

Aprendizado Profundo para Detecção de Movimentos Retóricos

Bruno Vinicius Veronez de Jesus¹, Arnaldo Candido Junior¹

¹ Institute of Biosciences, Humanities and Exact Sciences

São Paulo State University (UNESP)

São José do Rio Preto, SP – Brazil

{bvv.jesus, arnaldo.candido}@unesp.br

Abstract. *The lack of tools to support scientific writing in Portuguese hinders the production of cohesive texts. This paper describes the fine-tuning of the BERTimbau language model for the automatic identification of rhetorical moves based on Swales's model. The methodology involves using CorpusDT, a corpus of abstracts in the field of Computer Science, for sentence classification. The results from the classifier demonstrate high effectiveness, with an F1-score of 0,94 on the validation set, indicating the potential of this approach to improve the quality of scientific publications in Portuguese.*

Resumo. *A carência de ferramentas para auxílio à escrita científica em português dificulta a produção de textos coesos. Este artigo descreve o processo de fine-tuning do modelo de linguagem BERTimbau para a identificação automática de movimentos retóricos fundamentados no modelo de Swales. A metodologia utiliza o CorpusDT, um corpus de resumos da área de Computação, para a tarefa de classificação de sentenças. Os resultados do classificador demonstram alta eficácia, com um F1-Score de 0,94 no conjunto de validação, indicando o potencial da abordagem para aprimorar a qualidade das publicações científicas em língua portuguesa.*

1. Introdução

A produção de textos científicos exige não apenas domínio técnico do assunto, mas também uma organização retórica coerente, frequentemente realizada de maneira intuitiva pelos autores. Mesmo sem conhecimento teórico explícito, a maioria dos pesquisadores estrutura seus trabalhos a partir de padrões como a contextualização do tema, a identificação de uma lacuna, a definição do propósito, a descrição do método, a apresentação de resultados e as conclusões (Swales, 2004).

Esses elementos, conhecidos como movimentos retóricos, foram sistematizados pelo linguista John Malcolm Swales em sua obra seminal *Genre Analysis* (Swales, 1990), e sua aplicabilidade em textos acadêmicos em língua portuguesa já foi confirmada por diversas pesquisas (Dias and Bezerra, 2013). A articulação lógica desses movimentos é o que orienta o leitor e demonstra a eficácia do texto, desde a justificativa do estudo até suas implicações práticas.

Entretanto, a ausência de orientações sobre a aplicação desses movimentos resulta em textos com omissões relevantes, como a não explicitação da lacuna de pesquisa ou a desconexão entre resultados e objetivos. Essa deficiência é agravada pela escassez de

ferramentas automatizadas em língua portuguesa capazes de analisar textos para além de aspectos superficiais, desconsiderando a dimensão retórica essencial para a clareza do texto, fator que, segundo estudos, impacta diretamente a visibilidade e o número de citações de um trabalho (Ryba et al., 2019).

Como destacado por Benevides-Pereira (2012), a sobrecarga de tarefas dos docentes frequentemente limita o tempo disponível para a leitura e o acompanhamento individualizado dos alunos. Isso mostra a importância de ferramentas capazes de suprir a necessidade de alunos e pesquisadores e evitar uma sobrecarga de trabalho para os orientadores.

Para superar essas limitações, este estudo investiga a aplicação de um modelo de linguagem de estado da arte para a detecção de movimentos retóricos. O objetivo central é desenvolver um classificador de alta performance, capaz de avaliar automaticamente sentenças de textos científicos e identificar a presença de elementos retóricos presentes no texto, como contextualização, lacuna e propósito. Para isso, o modelo de linguagem BERTimbau (Souza et al., 2020) foi adaptado por meio de *fine-tuning* para a tarefa de classificação de padrões retóricos. A metodologia envolveu o treinamento e a validação do modelo com métricas de precisão e *recall*, utilizando um corpus de referência da área. Dessa forma, a proposta visa contribuir para o campo de Processamento de Linguagem Natural aplicado à redação acadêmica em português, oferecendo um modelo consolidado para análise retórica.

