Reconhecimento e Compartilhamento de Padrões Textuais em Notícias Falsas

Leonardo Emerson A. Alves[1], Jonice Oliveira[1] (orientadora), Sirius Thadeu F. da Silva[1] (coorientador)

[1]Universidade Federal do Rio de Janeiro (UFRJ)
– Rio de Janeiro – RJ – Brasil

{leonardoemerson,sirius}@ufrj.br, jonice@ic.ufrj.br

Abstract. *This research proposes a methodology for characterizing, describing the evolution, and identifying patterns of fake news written in Brazilian Portuguese. The fake news characterization was done through textual analysis of news collected from 2013 to 2021, using natural language processing and topic modeling techniques. The main difference between this research and the others is the use of an unbalanced corpus. Thus, an approach focused on unsupervised machine learning was defined using the modeling coherence metric to obtain optimized results.*

Resumo. *Esta pesquisa propõe uma metodologia para a caracterização, descrição da evolução e identificação de padrões de notícias falsas escritas em português-brasileiro. A caracterização das notícias falsas é realizada por meio da análise textual de notícias coletadas entre 2013 e 2021, com o uso de técnicas de processamento de linguagem natural e modelagem de tópicos. O principal diferencial dessa pesquisa consiste na abordagem de um corpus não-balanceado. Dessa forma, foi definida uma abordagem focada em aprendizado de máquina não-supervisionado com a utilização da métrica de coerência das modelagens para obter a otimização dos resultados.*

1. Introdução

O aumento do número de usuários nas mídias sociais transformou a interação humana. Com o crescimento do número de usuários ativos e suas interações online nas redes, a velocidade do fluxo de informações e conteúdos que atingem diversos públicos cresceu nos últimos anos. Nesse contexto ocorre a propagação de informações falsas [Vosoughi et al. 2018] [Guo et al. 2019] [Su et al. 2020] que, por sua vez, trazem inconsistências à integridade dos meios de informação e atingem todos os setores da sociedade, prejudicando tanto a tomada de decisão dos órgãos governamentais quanto o cotidiano da sociedade.

Os cidadãos que não verificam com precisão a origem das notícias consumidas podem ter suas opiniões moldadas ou reforçadas por conteúdos falsos [Gelfert 2021]. Dessa forma, ocorrem mudanças comportamentais nesses indivíduos [Bastick 2021], o que traz impactos em processos sociais, tais como: administração de crises de saúde, eleições, consultas públicas de opinião, violações de direitos humanos e ameaças ao Estado de Direito [Colomina et al. 2021].

Bodaghi e Oliveira [2022] mostram que os comportamentos associados com a disseminação de desinformação e de conteúdos verídicos são distintos. Sendo assim, soluções computacionais precisam estar preparadas para a rápida identificação, visando uma antecipação à propagação em larga escala. A maior parte das soluções computacionais, baseadas em aprendizado de máquina, precisam de grandes bases de dados para reconhecer padrões de diferentes categorias (por exemplo, conteúdo falso,

conteúdo verdadeiro). Para um bom aprendizado, as categorias precisam estar homogeneamente representadas, com um número de documentos similar representando cada categoria (o que chamamos de bases balanceadas). Neste ponto, surge um primeiro grande desafio: No cenário de identificação da desinformação, onde o tempo é um fator determinante, muitas vezes não há tempo hábil para a criação de bases balanceadas onde os algoritmos podem identificar os padrões.

A presente pesquisa propõe uma metodologia computacional para combater a desinformação, com foco no estudo de um *corpus* não-balanceado de notícias com texto em língua portuguesa denominado FAKEPEDIA [Charles et al. 2022]. Neste trabalho foram investigados o avanço temporal das notícias falsas e identificados os principais padrões textuais e suas influências na construção das mesmas. Para que outras soluções colaborativas possam ser criadas para o combate à desinformação, os padrões encontrados foram representados através de dicionários de tópicos que caracterizam as notícias estudadas de acordo com os períodos temporais de disseminação. Um dos objetivos específicos do trabalho consistiu no teste da hipótese de que os períodos temporais, as categorias temáticas e as entidades nomeadas presentes nas notícias falsas são aspectos importantes na identificação de padrões nos textos dessas notícias.

2. Revisão de Literatura

Os principais trabalhos analisados na revisão de literatura estão descritos na Tabela 1.

