



Processamento de Linguagem Natural com Transformers

Processamento de Linguagem Natural com Transformers

Definindo a Arquitetura BERT



BERT, que é a sigla para "Bidirectional Encoder Representations from Transformers", é um modelo de aprendizado profundo (Deep Learning) que usa uma arquitetura Transformer para entender a linguagem natural. Desenvolvido pelo Google AI Language, BERT causou uma grande mudança no campo de Processamento de Linguagem Natural (PLN) e aprimorou significativamente a performance de uma variedade de tarefas de PLN.

BERT é notável por seu treinamento bidirecional. Modelos de linguagem tradicionais treinam o texto da esquerda para a direita ou da direita para a esquerda, enquanto BERT faz ambos ao mesmo tempo. Isso é crítico porque a compreensão de uma palavra pode depender tanto do contexto que vem antes quanto do contexto que vem depois. Essa abordagem bidirecional oferece ao BERT uma compreensão muito mais rica da gramática e do significado das palavras.

O BERT é treinado em dois tipos de tarefas. Primeiro, o "Masked Language Model" (MLM), onde 15% das palavras são mascaradas (ou seja, escondidas) e o modelo tenta prever o que são essas palavras com base no contexto restante. A segunda tarefa é a "Next Sentence Prediction" (NSP), na qual o modelo aprende a prever se uma sentença é o próximo seguimento lógico de outra sentença.

Esse treinamento pode ser feito em uma grande quantidade de texto não rotulado, como a Wikipédia, e prepara o modelo para ser afinado em tarefas específicas, como classificação de texto, resposta a perguntas, resumo e tradução.

O BERT representa uma grande inovação em relação aos modelos de linguagem anteriores, pois ele pode capturar informações complexas de contexto de uma maneira que os outros modelos não conseguem. Isso se traduz em uma melhor performance em uma série de tarefas de PLN.

Recomendamos a leitura do paper original sobre o BERT:

<https://arxiv.org/abs/1810.04805>