

Mineração de Emoções Multirrotulo Em Textos Curtos

Ramon N. Mendes, Syanne. K. M. Tavares, Luiz Nicollas M. Campos, Fabíola P. O. Araújo

¹Instituto de Ciências Exatas e Naturais – Universidade Federal do Pará (UFPA)
CEP 66.075-110 – Pará – PA – Brazil

{ramon.mendes, syanne.tavares, luiz.campos}@icen.ufpa.br, fpoliveira@ufpa.br

Resumo. *Este artigo apresenta os resultados parciais da investigação de uma abordagem para o reconhecimento de múltiplas emoções em textos curtos em português brasileiro. Para isso, propõe-se a construção de um corpus com tweets coletados por Web Scraper e rotulados com base na teoria da roda de emoções de Plutchik. Além disso, são apresentadas as etapas realizadas no pré-processamento do corpus, no desenvolvimento e na análise comparativa entre os modelos SVM e BERT na detecção de emoções em textos gerados por um LLM. As avaliações demonstraram que o modelo SVM obteve quase 20% a mais de precisão do que o BERT.*

Abstract. *This article presents the partial results of the investigation of an approach for recognizing multiple emotions expressed in short texts in Brazilian Portuguese. To this end, we propose the construction of a corpus with tweets collected by Web Scraper and labeled based on Plutchik's theory of the wheel of emotions. Furthermore, the steps taken in pre-processing the corpus, development and comparative analysis between the SVM and BERT models in detecting emotions in texts generated by an LLM are presented. Evaluations demonstrated that the SVM model achieved almost 20% more accuracy than BERT.*

1. Introdução

A mineração de emoções é o campo de estudo que analisa o aspecto emocional expresso sobre uma diversidade de entidades e atividades humanas [Liu 2012]. Sendo as emoções fatores implícitos na tomada de decisão de qualquer indivíduo, pesquisas procuram entender a proximidade entre decisões cibernéticas e humanas [Pires 2023] e um estudo que vem ganhando destaque nesse contexto é a IA emocional, onde máquinas são treinadas para sentir, detectar, interpretar e analisar emoções. [Mantello et al. 2023].

No processo exploratório da pesquisa, observou-se que a grande quantidade das publicações realizadas na área se limitam em analisar sentimentos classificando-os quando à sua polaridade, podendo ser positiva, negativa ou neutra [Cardozo and Freitas 2021, Paes et al. 2022, Silva and Faria 2023, Seno et al. 2023]. Todavia, sabe-se que as emoções se comportam de forma bem mais complexa. Além disso, apesar da literatura científica ter visto uma proliferação de estudos sobre detecção de emoções em língua inglesa, a quantidade de trabalhos que exploram a temática em textos em português brasileiro (Pt-BR) é ainda escassa [de Oliveira and de Melo 2021, Pereira 2021] e a mineração multilíngue ainda enfrenta desafios em alcançar altas precisões e confiabilidade [Santos et al. 2014, Moreira et al. 2024].

Outra abordagem possível adotada por pesquisas utiliza recursos da aprendizagem profunda, como [Hammes and Freitas 2021], que utilizaram os modelos BERTimbal-base

e BERTimbal-large para classificar 27 emoções de sentenças do *dataset multi-label* Go-Emotions traduzidos para o português. Nele, os autores comentam sobre a indisponibilidade de *datasets* anotados para tarefa de classificação de emoções em português e da perda de qualidade do *dataset* devida a tradução, que pode levar a perda do sentido expresso pelas sentenças e interferência na classificação correta das emoções.

Nessa perspectiva, este trabalho propõe a construção de uma base de dados contendo *tweets* em português rotulados de acordo com suas múltiplas classes de emoções, o treinamento e comparação de modelos para identificar essas emoções no texto. Para alcançar esse objetivo, realizou-se a coleta automática de um grande volume de informações da rede social X, empregando técnicas de processamento de linguagem natural (PLN) e aprendizado de máquina. Além disso, foi realizada uma análise comparativa entre os modelos de máquinas de vetores de suporte (*Support Vector Machine*) e o modelo de aprendizagem profunda BERT (*Bidirectional Encoder Representations from Transformers*) na classificação emocional dos textos da base de dados e em dados não vistos gerados com auxílio de um *Large Language Model* (LLM).

