidade com que novos temas surgem em discussão.

Foram observadas algumas estratégias presentes em poucos trabalhos, mas que mostraram implicações relevantes nos resultados: a classificação numérica, indicando a intensidade do posicionamento; a remoção da classe “neutro”, tornando a classificação binária e ignorando possíveis postagens fora do tópico classificadas como neutras; e a classificação de subjetividade, que diz respeito a identificar quais postagens de fato possuem um posicionamento.

A forma de avaliação mais comum se deu pela métrica Macro-F1 e se mostrou bastante relevante à avaliação individual de cada classe e de cada tópico, em casos onde mais de um tópico foi utilizado. Um ponto de destaque para a utilização desta estratégia é a possibilidade de comparação com os resultados vistos em trabalhos anteriores.

Tabela 6: Síntese numérica da revisão sistemática

Categoria	Subcategoria	Qtd. de Trabalhos
Conjuntos de Dados	Enriquecidos ou Construídos (Coleta Própria)	9
	Prontos (Existentes de pesquisas anteriores)	4
Redes Sociais	Twitter (X)	11
	Outras (Facebook, CNN News, 4Forums)	2
Modelos e Algoritmos	SVM e variações (ex: SVM-Light)	7
	Redes Neurais (LSTM / BiLSTM / GRU)	6
	BERT (Embeddings / Fine-tuning)	4
	Modelos Baseados em Grafos	3
	Clusterização (Mean Shift)	2
Características	Texto (Histórico de Postagens - HP)	12
	Interações (Retweets, Likes, Favoritos)	8
	Conexões (Seguidores, Amizades)	3
	Similaridade entre Usuários	2
	Tópicos das Postagens (Extração via LDA/HLDA)	2
	Informações de Perfil	2

O foco nas possíveis análises derivadas da detecção de posicionamento de usuários como, por exemplo, a identificação das características geográficas da população de cada classe, a variação do posicionamento ao longo do tempo junto de acontecimentos relevantes no período analisado, fontes externas com as quais os usuários interagiram, entre outras possibilidades, pode trazer informações bastante relevantes e foi pouco empregada nos trabalhos desta revisão.

Por fim, é importante ressaltar as principais limitações desta revisão. A primeira diz respeito à incerteza quanto à variação dos termos utilizados para a detecção de posicionamento em nível de usuário. Durante a pesquisa exploratória prévia, não foram identificados outros termos amplamente utilizados além daqueles selecionados como palavras-chave para as buscas finais. Ainda assim, é possível que existam estudos relevantes que empreguem diferentes terminologias. Além disso, a revisão foi limitada pelo número reduzido de estudos incluídos e pelas restrições das ferramentas de busca disponíveis em algumas das bases consultadas.

Referências

- Abeyasinghe, B., Vulupala, G. R., Bourgeois, A. G. and Sunderraman, R. (2022). Unsupervised user stance detection on tweets against web articles using sentence transformers, *2022 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pp. 1163–1169. <https://doi.org/10.1109/IPDPSW55747.2022.00190>.
- Abeyasinghe, B., Vulupala, G. R. and Sunderraman, R. (2022). Misinformation in social media platforms and web articles: a dataset to infer user stance, *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*, pp. 269–273. <https://doi.org/10.1109/ICSC52841.2022.00051>.
- AlDayel, A. and Magdy, W. (2019). Your stance is exposed! analysing possible factors for stance detection on social media, *Proc. ACM Hum.-Comput. Interact.* 3(CSCW). <http://s://doi.org/10.1145/3359307>.
- ALDayel, A. and Magdy, W. (2021). Stance detection on social media: State of the art and trends, *Information Processing & Management* 58(4): 102597. <https://doi.org/10.1016/j.ipm.2021.102597>.
- Allaway, E. and McKeown, K. (2020). Zero-Shot Stance Detection: A Dataset and Model using Generalized Topic Representations, in B. Webber, T. Cohn, Y. He and Y. Liu (eds), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Online, pp. 8913–8931. <https://aclanthology.org/2020.emnlp-main.717>.
- Baziotis, C., Pelekis, N. and Doukeridis, C. (2017). DataStories at SemEval-2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis, *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Association for Computational Linguistics, Vancouver, Canada, pp. 747–754. <https://aclanthology.org/S17-2126>.
- Bechini, A., Ducange, P., Marcelloni, F. and Renda, A. (2021). Stance analysis of twitter users: The case of the vaccination topic in italy, *IEEE Intelligent Systems* 36(5): 131–139. <https://doi.org/10.1109/MIS.2020.3044968>.
- Biber, D. and Finegan, E. (1988). Adverbial stance types in english, *Discourse Processes* 11(1): 1–34. <https://doi.org/10.1080/01638538809544689>.
- Cambria, E., Poria, S., Bajpai, R. and Schuller, B. (2016). SenticNet 4: A semantic resource for sentiment analysis based on conceptual primitives, pp. 2666–2677. Disponível em <https://aclanthology.org/C16-1251/>.
- Chen, W.-F. and Ku, L.-W. (2018). We like, we post: A joint user-post approach for facebook post stance labeling, *IEEE Transactions on Knowledge and Data Engineering* 30(10): 2013–2023. <https://doi.org/10.1109/TKDE.2018.281087>.