2. Trabalhos Relacionados

A análise automática da estrutura retórica de textos científicos é uma área de pesquisa consolidada que busca identificar os papéis discursivos de cada sentença (e.g., contexto, método, resultado). Uma das abordagens seminais nesse campo é o modelo de *Argumentative Zoning* (AZ), proposto por Teufel and Moens (2002). O objetivo principal do trabalho era aprimorar a sumarização de artigos científicos, desenvolvendo uma estratégia de seleção de conteúdo que destacasse a nova contribuição de um artigo e a situasse em relação a trabalhos anteriores. Para isso, os autores propuseram um esquema de anotação com sete categorias retóricas não hierárquicas: *AIM* (objetivo da pesquisa), *OWN* (trabalho próprio), *BACKGROUND* (contexto geral), *OTHER* (trabalho de outros pesquisadores), *BASIS* (continuação de outros trabalhos), *CONTRAST* (contraste ou crítica a outros trabalhos) e *TEXTUAL* (estrutura do texto). A abordagem metodológica consistiu no treinamento de um classificador Naive Bayes sobre um corpus de 80 artigos da área de Linguística Computacional. O modelo se baseava em um conjunto de 16 atributos extraídos via regras, incluindo características posicionais (localização da sentença no artigo), lexicais (palavras do título, TF*IDF), sintáticas (tempo e voz verbal) e, notavelmente, três novos atributos de metadiscursivo: expressões formulaicas (frases-chave), agentes (e.g., “nós”, “outros autores”) e ações (classes semânticas de verbos como “apresentar” ou “criticar”). A complexidade da tarefa foi evidenciada pela concordância entre anotadores humanos, que atingiu um Kappa¹ de 0,71. O classificador automático alcançou um Kappa de 0,45 e uma acurácia de 73%, com um F1-Score macro de 0,50,

¹O coeficiente Kappa (Landis and Koch, 1977) mede a concordância entre anotadores, ajustada para o acaso. Valores próximos de 1 indicam alto acordo; valores próximos de 0 indicam que a concordância não é maior do que o esperado por sorte.

estabelecendo um importante benchmark para a área. A abordagem de Teufel e Moens demonstrou a viabilidade da classificação retórica a partir de características de superfície e estabeleceu uma base metodológica que influenciou diversos trabalhos subsequentes.

No contexto da língua portuguesa, em que os recursos computacionais para essa finalidade são mais escassos, um dos trabalhos pioneiros foi a tese de Feltrim (2004), que desenvolveu o ambiente SciPo, um sistema de auxílio à escrita que incluía o classificador estatístico AZPort. Diferentemente de abordagens baseadas em aprendizado profundo que extraem características de forma automática, o AZPort fundamentava-se em um classificador Naive Bayes que utilizava um conjunto de oito atributos extraídos por meio de regras, incluindo características lexicais (presença de expressões-padrão), posicionais (localização da sentença no resumo), e sintáticas superficiais (tempo e voz do primeiro verbo finito). O classificador apresentou uma acurácia de 72% e um Kappa de 0,65. A análise de Feltrim destacou que o atributo com maior poder de distinção era a presença de expressões-padrão, evidenciando a forte dependência do modelo em marcadores discursivos explícitos. Fundamentalmente, esse trabalho também resultou na construção do CorpusDT. Este corpus, composto por resumos da Ciência da Computação anotados manualmente com seis principais movimentos retóricos (Contexto, Lacuna, Propósito, Método, Resultado e Conclusão), permanece como um recurso valioso e um dos únicos disponíveis em língua portuguesa.

A tarefa de anotar e classificar movimentos retóricos, no entanto, apresenta desafios intrínsecos. A fronteira entre um movimento e outro é frequentemente sutil, o que pode levar a ambiguidades na classificação. Um estudo de referência que evidencia essa complexidade é o trabalho de Dayrell et al. (2012) que, ao desenvolver o corpus MAZEA para análise de resumos em inglês, investigou a concordância entre anotadores humanos. Os resultados reportaram um acordo classificado como apenas “moderado” (Kappa de 0,535) para o subcorpus de Ciências Físicas e Engenharia (Dayrell et al., 2012). A principal causa apontada para essa dificuldade é a natureza funcional e flexível dos movimentos, que, como define Swales, podem ser realizados por uma oração, uma sentença inteira ou várias sentenças (Swales, 2004). Essa constatação estabelece a sobreposição entre classes como um problema clássico da área, que desafia tanto a anotação manual quanto a classificação automática.

