



Processamento de Linguagem Natural com Transformers

Processamento de Linguagem Natural com Transformers

O Que São LLMs (Large Language Models) ?

LLMs, ou "Large Language Models" (Modelos de Linguagem de Grande Escala em tradução livre), são modelos de aprendizado de máquina treinados em enormes conjuntos de dados de texto para realizar tarefas de previsão de sequências e são amplamente usados em Processamento de Linguagem Natural (PLN). Esses modelos são projetados para entender e gerar texto de uma maneira que seja coerente e, idealmente, indistinguível do texto gerado por seres humanos.

Aqui estão algumas características e informações sobre os LLMs:

Arquitetura: A maioria dos LLMs modernos, como o GPT (Generative Pre-trained Transformer) da OpenAI, utiliza a arquitetura Transformer, que foi introduzida no famoso artigo de 2017 chamado "Attention is All You Need".

<https://arxiv.org/abs/1706.03762>

Pré-treinamento e Ajuste Fino: O treinamento de LLMs geralmente envolve duas fases. Primeiro, o modelo é pré-treinado em um grande conjunto de dados para aprender padrões gerais na linguagem (isso pode incluir gramática, fatos sobre o mundo, etc.). Depois, o modelo pode ser afinado em um conjunto de dados específico para uma tarefa particular (como tradução, geração de texto ou classificação). Aliás, é neste item que está uma das vantagens de usar um LLM. Não precisamos realizar a primeira fase, somente a segunda, o que requer um volume de dados bem menor.

Capacidade: Os LLMs contêm uma grande quantidade de parâmetros. Por exemplo, o GPT-3, um LLM criado pela OpenAI, possui 175 bilhões de parâmetros. O número de parâmetros indica a enorme capacidade desses modelos de armazenar e processar informações.

Aplicações: Os LLMs têm uma ampla gama de aplicações, desde responder perguntas, redigir texto, traduzir idiomas, ajudar no desenvolvimento de software, entre outras.

Desafios: Apesar de sua impressionante capacidade de geração de texto, os LLMs têm limitações. Eles não "entendem" o texto da maneira que os humanos entendem e, por vezes, podem gerar informações imprecisas ou tendenciosas. Além disso, sua confiabilidade e ética são áreas de pesquisa ativa, já que podem propagar preconceitos presentes nos dados em que foram treinados.

Interatividade: Uma das forças dos LLMs é sua capacidade de interagir em tempo real. Eles podem ser usados em chatbots, assistentes virtuais e outras aplicações que exigem interação em linguagem natural.

Os LLMs representam um avanço significativo na área de PLN e estão continuamente evoluindo, com novas versões e melhorias sendo lançadas regularmente.