

Detecção Automática de Fake News: Um Relato Agrupado de Dois Mapeamentos Sistemáticos Terciários

Daniel Oliveira de Freitas¹, Cassio Ceolin Junior¹
Guilherme Neri Bustamante Sá¹, Fábio Paulo Basso¹

¹Universidade Federal do Pampa (UNIPAMPA) – Alegrete, RS, Brasil

{danielodf2.aluno, cassioceolin.aluno}@unipampa.edu.br

{guilhermesa.aluno, fabiobasso}@unipampa.edu.br

Abstract. *The immeasurable growth of misinformation on social media and other platforms during pandemic situations such as COVID-19 can cause significant damage to physical and mental stability. To detect this false information, researchers have applied various techniques, including machine learning and deep learning, together with automatic data gathering techniques. This study aimed to systematically review, evaluate, and synthesize research articles that used different techniques for the automated detection of fake news in the scope of COVID-19. We identify publicly available datasets and review data processing, feature extraction, and classification techniques. We also discuss challenges and limitations found in the literature.*

Resumo. *O crescimento da desinformação em mídias sociais e outras plataformas durante pandemias como a COVID-19 pode causar danos significativos à estabilidade física e mental. Para detectar essas informações falsas, pesquisas têm aplicado técnicas de machine learning e deep learning, aliadas à coleta automática de dados. Este estudo tem como objetivo revisar, avaliar e sintetizar, de forma sistemática, artigos que empregam diferentes técnicas para a detecção automática de fake news no contexto da COVID-19. São identificados conjuntos de dados públicos e revisadas etapas de processamento, extração de características e classificação, além de desafios e limitações relatados na literatura.*

1. Introdução

O volume de informações digitais cresceu de forma exponencial nos últimos anos, intensificando a propagação de conteúdos falsos ou enganosos online. Identificar conteúdo confiável na Web é desafiador, motivando o uso de mineração de dados, PLN e mineração de texto para combater a desinformação e as *fake news* [Thota et al. 2018, Fernandes et al. 2023]. Pesquisas que integram inteligência artificial, *machine learning*, *deep learning* e análise de grandes volumes de dados expandiram aplicações que vão desde a análise de sentimentos até a classificação automática de notícias falsas, melhorando a confiabilidade das informações disponíveis [Torregrosa et al. 2022, Ignaczak et al. 2021].

Apesar dos avanços, ainda há necessidade de abordagens mais precisas, reproduzíveis e validadas em larga escala, apoiadas em conjuntos de dados públicos e protocolos padronizados [Wohlin et al. 2012, Kitchenham and Charters 2007]. Este artigo apresenta

uma síntese terciária: dois *mapeamentos sistemáticos da literatura (MSL)* independentes foram conduzidos e posteriormente agrupados em um único relato para consolidar o estado da arte na detecção automática de fake news durante a pandemia de COVID-19.

A documentação completa do protocolo, planilhas de triagem e listas integrais de estudos está disponível como material suplementar em repositório público (Zenodo)¹, facilitando verificação e reuso.

2. Mapeamento Sistemático Terciário

O estudo terciário foi estruturado a partir da integração de dois Mapeamentos Sistemáticos da Literatura (MSL) conduzidos de forma independente, ambos com o objetivo de mapear as principais técnicas e métodos aplicados à detecção automática de notícias falsas. Cada revisão seguiu protocolos padronizados compostos por planejamento, execução e análise dos resultados, conforme orientações de Kitchenham [Kitchenham and Charters 2007] e Wohlin [Wohlin et al. 2012].

Planejamento. As etapas contemplaram a formulação de questões de pesquisa, definição das strings de busca, escolha das bases científicas, além do estabelecimento dos critérios de inclusão, exclusão e qualidade. Esse planejamento foi realizado em dupla, de forma independente, e depois conciliado em reuniões para alinhamento metodológico, o que garante maior confiabilidade ao processo.