Tabela 1 - Comparação com os principais trabalhos (MD-Modelagem de Tópicos, EN - Entidades Nomeadas, AS-Análise de Sentimento, CLT - Características Linguísticas e Textuais, BB - Base Balanceada?, PT-BR - Texto em Português-Brasileiro)

Referência	MD	EN	AS	CLT	BB	Corpus	PT-BR
[Melo e Figueiredo 2021]	X	X	X		Não	Notícias e Mídias Sociais	X
[Pérez-Rosas et al. 2017]				X	Sim	Notícias	
[Reis 2020]				X	Sim	Notícias e Mídias Sociais	X
[Monteiro et al. 2018]				X	Sim	Notícias	X
[Newman et al. 2006]	X	X			Não	Notícias	
[Pritzkau 2022]	X			X	Sim	Notícias	
[Nwankwo et al. 2020]	X		X		Não	Mídias Sociais	

Ao observar os trabalhos correlatos, é possível identificar uma lacuna no estudo e caracterização de conjuntos de dados não-balanceados contendo apenas notícias falsas com o intuito de identificar os padrões temporais, categóricos e das entidades nomeadas presentes nos textos de desinformação.

3. Abordagem voltada a conjuntos desbalanceados

A abordagem proposta neste trabalho é composta por três processos: tratamento e aperfeiçoamento do *corpus* estudado (seção 3.1), análise temporal das notícias (seção 3.2) e modelagem de tópicos (seção 3.3). Tais seções estão dispostas nos capítulos 4, 5 e 6 de [Alves 2023]. Para cada um desses processos foram construídos artefatos de código contendo todas as etapas inerentes a descrição e execução da metodologia proposta.

3.1. Tratamento e Aperfeiçoamento do Corpus

O tratamento e aperfeiçoamento do *corpus* foi realizado com a necessidade de remover frações de dados textuais não relevantes para a pesquisa (*e-mails*, anúncios, telefones, sinais de pontuação e outros). O principal objetivo da limpeza consistiu na obtenção de um conjunto de dados textuais sem ruídos para os posteriores processos de análise e modelagem de tópicos que necessitam de baixo número de redundâncias nos dados textuais estudados.

O processo de tratamento foi realizado em três etapas: i) Identificação dos Padrões, ii) Construção de Expressões Regulares de Captura e iii) Tratamento Específico. Na primeira etapa foram realizadas consultas para identificar padrões textuais que não trouxessem significado para a análise e modelagem de tópicos. Na segunda etapa foram construídas expressões regulares para capturar e tratar cada um desses padrões. Na terceira etapa foram realizados os tratamentos de acordo com cada padrão. Adicionalmente, durante a pesquisa, foi identificada a necessidade de obter dados adicionais associados às notícias estudadas em decorrência da realização de análises temporais e categóricas. Desse modo, foram adicionadas ao *corpus*, tanto as datas de publicação, quanto as categorias temáticas dessas notícias aplicando a técnica de *Web Scraping* no site de origem das notícias.

3.2. Análise Temporal das Notícias

A análise temporal das notícias considerou diversos aspectos e características das notícias para a compreensão dos padrões de escrita nos conteúdos de desinformação. Os principais aspectos analisados foram:

- Frequência de palavras - buscando obter uma visão geral dos termos mais recorrentes nos textos estudados.
- Categorias temáticas - identificação das categorias das notícias falsas publicadas.
- Análise temporal da quantidade de publicações de notícias falsas - visando compreender quais foram os períodos com maior disseminação desse tipo de conteúdo.
- Análise temporal das categorias temáticas mais recorrentes nas notícias disseminadas - obtendo uma visão geral dos assuntos mais abordados nesse tipo de conteúdo ao longo do tempo.
- Análise temporal das entidades - foram estudadas as entidades nomeadas com maior recorrência de citações ao longo dos textos.
- Análise temporal das entidades por categoria - as entidades foram estudadas temporalmente por meio das categorias as quais foram inseridas nas veiculações de conteúdo desinformacional.

3.3. Modelagem de Tópicos

Como visto na Tabela 1, grande parte das soluções utilizam como recursos, os tópicos e características textuais (como entidades nomeadas) para o estudo computacional de notícias falsas. Além de realizar a análise de padrões, um dos objetivos deste trabalho foi colaborar com soluções nacionais - focadas em língua portuguesa - que necessitem de tópicos como mecanismos de entrada. Para isto, foi criado um dicionário de tópicos, que pode ser atualizado e reutilizado por outras soluções, em diferentes cenários. O dicionário de tópicos foi construído a partir de duas técnicas para modelagem de tópicos: Latent Dirichlet Allocation (LDA) e Latent Semantic Analysis (LSA). Foram gerados três dicionários de tópicos para cada técnica aplicada, totalizando seis dicionários de tópicos distintos. Os três dicionários para cada modelagem desempenharam as seguintes funções: caracterizar os tópicos do *corpus* por período de tempo (ano de publicação da notícia); descrever os tópicos do *corpus* por categoria (assunto da notícia publicada); e identificar os tópicos utilizando as amostras existentes no *corpus*, onde foi realizada uma modelagem geral das notícias falsas.