Atualmente, o estado da arte para a análise de textos científicos em inglês é dominado por modelos de linguagem baseados na arquitetura Transformer. O SciBERT (Beltagy et al., 2019), por exemplo, é uma versão do BERT pré-treinada especificamente em um vasto corpus científico, o que resultou em um desempenho superior em diversas tarefas de Processamento de Linguagem Natural no domínio acadêmico. A ausência de um modelo similar com pré-treinamento específico em textos científicos do português evidencia uma lacuna tecnológica e uma oportunidade de pesquisa.

A estratégia de especializar modelos de linguagem para domínios específicos, como o científico, tem se mostrado altamente eficaz. Um exemplo proeminente é o BioBERT (Lee et al., 2019), um modelo de representação de linguagem pré-treinado especificamente para a mineração de textos biomédicos. Os desenvolvedores do BioBERT partiram da hipótese de que a aplicação direta de modelos de uso geral produz resultados insatisfatórios em domínios especializados devido à diferença na distribuição de vocabulário e na complexidade dos textos. A metodologia adotada consistiu em inicializar o BioBERT

com os pesos do BERT (pré-treinado em um corpus geral) e continuar o pré-treinamento em um vasto volume de corpora da área biomédica, como resumos do PubMed² e artigos de texto completo do PubMed Central (PMC)³. Os resultados demonstraram que o BioBERT superou significativamente o desempenho do BERT e de outros modelos de estado da arte em tarefas como reconhecimento de entidade nomeada, extração de relações e resposta a perguntas. Essa abordagem confirma que o pré-treinamento em corpora de domínio é crucial para a compreensão de textos complexos e estabelece um precedente metodológico consolidado, reforçando a importância de adaptar modelos de linguagem a domínios e idiomas específicos, como o português científico.

Diante do exposto, este trabalho se posiciona como um avanço ao aplicar um modelo de linguagem treinado para a língua portuguesa, o BERTimbau, à tarefa de classificação de movimentos retóricos. Ao contrário de algumas abordagens anteriores baseadas em regras ou modelos estatísticos mais simples, buscamos estabelecer um novo *benchmark* de performance, aproveitando o poder de representação contextual dos Transformers para aprimorar a análise retórica automática na língua portuguesa.

3. Materiais e Método

Esta seção detalha o corpus utilizado, o ambiente experimental, o pré-processamento dos dados, a arquitetura do modelo e os procedimentos de treinamento e avaliação adotados no estudo.

3.1. Corpus

A base de dados central para este projeto foi o **CorpusDT**, composto por 52 resumos científicos em português da área de Ciência da Computação, anotados com papéis retóricos (Feltrim, 2004). A composição deste corpus, extraída do trabalho original de Feltrim, é detalhada na Tabela 1. As anotações originais, em formato XML, foram processadas para extrair o rótulo do componente estrutural de cada sentença (e.g., Contexto, Lacuna), que serviu como a classe-alvo para o nosso modelo.

Tabela 1. Número de trabalhos em cada área de pesquisa no CorpusDT Feltrim (2004).

Área de Pesquisa	Diss.	Tese
Banco de Dados	3	-
Inteligência Computacional	7	-
Engenharia de Software	15	1
Hipermídia	12	-
Sistemas Digitais	1	-
Sistemas Distribuídos e Programação Concorrente	10	2
Computação Gráfica e Processamento de Imagens	1	-
Total Parcial	49	3
Total Geral		52

²<https://pubmed.ncbi.nlm.nih.gov/>

³<https://pmc.ncbi.nlm.nih.gov/>

3.2. Ambiente Experimental

Os experimentos foram conduzidos em um computador pessoal equipado com uma GPU NVIDIA RTX 5070 com 12 GB de VRAM. O desenvolvimento foi realizado na linguagem Python, com o suporte principal das bibliotecas PyTorch e Hugging Face Transformers para a implementação e treinamento do modelo de aprendizado profundo.

3.3. Pré-processamento e Aumento de Dados

Considerando o tamanho limitado do corpus original, aplicou-se uma estratégia de aumento de dados (*data augmentation*) para expandir o conjunto de treinamento e aprimorar a capacidade de generalização do modelo. Foram combinadas múltiplas técnicas para gerar variações sintáticas realistas das sentenças originais, preservando seu sentido retórico. As principais técnicas empregadas foram:

- **Retrotradução (*Back-translation*):** Cada sentença do corpus original foi traduzida para o inglês e, em seguida, traduzida de volta para o português. Essa abordagem é altamente eficaz para gerar paráfrases semanticamente equivalentes.
- **Substituição de Sinônimos:** Palavras-chave foram substituídas por sinônimos de um dicionário pré-definido (e.g., “mostrar” por “demonstrar”), mantendo a capitalização original.
- **Inserção e Troca Aleatória:** Para aumentar a variabilidade sintática, conectivos semanticamente neutros (e.g., “assim”, “além disso”) foram inseridos em posições aleatórias e pares de palavras dentro de uma mesma sentença foram trocados de posição (*random swap*).
- **Deleção Aleatória:** Palavras foram removidas com uma baixa probabilidade ($p=0,1$) como forma de regularização, forçando o modelo a não depender excessivamente de palavras-chave específicas e a focar no contexto geral.