Execução. As buscas principais foram realizadas entre 2019 e 2023 nas bases IEEE Xplore, Scopus, ACM Digital Library, ScienceDirect e SpringerLink, com ajustes de sintaxe por base. Para ampliar a cobertura de PLN, incluímos a ACL Anthology, e, para contemplar produção nacional, uma varredura nos anais de eventos da SBC (p. ex., BRACIS, STIL, KDMiLe). Adicionalmente, efetuamos uma varredura complementar focal (2024–2025) para capturar avanços recentes sem alterar o protocolo central; os itens e filtros usados estão descritos no material suplementar (Zenodo).

Seleção de estudos. O primeiro mapeamento identificou 152 artigos, dos quais 47 foram incluídos após critérios de triagem. O segundo encontrou 136 artigos, resultando em 39 incluídos após a aplicação de critérios de qualidade. No total, 86 estudos compuseram a síntese terciária. A triagem foi feita em duas etapas: leitura de título/resumo e, posteriormente, leitura completa dos artigos. Divergências foram resolvidas por consenso entre os pesquisadores.

Questões de pesquisa. Ambas as revisões buscaram responder:

- Quais são as características do domínio de fake news em diferentes contextos?
- Quais mecanismos de mineração e pré-processamento são mais empregados?
- Quais técnicas de classificação e extração de atributos predominam?
- Quais as principais ameaças à validade relatadas?
- Quais tendências de pesquisa emergem no campo da detecção automática de fake news?

Crítérios de inclusão e exclusão. Foram incluídos estudos secundários e terciários que tratassem de mineração de fake news com técnicas de Web Scraping, Web Crawling ou Web/Text Mining. Foram excluídos artigos com menos de 6 páginas, publicados antes de 2014, não escritos em inglês, duplicados ou sem acesso via portal da CAPES.

¹<https://doi.org/10.5281/zenodo.16934158>

Cr terios de qualidade. Cada artigo selecionado foi avaliado de acordo com cr terios de clareza metodol gica, descri  o da string de busca, n mero de estudos identificados e justificativa para exclus es. Apenas estudos avaliados como de boa ou  tima qualidade foram inclu dos na s ntese.

Integra  o dos mapeamentos. Ap s a sele  o individual, os resultados de ambos os MSL foram cruzados. Essa integra  o permitiu comparar achados, identificar complementaridades e mitigar vieses decorrentes de uma  nica revis o. Assim, a abordagem terci ria garante maior abrang ncia e robustez   s ntese produzida.

3. Resultados

Conjuntos de dados: A maioria dos estudos usa corpora de redes sociais e portais de not cias digitais. Apesar da abund ncia, os conjuntos s o heterog neos em curadoria, balanceamento e dom nio, dificultando compara  es diretas entre algoritmos [Thompson et al. 2022, Mridha et al. 2021].

Pr -processamento e atributos: Foram relatadas t cnicas tradicionais de limpeza textual, tokeniza  o, remo  o de *stopwords* e normaliza  o. Representa  es cl ssicas como Bag-of-Words (BoW) e TF-IDF seguem competitivas para modelos lineares, mas observa-se avan o na ado  o de representa  es contextuais, como embeddings baseados em Word2Vec, FastText, GloVe e, mais recentemente, transformadores (BERT e variantes) [Torregrosa et al. 2022, Fernandes et al. 2023].

Classifica  o: Regress o Log stica, SVM e Random Forest aparecem com frequ ncia e t m bom desempenho com atributos bem constru dos. Resultados superiores emergem com redes profundas (CNN, LSTM) e transformadores; abordagens multimodais (texto+imagem+metadados) s o promissoras [Ignaczak et al. 2021, Thota et al. 2018].

M tricas: Acur cia, precis o, revoca  o, F1 e AUC foram as mais relatadas, por m sem padroniza  o e raramente com testes de signific ncia, o que dificulta compara  es e reduz confian a.

