

DESENVOLVIMENTO DE FERRAMENTAS PARA AUXILIAR NA IDENTIFICAÇÃO DE NOTÍCIAS FALSAS

Aluna: Giovanna Pedrino Belasco

Colaborador: Jenifer Galvão Moraes

Orientador: Luciano Digiampietri

Universidade de São Paulo

giovannabelasco@usp.br

Objetivos

Com o aumento do fluxo de informação e a falta de tempo das pessoas, é difícil averiguar quais notícias são verdadeiras. Apesar de haver recomendações para evitar o espalhamento de *fake news* [1], é difícil a identificação por seres humanos.

Para isso, há diversos procedimentos e classificadores já desenvolvidos para a detecção automática de notícias falsas. Em vista disso, os objetivos principais dessa pesquisa são: construção de uma base de dados contendo notícias falsas e verdadeiras em português e a análise da acurácia de classificadores já existentes para a classificação das notícias da base de dados fact.br [2].

Métodos e Procedimentos

Para o desenvolvimento da pesquisa, a metodologia consistiu na realização de um estudo de artigos científicos relacionados ao tema, seguido de um estudo da linguagem de programação Python, pois essa foi utilizada na implementação da solução proposta, devido à grande disponibilidade de bibliotecas que facilitam a extração de características de dados textuais, bem como dos diversos classificadores disponíveis.

Posteriormente, foi construída uma base de dados própria contendo notícias em português,

retiradas de sites jornalísticos e classificadas por estes entre falsas e verdadeiras.

Por fim, foram utilizadas as notícias verdadeiras e falsas da base de dados factck.br para análise da acurácia dos classificadores.

Resultados

O conjunto de dados efetivamente utilizado é desbalanceado, contendo 88,7% de notícias falsas, assim, todo classificador que classifique todas as notícias como falsas terá uma acurácia de 88,7%.

Os classificadores utilizados foram: Dummy que classifica todas as notícias como pertencendo à classe majoritária (utilizado como *baseline*); um classificador baseado em regressão logística; um classificador do tipo SVM, uma rede neural MLP; e um classificador bayesiano (*Multinomial Naive Bayes*). As notícias foram representadas utilizando n-gramas de palavras, com n variando de 1 a 3, bem como a combinação de uni-bi-trigramas. A ponderação dos n-gramas foi feita por contagem e pelo uso do valor de TF-IDF. Por fim, verificou-se três abordagens relacionadas ao número de atributos: utilizar todos, selecionar os 300 principais atributos (utilizando o seletor *KBest*) e projetar os atributos em 300 novas dimensões, utilizando análise de componentes principais (PCA).

A Tabela 1 apresenta os melhores resultados para cada classificador (resultados obtidos pela

média dos resultados utilizando validação cruzada com 3 subconjuntos [*3-fold cross-validation*]). Destaca-se que o classificador baseado na rede neural MLP atingiu a melhor acurácia (de 89,2%), bem como o melhor resultado para a medida F1 (0,528). Estes resultados foram obtidos utilizando-se trigramas de palavras ponderados de acordo com seu valor TF-IDF e sem nenhuma estratégia de redução de dimensionalidade. Já o melhor resultado para a medida área sob a curva ROC (AUC) foi obtido pelo classificador SVM utilizando unigramas ponderados pelo valor TF-IDF.

Tabela 1: Resultados dos classificadores

Classificador	Modelo de Representação	Ponderação	Redução de Dimensionalidade	Acurácia	F1-macro	AUC
Dummy				0,887	0,470	0,500
MLP	trigrama	TF-IDF	nenhuma	0,892	0,528	0,720
Multinomial NB	trigrama	Contagem	nenhuma	0,890	0,509	0,461
RegLog	trigrama	TF-IDF	nenhuma	0,884	0,512	0,718
SVM	unigrama	TF-IDF	nenhuma	0,887	0,47	0,766

O conjunto de dados construído no presente projeto, em parceria com a aluna Jenifer Galvão Moares pode ser encontrado no seguinte link: <https://www.each.usp.br/digiampietri/dados/FakeNewsDataset20220505.xlsx>

Conclusões

Este projeto teve por objetivo iniciar uma pesquisa na classificação automática de notícias falsas em português.

Apesar dos resultados superarem o *baseline* utilizado, observa-se que há ainda um grande desafio em se identificar com 100% de precisão se uma notícia é falsa ou não e, provavelmente, este desafio só será superado utilizando bases de dados externas e não apenas o conteúdo textual da notícia. Destaca-se que o conjunto de dados utilizado não era balanceado, possuindo mais de 88% de notícias falsas. Já a base de dados construída neste projeto, cujos resultados não constam neste resumo, é balanceada (contendo 50% de notícias verdadeiras e 50% de notícias falsas), totalizando 104 notícias no

total. Essa diferença de balanceamento pode implicar em diferenças significativas em algumas medidas de desempenho.

Como trabalhos futuros, além da realização de testes em diferentes conjuntos de dados, pretende-se explorar a extração de outras características textuais, por exemplo, utilizando *word embeddings*, transformadores, extraindo características sintáticas e semânticas.

Bibliografia

- Hunt, Elle (2016). What is fake news? How to spot it and what you can do to stop it. **The Guardian**. Disponível em: <https://www.theguardian.com/media/2016/dec/18/what-is-fake-news-pizzagate>
- Moreno, João; Bressan, Graça. FACTCK.BR: A New Dataset to Study Fake News. *In: Simpósio Brasileiro de Sistemas Multimídia e Web (WEBMEDIA)*, 25. , 2019, Rio de Janeiro. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2019. p. 525-527.
- Zhou, Xinyi; Zafarani, Reza (2019). Fake News Detection: An Interdisciplinary Research. *In Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*, Ling Liu and Ryen White (Eds.). ACM, New York, NY, USA, 1292-1292. DOI: <https://doi.org/10.1145/3308560.3316476>
- Resende, Gustavo; Messias, Johnnatan; Silva, Márcio; Almeida, Jussara; Vasconcelos, Marisa; Benevenuto, Fabrício (2018). A System for Monitoring Public Political Groups in WhatsApp. *In Proceedings of the 24th Brazilian Symposium on Multimedia and the Web (WebMedia '18)*. ACM, New York, NY, USA, 387-390.