A organização do artigo ocorre da seguinte maneira: na Seção 2, descreve-se a metodologia, incluindo a preparação do corpus, a etapa de pré-processamento, o modelo treinado para a classificação e as métricas de avaliação usadas; na Seção 3, apresenta-se os resultados parciais da análise realizada e discute-se suas implicações; Por fim, na Seção 4, as considerações finais desta pesquisa são apresentadas.

2. Metodologia

No presente trabalho, as emoções foram identificadas de acordo com o modelo do psicólogo norte-americano Robert Plutchik, elaborado em 1980. A roda das emoções é um recurso gráfico em formato de flor de oito pétalas¹. Ela se diferencia das demais teorias ao propor relações de tipologia, antagonismo e intensidade entre as emoções, além de identificá-las. Quanto à tipologia, podem ser 8 emoções primárias (alegria, confiança, medo, surpresa, tristeza, nojo, raiva e antecipação), ou emoções secundárias, que ficam entre as pétalas, geradas pela combinação das emoções primárias adjacentes.

A preparação do corpus é essencial para o desenvolvimento do modelo de classificação de emoções. O conjunto de dados, com 12.160 publicações da rede social X, foi obtido via *Web Scraper* em Python, utilizando sinônimos de emoções primárias e secundárias. As emoções secundárias caracterizaram o conjunto como multirrótulo.

A rotulação de cada *tweet* teve como base as palavras sinônimas utilizadas no *Web Scraper*. Por exemplo, o texto "Do nada bate um desânimo!" foi coletado por meio do sinônimo "Desânimo" e foi classificado como Tristeza. Já "Não tenho provas, tenho convicção" foi coletado por meio de "Convicção" e foi classificado como confiança.

O pré-processamento do texto é essencial na análise de emoções, pois prepara e limpa os dados para um processamento mais eficiente. Primeiramente, são removidos caracteres especiais e acentos, especialmente em idiomas como o português. Depois, ocorre a tokenização, onde o texto é dividido em palavras (*tokens*), descartando caracteres e números restantes. Em seguida, retiram-se as *stopwords*, palavras comuns que pouco

¹https://github.com/MiningEmotion/EmotionsMiningPTBR/blob/main/imagens/Roda_Das_Emocoes.png

contribuem para o contexto. Por fim, utiliza-se o *stemming* para reduzir palavras ao seu radical, normalizando variações verbais e nominais.

Após a limpeza dos dados textuais descrita acima, aplicou-se a técnica de vetorização com a técnica *TF-IDF* (*Term frequency - inverse document frequency*), que considera a frequência e importância das palavras para que se possa aplicar um classificador e determinar a predição final do texto [LANDIM and TRESSO 2023].

3. Modelos de Aprendizagem de Máquina

Foram utilizados dois modelos de aprendizado de máquina para as fases de treino e teste, ambos com as mesmas métricas para uma melhor comparação. Os modelos empregados foram o *Support Vector Machine* (SVM) e o *Bidirectional Encoder Representations for Transformers* (BERT).

O modelo SVM linear foi utilizado em conjunto com o *MultiOutputClassifier*² que consiste em treinar um classificador separado por cada rótulo, adaptando o classificador base que não suporta nativamente uma saída com mais de um rótulo, como o SVM, para um modelo de classificação multirrótulo.

O BERT é baseado na arquitetura *transformer* [Devlin et al. 2018], ao contrário dos modelos direcionais, que lêem a entrada de texto sequencialmente, da esquerda para a direita ou da direita para a esquerda, o codificador *transformer* lê toda a sequência de palavras de uma vez, por isso é considerado bidirecional, também utilizado na sua versão multirrótulo e treinado com 4 épocas .

