

2023-2024学年第一学期本科生课程

《神经网络与深度学习》

课程实验作业(四)

主讲人：戴金晟(副教授，博士生导师)

daijincheng@bupt.edu.cn

神经网络与深度学习课程组



北京邮电大学
Beijing University of Posts and Telecommunications

AFQMC数据集

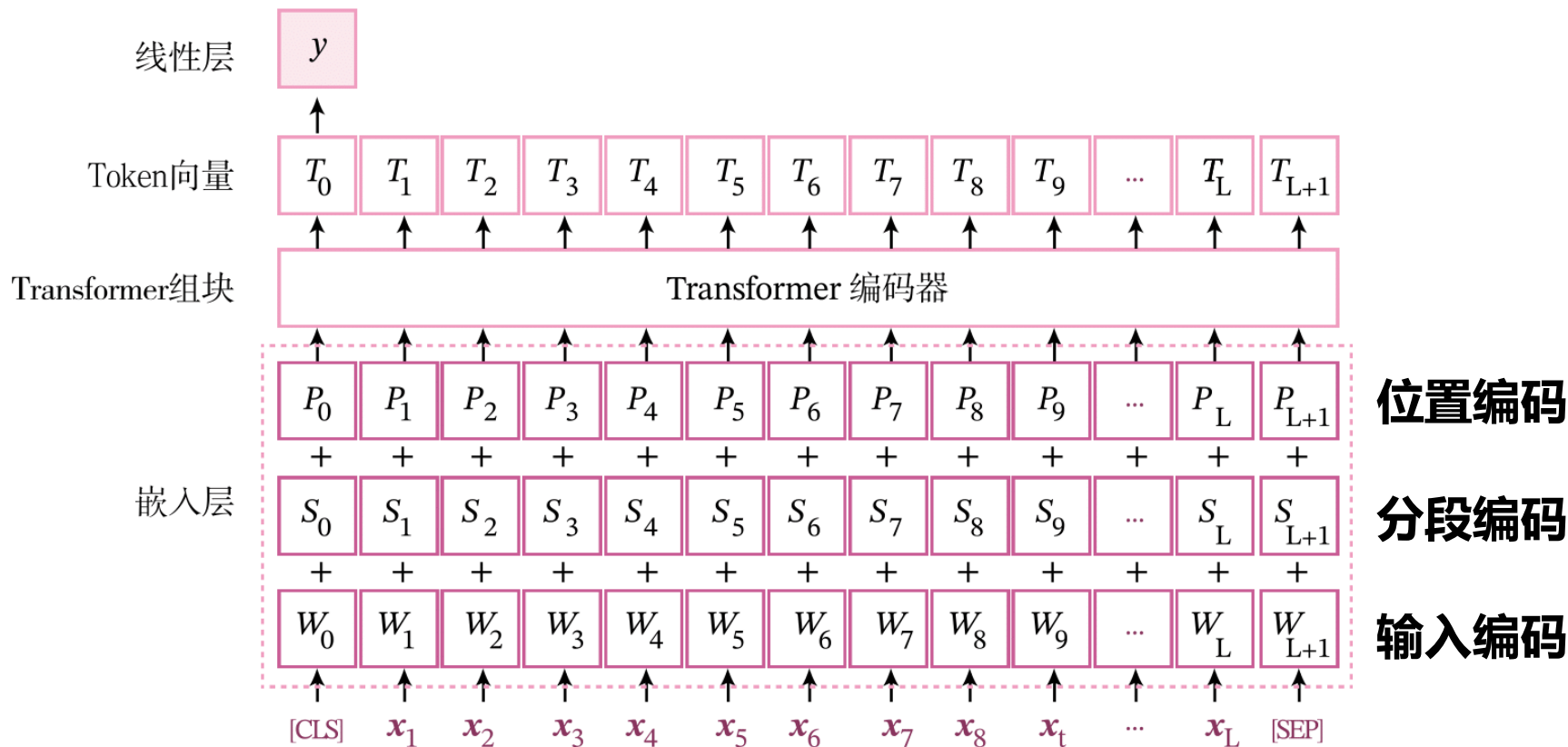
□ AFQMC数据集：

- AFQMC数据集是一个蚂蚁金融语义相似度数据集，用于问题相似度计算，数据集包括训练集、验证集、测试集3个文件，分别包含34334、4316以及3861条数据，每条数据有三个属性，分别是句子1、句子2、句子相似度标签。相似度标签为1表示两个句子含义类似，标签为0则表示含义不同。
- 下载地址：<https://github.com/CLUEbenchmark/CLUE>

数据示例

```
{"sentence1": "双十一花呗提额在哪", "sentence2": "里可以提花呗额度", "label": "0"}
```

文本语义匹配模型结构



- 1、Transformer 组块中，多头注意力层以及 Add&Norm 层需要自行编程实现
- 2、为了后续任务中验证增加 Transformer 层数的效果，初始时设置层数为2层

搭建嵌入层

□输入编码

- 把输入的词转化成向量的形式。
- 输入编码看作是一个查表的操作，对于每个单词，要将这些符号转换为向量形式。一种简单的转换方法是通过一个嵌入表来将每个符号直接映射成向量表示
- 在Transformer的输入编码的实现中，初始化使用的是随机正态分布，均值为0，标准差为 $\frac{1}{\sqrt{\text{emb_size}}}$ ，其中 emb_size 为词向量的维度。所有的向量都乘以了 $\sqrt{\text{emb_size}}$

