

Avaliação Automática de Redações Dissertativo-Argumentativas no Modelo do ENEM

Trabalho de Formatura Supervisionado (MAC0499)

Aluno: José Lucas Silva Mayer
Orientador: Denis Deratani Mauá
Co-orientador: Igor Cataneo Silveira

Abril, 2023

Resumo

O Exame Nacional do Ensino Médio (ENEM) é uma das principais formas de acesso ao ensino superior no Brasil e abrange diversas áreas do conhecimento. Uma parte dele é composta pela prova de redação, que é aplicada para avaliar a habilidade de produção textual dos alunos sobre um tema relevante em âmbito nacional. A correção das redações, cujo modelo é o dissertativo-argumentativo, é realizada manualmente por avaliadores, o que torna o processo custoso e demorado. Nesse sentido, o uso de técnicas de inteligência artificial para automatizar a correção desses textos pode representar uma alternativa viável para agilizar a tarefa. Diante disso, este tema proposto tem como objetivo explorar o uso de técnicas de aprendizado profundo para a classificação automática de redações dissertativo-argumentativas do ENEM, com base em modelos pré-treinados de processamento de linguagem natural para a língua portuguesa. Visa-se, com isso, definir um projeto cujo intuito é investigar arquiteturas de redes neurais para ajustar os modelos e avaliar os resultados com métricas de precisão conhecidas. Essa pesquisa poderá ser um avanço para o desenvolvimento de sistemas automáticos de correção de redações em língua portuguesa, bem como para aprimorar e facilitar o processo de avaliação dos estudantes no ENEM.

Palavras-chave: Redes Neurais • Processamento de Linguagem Natural • Correção Automática • Redações.

1 Introdução

Uma das principais habilidades exigidas dos estudantes do ensino básico é a produção textual. O Exame Nacional do Ensino Médio (ENEM), prova aplicada nacionalmente com o intuito de analisar o conhecimento construído no período de formação básica dos alunos, exige que os estudantes redijam uma redação no modelo dissertativo-argumentativo, a partir de uma proposta de tema e de textos de apoio norteadores. A avaliação desses textos é realizada com base em cinco competências cujas notas podem valer de 0 a 200, de modo que ao menos dois professores avaliam independentemente cada redação, conforme a cartilha de orientação do exame em 2022 [1].

O modelo de correção adotado pelo ENEM, no entanto, demanda muito tempo e muitos recursos, e a quantidade de redações que precisam ser avaliadas por cada corretor pode comprometer a qualidade do processo. Nesse contexto, a utilização de técnicas de inteligência artificial tem se mostrado uma alternativa viável para automatizar a correção de redações.

Um dos principais desafios para a tarefa é a escassez de dados em língua portuguesa disponíveis para o treinamento de modelos de aprendizado de máquina, impactando na precisão dos sistemas desenvolvidos. Recentemente, organizou-se uma coletânea aberta de redações no modelo do ENEM e suas avaliações correspondentes, o que abre novas possibilidades para a pesquisa na área.

Nesse contexto, esta proposta de trabalho sugere o uso de técnicas de aprendizado profundo para avaliar redações de acordo com as competências do ENEM, baseando-se em modelos conhecidos de processamento de linguagem natural, pré-treinados com foco na língua portuguesa. As próximas seções, em ordem, detalham as justificativas, os objetivos, a metodologia proposta e, por fim, o cronograma esperado do projeto.

2 Justificativa

A avaliação de textos dissertativo-argumentativos no ENEM é um processo que demanda muito tempo devido ao modelo de correção utilizado e à falta de equilíbrio entre o número de redações e o número de corretores disponíveis, como mencionado por Lesme (2021) [2]. Por conta disso, a classificação automática de redações no modelo ENEM já foi objeto de estudos anteriores na área de inteligência artificial, incluindo, dentre outros, o desenvolvimento de um classificador de Bayes por Bazelato e Amorim (2013) [3] e a extração e análise automática do volume de aspectos textuais das dissertações por Veloso e Amorim (2017) [4].

Como o ENEM segue diretrizes bem definidas, a tarefa de correção tem padrões intrínsecos que podem ser detectados e compreendidos com o uso do aprendizado profundo. No entanto, a maioria dos trabalhos anteriores foi baseada em técnicas de inteligência artificial mais elementares, que não lidam com grandes volumes de textos para classificação. Isso ocorre porque a correção automática de redações em português é uma área que enfrenta o desafio da falta

de dados na linguagem, como mencionado por Veloso e Amorim (2017) [4], o que limita a precisão do treinamento.

Um estudo recente de Marinho *et al.* (2021) [5] organizou uma coletânea aberta de redações no modelo do ENEM e suas avaliações correspondentes, atribuídas por profissionais da área de acordo com os critérios de correção do exame. Assim, com esse maior conjunto de dados, é possível investigar o uso de técnicas mais avançadas de aprendizado supervisionado na correção automática e comparar o desempenho de modelos mais robustos com as abordagens anteriormente realizadas nesse campo.

