

GPT 经验篇

来自：AiGC面试宝典

 宁静致远

2024年01月27日 19:14



扫码
查看更

- GPT 经验篇
 - 一、gpt源码past_key_value是干啥的？
 - 二、gpt onebyone 每一层怎么输入输出？
 - 三、bert和gpt有什么区别
 - 四、文本生成的几大预训练任务？
 - 五、讲讲T5和Bart的区别，讲讲bart的DAE任务？
 - 六、讲讲Bart和Bert的区别？
 - 七、gpt3和gpt2的区别？
 - 致谢

一、gpt源码past_key_value是干啥的？

在GPT (Generative Pre-trained Transformer) 中，past_key_value是用于存储先前层的注意力权重的结构。在进行推理时，过去的注意力权重可以被重复使用，避免重复计算，提高效率。

二、gpt onebyone 每一层怎么输入输出？

在GPT One-by-One中，每一层的输入是上一层的输出。具体而言，输入是一个序列的嵌入表示（通常是词嵌入），并通过自注意力机制和前馈神经网络进行处理，得到输出序列的表示。

三、bert和gpt有什么区别

BERT (Bidirectional Encoder Representations from Transformers) 和GPT (Generative Pre-trained Transformer) 是两种不同类型的预训练语言模型。主要区别在于：

BERT是一个双向编码器，它预测输入序列中的缺失部分，因此可以用于多种任务，如文本分类、命名实体识别等。

GPT是一个单向解码器，它生成文本的下一个单词，因此主要用于生成型任务，如文本生成、对话生成等。

四、文本生成的几大预训练任务？

- GPT (Generative Pre-trained Transformer) 系列：包括GPT、GPT-2、GPT-3等。这些模型使用Transformer架构进行预训练，在大规模语料上学习语言模型，能够生成连贯、具有语义的文本。
- BART (Bidirectional and Auto-Regressive Transformer)：BART是一种基于Transformer的生成式预训练模型。它通过自回归解码器实现文本生成，通过自编码器预训练目标来重构输入文本，能够生成流畅、连贯的文本。
- T5 (Text-to-Text Transfer Transformer)：T5是一种通用的文本生成模型，使用了编码器-解码器结构。它将不同的自然语言处理 (NLP) 任务转换为文本到文本的转换任务，可用于机器翻

译、摘要生成、问题回答等多个NLP任务。

- XLNet: XLNet是一种基于Transformer架构的预训练模型，采用了自回归和自编码器的组合方式进行训练。它在语言建模任务上引入了全局的上下文信息，能够生成更加准确和连贯的文本。
- UniLM (Unified Language Model) : UniLM是一种多任务学习的预训练模型，将不同的自然语言处理任务转化为统一的生成式任务。它可以用于文本摘要、问答系统、机器翻译等多个任务。

五、讲讲T5和Bart的区别，讲讲bart的DAE任务？

T5 (Text-to-Text Transfer Transformer) 和Bart (Bidirectional and Auto-Regressive Transformer) 是两个常见的预训练模型，它们之间的区别如下：

- T5是一种基于Transformer的通用文本生成模型。T5的训练目标是将不同的自然语言处理 (NLP) 任务统一为文本到文本的转换任务。它采用了编码器-解码器结构，通过输入一个自然语言文本，输出另一个相关的自然语言文本，可以应用于机器翻译、摘要生成、问题回答等多个NLP任务。
- Bart是建立在T5模型基础上的一个变种，它专注于生成式任务。Bart模型使用了自回归解码器，通过训练一个自编码器来重构原始文本，同时采用了标准的语言模型预训练目标，从而使生成的文本更加流畅和连贯。Bart的主要应用领域包括文本生成、摘要生成、对话系统等。

在任务类型上，T5更加通用，适用于多种NLP任务的文本转换，而Bart则更加专注于生成式任务，并且在生成文本的质量和连贯性上有所优化。

关于Bart的DAE (Denoising AutoEncoder) 任务，它是Bart模型的一种预训练目标。DAE任务要求模型从输入的有噪声的文本中恢复原始的无噪声文本。通过在训练过程中向输入文本中添加噪声，并要求模型重建无噪声的文本，Bart可以学习到更好的文本表示和重构能力，从而提高生成文本的质量和准确性。

六、讲讲Bart和Bert的区别？

Bart和Bert是两个不同的预训练模型，它们之间的区别如下：

- Bart是一种基于Transformer的生成式预训练模型，主要应用于文本生成、摘要生成、对话系统等任务。Bart采用了自回归解码器，通过自编码器预训练目标来重构输入文本，从而生成流畅、连贯的文本。
- Bert (Bidirectional Encoder Representations from Transformers) 是一种双向的预训练模型，用于生成文本的上下文表示。与Bart不同，Bert采用了双向的Transformer编码器，通过将上下文的信息融合到表示中，提供了更全面的语境理解能力。Bert主要应用于词嵌入、文本分类、命名实体识别等任务。

总体上说，Bart侧重于生成式任务和文本生成，而Bert侧重于上下文表示和语境理解。它们在模型结构和应用场景上存在一定的差异。

七、gpt3和gpt2的区别？

GPT-3和GPT-2是由OpenAI开发的两个语言模型。它们的区别主要在于规模和功能上的不同。

GPT-3是目前最大的语言模型，具有1750亿个参数，而GPT-2则有15亿个参数。

由于GPT-3规模更大，它在自然语言处理任务上的表现更好，并且能够生成更连贯、更具逻辑性的文本。GPT-3还支持零样本学习，即可以在没有对特定任务进行显式训练的情况下执行各种语言任务。

另一个区别是GPT-3在文本生成方面的能力更强大，可以生成更长的文本，而GPT-2的生成长度有一定的限制。此外，GPT-3的使用需要更高的计算资源和成本。