Ambos os modelos foram treinados com o conjunto de dados coletados e testados com dados gerados por um LLM, onde foram geradas 10 frases de cada emoção secundária para testar o modelo treinado, categorizando um texto com múltiplas classes simultaneamente. O *prompt* utilizado na criação das frases foi "Gere 10 frases com as emoções primárias : 'alegria' e 'confiança' e com a emoção secundária 'amor' que imite a mesma linguagem informal de tweets".

4. Resultados e Discussões

Visando avaliar os resultados e o desempenho dos modelos, foram adotadas as seguintes métricas para a análise da classificação multirrótulo: acurácia, precisão, revocação e F1. A seleção das métricas se deu considerando a especificidade de cada uma delas.

De acordo com a Figura 1, o modelo SVM demonstrou uma precisão significativa de 85,18%, indicando uma boa capacidade de classificação. No entanto, o BERT ao ser confrontado com as mesmas frases não vistas geradas pelo LLM, a acurácia caiu drasticamente para 66,67%, sinalizando dificuldades na classificação desses dados. Além disso, a acurácia, revocação e métrica F1 também diminuíram consideravelmente para as frases geradas, indicando uma tendência do segundo modelo em cometer mais erros de classificação e em perder instâncias de emoções.

A diminuição nas métricas pode ser atribuída à natureza dos modelos. O SVM tenta encontrar um hiperplano com a maior margem de separação, assegurando que os dados de cada classe sejam classificados corretamente com maior probabilidade. Já o BERT

²<https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.MultiOutputClassifier.html>

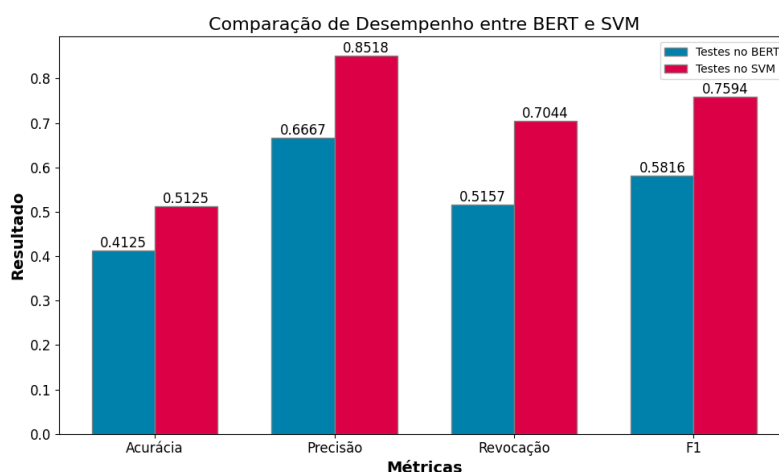


Figura 1. Comparação de Desempenho entre BERT e SVM. (Fonte: Autores).

utiliza o treinamento bidirecional do *transformer*, permitindo que o modelo aprenda o contexto de uma palavra com base em seu entorno completo.

No modelo SVM, quanto menor o texto, menor será o processamento exigido pela máquina. Em contrapartida, o BERT, sendo um modelo de linguagem mais robusto, necessita de textos mais extensos para utilizar todo o seu potencial. Conforme demonstrado por [Devlin et al. 2018], ao empregar características semânticas dos revisores junto com textos curtos, observou-se uma melhoria na acurácia da classificação do modelo BERT, enquanto o SVM não apresentou mudanças significativas nos valores das métricas.

5. Considerações Finais e Direções Futuras

Este trabalho contribuiu na investigação referente à classificação multirrotulo de emoções em textos curtos com a disponibilização de códigos, a base de dados *tweetEmotionsPTBR* e as frases geradas por um LLM, já classificados em português brasileiro, no repositório *EmotionMiningPTBR*³ do Github.