Essas técnicas foram aplicadas de forma estocástica, visando principalmente as classes minoritárias para alcançar uma distribuição mais equilibrada. O processo resultou em um conjunto de dados final para treinamento composto por 1.409 exemplos. Em seguida, este corpus aumentado foi dividido de forma estratificada em 80% para treino (1.127 exemplos) e 20% para validação (282 exemplos).

3.4. Arquitetura do Modelo e Fine-Tuning

A arquitetura do classificador foi construída a partir do modelo de linguagem pré-treinado `adalbertojunior/distilbert-portuguese-cased` (Adalberto Ferreira Barbosa Junior, 2024), uma versão destilada da arquitetura BERT. Para a tarefa específica de classificação de movimentos retóricos, um cabeçalho de classificação (*classification head*) foi adicionado ao topo do modelo base. O mecanismo de *fine-tuning* opera da seguinte forma: para cada sentença de entrada, o modelo gera uma representação vetorial contextualizada. Utiliza-se especificamente o vetor de saída da última camada oculta associado ao token especial `[CLS]`, que funciona como uma representação semântica agregada de toda a sentença. Este vetor, com **768 dimensões**, é então alimentado em uma camada de regularização Dropout (com taxa de 0,1, para mitigar o sobreajuste) e, subsequentemente, em uma camada linear densa. Esta camada final projeta o vetor de 768 dimensões para um espaço de 6 dimensões, correspondendo às seis classes retóricas do estudo (*Contexto, Lacuna, Propósito, Método, Resultado e Conclusão*). Por fim, uma função de ativação Softmax é aplicada à saída para produzir uma distribuição de probabilidade sobre as classes, indicando a predição final do modelo.

3.5. Treinamento e Avaliação

O modelo foi treinado com o objetivo de minimizar a função de perda *Cross-Entropy Loss*, com pesos aplicados para mitigar os efeitos do desbalanceamento entre as classes. Para otimizar a velocidade e evitar sobreajuste (*overfitting*), foram empregadas as técnicas de precisão mista (FP16) e parada antecipada (*Early Stopping*), que monitora a performance no conjunto de validação e interrompe o processo quando não há melhora. A performance final do classificador foi avaliada com um conjunto padrão de métricas: acurácia, precisão, *recall* e F1-Score (ponderado e macro). Para a análise qualitativa dos erros, foram geradas visualizações como curvas de aprendizado e matrizes de confusão.

4. Resultados

A aplicação da estratégia de aumento de dados, detalhada na Seção 3.3, resultou em um corpus final de 1.409 exemplos. Uma análise deste conjunto de dados revelou uma distribuição de classes mais balanceada, conforme detalhado na Tabela 2, com a classe “P” (Propósito) sendo a menos frequente (12,6%) e a “C” (Conclusão) a mais prevalente (18,8%). Adicionalmente, uma análise sobre o comprimento dos textos indicou que 95% das sentenças possuíam até 43 tokens, valor que foi adotado como comprimento máximo de sequência (MAX_LEN) para otimizar o treinamento.

Tabela 2. Distribuição das classes no conjunto de dados aumentado.

Classe	Nº de Exemplos	%
C (Conclusão)	265	18,8
R (Resultado)	258	18,3
M (Método)	252	17,9
B (Contexto)	244	17,3
G (Lacuna)	212	15,0
P (Propósito)	178	12,6
Total	1.409	100,0

