BERT从零详细解读

右下角【联系我】

扫码关注微信公众号

文章周更

知识分享

一起进步

求关注，求点赞，求一切！！

1.BERT整体模型架构

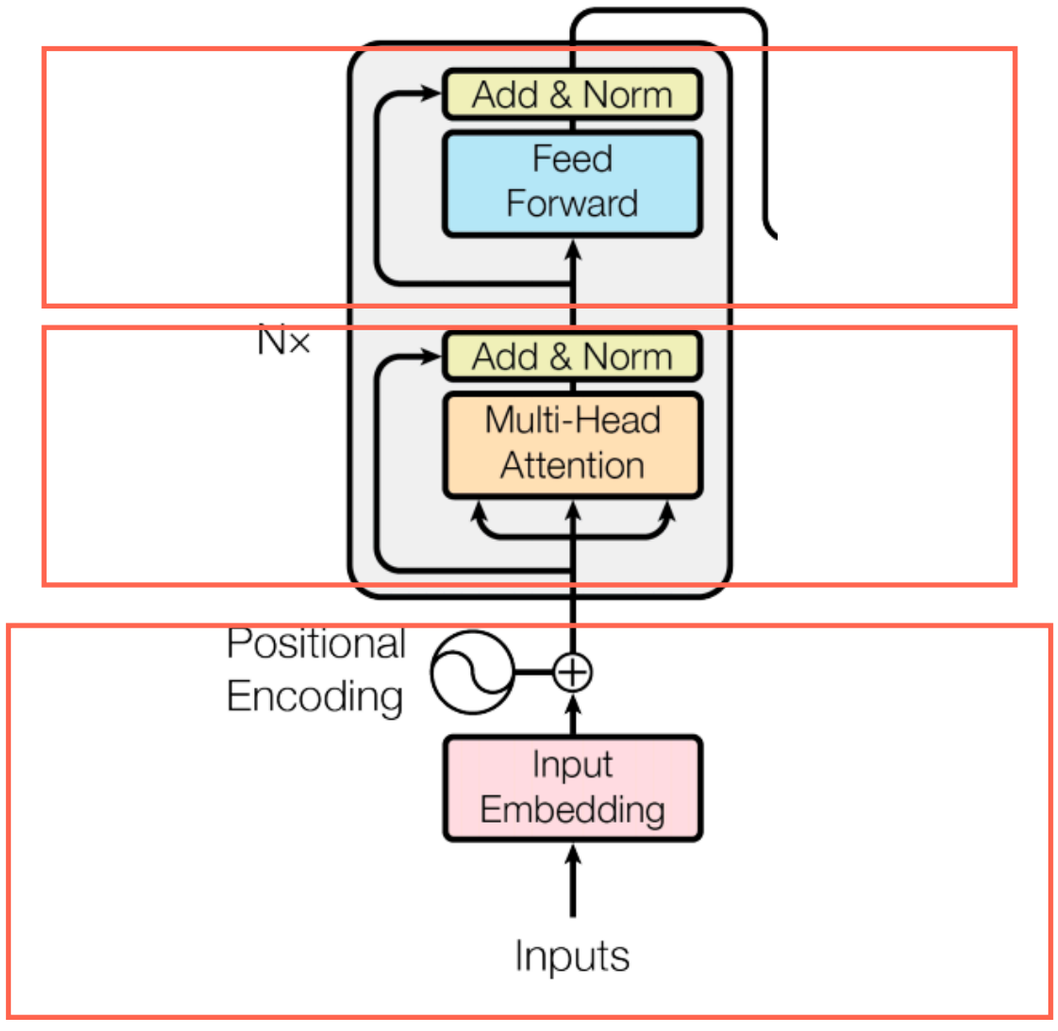
2.如何做BERT预训练：参数+MLM+NSP

3.如何微调BERT，提升BERT在下游任务中的效果

4. 微调BERT做文本分类的代码解析。

1.BERT整体模型架构

BERT基础架构-encoder



3 前馈神经网络

2 注意力机制

1 输入部分

Encoder

Encoder

Encoder

Encoder

