

FERRAMENTAS PARA AUXILIAR NA IDENTIFICAÇÃO DE NOTÍCIAS FALSAS

Aluna: Jenifer Galvão de Moraes

Coautora: Giovanna Pedrino Belasco

Orientador: Prof. Dr. Luciano Antonio Digiampietri

Escola de Artes Ciências e Humanidades/Universidade de
São Paulo

jenifer.galvao@usp.br

Objetivos

O objetivo principal desta pesquisa foi a implementação de um conjunto de ferramentas que possam auxiliar na identificação automática de notícias falsas (fake news), por meio de uma análise das características textuais desse tipo de notícia.

Nossos objetivos específicos foram o estudo bibliográfico de artigos acadêmicos correlatos ao tema, o estudo da linguagem Python, a construção de uma base de dados de notícias verdadeiras e falsas em português e, posteriormente, a identificação e implementação de modelos usando os classificadores que obtiveram os melhores resultados em pesquisas correlatas.

Métodos e Procedimentos

A partir da revisão da literatura correlata, foram especificadas, implementadas e testadas ferramentas para auxiliar na extração de diferentes características dos textos de notícias verdadeiras e falsas para serem usadas por diferentes algoritmos de classificação.

As notícias coletadas para a formação do conjunto de dados foram extraídas de sites jornalísticos que fazem a verificação manual dessas notícias.

Além da verificação, esses sites deixam disponível o meio pelo qual essa notícia foi publicada (em sua maioria, redes sociais e sites), o que auxiliou para que encontrássemos as fontes de propagação dessas notícias.

O conjunto de dados construído neste projeto é composto por 104 notícias, 52 verdadeiras e 52 falsas, sendo, assim, um conjunto totalmente balanceado. Este conjunto de dados pode ser encontrado no seguinte endereço eletrônico:

<https://www.each.usp.br/digiampietri/dados/FakeNewsDataset20220505.xlsx>

A avaliação do desempenho da solução desenvolvida considerou medidas amplamente usadas em sistemas de classificação, tais como precisão, revocação, medida F1 e área sob a curva ROC (AUC). A avaliação considerou a validação cruzada com três subconjuntos.

Resultados

Os seguintes classificadores foram testados: Dummy, que classifica todas as notícias como pertencendo à classe majoritária (este classificador foi utilizado como *baseline*); um classificador baseado em regressão logística (RegLog); um classificador do tipo SVM, uma rede neural MLP; e um classificador bayesiano (*Multinomial Naive Bayes*). As notícias foram representadas (isto é, suas características foram extraídas) utilizando n-gramas de palavras, considerando n de 1 a 3 e a combinação destes três n-gramas, o peso dado a cada característica (a cada n-grama) foi calculado de duas maneiras: frequência dos n-gramas e TF-IDF; por fim, as características foram apresentadas aos classificadores considerando três possibilidades: todas as características, características selecionadas

por um seletor de atributos e projeção das características utilizando análise de componentes principais (PCA). Excetuando-se o classificador Dummy, ao todo foram realizados 96 testes, correspondendo a quatro classificadores, utilizando quatro representações de texto, com duas estratégias de valoração/ponderação e três possibilidades relacionadas à redução de dimensionalidade. A tabela 1 sumariza os resultados, exibindo o melhor resultado em termos de medida F1 de todos os classificadores. Os melhores valores atingidos para cada métrica encontram-se destacados.

Tabela 1 - Resultados da pesquisa

| Classificador | Modelo de Representação | Ponderação | Redução de Dimensionalidade | F1-macro | Acurácia | AUC |
|----------------|-------------------------|------------|-----------------------------|--------------|--------------|--------------|
| Dummy | | | | 0,325 | 0,482 | 0,500 |
| MLP | trigrama | TF-IDF | nenhuma | 0,833 | 0,836 | 0,933 |
| Multinomial NB | trigrama | Contagem | nenhuma | 0,793 | 0,800 | 0,916 |
| RegLog | trigrama | TF-IDF | nenhuma | 0,767 | 0,781 | 0,923 |
| SVM | unigrama | TF-IDF | nenhuma | 0,818 | 0,820 | 0,856 |

O melhor resultado obtido, para as três métricas avaliadas, foi obtido pelo classificador MLP com o texto representado por trigramas de palavras, pontuados de acordo com a medida TF-IDF e sem o uso de nenhuma abordagem de redução de dimensionalidade. Os resultados atingidos por esta combinação foram: 83,3% para a Medida F1; 83,6% de acurácia e 93,3% de AUC.

Destaca-se, pela tabela 1, que o uso de estratégias de redução de dimensionalidade não estão presentes entre os melhores resultados de nenhum dos algoritmos. A maioria dos melhores resultados foi obtida com o uso de trigramas de palavras (apenas a solução utilizando SVM apresentou seus melhores resultados utilizando unigramas). Em termos de valoração dos n-gramas, o classificador Multinomial Naive Bayes obteve seus melhores resultados usando contagem, já os demais utilizaram TF-IDF.

Conclusões

Com base nos resultados, observa-se que a representação de texto utilizando TF-IDF foi a que obteve os melhores resultados em todas as métricas utilizadas. Em termos de classificadores, o MLP teve o melhor desempenho em termos de acurácia, AUC e medida F1 (resultados compatíveis com a literatura correlata, mesmo para outros idiomas), mostrando apresentar bom desempenho em relação à classificação tanto de notícias falsas como verdadeiras (o melhor equilíbrio entre precisão e revocação dentre os algoritmos utilizados).

Conclui-se que os resultados atingidos foram satisfatórios, conseguindo-se classificar corretamente mais de 80% das notícias analisadas. Como trabalhos futuros pretende-se utilizar diferentes técnicas para a extração de características do texto, como *word embeddings* e transformadores.

Bibliografia

1. FERREIRA, G.E.; SANTOS, B.L.; DO Ó, M.T.; BRAZ, R.R.; DIGIAMPIETRI, L.A.. Social bots detection in Brazilian presidential elections using natural language processing. **Anais do Simpósio Brasileiro de Sistemas de Informação (SBSI)**, 2021, Uberlândia. Sociedade Brasileira de Computação, 2021 .
2. VICARIO, M.D.; QUATTROCIOCCHI, W.; SCALA, A.; ZOLLO, F.. (2019). Polarization and Fake News: Early Warning of Potential Misinformation Targets. **ACM Trans. Web** 13, 2, Article 10 (March 2019), 22 pages.
3. ZHOU, X.; ZAFARANI, R.. (2019). Fake News Detection: An Interdisciplinary Research. **In Proceedings of The 2019 World Wide Web Conference (WWW '19)**, Ling Liu and Ryen White (Eds.). ACM, New York, NY, USA, 1292-1292.
4. ZHOU, X.; ZAFARANI, R., SHU, K.; LIU, H.. (2019). Fake News: Fundamental Theories, Detection Strategies and Challenges. **In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)**. ACM, New York, NY, USA, 836-837.