A seleção dos hiperparâmetros foi um passo importante para otimizar o desempenho do modelo. Por meio de um ajuste experimental, foram avaliadas diferentes configurações para a taxa de aprendizado, tamanho do lote e regularização, buscando maximizar o F1-Score no conjunto de validação e evitar o sobreajuste. Testes com taxas de aprendizado entre 1e-5 e 5e-5 indicaram que um valor de 1,5e-5, combinado com um agendador do tipo cosseno (*cosine scheduler*), proporcionou uma convergência mais estável. Diferentes tamanhos de lote foram explorados, e um *batch size* de 8 com 2 passos de acumulação de gradiente (resultando em um lote efetivo de 16) ofereceu o melhor equilíbrio entre eficiência computacional e generalização. Para o número de épocas, foi definido um limite superior de 20, com a implementação de *Early Stopping* com tolerância de 3 épocas monitorando o F1-Score de validação, garantindo que o treinamento fosse interrompido no ponto ótimo. Adicionalmente, um *weight decay* de 0,05 foi aplicado para regularização. A configuração final, detalhada na Tabela 3, representa a combinação que produziu o modelo mais robusto.

Com os hiperparâmetros definidos, o modelo foi submetido ao treinamento final. A evolução do treinamento, visualizada na Figura 1, demonstra uma rápida convergência.

As curvas de perda (*loss*), representadas na Figura 2, e a comparação entre acurácia e F1-Score indicam que o modelo aprendeu eficientemente, sem sinais de *overfitting* significativo. Conforme planejado, a estratégia de *Early Stopping* interrompeu o treinamento na época 12, selecionando o modelo da época 9 como o de melhor desempenho, que atingiu um F1-Score de 0,94 no conjunto de validação.

Tabela 3. Hiperparâmetros utilizados no novo treinamento.

Hiperparâmetro	Valor
Modelo Base	adalbertojunior/distilbert-portuguese-cased
MAX_LEN	43 tokens
Batch Size	8 (Efetivo de 16 com <i>gradient accumulation</i>)
Learning Rate	1,5e-5
Épocas	20 (com <i>Early Stopping</i> , paciência de 3)
Otimizador	AdamW
Weight Decay	0,05
LR Scheduler	Cosine

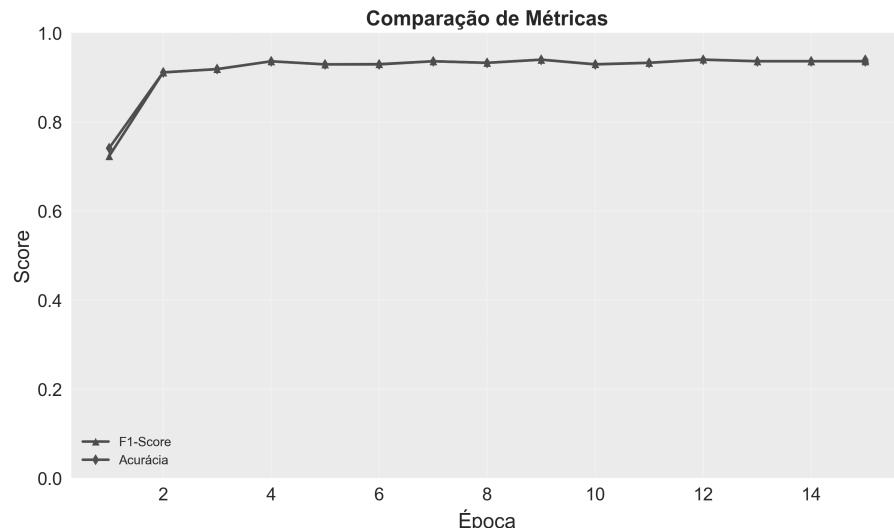


Figura 1. Comparativo entre Acurácia e F1-Score na validação.

A avaliação final do melhor modelo no conjunto de validação alcançou uma acurácia geral de 93,97% e um F1-Score ponderado de 0,9400 (macro de 0,9403). A Tabela 4 detalha o desempenho por classe.

Para uma análise qualitativa dos erros, a matriz de confusão (Figura 3) revela que as classificações incorretas ocorreram majoritariamente entre classes semanticamente próximas. Este padrão de erros é um problema clássico na classificação de movimentos retóricos (Swales, 1990; Teufel and Moens, 2002): a sobreposição semântica entre categorias adjacentes no fluxo retórico. Essa dificuldade também é observada em inglês, como demonstram os desafios de anotação e os índices de concordância apenas moderados em corpora como o MAZEA (Dayrell et al., 2012). A confusão mais notável no nosso

