

Universidade Federal do Agreste de Pernambuco (UFAPE)  
Curso de Ciência da Computação

# Modelo BERT: Estrutura e Aplicações no PLN

Aluno: João Guilherme Benjamin Alves de Rezende

Professor: Ryan Ribeiro

28 de novembro de 2024

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Como o BERT Funciona</b>	<b>2</b>
2.1	Pré-treinamento . . . . .	2
2.2	Fine-tuning . . . . .	2
<b>3</b>	<b>Topologia da Arquitetura</b>	<b>2</b>
<b>4</b>	<b>Aplicações e Resultados</b>	<b>2</b>
4.1	Visualizações . . . . .	3
<b>5</b>	<b>Conclusão</b>	<b>4</b>
<b>6</b>	<b>Referências</b>	<b>5</b>

# 1 Introdução

O modelo **BERT** (Bidirectional Encoder Representations from Transformers) é um dos avanços mais significativos em Processamento de Linguagem Natural (PLN). Ele é baseado na arquitetura Transformer e foi projetado para compreender o contexto bidirecional em textos, possibilitando tarefas como análise sintática, tradução automática, classificação de texto, e muito mais.

## 2 Como o BERT Funciona

O BERT utiliza duas etapas principais:

### 2.1 Pré-treinamento

Durante esta fase, o modelo é treinado em tarefas como:

- **Masked Language Model (MLM):** Algumas palavras do texto são mascaradas, e o modelo tenta prever essas palavras com base no contexto ao redor.
- **Next Sentence Prediction (NSP):** Dados dois segmentos de texto, o modelo aprende a identificar se um segue logicamente o outro.

### 2.2 Fine-tuning

Nesta etapa, o BERT é ajustado para tarefas específicas, como classificação de texto ou análise de sentimentos, adicionando camadas densas ao modelo pré-treinado.

## 3 Topologia da Arquitetura

O BERT é composto por várias camadas Transformer. Cada camada inclui:

- **Mecanismo de Self-Attention:** Permite que o modelo avalie relações entre palavras em uma frase.
- **Feedforward Neural Network:** Camadas densas aplicadas após a atenção.
- **Embedding Posicional:** Adiciona informações sobre a posição das palavras na sequência.

A versão *base* do BERT possui 12 camadas, enquanto a versão *large* possui 24 camadas.

## 4 Aplicações e Resultados

Na Lista 2, utilizamos o BERT para tarefas como:

- Tokenização e extração de embeddings.
- Análise de similaridade entre tokens.

- Visualização de distribuições e frequências.
- Comparação com abordagens baseadas em Stanza e spaCy.

## 4.1 Visualizações

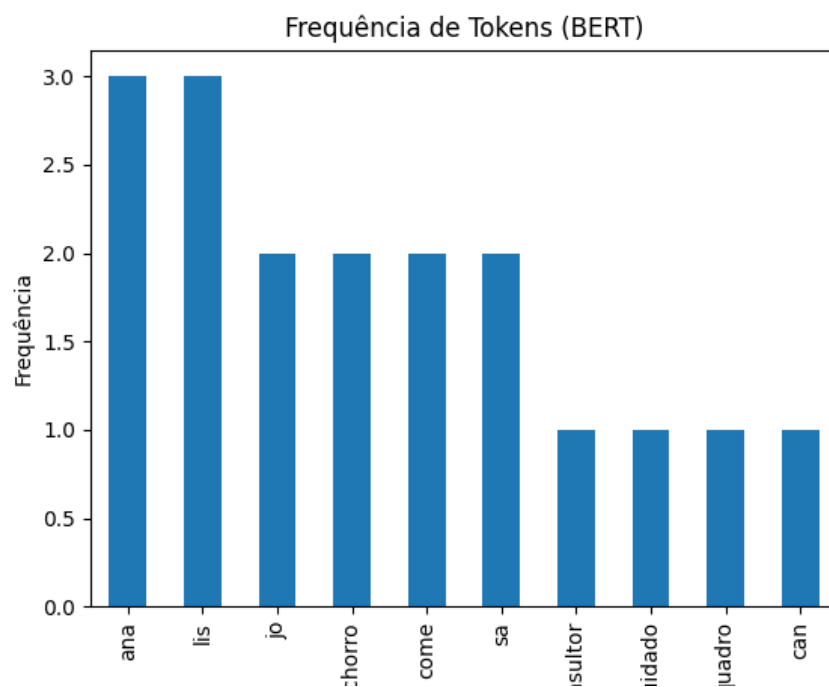


Figura 1: Frequência dos tokens gerados pelo BERT.

