



APLICAÇÃO DE MODELOS DE LINGUAGEM NA DETECÇÃO DE IRONIA/SARCASMO

*CLAUDIO LUIS DA SILVA MACHADO JUNIOR¹; *LARISSA ASTROGILDO DE FREITAS²

¹*Universidade Federal de Pelotas – clsmachado@inf.ufpel.edu.br*

²*Universidade Federal de Pelotas – larissa@inf.ufpel.edu.br*

1. INTRODUÇÃO

Processamento de Linguagem Natural (PLN) é o campo da computação que se concentra na análise da maneira como dados na linguagem natural são expressos (BIRD et al., 2009).

Uma das tarefas de PLN é a detecção de ironia/sarcasmo (em inglês, *Irony Detection* - ID), que tem como objetivo classificar textos como irônicos ou não irônicos (CARVALHO et al., 2009). Ironia e sarcasmo são conceitos que possuem a mesma finalidade, isto é, de expressar o oposto do que realmente está tentando ser passado por quem estiver se comunicando. Mesmo que muito parecidos, ainda há certos detalhes que os diferenciam, porém neste trabalho consideramos os dois conceitos como equivalentes.

O objetivo deste artigo reside na detecção automatizada de textos que contêm ironia, ou seja, classificação binária em contextos em que a leitura é a principal forma de consumo, com ênfase na busca pela máxima precisão. Para aprimorar a compreensão, serão apresentados exemplos e comparações de modelos específicos para o português brasileiro, abrangendo a utilização de *transformers* (VASWANI et al., 2017).

O artigo segue uma estrutura que começa com a seção da **Metodologia**, que aborda o processo de agrupamento de dados e em seguida oferece um resumo do treinamento do modelo e discute como as avaliações de desempenho do modelo são realizadas. A seção de **Resultados e Discussão**, que utiliza um exemplo anterior de aplicação de detecção de ironia/sarcasmo em pesquisa científica para demonstrar a eficácia das técnicas disponíveis atualmente. Por fim, a seção de **Conclusões** que oferece *insights* sobre o significado desse avanço tecnológico para áreas dependentes dessas capacidades.

2. METODOLOGIA

O primeiro passo da metodologia é o pré-processamento de textos, que envolve a limpeza de textos, que envolve fazer com que caracteres especiais sejam removidos, bem como, a normalização do texto. Neste trabalho utilizamos o *dataset* IDPT 2021 (CORRÊA et al. 2021), que foi manualmente anotado por estudantes de Ciência da Computação. O *dataset* é dividido em duas colunas, uma composta por “text” que cobre os *tweets* em si, e outra para a “prediction”, que determina se o que está escrito no *tweet* é uma forma de ironia/sarcasmo. O *dataset* de treino é constituído por 12.736 *tweets* irônicos e 2.476 não-irônicos, enquanto o de teste é constituído por 177 *tweets* irônicos e 123 não-irônicos.

O próximo passo é a escolha de modelos pré-treinados. Existem uma grande quantidade de modelos, a maioria derivada do modelo BERT (*Bidirectional Encoder Representations from Transformers*) (HOREV, 2018). Neste trabalho utilizamos o Albertina (RODRIGUES, 2023) em sua versão *Base* (reduzida) e

*Agradecemos à FAPERGS e à NVIDIA Corporation pelo financiamento parcial deste trabalho.



Large (inteira), modelo específico para o português, tanto o de Portugal quanto do Brasil.

Após o modelo ser escolhido, passamos então para o passo de *fine-tuning*, que se refere a usar pesos de um network já existente e treinado como valores iniciais para treinar um novo network (BARRETO, 2023). Neste trabalho utilizamos os hiperparâmetros de ajuste para otimizar o desempenho (SHAHUL, 2023), como a **taxa de aprendizado**, que controla o tamanho dos passos que o algoritmo de otimização dá durante o treinamento; o **tamanho do lote**, que determina o número de exemplos usados em cada iteração durante o treinamento; o **número de épocas**, que representa quantas vezes o modelo passará por todo o conjunto de dados de treinamento; entre outros. Como mencionado anteriormente, também utilizamos a versão *Base* ou *Large* do modelo Albertina, as quais controlam a quantidade de camadas de *encoders* (AKHTAR, 2021). Na Tabela 1 são apresentados os hiperparâmetros utilizados nos experimentos.

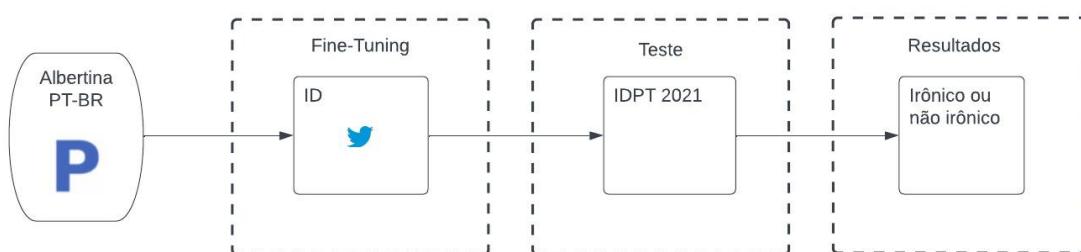
Tabela 1: Hiperparâmetros utilizados

	Base	Large
Tamanho do lote	8	2
Épocas	3	3
Taxa de aprendizado	1e-5	1e-5
Otimizador	AdamW	AdamW

Como último passo, temos a avaliação dos resultados obtidos. O que encontramos no experimento deste artigo, são resultados que medem a proporção de exemplos: classificados corretamente, ou sua **acurácia**; irônicos classificados corretamente em relação ao total de exemplos classificados como irônicos pelo modelo, ou a **precisão**; irônicos classificados corretamente em relação ao total de exemplos que são realmente irônicos, sua **sensibilidade**; e a combinação da precisão com a sensibilidade em uma única pontuação, também chamada de **medida-f** (BROWNLEE, 2016).

