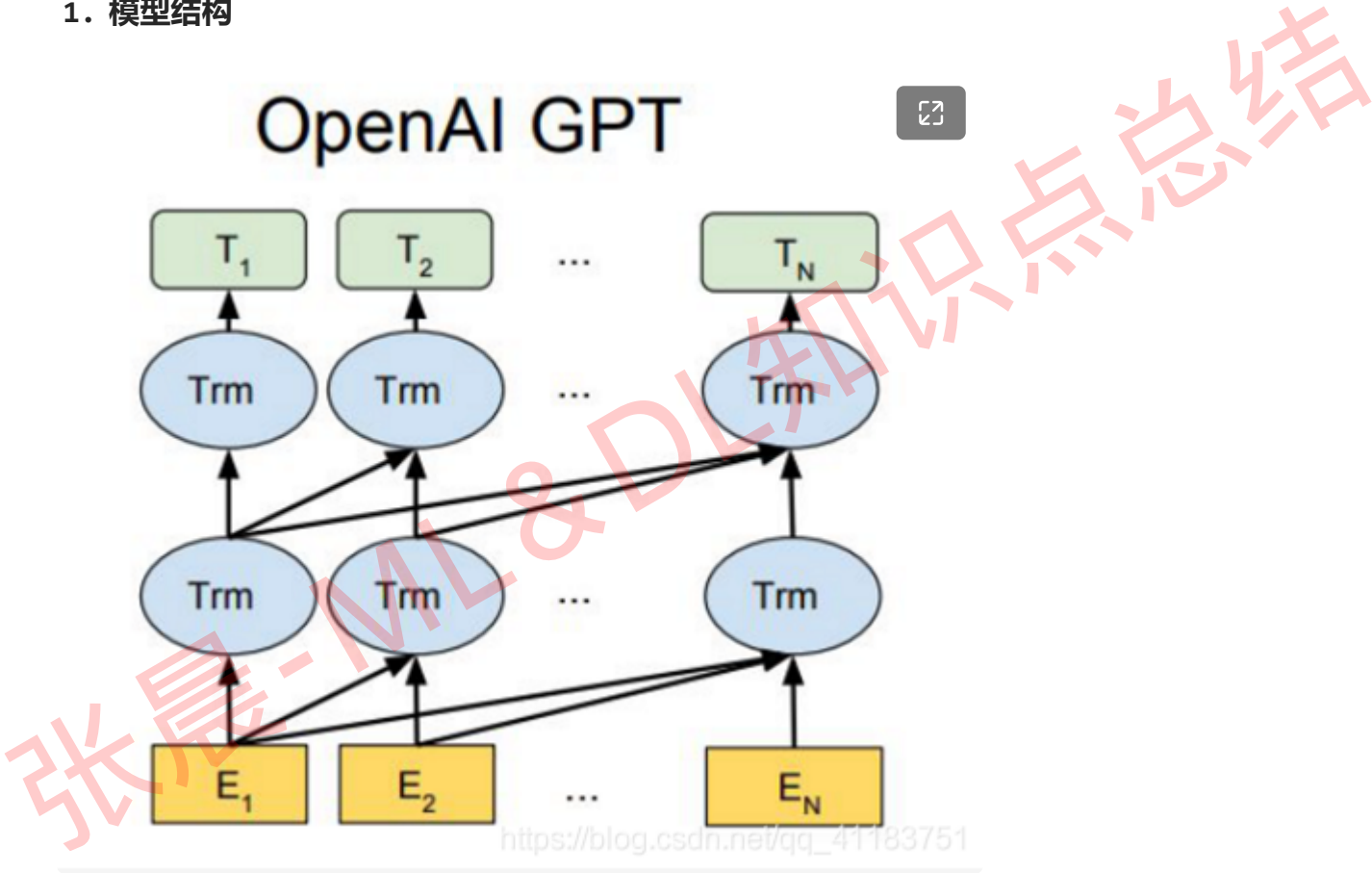
 加入语雀，获得更好的阅读体验

[注册](#) 或 [登录](#) 后可以收藏本文随时阅读，还可以关注作者获得最新文章推送

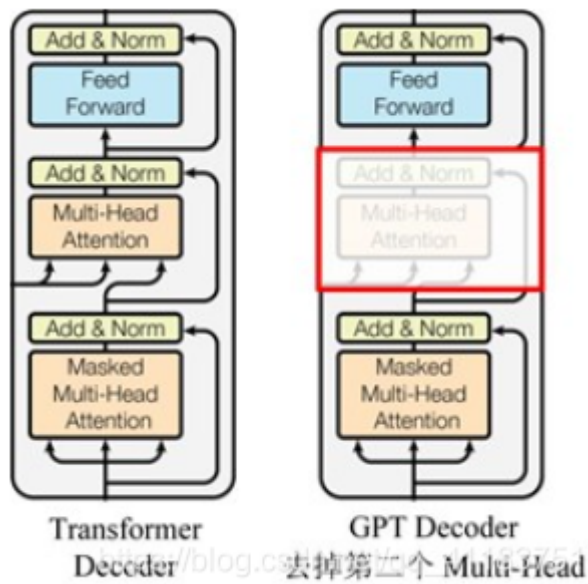
立即加入

## 7. GPT

### 1. 模型结构



GPT 使用 Transformer 的 Decoder 结构，并对 Transformer Decoder 进行了一些改动（掩码多头自注意力+前馈）



$$h_0 = UW_e + W_p$$
$$h_i = \text{transformer\_block}(h_{i-1}) \forall i \in [1, n]$$
$$P(u) = \text{softmax}(h_n W_e^T)$$