Foi realizada uma etapa prévia de seleção e preparação dos dados, separando as porções de dados de acordo com os períodos temporais e categorias temáticas. O triênio 2013-2015 foi separado como um conjunto único, devido ao fato desse período possuir uma quantidade menor de notícias em relação aos outros anos. Os demais anos foram separados individualmente. As categorias temáticas foram separadas em: política, assuntos nacionais, assuntos internacionais, religião, ciência, entretenimento, esporte, saúde e tecnologia.

Um conjunto de etapas para o aprimoramento das técnicas foi desenvolvido e detalhado em [Alves 2023]. Após a organização dos conjuntos, os documentos foram *tokenizados* utilizando a biblioteca para modelagem de tópicos *gensim* [Řehůřek e Sojka 2010]. As acentuações foram removidas e foi realizado o tratamento em relação aos bigramas e trigramas presentes. Adicionalmente, foi realizada a lematização dos textos, visto que beneficia a modelagem por meio do método LDA, onde sua utilização otimiza os resultados relativos à modelagem de tópicos em linguagens morfologicamente ricas (o que inclui a língua portuguesa) [May et al. 2019], bem como beneficia o método LSA, visto que ao processar um *corpus* lematizado, o método LSA tende a ser computacionalmente mais eficiente [Zipitria et al. 2006].

O método LDA foi aplicado com uma modelagem inicial de 10 tópicos para cada divisão do *corpus,* para a obtenção de um valor inicial da métrica de coerência. Essa métrica indica quantitativamente a ocorrência mútua entre os termos associados ao tópico estudado que, por sua vez, se traduz em um valor que indica o pertencimento dos termos a um mesmo tema. [Röder et al. 2015] mostra que a métrica de coerência tem correlação com resultados advindos de observações humanas. Posteriormente, foi realizada uma modelagem de tópicos por meio da otimização da métrica de coerência, onde foram utilizados hiperparâmetros do modelo de Dirichlet, tais como: o número de tópicos (K), a densidade de tópicos em cada documento (α) e a densidade de palavras em cada tópico (β). O ajuste foi realizado com a utilização de testes em sequência para cada hiperparâmetro, considerando dois conjuntos de validação do *corpus*. Um dos conjuntos de validação consiste em 75% dos dados e o outro consiste em 100% dos

dados do *corpus* utilizado.

Na modelagem de tópicos com o uso do método LSA, o número *K* de tópicos foi o principal hiperparâmetro considerado para o treinamento. Dessa forma, foram treinados 17 modelos distintos para cada conjunto de documentos e, para cada modelo, foram testados valores para *K* no intervalo [4, 20]. O valor de coerência de cada modelo foi obtido e o resultado final considerado para cada conjunto foi o modelo com o valor de tópicos que resultou em máxima coerência.

4. Experimentos e Resultados

Nas análises que compreendem os períodos temporais e as categorias temáticas do *corpus* foram identificadas relações entre acontecimentos históricos e seus períodos de propagação e temáticas mais frequentes entre as notícias falsas. Acontecimentos históricos implicam na alta construção de notícias falsas [Alves 2023], como pôde ser visto ao considerarmos o período de surgimento da pandemia de COVID-19 e sua relação com o aumento no número de notícias falsas, bem como o aumento de notícias da categoria relacionada à saúde. Outro acontecimento histórico que destaca essa implicação pode ser notado ao considerar o período de eleições de 2018. Foi identificado que a maior parte das notícias presentes no *corpus* esteve associada com assuntos nacionais e política, sendo que as notícias falsas relacionadas com o tema sobre saúde obtiveram grande proporção a partir do momento da deflagração da pandemia de COVID-19. Adicionalmente, a análise voltada para as entidades mais recorrentes reafirmou o conteúdo dos assuntos tratados, onde sempre se destacaram agentes e instituições relacionadas com os assuntos mais abordados nos conteúdos falsos.