嵌入表已提前给出，也可通过 word2vec 或者 torchtext 自建来得到嵌入表

搭建嵌入层

□分段编码

- 使得模型能够接受句子对进行训练，用编码的方法使得模型能够区分两个句子。这里指定0来标记句子0，用1标记句子1。对于句子0，创建标记为0的向量，对于句子1，创建标记为1的向量。分段编码的实现跟输入编码类似，不同在于词表大小为2。

什	么	花	一	年	四	季	都	开
---	---	---	---	---	---	---	---	---

0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---

什	么	花	一	年	四	季	都	是	开	的
---	---	---	---	---	---	---	---	---	---	---

1	1	1	1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---	---	---	---

搭建嵌入层

□ 位置编码

- 为了使得自注意力模块可以感知序列的顺序信息，Transformer 给编码层输入添加了一个额外的位置编码。位置编码的目的是在让自注意力模块在计算时引入词之间的距离信息。
- 本实验我们用三角函数（正弦或者余弦）来编码位置信息。假设位置编码的维度为 D ，则其中每一维的值的计算如下：

偶数维 $PE(pos, 2i) = \sin\left(\frac{t}{10000^{2i/D}}\right)$

$0 \leq i \leq \frac{D}{2}$ 为编码向量的维数

奇数维 $PE(pos, 2i + 1) = \cos\left(\frac{t}{10000^{2i/D}}\right)$

t 为当前词在句子中的位置

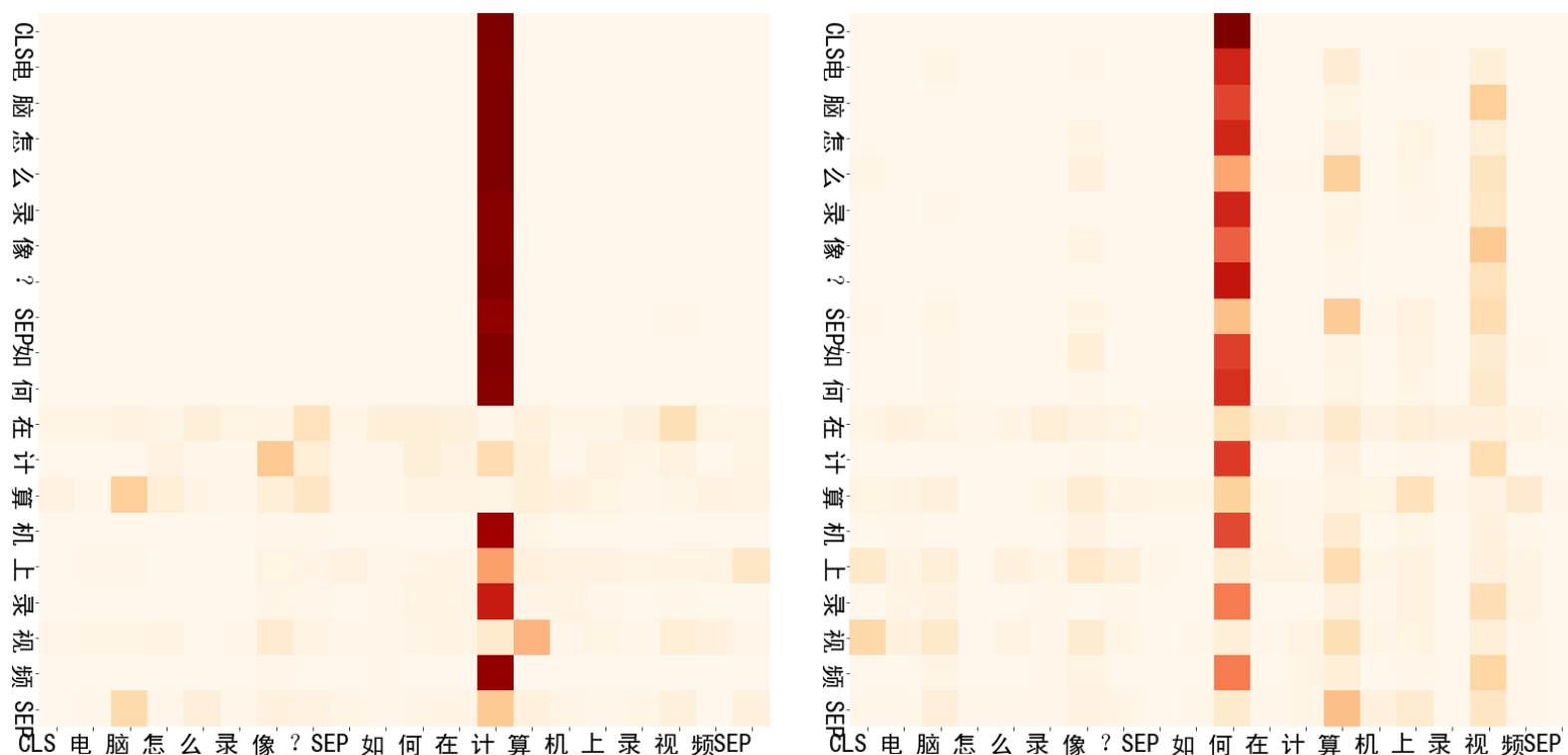
搭建嵌入层

- 最后，把输入编码，分段编码和位置编码进行相加，并对加和后的向量进行层规范化和 Dropout 操作

进阶任务：说明自己的模型中层规范化操作的位置是pre_norm 还是 post_norm，然后尝试另一种层规范化操作，对比二者在具体训练中的区别并分析原因

可视化分析注意力机制权重

□ 输入一条样本，提取出多头注意力权重，利用 heatmap 对注意力进行可视化，并**进行结果分析**



颜色越深，表示的权重越高