**深度学习第五次实验作业报告**

课程**/**2023深度学习 学号**/**SA23229086 学生**/**郎文翀

大数据学院计算机技术班

2023年2月3日星期六

# 1 实验要求

实现 Transformer 模型用于英译中机器翻译，模型中块的数量、模型维度甚至是数据规模可以自己调整以适应个人电脑。使用 BLEU 值作为评价指标。

# 2 实验设计

## 2.1 设计需求

本次实验选择Pytorch实现，并使用GPU版本进行训练，最后采用tensorbaord完成绘图与实验结果分析，具体程序设计需求如下：

数据集处理：数据集使用助教提供的yelp3数据集，由于并未提前划分，因此需要手动将数据集按照7：2：1划分为训练集，验证集和测试集。

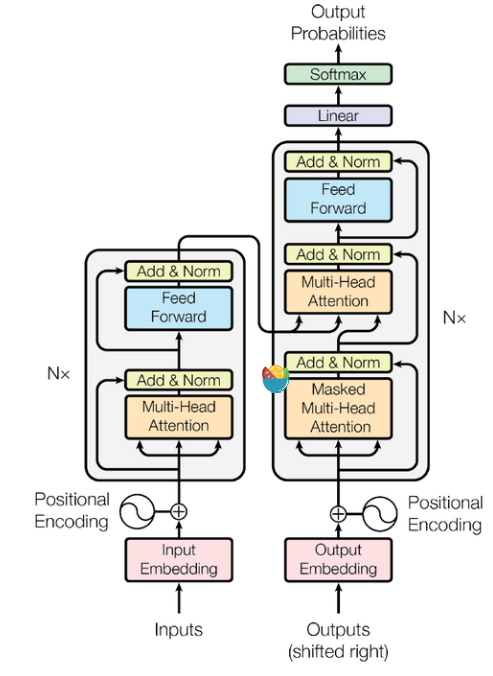
模型设计：本次实验需要参考Transformer论文自行设计并实现模型。具体需要首先对首先对文本数据集进行清洗后对不同的单词进行划分编码，得到编码后的文本后再输入至模型中进行训练，基于Transformer论文模型原理按次序实现中英翻译。

模型参数设置：尝试设置不同的超参数包括模型深度，编码向量宽度，Encoder，Decoder数量以及注意力头的数量等。

分批次训练：每Epoch数据分为128一组进行训练。

## 2.2 Transformer论文分析

由于本次实验参考论文《Attention Is All You Need》实现，因此首先简单讲解论文中的基于attention机制的transformer模型架构。在Transformer模型结构中，模型设计分层清晰，主要分为四个阶段，具体如下图演示，包括Embedding,Position Encoding，Encoding，Dcodinge四个阶段。其中input与output都是编码格式，需要进一步结合对应的文本编解码map进行翻译。



1.Embedding：由于文本编码后每一个单词对应一个简单的独一无二的数字编码(这里本次实验选择bpe进行编码，因此高频词文本会对应更小的数值编码)，为了进一步增大不同单词为文本之间的特征差异，需要将序列中的每一个单词或者标记映射到高维空间以便嵌入后模型更易捕捉单词之间的语义差异关系。

2.Position Encoding:由于在实现文本翻译时是对文本序列进行处理，因此单词的位置关系很重要，同时在编码与解码过程中也要重视位置前后关系导致的注意力处理差异(Mask作用)。通过位置编码层可以将位置信息嵌入到Embedding后的词向量中，使得模型能够区分不同位置的单词或标记。

3.Encoding：在编码阶段，需要经过多次重复的编码块，每一个编码块中都包含一次多头注意，残差，归一化的过程。在注意力处理过程中，注意力头会自行根据学习的可训练参数生成句子中不同词语之间的语义关系(具体通过K,Q,V实现)。

4.Decoding:在解码阶段同样由N个重复的解码块完成，其中中在解码阶段包含两次注意过程，首先是Masked多头注意过程，这是因为需要考虑在文本翻译过程中当前文本的翻译只能结合前面位置的文本信息，而此时还未能“看到”后续的文本信息，因此在此时的Masked多头注意需要通过掩码将当前文本后续位置的单词遮住。之后再经过一次和Encoding过程中相同的多头注意与前馈线性映射解码到指定维度。