Através dos resultados, o SVM obteve maior desempenho em relação ao BERT na maioria das métricas devido à natureza do modelo. Todavia, as métricas não desempenharam conforme esperado pelo modo como os modelos foram treinados, exclusivamente com conteúdo da rede social X, carecendo de uma quantidade significativa de exemplos que contenham uma linguagem mais elaborada e nuances figurativas. Sendo assim, foram realizados testes iniciais com conjuntos de dados maiores e balanceados, treinando os modelos com as publicações de redes sociais e frases geradas por LLM. Notou-se uma melhora significativa nas métricas do modelo BERT, superando o SVM.

Como direções futuras, é necessário a construção de um conjunto de dados mais diversificado linguisticamente, proveniente de uma variedade de fontes. Isso ajuda a evitar que os modelos fiquem excessivamente adaptados a um único estilo de linguagem. Além disso, seria importante a elaboração de um *Gold-Standard Corpus* com a revisão dos rótulos feita por especialistas na teoria da roda das emoções.

³<https://github.com/MiningEmotion/EmotionsMiningPTBR>

Referências

- Cardozo, L. and Freitas, L. (2021). Análise de sentimentos: Avaliando o desempenho de pré-processamento e de algoritmos de aprendizagem de máquina sobre o dataset tweetsentbr. In *Anais do X Brazilian Workshop on Social Network Analysis and Mining*, pages 169–174, Porto Alegre, RS, Brasil. SBC.
- de Oliveira, M. and de Melo, T. (2021). An empirical study of text features for identifying subjective sentences in portuguese. In Britto, A. and Valdivia Delgado, K., editors, *Intelligent Systems*, pages 374–388, Cham. Springer International Publishing.
- Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.
- Hammes, L. and Freitas, L. (2021). Utilizando bertimbau para a classificação de emoções em português. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 56–63, Porto Alegre, RS, Brasil. SBC.
- LANDIM, G. P. P. and TRESSO, G. J. (2023). Identificação de sentimentos em textos utilizando o modelo term frequency-inverse document frequency.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5:1–167.
- Mantello, P., Ho, M.-T., Nguyen, M.-H., and Vuong, Q.-H. (2023). Machines that feel: behavioral determinants of attitude towards affect recognition technology—upgrading technology acceptance theory with the mindsponge model. *Humanities and Social Sciences Communications*, 10(1):1–16.
- Moreira, B. G., Pfitscher, R. J., Camargo, L. C., and Garcia, T. R. (2024). Análise de sentimentos e emoções com o uso da ferramenta orange datamining: uma avaliação a partir da tradução dos textos do português. *Anais do Computer on the Beach*, 15:352–354.
- Paes, V., Araújo, D., Brito, K., and Andrade, E. (2022). Análise de sentimento em tweets relacionados ao desmatamento da floresta amazônica. In *Anais do XI Brazilian Workshop on Social Network Analysis and Mining*, pages 61–72, Porto Alegre, RS, Brasil. SBC.
- Pereira, D. A. (2021). A survey of sentiment analysis in the portuguese language. *Artif. Intell. Rev.*, 54(2):1087–1115.
- Pires, S. F. S. (2023). Inteligência artificial e emoções: Quão próximas estão as decisões cibernéticas das humanas? *A sociedade do conhecimento e suas tecnologias: estudos em Ciências Exatas e Engenharias: Volume 10*.
- Santos, A., Becker, K., and Moreira, V. (2014). Um estudo de caso de mineração de emoções em textos multilíngues. In *Anais do III Brazilian Workshop on Social Network Analysis and Mining*, pages 140–151, Porto Alegre, RS, Brasil. SBC.
- Seno, E., Anno, F., Lazarini, L., and Caseli, H. (2023). Classificação de polaridade orientada aos alvos de opinião em comentários sobre debate político em português. In *Anais do XIV Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana*, pages 84–93, Porto Alegre, RS, Brasil. SBC.

Silva, S. and Faria, E. (2023). Análise de sentimentos expressos no twitter em relação aos candidatos da eleição presidencial de 2022. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pages 79–90, Porto Alegre, RS, Brasil. SBC.