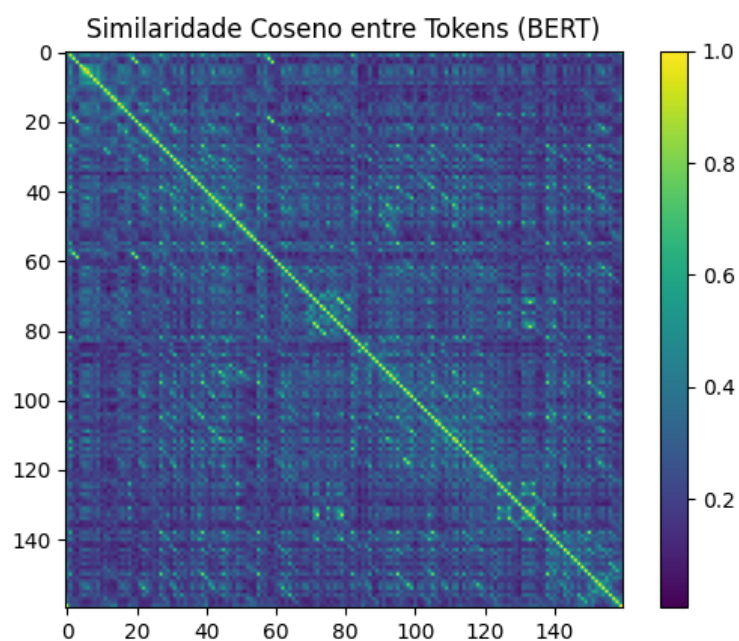


Figura 2: Matriz de similaridade coseno entre tokens.

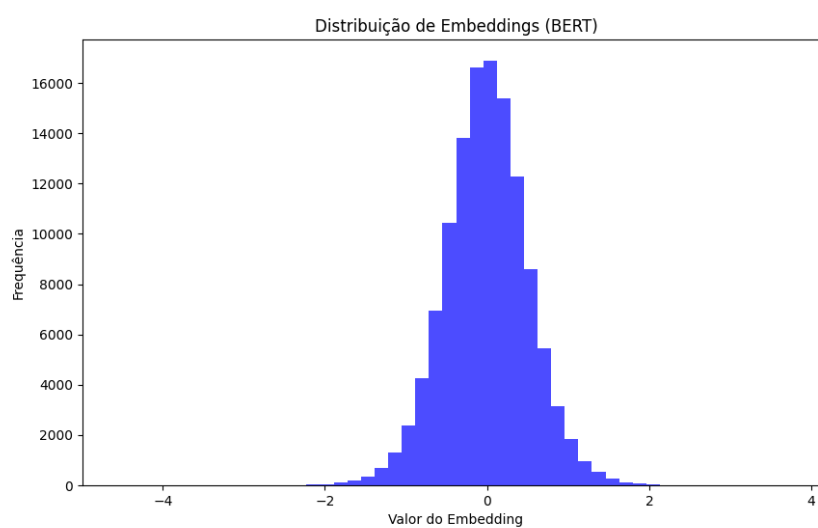


Figura 3: Distribuição dos valores de embeddings extraídos.

## 5 Conclusão

O BERT é uma ferramenta poderosa que revolucionou o PLN, oferecendo uma compreensão profunda do contexto linguístico. A integração deste modelo em nossas análises mostrou melhorias significativas na extração de informações e na compreensão semântica, como evidenciado pelos gráficos e resultados apresentados.

## 6 Referências

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv:1810.04805.
- <https://huggingface.co/docs/transformers>
- <https://pytorch.org/>