Tabela 2 - Coerência ótima das abordagens LDA e LSA por período temporal

Período Temporal	LDA	LSA
2013-2015	0,5511	0,6807
2016	0,4952	0,4814
2017	0,4802	0,5397
2018	0,5172	0,6170
2019	0,4735	0,4893
2020	0,4929	0,4990
2021	0,5108	0,4759

Tabela 3 - Coerência ótima das abordagens LDA e LSA por categoria temática

Categoria	LDA	LSA
Política	0,6615	0,4335
Assuntos Nacionais (Brasil)	0,6033	0,4979
Saúde	0,4963	0,4476
Entretenimento	0,6090	0,4520
Tecnologia	0,5481	0,6435
Ciência	0,5096	0,7869
Esporte	0,5800	0,6926
Assuntos internacionais (Mundo)	0,5115	0,5302
Religião	0,5738	0,6038

Considerando os resultados advindos da modelagem de tópicos é possível observar nas tabelas 2 e 3 que o método LDA tende a apresentar melhores resultados em conjuntos de dados com maiores quantidades de documentos, enquanto que o método LSA apresentou melhores resultados em conjuntos com menor quantidade de documentos, e particularmente onde os documentos possuem temáticas mais dispersas.

Observando os ganhos de coerência obtidos após a otimização das modelagens de tópicos com a abordagem de divisão (temporal e categórica) proposta, foi possível constatar que considerar o período temporal, a categoria (temática) e as entidades nomeadas (análise) como fatores determinantes para a identificação de padrões em notícias falsas trouxe benefícios para o processo. Isso comprova nossa hipótese, visto que considerando esses fatores no processo de divisão dos dados para modelagem de tópicos foram obtidos ganhos médios de coerência consideráveis, fornecendo modelagens com melhores resultados. Adicionalmente, foi possível observar que a análise das entidades nomeadas facilitou a interpretação do contexto em relação aos tópicos gerados no processo.

5. Contribuições e Trabalhos Futuros

Como **contribuições científicas** na área de Sistemas Colaborativos, podemos ressaltar que este trabalho lida com "o desafio da diferença de credibilidade entre os participantes de sistemas colaborativos e com a tarefa de identificar notícias falsas e seus disseminadores", mencionado por Ponciano e Andrade [2018]. Em Sistemas Colaborativos, este TCC aborda a comunicação (disseminação de desinformação), cooperação (redes sociais) e coordenação (auxílio no combate às fake news). Foi desenvolvida no presente trabalho, uma metodologia para análise de padrões textuais de bases não-balanceadas de notícias falsas. Muitos dos trabalhos que foram estudados na revisão de literatura abordam o problema da desinformação utilizando conjuntos de documentos balanceados e, principalmente, com textos em língua inglesa. O *framework* desenvolvido pode ser utilizado em qualquer cenário que justifique a análise de desinformação, principalmente quando há um número desproporcional (base desbalanceada) de classes de documentos. Esta pesquisa se enquadra na área de Computação Social, tangenciando temas como: privacidade, segurança e suas implicações. O presente trabalho gerou um artigo [Alves et al., 2023] que foi selecionado e apresentado na Semana de Integração Acadêmica da UFRJ (SIAC/UFRJ).

Entre as principais **contribuições tecnológicas**, é possível listar: o framework e o dicionário construído a partir de duas técnicas tradicionais de modelagem de tópicos com uma estratégia de otimização de resultados, disponíveis em um repositório[1] digital. No **campo social** destacamos o apoio ao combate à desinformação, identificando as temáticas e termos mais utilizados em *fake news*. Os artefatos produzidos nesta pesquisa podem auxiliar nas estratégias públicas de comunicação.

Existem vários desdobramentos que vislumbramos como trabalhos futuros, mas destacamos o "human-in-the-loop", inserindo a colaboração humana nas inferências. Com isto podemos aproveitar informações contextuais (inclusive, características regionais de linguagem) e aperfeiçoar os mecanismos de detecção de desinformação.

[1] Disponível em: https://github.com/leonardoemerson/TCC-Leonardo-Emerson

Referências

Alves, L. E. A. (2023). Caracterização, evolução e identificação de padrões em notícias falsas: uma abordagem voltada à modelagem de tópicos. Trabalho de Conclusão de Curso. (Graduação em Ciência da Computação) - Universidade Federal do Rio de Janeiro. Disponível em: https://pantheon.ufrj.br/handle/11422/21240. Acessado em 06/01/2024.

