

Comentário crítico *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*

Aluno: Maria Fernanda Bosco

RA: 183544

O artigo que introduz o BERT, representando um marco no processamento de linguagem natural. Sua principal contribuição foi demonstrar que pré-treinar modelos de linguagem em larga escala, usando apenas a parte encoder da arquitetura do Transformer, permite transferir conhecimento para uma ampla gama de tarefas de NLP com excelentes resultados. Isso consolidou a abordagem de *pre-training* + *fine-tuning* como paradigma dominante na área.

Entre as inovações técnicas, destacam-se o uso do Masked Language Modeling (MLM), que possibilitou o aprendizado bidirecional de representações textuais. Além disso, também houve a introdução da tarefa de Next Sentence Prediction (NSP), que buscava capturar relações intersentenciais e tornou o modelo útil em tarefas de pares de frases, como QA.

Os resultados foram muito bons, com o BERT estabelecendo *state-of-the-art* em 11 benchmarks, mostrando que um único modelo pré-treinado poderia ser adaptado a tarefas muito distintas com ajustes mínimos.

Por outro lado, algumas limitações também se destacaram. O custo computacional do pré-treinamento é elevado, restringindo o acesso a instituições com grande poder computacional. Outro ponto é que o BERT, embora poderoso para entendimento, não é adequado para geração de texto.

Em síntese, o artigo do BERT foi revolucionário por popularizar a transferência de conhecimento em NLP por meio de embeddings contextuais bidirecionais, estabelecendo as bases para a onda de modelos subsequentes. Apesar de seus custos e limitações, sua influência permanece central e incontornável no avanço da área.