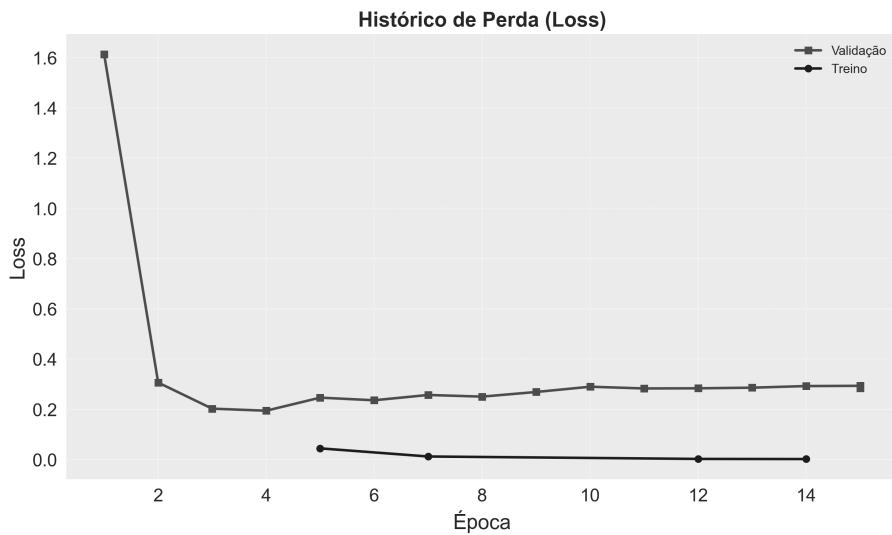


Figura 2. Evolução das perdas de treino e validação.

experimento ocorreu com a classe “R” (**Resultado**), onde 7,7% dos seus exemplos foram incorretamente classificados como “M” (**Método**). Essa sobreposição é esperada, pois frases que descrevem resultados frequentemente contêm elementos metodológicos, como em: “Os dados revelaram uma correlação significativa ($p < 0,05$) após a aplicação do teste T ”, um exemplo hipotético em que o resultado está intrinsecamente ligado ao método estatístico. De forma similar, a confusão entre “B” (**Contexto**) e “G” (**Lacuna**), e entre “P” (**Propósito**) e “C” (**Conclusão**), embora menos frequente, reflete a dificuldade inerente em delimitar precisamente onde a contextualização termina e a identificação da lacuna começa, ou onde o propósito do estudo é reafirmado na conclusão. Portanto, os erros não são aleatórios, mas seguem um padrão lógico que destaca a complexidade da tarefa, o que, reforça a capacidade do modelo em capturar as distinções mais evidentes.

Tabela 4. Relatório de classificação no conjunto de validação.

Classe	Precisão	Recall	F1-Score	Suporte
B	0,9783	0,9184	0,9474	49
C	0,9623	0,9623	0,9623	53
G	0,9318	0,9762	0,9535	42
M	0,8679	0,9200	0,8932	50
P	0,9444	0,9444	0,9444	36
R	0,9600	0,9231	0,9412	52
Média Macro	0,9408	0,9407	0,9403	282
Média Pond.	0,9411	0,9397	0,9400	282