3 Objetivos

Baseando-se na quantidade de dados de redações no modelo do ENEM disponibilizadas abertamente por Marinho *et al.* (2021) [5], assim como nas motivações anteriormente apresentadas, almeja-se, com a proposição desse trabalho:

- Utilizar técnicas de aprendizado profundo para classificar redações em notas de acordo com as competências do ENEM.
- Investigar arquiteturas de redes neurais para o ajuste fino de modelos pré-treinados que possibilitem a avaliação automática de redações.
- Avaliar os modelos de correção automática gerados, baseando-se em suas medidas de precisão.
- Adquirir familiaridade com bibliotecas de aprendizado de máquina e de avaliação estatística de dados.

4 Metodologia

Para realizar a avaliação automática de redações no modelo do ENEM, pretende-se utilizar uma abordagem baseada em técnicas de aprendizado profundo, a partir do ajuste fino da arquitetura de redes neurais conhecida como BERT (**B**idirectional **E**ncoder **R**epresentations from **T**ransformers), apresentada pelo artigo de Devlin *et al.* (2018) [6].

O BERT é um modelo pré-treinado de grande escala, especializado em tarefas envolvendo o processamento de linguagem natural, que utiliza uma arquitetura baseada em um *transformer* para capturar relações de dependência entre palavras em textos. Essa arquitetura é especialmente adequada para lidar com a complexidade da língua portuguesa, devido à existência de uma ordem livre de palavras e dependências sintáticas complexas dentro da linguagem.

Para o desenvolvimento do projeto, objetiva-se utilizar o BERTimbau, a versão pré-treinada do BERT em português desenvolvida por Souza *et al.* (2019) [7]. Esse modelo foi construído com base em uma grande quantidade de textos em português, permitindo a captura de características específicas da língua que possam aparecer nas redações a serem avaliadas.

A partir do BERTimbau, pretende-se criar uma arquitetura capaz de refinar o *transformer* pré-treinado para cumprir a tarefa de correção automática dos textos de acordo com as cinco competências exigidas pelo ENEM. Essa arquitetura será treinada no conjunto de dados contendo as redações e suas respectivas notas, disponibilizado por Marinho *et al.* (2021) [5]. A avaliação do desempenho da arquitetura será realizada por meio de métricas como acurácia, *Quadratic Weighted Kappa* (QWK) (Cohen, 1968 [8]), *F1-score*, etc.

Para a implementação da arquitetura de redes neurais baseada no BERTimbau, pretende-se utilizar a biblioteca do Tensorflow, que fornece uma ampla gama de ferramentas para o treinamento e avaliação de modelos de aprendizado profundo. Para a etapa de manipulação do *dataset*, pretende-se utilizar, também, a biblioteca Pandas, que oferece uma forma fácil de lidar e extrair informações de um grande conjunto de dados. Além disso, para acelerar o processo de treinamento, serão utilizadas máquinas com GPU disponibilizadas pela ferramenta Google Colab, que permitem uma execução mais rápida e eficiente de tarefas computacionalmente intensivas.

Por fim, a análise dos resultados obtidos será realizada para identificar arquiteturas que tenham um bom desempenho na classificação dos textos dissertativo-argumentativos e para avaliar o uso das técnicas de aprendizado profundo em relação a outras abordagens já realizadas.

5 Cronograma

Tarefas	Meses								
	Abril	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro
Revisão bibliográfica	×	×	×						
Estudo dos modelos de linguagem pré-treinados (BERT e BERTimbau) e do <i>dataset</i> de redações.	×	×	×						
Elaboração da arquitetura de ajuste fino para avaliar as competências em separado.		×	×	×	×				
Aprimoramento da estrutura dos avaliadores para o cálculo da nota final.					×	×	×		
Avaliação das arquiteturas de aprendizado e análise dos resultados obtidos.						×	×		
Elaboração da monografia.						×	×	×	
Produção do pôster e preparação da apresentação final.								×	×

Referências

- [1] BRASIL. *A redação no Enem 2022: cartilha do participante*. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). Brasília, 2022.
- [2] LESME, A. *Enem 2021: corretores podem corrigir até 200 redações por dia*. Brasil Escola, 2021. Disponível em: <https://vestibular.brasilecola.uol.com.br/enem/enem-2021-corretores-podem-corrigir-ate-200-redacoes-por-dia/351641.html>. Acesso em: 24 de abril de 2023.
- [3] BAZELATO, B. S.; AMORIM, E. C. F. *A Bayesian Classifier to Automatic Correction of Portuguese Essay*. XVIII Conferência Internacional sobre Informática na Educação. Porto Alegre, 2013. p. 779-782.
- [4] VELOSO, A.; AMORIM, E. C. F. *A multi-aspect analysis of automatic essay scoring for Brazilian Portuguese*. Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia, 2017. p. 94-102.
- [5] MARINHO, J. C.; ANCHIÊTA, R. T.; MOURA, R. S. *Essay-BR: a Brazilian Corpus of Essays*. Anais do III Dataset Showcase Workshop. Rio de Janeiro, 2021. Porto Alegre: Sociedade Brasileira de Computação. p. 53-64.
- [6] DEVLIN, J.; LEE, K.; TOUTANOVA, K.; CHANG, M. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv preprint arXiv:1810.04805, 2019.
- [7] SOUZA, F.; LOTUFO, R.; NOGUEIRA, R. *BERTimbau: Pretrained BERT Models for Brazilian Portuguese*. Intelligent Systems. Cham, 2020. Springer International Publishing. p. 403-417.
- [8] COHEN, J. *Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit*. Psychological bulletin, volume 70, issue 4. 1968.