Encoder

Encoder

Encoder

Encoder

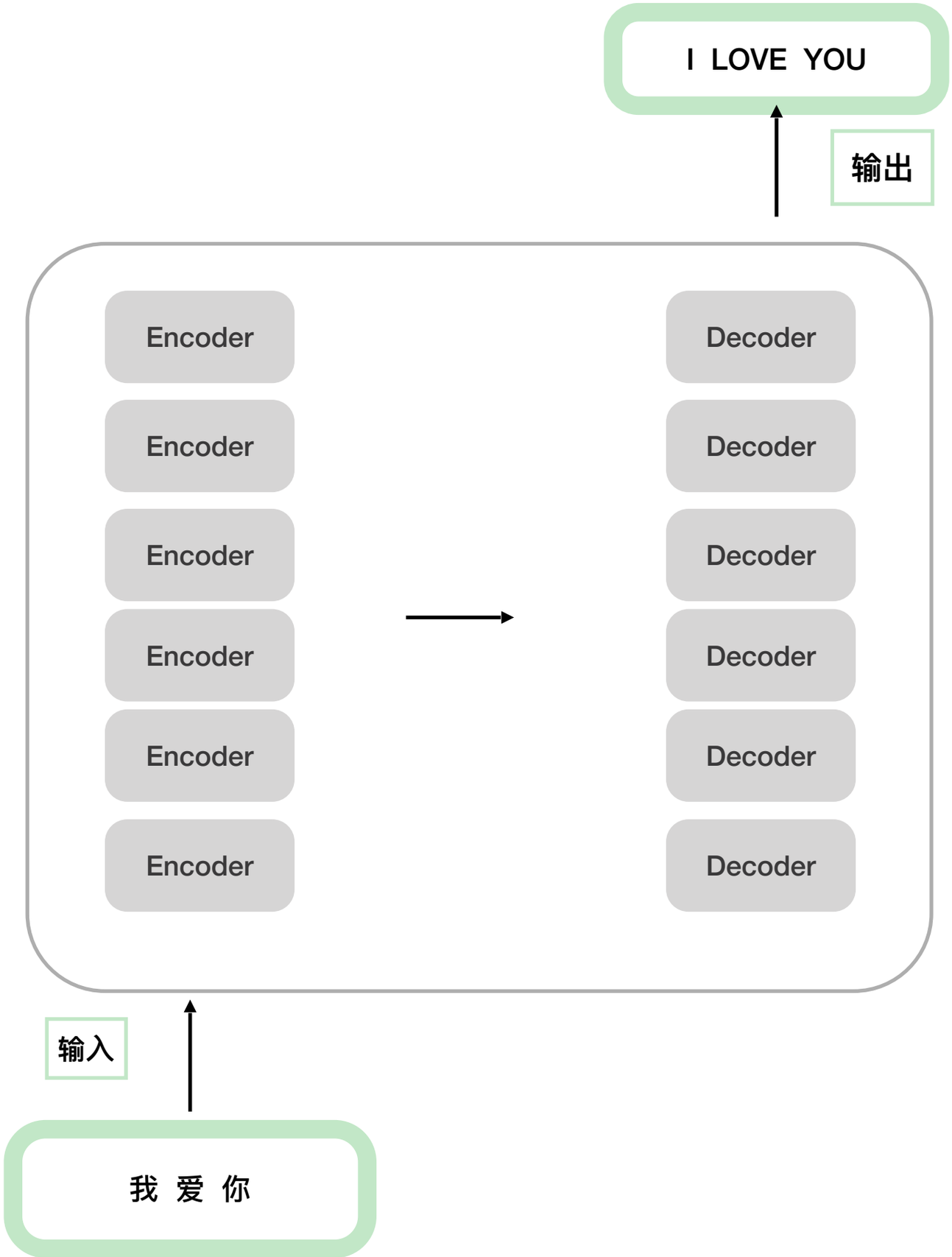
Encoder

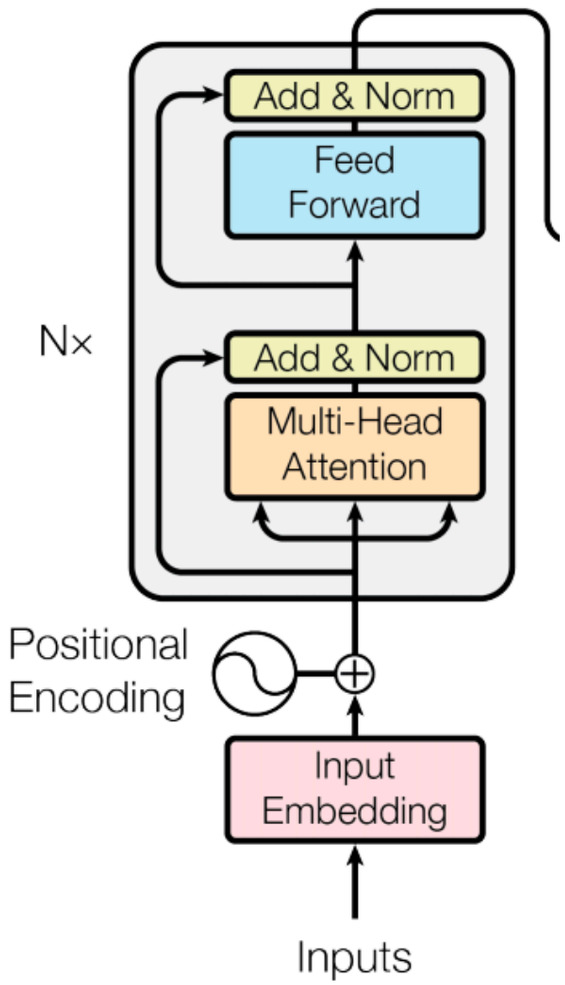
Encoder

Encoder

Encoder

BERT base 12层


I LOVE YOU
输出
Encoder
Encoder
Encoder
Encoder
Encoder
Encoder
Decoder
Decoder
Decoder
Decoder
Decoder
Decoder
输入
我 爱 你