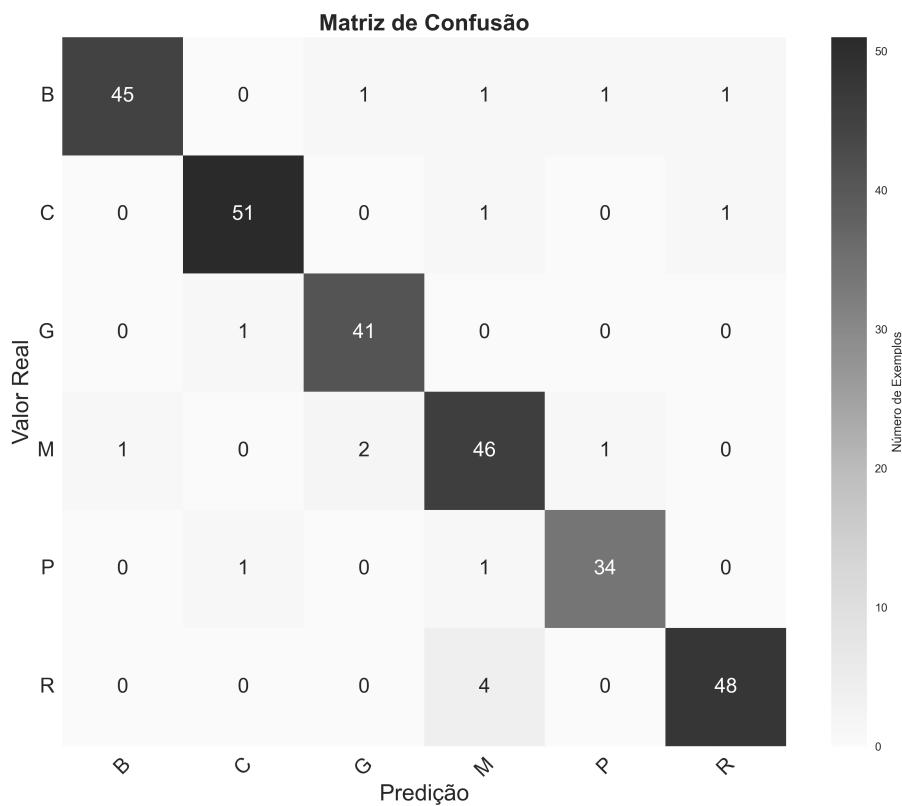


Figura 3. Matriz de confusão (valores absolutos).

5. Conclusão

Este trabalho apresentou a criação de um classificador para a detecção de movimentos retóricos, baseado no modelo DistilBERT e no CorpusDT. Segundo nosso conhecimento, os resultados obtidos foram superiores aos modelos de detecção de movimentos retóricos existentes para a língua portuguesa. Os resultados do classificador demonstram uma eficácia robusta, com um F1-Score de 0,94 no conjunto de validação, indicando o potencial da abordagem para uso em sistemas destinados a aprimorar a qualidade das publicações científicas, com foco na língua portuguesa.

Como trabalhos futuros, pretende-se incorporar o classificador desenvolvido em uma plataforma para ensino de escrita científica. Também pretende-se verificar outras variações do Transformer com foco na classificação de textos. Outro passo importante será expandir o corpus de treinamento através da anotação de novos artigos em português de diferentes áreas do conhecimento, como biologia e física, utilizando como base a metodologia e o esquema de anotação de movimentos retóricos do projeto MAZEA (Dayrell et al., 2012). Como o corpus original do MAZEA está em inglês, será necessário adaptá-lo e traduzi-lo para o português a fim de manter a consistência linguística e assegurar a validade das anotações.

Referências

Adalberto Ferreira Barbosa Junior. distilbert-portuguese-cased (revision df1fa7a), 2024. URL <https://huggingface.co/adalbertojunior/distilbert-portuguese-cased>.

- Iz Beltagy, Kyle Lo, and Arman Cohan. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.
- Ana Maria Teresa Benevides-Pereira. Considerações sobre a síndrome de burnout e seu impacto no ensino. *Boletim de Psicologia*, 62:155 – 168, 12 2012. ISSN 0006-5943. URL http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S0006-59432012000200005&nrm=iso.
- Carmen Dayrell, Arnaldo Candido Jr, Gabriel Lima, Danilo Machado Jr, Ann A Copestake, Valéria Delisandra Feltrim, Stella EO Tagnin, and Sandra M Aluísio. Rhetorical move detection in english abstracts: Multi-label sentence classifiers and their annotated corpora. In *LREC*, pages 1604–1609, 2012.
- Fernanda Goulart Ritti Dias and Benedito Gomes Bezerra. Análise retórica de introduções de artigos científicos da área da saúde pública. *Horizontes de Linguística Aplicada, Brasília, ano*, 12:163–182, 2013.
- Valéria Delisandra Feltrim. *Uma abordagem baseada em corpus e em sistemas de crítica para a construção de ambientes Web de auxílio à escrita acadêmica em português*. PhD thesis, Universidade de São Paulo, 2004.
- J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.
- Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 09 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz682. URL <https://doi.org/10.1093/bioinformatics/btz682>.
- Ren Ryba, Zoë A. Doubleday, and Sean D. Connell. How can we boost the impact of publications? try better writing. *Proceedings of the National Academy of Sciences*, 116(2):341–343, 2019. doi: 10.1073/pnas.1819937116. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1819937116>.
- Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In *Proceedings of the 9th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 403–417. Springer, 2020.
- John M Swales. *Genre analysis: English in academic and research settings*. Cambridge University Press, 1990.
- John M Swales. *Research genres: Explorations and applications*. Cambridge University Press, 2004.
- Simone Teufel and Marc Moens. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational linguistics*, 28(4):409–445, 2002.