

GPT 和 BERT

GPT (Generative Pre-trained Transformer)和 BERT (Bidirectional Encoder Representations from Transformers) 是自然语言处理(NLP)领域的两大代表，我们通过了解它们的特点和用途，也就能了解 NLP 的一些基本概念。

应用场景

- GPT：目标是从给定的前缀文本中预测下一个单词，适合于**生成式**任务。由于其强大的文本生成能力，广泛应用于如文章创作、故事写作、对话系统等需要生成连贯、上下文相关的文本的任务中。
- BERT：除了掩码语言模型外，还包含了一个额外的任务——下一句预测(NSP)，旨在理解两个句子之间的关系。这使得BERT更适用于需要理解或分析文本的任务，如**文本分类**、**情感分析**、**命名实体识别(NER)**、**问答系统(特别是理解问题和查找答案方面)**、**语句相似度比较**等。此外，BERT还可以作为特征提取器，将文本转换为固定长度的向量表示，供其他机器学习模型使用。

模型架构

- GPT：主要使用了Transformer的解码器部分，专注于从左到右的语言模型训练。它在生成文本时只能依据已生成的部分来预测下一个词。
- BERT：则采用了Transformer的编码器部分，并且是双向的，可以同时考虑一个词语在其输入句子中的左右上下文信息。这种双向性通过掩码语言模型(MLM)技术实现。BERT的输入由三种嵌入组成：词嵌入(Token Embeddings)、段落嵌入(Segment Embeddings)和位置嵌入(Position Embeddings)。这使得BERT能够处理句子间关系以及序列数据的位置信息。

参数量

- GPT系列
 - GPT-1: 约1.17亿，即0.117B。
 - GPT-2: 小型为1.5亿参数，即0.15B；中型为7.62亿参数，即0.762B；大型为15亿参数，即1.5B。
 - GPT-3: 1750亿参数，即175B。
- BERT系列
 - BERT Base: 约1.1亿参数，即0.11B。
 - BERT Large: 约3.4亿参数，即0.34B。