$W_p$  是单词位置的 Embedding,  $W_e$  是单词的 Embedding。用  $\text{voc}$  表示词汇表大小,  $\text{pos}$  表示最长的句子长度,  $\text{dim}$  表示 Embedding 维度, 则  $W_p$  是一个  $\text{pos} \times \text{dim}$  的矩阵,  $W_e$  是一个  $\text{voc} \times \text{dim}$  的矩阵。  
 $h_0$  代表此时输入的序列特征, 得到输入  $h_0$  之后, 需要将  $h_0$  依次传入 GPT 的所有 Transformer Decoder 里, 最终得到  $h_t$ 。

2. 预训练部分

(1) 给定无标签语料库  $U = \{u_1, \dots, u_n\}$ , 使用标准语言建模目标来最大化以下可能性:

$$L_1(U) = \sum_i \log P(u_i | u_{i-k}, \dots, u_{i-1}; \Theta)$$

$k$  是基于语境窗口的大小, 条件概率  $P$  表示在参数下采用神经网络建模的可能性 (下一个单词仅和前  $k$  个单词有关)。训练参数可以通过梯度下降获得。

3. 微调部分

(1) 拿简单分类任务举例，假设我们有带标签的数据集 $C$ ，对于输入序列 $[x^1, x^2, \dots, x^m]$ 以及标签 $y$ ，首先将输入序列输入到预训练模型中，得到 transformer 最后一层的输出 $h_l^m$ ，然后再经过全连接层与softmax，得到预测的概率。简单分类任

$$P(y|x^1, \dots, x^m) = \text{softmax}(h_l^m W_y).$$

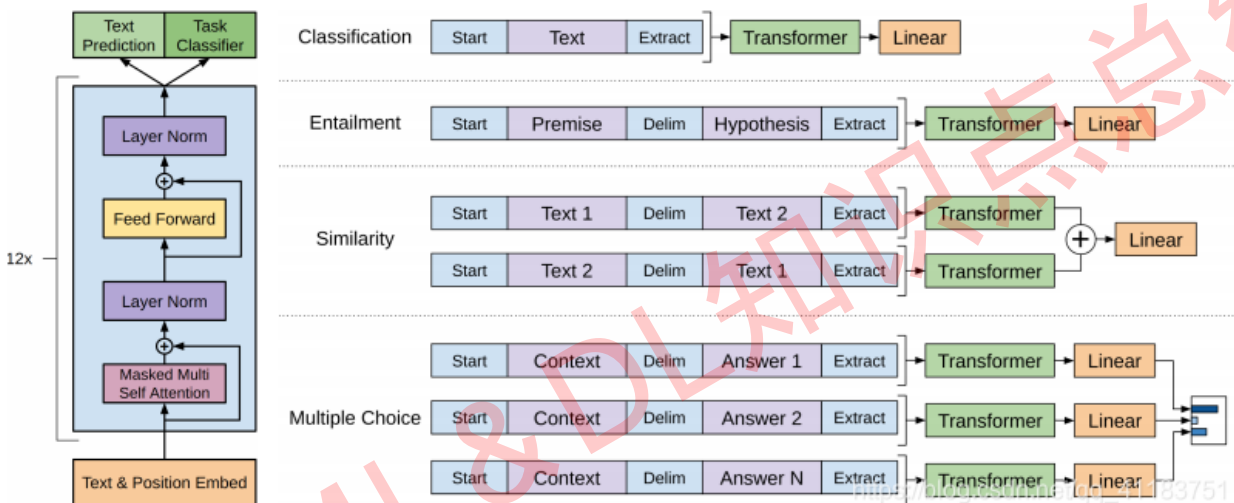
务的目标函数是

$$L_2(C) = \sum_{(x,y)} \log P(y|x^1, \dots, x^m).$$

(2) 具体微调时结合了语言模型的部分

$$L_3(C) = L_2(C) + \lambda * L_1(C)$$

不同的任务有不同的输入构造方式，但大同小异，如下图所示



#### 4. 下游NLP任务

- 分类
- 文本蕴含
- 文本相似度计算
- 常识推理和问答

#### 5. 损失函数

$$L_3(C) = L_2(C) + \lambda * L_1(C)$$

#### 6. GPT与ELMO的区别

- ELMO采用的是双向语言模型，GPT是单向语言模型
- ELMO采用LSTM建立语言模型，GPT采用transformer的decoder部分
- GPT在fine-tune的时候，最终目标里加入了原语言模型的训练目标最为辅助，提升了收敛速度以及更好的效果

## 7. GPT与BERT的区别

- bert采用transformer中的encoder部分；GPT采用transformer中的decoder部分
- bert是双向模型，GPT是单向模型
- GPT的预训练部分任务是语言模型，BERT预训练部分的任务是MML和NSP

## 8. GPT的优缺点

- 优点是利用了transformer的中attention的优势，采用attention的方式更好对长距离依赖关系进行建模
- 缺点是不是双向模型

☐ [https://blog.csdn.net/qq\\_41183751/article/details/111603365](https://blog.csdn.net/qq_41183751/article/details/111603365)

<[https://blog.csdn.net/qq\\_41183751/article/details/111603365](https://blog.csdn.net/qq_41183751/article/details/111603365)>

☐ <https://zhuanlan.zhihu.com/p/65886515> <<https://zhuanlan.zhihu.com/p/65886515>>

张晨-ML&DL知识点总结