NLP natural language processing

NLU natural language understanding

Transformer是基础。语言大模型架构：①纯编码器 ②纯解码器（有自回归）decoder only 【gpt所用】 ③编码器-解码器架构

①纯编码器：自注意力 处理数据时 注意到序列中的元素与其他元素的关系 关联度（权重）：如语义关联度 在自注意力机制中，对于输入序列中的每个元素，模型会计算一个查询向量（Query）、一个键向量（Key）和一个值向量（Value）。查询向量用于询问序列中每个元素与其他元素的关系，键向量用于索引每个元素的特征，而值向量则包含了每个元素的信息。通过将查询向量与所有键向量进行点积计算，可以得到注意力权重，这些权重用于加权求和值向量，从而得到最终的输出向量。

前馈网络 单向的 深度神经网络的基础 速度慢：encoder decoder 的出现

MLM（Masked Language Model，掩码语言模型） 尤其应用于BERT

③编码器-解码器 cross attention（交叉注意力） 掩码自注意力

Teacher Forcing是指在训练序列生成模型时，将前一个时间步的真实输出（即标签文本）作为下一个时间步的输入，而不是使用模型在训练过程中的实际输出。

**与自回归的区别**：自回归模式下，模型在每个时间步都使用自己前一步的输出作为输入，而Teacher Forcing则是使用真实标签的输出。 自回归：自己前一步输出做为输入

RNN生成文本的连贯性低于自回归模型（如GPT），多用于理解任务而非生成任务。

#### ****LSTM****引入遗忘门、输入门、输出门，解决传统RNN的梯度消失问题。细胞状态（cell state）长期保存关键信息，适用于长序列任务（如机器翻译）

**与Transformer对比**： transformer 计算量更大，但优化手段可以降低成本

**RNN优势**：参数共享、内存占用低，适合实时处理（如自动驾驶激光雷达点云分割）。

**Transformer优势**：自注意力机制全局建模，适合长文本生成和大规模预训练。

Bert（mask掩码） 编码器的叠加 (transformer再到bert：预训练语言模型--》词向量)区别：trans是翻译等多模态任务 bert是文本分类回答等理解任务 预训练 【图像cv 语言nlp】

1. 预训练 定义：用已经通过大量数据训练好的模型A，去完成新的小（或大的也可以，节省时间）数据任务模型B，即使用了浅层参数。【但任务A和B相似，所以浅层参数相似】

：机器学习 偏数学 （《统计学习方法》） 本质：线性变换

深度学习 ai 大数据支持 本质：非线性变换（激活函数Relu）

1. 统计语言模型 统计方法解决问题，词出现的概率问题 n元统计语言模型

Transformer nlp预训练就是生成词向量