- Cotfas, L.-A., Delcea, C., Roxin, I., Ioanăș, C., Gherai, D. S. and Tajariol, F. (2021). The longest month: Analyzing covid-19 vaccination opinions dynamics from tweets in the month following the first vaccine announcement, *IEEE Access* **9**: 33203–33223. <https://doi.org/10.1109/ACCESS.2021.3059821>.
- 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. <https://doi.org/10.1016/j.eswa.2018.09.009>.
- Darwish, K., Magdy, W. and Zanoua, T. (2017). Improved stance prediction in a user similarity feature space, *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17*, Association for Computing Machinery, New York, NY, USA, p. 145–148. <https://doi.org/10.1145/3110025.3110112>.
- Dong, R., Sun, Y., Wang, L., Gu, Y. and Zhong, Y. (2017). Weakly-guided user stance prediction via joint modeling of content and social interaction, *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM '17*, Association for Computing Machinery, New York, NY, USA, p. 1249–1258. <https://doi.org/10.1145/3132847.3133020>.
- Han, X., Wang, J., Zhang, M. and Wang, X. (2020). Using social media to mine and analyze public opinion related to covid-19 in china, *International Journal of Environmental Research and Public Health* **17**(8). <https://www.mdpi.com/1660-4601/17/8/2788>.
- Hashemian, M., Rahimi, S. K. and Dastjerdi, A. B. (2023). An influences-adapted two-phase approach to stance detection in the diachronic perspective, *Expert Systems with Applications* **231**: 120773. <https://doi.org/10.1016/j.eswa.2023.120773>.
- Hisamitsu, S., Cho, S., Jin, H., Toyoda, M. and Yoshinaga, N. (2022). Diachronic analysis of users' stances on covid-19 vaccination in japan using twitter, *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 237–241. <https://doi.org/10.1109/ASONAM55673.2022.10068695>.
- Jia, P., Du, Y., Hu, J., Li, H., Li, X. and Chen, X. (2022). An improved bilstm approach for user stance detection based on external commonsense knowledge and environment information, *Applied sciences* **12**(21): 10968. <https://doi.org/10.3390/app122110968>.
- Küccük, D. and Can, F. (2020). Stance detection: A survey, *ACM Comput. Surv.* **53**(1). <https://doi.org/10.1145/3369026>.
- Küçük, Dilek and Can, Fazli (2020). Stance detection: A survey, *ACM Comput. Surv.* **53**(1). <https://doi.org/10.1145/3369026>.
- Lai, M., Patti, V., Ruffo, G. and Rosso, P. (2020). Brexit: Leave or remain? the role of user's community and diachronic evolution on stance detection, *arXiv.org*. <https://doi.org/10.3233/JIFS-179895>.
- McGregor, S. C. (2019). Social media as public opinion: How journalists use social media to represent public opinion, *Journalism* **20**(8): 1070–1086. <https://doi.org/10.1177/1464884919845458>.
- Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X. and Cherry, C. (2016). A dataset for detecting stance in tweets, in N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk and S. Piperidis (eds), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, European Language Resources Association (ELRA), Portorož, Slovenia, pp. 3945–3952. Disponível em <https://aclanthology.org/L16-1623/>.
- Newman, N. (2011). Mainstream media and the distribution of news in the age of social media.
- Qazi, A., Qazi, J., Naseer, K., Zeeshan, M., Hardaker, G., Maitama, J. and Haruna, K. (2020). Analyzing situational awareness through public opinion to predict adoption of social distancing amid pandemic covid-19, *Journal of Medical Virology* **92**. <https://doi.org/10.1002/jmv.25840>.
- Samih, Y. and Darwish, K. (2021). A few topical tweets are enough for effective user stance detection, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, Association for Computational Linguistics, Online, pp. 2637–2646. <https://aclanthology.org/2021.eacl-main.227>.
- Sridhar, D., Foulds, J., Huang, B., Getoor, L. and Walker, M. (2015). Joint models of disagreement and stance in online debate, in C. Zong and M. Strube (eds), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Beijing, China, pp. 116–125. <https://doi.org/10.3115/v1/P15-1012>.
- Stefanov, P., Darwish, K., Atanasov, A. and Nakov, P. (2020). Predicting the topical stance and political leaning of media using tweets, in D. Jurafsky, J. Chai, N. Schuster and J. Tetreault (eds), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, pp. 527–537. <https://aclanthology.org/2020.acl-main.50>.
- Zhang, H., Kwak, H., Gao, W. and An, J. (2023). Wearing masks implies refuting trump?: Towards target-specific user stance prediction across events in covid-19 and us election 2020, *Proceedings of the 15th ACM Web Science Conference 2023, WebSci '23*, Association for Computing Machinery, New York, NY, USA, p. 23–32. <https://doi.org/10.1145/3578503.3583606>.
- Zhu, L., He, Y. and Zhou, D. (2020). Neural opinion dynamics model for the prediction of user-level stance dynamics, *Information processing & management* **57**(2): 102031. <https://doi.org/10.1016/j.ipm.2019.03.010>.