Alves, L.E.A et al. (2023). Caracterização, evolução e identificação de padrões em notícias falsas via modelagem de tópicos (id: 2845). Semana de Integração Acadêmica da UFRJ (12.:2023): CCMN.

Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, v. 116, p. 106633.

Bodaghi, A, and Oliveira, J. (2022) The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter. Expert Systems with Applications 189 : 116110.

Charles, A., Ruback, L. and Oliveira, J. (2022). Fakepedia Corpus: A Flexible Fake News Corpus in Portuguese. International Conference on Computational Processing of the Portuguese Language (pp. 37-45). Springer International Publishing.

Colomina, C., Margalef, H. S. and Youngs, R. (2021). The impact of disinformation on democratic processes and human rights in the world. Brussels: European Parliament.

Gelfert, A. (2021). Fake News, False Beliefs, and the Fallible Art of Knowledge Maintenance. In: Bernecker, S.; Flowerree, A. K.; Grundmann, T.[Eds.]. *The Epistemology of Fake News*. Oxford University Press. p. 0.

Guo, B., Ding, Y., Yueheng, S., Ma, S. and Li, K. (2019). *The Mass, Fake News, and Cognition Security*.

May, C., Cotterell, R. and Van Durme, B. (2019). An Analysis of Lemmatization on Topic Models of Morphologically Rich Language. arXiv. Disponível em http://arxiv.org/abs/1608.03995. Acessado em 11/01/2024.

Melo, Tiago de; Figueiredo, Carlos M. S. Comparing News Articles and Tweets About COVID-19 in Brazil: Sentiment Analysis and Topic Modeling Approach. JMIR Public Health and Surveillance, v. 7, n. 2, p. e24585, 2021.

Monteiro, R. A., Santos, R. L. S., Pardo, T. A. S., et al. (2018). Contributions to the Study of Fake News in Portuguese: New Corpus and Automatic Detection Results. [A. Villavicencio, V. Moreira, A. Abad, et al., Eds.]In *Computational Processing of the Portuguese Language*. , Lecture Notes in Computer Science. Springer International Publishing.

Newman, D., Chemudugunta, C., Smyth, P. and Steyvers, M. (2006). Analyzing Entities and Topics in News Articles Using Statistical Topic Models. [S. Mehrotra, D. D. Zeng, H. Chen, B. Thuraisingham, & F.-Y. Wang, Eds.]In *Intelligence and Security Informatics*. , Lecture Notes in Computer Science. Springer.

Nwankwo, E., Okolo, C., Habonimana, C. and Beach, C.-L. (2020). Topic Modeling Approaches for Understanding COVID-19 Misinformation Spread in Sub-Saharan Africa.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A. and Mihalcea, R. (2017). Automatic Detection of Fake News. arXiv. Disponível em http://arxiv.org/abs/1708.07104. Acessado em 11/01/2024.

Ponciano, L. and Andrade, N. (2018). Perspectivas em Computação Social. Computação Brasil, Raquel Prates and Thais Castro (Eds.) 36. p. 30–33.

Pritzkau, A., Blanc, O., Geierhos, M. and Schade, U. (2022). NLytics at CheckThat! 2022: Hierarchical multi-class fake news detection of news articles exploiting the topic structure.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora.

Reis, J. C. S. and Benevenuto, F. (2021). Towards Automatic Fake News Detection in Digital Platforms: Properties, Limitations, and Applications. In *Anais do Concurso de Teses e Dissertações (CTD)*. SBC. Disponível em https://sol.sbc.org.br/index.php/ctd/article/view/15754. Acessado em 11/01/2024.

Röder, M., Both, A. and Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. , WSDM '15. Association for Computing Machinery. Disponível em https://doi.org/10.1145/2684822.2685324. Acessado em 11/01/2024.

Su, Q., Wan, M., Liu, X. and Huang, C.-R. (2020). Motivations, Methods and Metrics of Misinformation Detection: An NLP Perspective. *Natural Language Processing Research*, v. 1, n. 1–2, p. 1–13.

Vosoughi, S., Roy, D. and Aral, S. (2018). The spread of true and false news online. *Science*, v. 359, n. 6380, p. 1146–1151.

Zipitria, I., Arruarte, A. and Elorriaga, J. A. (2006). Observing Lemmatization Effect in LSA Coherence and Comprehension Grading of Learner Summaries. [M. Ikeda, K. D. Ashley, & T.-W. Chan, Eds.]In *Intelligent Tutoring Systems*. , Lecture Notes in Computer Science. Springer.