## 2.3 模型设计与实现

具体的模型代码实现过程中，首先使用bpe对数据集中的每一个单词进行编码，这里设置为最多编码320000个单词，再每一个句子的开头与结尾设置两个标志符拼接，对于空格以及余下的低频单词统一使用pad标识符表示。之后即输入至参考Transformer论文模型进行训练。这里再实验过程中对比了不同注意力头，不同编码解码深度训练后的最佳模型再测试集的BLEU分数。由于硬件资源有限，本次实验训练时间较长，因此仅在News Commentray v15上进行了训练与测试。

# 3 实验分析

## 3.1 baseline模型

我们选取embedding向量size为32000，Encoding与Decoding的深度为4，batch\_size为128，学习率为3e-4，训练40个epoch（设置early stop，经过实验发现在20epoch后模型基本收敛会自动停止），每epoch后学习率衰减为原学习率的0.5作为baseline模型进行分析，以下是验证集损失/BELU分数等展示图，具体日志请查看logs文件夹。

每 Epoch 验证集损失/BELU得分

|  |  |
| --- | --- |
|  |  |

观察上图发现模型在12Epoch后验证集损失不降反升说明模型在第12Epoch后已收敛达到最佳性能。此时模型的在训练集的BELU得分为28.6，虽然在后续的训练Epoch BELUE损失仍然在提升但是可能时在训练集上过拟合导致，我们保存在验证集上表现最好的第12 Epoch对应的模型并查看模型在测试集上最佳表现，此时的BLEU得分为26.456。

## 3.2 增加注意力头

之后我们在baseline基础上将注意力头的数量翻倍得到如下结果

每Epoch 验证集损失/BELU得分

|  |  |
| --- | --- |
|  |  |

结合上图发现增加注意力头数量后模型的损失略微下降，BLEU得分小幅提升推测这可能时因为增加注意力头数量不够多，如果扩大注意力头数量为baseline的4倍以上可能得到更加性能的模型，此次训练的模型在测试集的BLEU得分为26.62。

## 3.3 增加编码/解码块数

在baseline中我们对词汇w与语句s的RNN encoder均仅使用了一层双向RNN，因此推测可能提升双向RNN深度可能有助于进一步加大潜在编码向量的差异距离从而提升模型的性能，因此这里我们尝试将rnn\_layer\_size翻倍。

