

Bert

BERT 是一个预训练语言模型，旨在通过大量文本数据进行训练，学习语言的上下文信息。它被设计为一种“双向”模型，这意味着它不仅可以从左到右处理文本，还能从右到左处理，从而更好地捕捉到词语之间的双向关系。BERT 在多个自然语言处理（NLP）任务中都表现出了出色的效果，比如文本分类、命名实体识别（NER）、问答系统等。

1.BERT的核心思想：

双向Transformer：

BERT 采用的是仅使用 Transformer 编码器（Encoder）的结构，不使用解码器。它的核心思想是：在预训练时，通过双向（双向注意力）机制让模型能够同时关注左右两侧的上下文信息。

GPT中解码器中的掩码注意力机制导致了其单向性，而BERT仅使用Transformer的编码器结构，编码器中没有掩码注意力机制，所以BERT具有双向性。

传统的语言模型大多是单向的，即从左到右（或右到左）处理文本。但语言的意义往往依赖于上下文，因此单向模型容易出现理解偏差。BERT 的关键创新就是“**双向**”理解：它可以同时从左到右和从右到左来获取上下文信息，从而获得更全面的语言理解。

2.BERT的预训练与微调：

预训练任务1：MLM（Masked Language Model）

BERT随机遮挡输入文本中的部分单词，让模型去预测被遮挡的词是什么。

这样可以强迫 BERT 关注整个句子的上下文，从而形成强大的语义理解能力。

具体的Mask方式：

80% 的时间：用 [MASK] 代替单词。（如 "sat" → [MASK]）

10% 的时间：用一个随机的单词替换。（如 "sat" → "dog"）

10% 的时间：保持原单词不变。（如 "sat" → "sat"）

这种随机替换方式可以让BERT学会区分正确和错误的单词。

预训练任务 2：NSP（Next Sentence Prediction）

让 BERT 学会理解 句子之间的关系，从而增强跨句子的理解能力（这对于 QA、阅读理解等任务至关重要）。

训练过程：

1.输入两句话

2.BERT需要判断B是否是A的下一句：

正样本（IsNextSentence, 50%） 负样本（NotNextSentence, 50%）

NSP的作用：

增强 BERT 在长文本（如文章级别文本）中的理解能力。

让 BERT 学会推理两个句子之间的关系（适用于 QA、摘要生成等任务）。

微调：

预训练完BERT之后，相当于它已经掌握了通用的语言知识，但还没有针对具体任务进行优化。

微调阶段，我们会在**具体任务的数据集**上训练 BERT，使其适应不同的 NLP 任务，如：

- **文本分类**（如情感分析、垃圾邮件检测）
- **问答系统**（如 SQuAD 数据集）
- **命名实体识别（NER）**
- **文本生成**

只需在少量任务数据上训练少量 epoch（通常 2-4 轮），就能达到极佳效果。

保留 BERT 预训练得到的知识，只对任务相关部分进行调整

微调的具体流程：

1.加载预训练好的BERT模型

2.添加任务特定的输出层：

例如：

- 文本分类任务 → 连接一个全连接层（softmax 预测类别）。
- 问答任务 → 连接一个双线性层，预测答案的起始和结束位置。

3.在目标任务数据集上训练：

使用 **带监督的任务数据集**（如 IMDB 情感分类数据）。

采用 **微调（fine-tuning）** 方法训练模型，而不是重新训练整个 BERT。

4.在测试集上评估性能