## AI面试百题训练营1207题目

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 难度 | 题目 | 注意事项 |
| 1 | 中等 | RNN 发生梯度消失的原因是什么？ | 分析RNN反向传播的公式 |
| 2 | 中等 | RNN 中使用 ReLU 可以解决梯度消失问题吗？ | 根据RNN梯度消失的原因来分析 |
| 3 | 中等 | LSTM为什么能解决梯度消失/爆炸的问题 ？ | LSTM的哪个机制起到了作用 |
| 4 | 中等 | GRU 和 LSTM 的区别 | 结构上有哪些不同 |
| 5 | 中等 | LSTM算法的不足之处有哪些？ | 从LSTM的结构和训练等方面来分析 |

## AI面试百题训练营1208题目

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 难度 | 题目 | 注意事项 |
| 1 | 中等 | 写出Attention的公式  Attention机制，里面的q,k,v分别代表什么 | 理解Attention 的公式 |
| 2 | 中等 | Transformer 中使用多头注意力的好处是什么 | 从信息抽取和表达的维度来理解 |
| 3 | 中等 | Attention 中 self-attention 的时间复杂度 | 理解self-attention 的计算过程 |
| 4 | 中等 | Transformer 中的 encoder 和 decoder 的异同点 | 从结构和作用来分析 |
| 5 | 中等 | Bert 和 GPT 的异同点 | 从原理，结构，语言模型，输入参数等维度来分析 |

## AI面试百题训练营1209题目

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 难度 | 题目 | 注意事项 |
| 1 | 中等 | Self-attention对比RNN和CNN在处理NLP任务时分别有哪些优势 | 对结合NLP任务的特点对三者的结构进行比较 |
| 2 | 中等 | CNN 中的 1\*1 卷积有什么作用 | 从1\*1卷积的计算过程来分析 |
| 3 | 中等 | CNN 和 RNN 在 NLP 应用中各自的优缺点 | 结合NLP任务的特点对两者的结构进行比较 |
| 4 | 中等 | 为什么transformer块使用LayerNorm而不是BatchNorm？ | 从计算过程和NLP任务的特点来分析 |
| 5 | 中等 | Self-Attention 中的计算为什么要使用根号dk缩放 | 如果不进行缩放会造成什么后果 |

## AI面试百题训练营1210题目

|  |  |  |  |
| --- | --- | --- | --- |
| 序号 | 难度 | 题目 | 注意事项 |
| 1 | 中等 | 为什么self-attention model在长距离序列中如此强大 | 从结构上来分析，对比为什么RNN，CNN不适合长距离序列，而Transformer 适合 |
| 2 | 中等 | Bert 类模型中的绝对位置 embedding 和 相对位置 embedding 怎么理解，各自的优缺点和使用场景 | 从位置embedding的计算函数和应用场景来理解 |
| 3 | 中等 | Bert 的预训练任务有哪些，各自的作用是什么 | 对Bert生态有个大致了解 |
| 4 | 中等 | Roberta、Albert 分别对 Bert 做了哪些改进 | 从结构，训练机制和方式等方面来分析 |
| 5 | 中等 | XLNet 如何实现 Permutation Language Model。 | 通过对比Bert 来理解XLNet的结构 |