

大模型面经之bert和gpt区别

瓦力算法学研所 2024年07月28日 14:08 安徽

◇◇ 面试专栏 ◇◇

作者：喜欢卷卷的瓦力



瓦力算法学研所

我们是一个致力于分享人工智能、机器学习和数据科学方面理论与应用知识的公众号。我...
117篇原创内容

公众号

本篇介绍bert和gpt区别。

BERT和GPT是自然语言处理（NLP）领域中的两种重要预训练语言模型，它们在多个方面存在显著的区别。以下是对BERT和GPT区别的详细分析

一、模型基础与架构

- BERT：
 - 全称：Bidirectional Encoder Representations from Transformers。
 - 架构：基于Transformer的编码器部分进行堆叠构建，通过预训练和微调两个阶段来生成深度的双向语言表征。
 - 特点：使用了Transformer的encoder部分，通过双向语言模型预训练来学习上下文相关的词表示。
- GPT：
 - 全称：Generative Pre-trained Transformer。
 - 架构：基于Transformer的解码器部分，通过自回归语言模型预训练来学习生成连贯文本的能力。
 - 特点：采用了自回归语言模型的预训练方式，逐步生成下一个词语，以此生成连贯的文本。

二、训练方式与任务

- BERT:
 - 训练任务：主要包括掩码语言模型（Masked Language Model, MLM）和下一句预测（Next Sentence Prediction, NSP）。
 - MLM：在输入序列中随机掩盖一些词语，要求模型预测这些被掩盖的词语。
 - NSP：判断两个句子是否是连续的文本序列。
 - 训练方式：双向预训练，同时考虑前后文信息。
- GPT:
 - 训练任务：自回归语言模型预训练。
 - 训练方式：单向预训练，从左到右生成文本，只能依赖已生成的上文来预测下一个词语。

三、上下文理解能力

- BERT:
 - 由于采用了双向语言模型，BERT能够同时考虑前后文信息，因此在理解整个句子或段落时表现出色。
 - 适用于需要理解整个文本的任务，如分类、命名实体识别和句子关系判断等。
- GPT:
 - 作为单向模型，GPT在生成文本时只能依赖已生成的上文，因此在处理需要理解整个文本的任务时可能表现不足。
 - 但其生成文本的能力较强，适用于各种生成式的NLP任务。

四、应用领域

- BERT:
 - 因其强大的上下文理解能力，BERT在多种NLP任务中都有广泛应用，如情感分析、问答系统、命名实体识别等。
- GPT:
 - GPT的强项在于生成连贯、有逻辑性的文本，因此在文本生成、机器翻译、对话系统等任务中表现出色。

BERT和GPT在模型基础、训练方式、上下文理解能力和应用领域等方面都存在显著差异。BERT更适合于需要理解整个文本的任务，而GPT则更擅长于生成式的NLP任务。在实际应用中，可以根据具体任务的需求选择适合的模型。

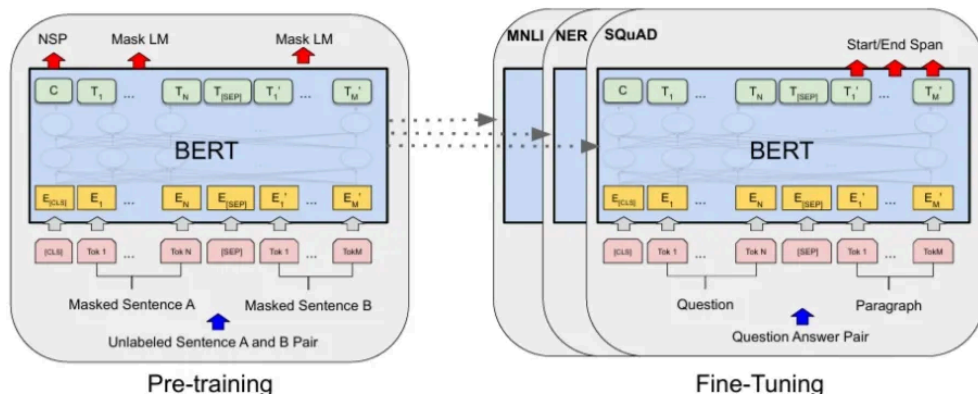


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers).

bert

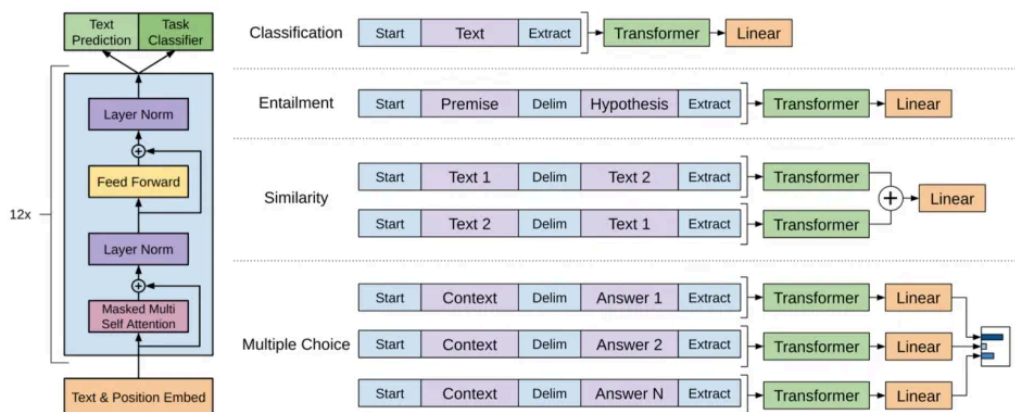


Figure 1: (left) Transformer architecture and training objectives used in this work. (right) Input transformations for fine-tuning on different tasks. We convert all structured inputs into token sequences to be processed by our pre-trained model, followed by a linear+softmax layer.

gpt1

想要获取技术资料的同学欢迎关注公众号，进群一起交流~



瓦力算法学研所

我们是一个致力于分享人工智能、机器学习和数据科学方面理论与应用知识的公众号。我...

117篇原创内容