Input=token emb + segment emb+ position emb

Input=token emb + segment emb+ position emb

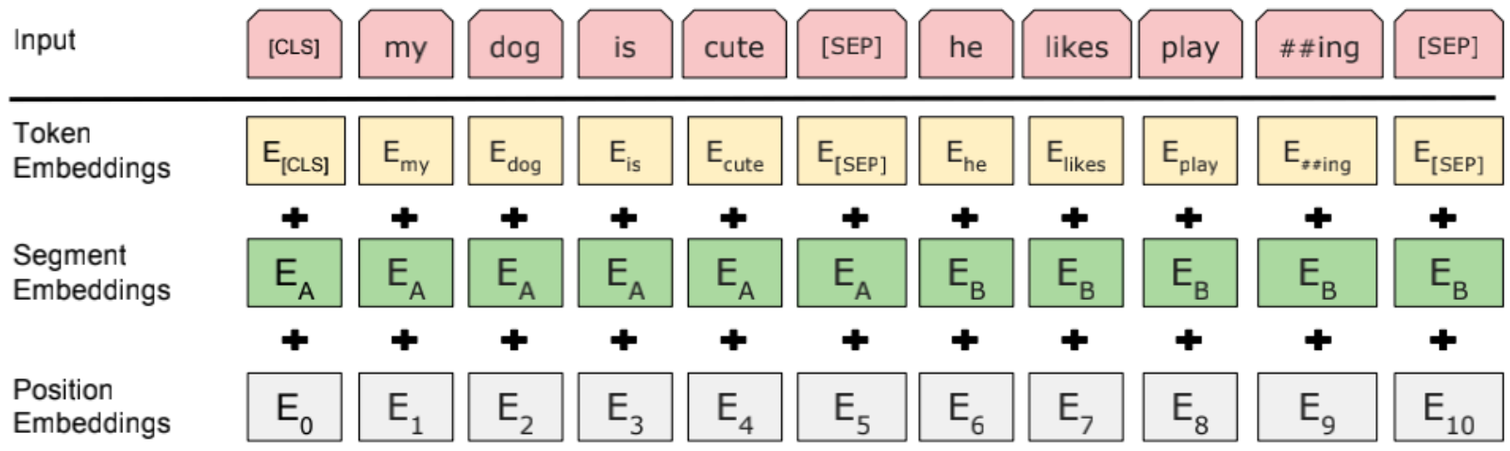


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

CLS向量不能代表语义信息

3.bert pretrain模型直接拿来用作 sentence embedding效果甚至不如word embedding，cls的emebdding效果最差（也就是你说的pooled output）。把所有普通token embedding做pooling勉强能用（这个也是开源项目bert-as-service的默认做法），但也不会比word embedding更好。

2.如何做预训练： MLM+NSP

AR

1. 一种是AR，也就是autoregressive，我们称之为自回归模型；只能考虑单侧的信息，典型的就是GPT

无监督目标函数

AE

1. 一种是AE，也就是autoencoding，我们称之为自编码模型；从损坏的输入数据中预测重建原始数据。可以使用上下文的信息

【我爱吃饭】

AR　　P(我爱吃饭) = P(我)P(爱|我)P(吃|我爱)P(饭|我爱吃)；

AE　　**mask之后：【我爱mask饭】**

P(我爱吃饭|我爱mask饭)=P(吃|我爱饭)

mask概率问题

10%替换成其他

随机mask15%单词

10%保持不变

80%替换为mask

mask代码实践

```
for index in mask_indices:
    # 80% of the time, replace with [MASK]
    if random.random() < 0.8:
        masked_token = "[MASK]"
    else:
        # 10% of the time, keep original
        if random.random() < 0.5:
            masked_token = tokens[index]
        # 10% of the time, replace with random word
        else:
            masked_token = random.choice(vocab_list)
```

NSP任务

NSP样本如下:

- 从训练语料库中取出两个连续的段落作为正样本
- 从不同的文档中随机创建一对段落作为负样本

3.提升BERT在下游任务中的效果

1. 谷歌中文bert

2. 基于任务数据做未微调

我们可以分为四步骤走：

比如做微博文本情感分析

1. 在大量通用语料上训练一个LM（Pretrain）； --中文谷歌BERT

2. 在相同领域 上继续训练LM（Domain transfer）； --在大量微博文本上继续训练这个BERT

3. 在任务相关的小数据上继续训练LM（Task transfer）； ---在微博情感文本上（有的文本不属于分析的范畴）

4. 在任务相关数据上做具体任务（Fine-tune）。-

1.动态mask

1.ngram-mask

参数：

数据增强/自蒸馏/外部知识融入