S ntese das duas revis es: A primeira revis o enfatizou m todos de PLN aplicados   COVID-19 e a an lise de sentimentos em redes sociais, revelando a vulnerabilidade da sa de p blica   desinforma  o. A segunda concentrou-se em t cnicas de minera  o de dados, com destaque para associa  o, agrupamento e classifica  o, identificando lacunas metodol gicas e refor ando a necessidade de dados abertos. Em conjunto, as revis es mostram que, apesar da evolu  o das t cnicas, ainda h  fragmenta  o na  rea e baixa reproduzibilidade entre estudos.

An lise cr tica: Os resultados apontam tr s grandes desafios: (i) falta de padroniza  o na avalia  o; (ii) dificuldade de generaliza  o entre dom nios; (iii) escassez de recursos de interpretabilidade. Essas limita  es refor am a import ncia de esfor os colaborativos para consolida  o de bases de dados p blicas e protocolos de avalia  o mais consistentes.

Tend ncias: Destacam-se como frentes de pesquisa: (i) uso de transformadores e aprendizado profundo robusto; (ii) adapta  o de dom nio e robustez temporal; (iii) interpretabilidade e explicabilidade dos modelos; (iv) integra  o de t cnicas fracas ou ativas para reduzir custos de anota  o; (v) an lise multimodal e multilingu stica em larga escala [Torregrosa et al. 2022, Fernandes et al. 2023].

Em comparação a revisões e mapeamentos anteriores, nossa síntese agrega dois MSL independentes com foco método-cêntrico (pré-processamento, representações, classificadores e práticas de avaliação), mostrando convergência na competitividade de SVM/BoW e no avanço de transformadores. Persistem lacunas recorrentes observadas também na literatura: padronização frágil de métricas, pouca análise de significância, vies de domínio e reprodutibilidade limitada.

4. Conclusão

Este artigo sintetizou dois mapeamentos sistemáticos independentes sobre detecção automática de fake news, reunindo evidências sobre conjuntos de dados, técnicas de pré-processamento, algoritmos de classificação e desafios recorrentes. Os resultados evidenciam tanto a consolidação de pipelines metodológicos quanto a heterogeneidade de práticas, que dificulta comparações e compromete a reprodutibilidade.

Conclui-se que, embora haja avanços relevantes, ainda são necessárias iniciativas colaborativas para construção de bases públicas amplas, padronização de protocolos de avaliação e integração de recursos de explicabilidade. O fortalecimento dessas direções poderá elevar a confiabilidade e a aplicabilidade dos sistemas de detecção de fake news, ampliando seu impacto social e científico. Como síntese, entregamos uma consolidação *método-cêntrica* com cobertura ampliada (ACL Anthology e eventos SBC) e material suplementar no Zenodo para reprodutibilidade.

Referências

- Fernandes, E., Moro, S., and Cortez, P. (2023). Data science, machine learning and big data in digital journalism. *Expert Systems with Applications*.
- Ignaczak, L. et al. (2021). Text mining in cybersecurity: A systematic literature review. *ACM Computing Surveys*.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, EBSE Technical Report, Keele University and University of Durham.
- Mridha, M. F. et al. (2021). A comprehensive survey on fake news detection: Approaches, challenges and opportunities. *Journal of Information and Telecommunication*, 5(2):1–27.
- Thompson, R. C., Joseph, S., and Adeliyi, T. T. (2022). A systematic literature review and meta-analysis of studies on online fake news detection. *Information*, 13(11).
- Thota, A., Tilak, P., Ahluwalia, S., and Lohia, N. (2018). Fake news detection: A deep learning approach. *SMU Data Science Review*, 1(3). Article 10.
- Torregrosa, J., Bello-Orgaz, G., Martínez-Cámara, E., Ser, J. D., and Camacho, D. (2022). A survey on extremism analysis using natural language processing: definitions, literature review, trends and challenges. *Journal of Ambient Intelligence and Humanized Computing*.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.