Na Figura 1 estão representadas as etapas por que passaram a metodologia.

Figura 1: Passos da metodologia



3. RESULTADOS E DISCUSSÃO

Neste experimento, concentrarmos nossa análise em uma comparação direta com os resultados obtidos por Corrêa et al. (2021) em sua pesquisa. No estudo mencionado, os pesquisadores utilizaram o modelo BERTimbau na tarefa de detecção de ironia/sarcasmo. Comparar o modelo Albertina com o BERTimbau

proporciona insights valiosos sobre a eficácia relativa de ambos e permite avaliar se o Albertina apresenta melhorias significativas ou se mantém competitivo em relação a uma referência estabelecida. Na Tabela 2 temos o resultado e comparação entre modelos.

Tabela 2: Comparação de resultados entre modelos BERTimbau e Albertina *Base* (B) e *Large* (L).

Modelo	Acurácia	Precisão	Sensibilidade	Medida F
BERTimbau	0.41	0.36	0.41	0.25
Albertina	0.40	0.40	0.99	0.57
BERTimbau	0.41	0.16	0.40	0.22
Albertina	0.41	0.41	1.0	0.58

Podemos, então, comprovar que a partir destes resultados o avanço das porcentagem de precisão, sensibilidade e medida-f no modelo Albertina *Base*. Já no modelo Albertina *Large*, os avanços continuam consistentes, sendo notável apenas a diferença da acurácia, ganhando 1% do modelo Albertina *Base* e se igualando ao experimento realizado com o modelo BERTimbau.

Essa comparação direta com o modelo de referência permitiu avaliar de forma sólida a eficácia do modelo Albertina na detecção de ironia/sarcasmo. Ainda, é importante salientar que os hiperparâmetros utilizados neste experimento foram em menor escala a fim de conseguir resultados por conta de limitações de hardware.

4. CONCLUSÕES

Seguindo as etapas da Figura 1 e com base nos resultados obtidos na Tabela 2, podemos concluir que o modelo utilizado Albertina supera o desempenho do BERTimbau na tarefa de detecção de ironia/ sarcasmo, quando utilizados os hiperparâmetros especificados na Tabela 1. Essa constatação é promissora e sugere que estamos avançando na direção certa para aprimorar a capacidade de compreensão de nuances na linguagem humana.

No entanto, esses resultados representam apenas um passo inicial em direção à excelência na detecção de ironia e sarcasmo. O próximo estágio de pesquisa envolverá esforços contínuos para aprimorar ainda mais o desempenho, explorando outros modelos e afinando-os para se aproximarem cada vez mais dos valores reais . Isso pode incluir a otimização de hiperparâmetros, o treinamento em conjuntos de dados mais diversos e a incorporação de técnicas avançadas de PLN.

A capacidade de detectar ironia/sarcasmo tem um papel crucial na prevenção na propagação de desinformação, e também uma melhor compreensão intercultural, contribuindo para um ambiente de comunicação mais informado e harmonioso.

Enquanto existe o avanço da tecnologia neste caminho, dá para se esperar uma significativa elevação na qualidade de interações e capacidade de máquinas e sistemas compreenderem o que faz a comunicação humana tão complexa, beneficiando a sociedade graças à evolução das técnicas de PLN.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Bird, S., Klein, E., & Loper, E. (2009). **Natural language processing with Python: analyzing text with the natural language toolkit.** " O'Reilly Media, Inc.".
- Carvalho, Paula, et al. (2009) **Clues for Detecting Irony in User-Generated Contents: Oh...!! It's “so easy” ;-)** *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion.*
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). **Attention is all you need.** *Advances in neural information processing systems*, 30.
- Anchiête, R. T., Neto, F. A. R., Marinho, J. C., do Nascimento, K. V., & Moura, R. S. (2021). **PiLN IDPT 2021: Irony Detection in Portuguese Texts with Superficial Features and Embeddings.** In *IberLEF@ SEPLN* (pp. 917-924).
- Horev, R. (2017). **BERT Explained: State of the art language model for NLP.** Acesso em 8 de setembro de 2023. Online. Disponível em: <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>
- Rodrigues, J. et al. (2023). **Advancing Neural Encoding of Portuguese with Transformer Albertina PT.** arXiv preprint arXiv:2305.06721.
- Barreto, S. (2023). **What Is Fine-Tuning in Neural Networks?**. Acesso em 13 de setembro de 2023. Online. Disponível em: <https://www.baeldung.com/cs/fine-tuning-nn>
- Akhtar, Z. (2021). **BERT base vs BERT large.** Acesso em 9 de setembro de 2023. Online. Disponível em: <https://iq.opengenus.org/bert-base-vs-bert-large>
- Brownlee, J. (2016). **Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End.** *Machine Learning Mastery*.