每Epoch 验证集损失/BLEU得分

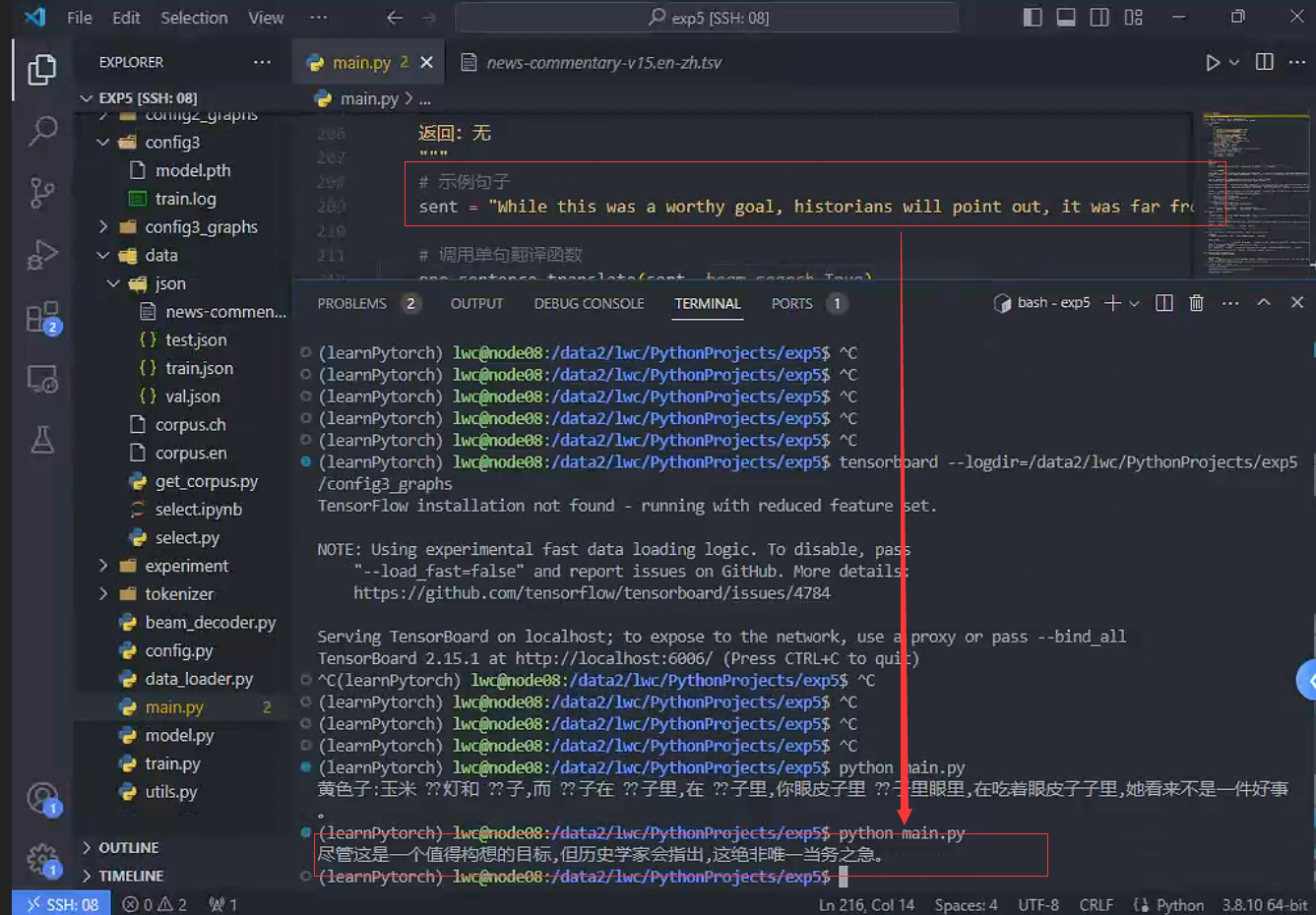
|  |  |
| --- | --- |
|  |  |

结合上图发现模型在第5轮即达到了最佳性能，推测增加编码/解码层数可以有效提升模型表达能力，加速模型的训练收敛速度。此时模型在测试集的BLEU得分与baseline基本相同。

## 3.5 文本翻译测试

这里我们以一段文本作为输入测试最终收敛后的模型翻译结果，具体的数据集的翻译结果请查看logs文件夹下不同配置的output.txt文件。

测试文本：" While this was a worthy goal, historians will point out, it was far from the only imperative. “



# 4. Requirements.txt

本次实验使用的虚拟环境可能包含一些不必要的库，基于2张RTX 3090训练完成实验

|  |
| --- |
| absl-py==2.0.0  autopep8==2.0.4  cachetools==5.3.2  certifi==2023.11.17  charset-normalizer==3.3.2  click==8.1.7  colorama==0.4.6  contourpy==1.2.0  cycler==0.12.1  fonttools==4.45.1  gensim==4.3.2  google-auth==2.23.4  google-auth-oauthlib==1.1.0  grpcio==1.59.3  idna==3.6  importlib-metadata==6.8.0  importlib-resources==6.1.1  Jinja2==3.1.2  joblib==1.3.2  kiwisolver==1.4.5  lxml==5.1.0  Markdown==3.5.1  MarkupSafe==2.1.3  matplotlib==3.8.2  nltk==3.8.1  numpy==1.26.2  oauthlib==3.2.2  opencv-python==4.8.1.78  packaging==23.2  pandas==2.1.4  Pillow==10.1.0  portalocker==2.8.2  protobuf==4.23.4  psutil==5.9.7  pyasn1==0.5.1  pyasn1-modules==0.3.0  pycodestyle==2.11.1  pyparsing==3.1.1  python-dateutil==2.8.2  pytz==2023.3.post1  regex==2023.10.3  requests==2.31.0  requests-oauthlib==1.3.1  rsa==4.9  sacrebleu==2.4.0  scikit-learn==1.3.2  scipy==1.11.4  sentencepiece==0.1.99  six==1.16.0  smart-open==6.4.0  tabulate==0.9.0  tensorboard==2.15.1  tensorboard-data-server==0.7.2  threadpoolctl==3.2.0  tomli==2.0.1  torch==1.7.1+cu110  torch-geometric==2.0.4  torchaudio==0.7.2  torchvision==0.8.2+cu110  tqdm==4.66.1  typing\_extensions==4.8.0  tzdata==2023.3  urllib3==2.1.0  Werkzeug==3.0.1  zipp==3.17.0 |