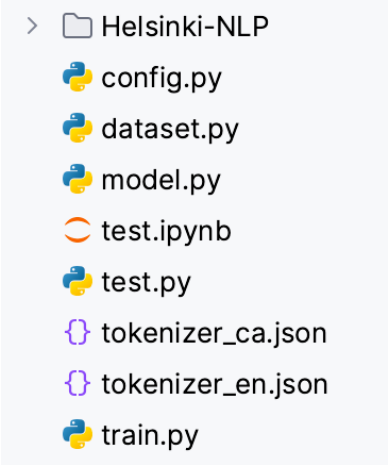
Project7报告

1. **代码结构**

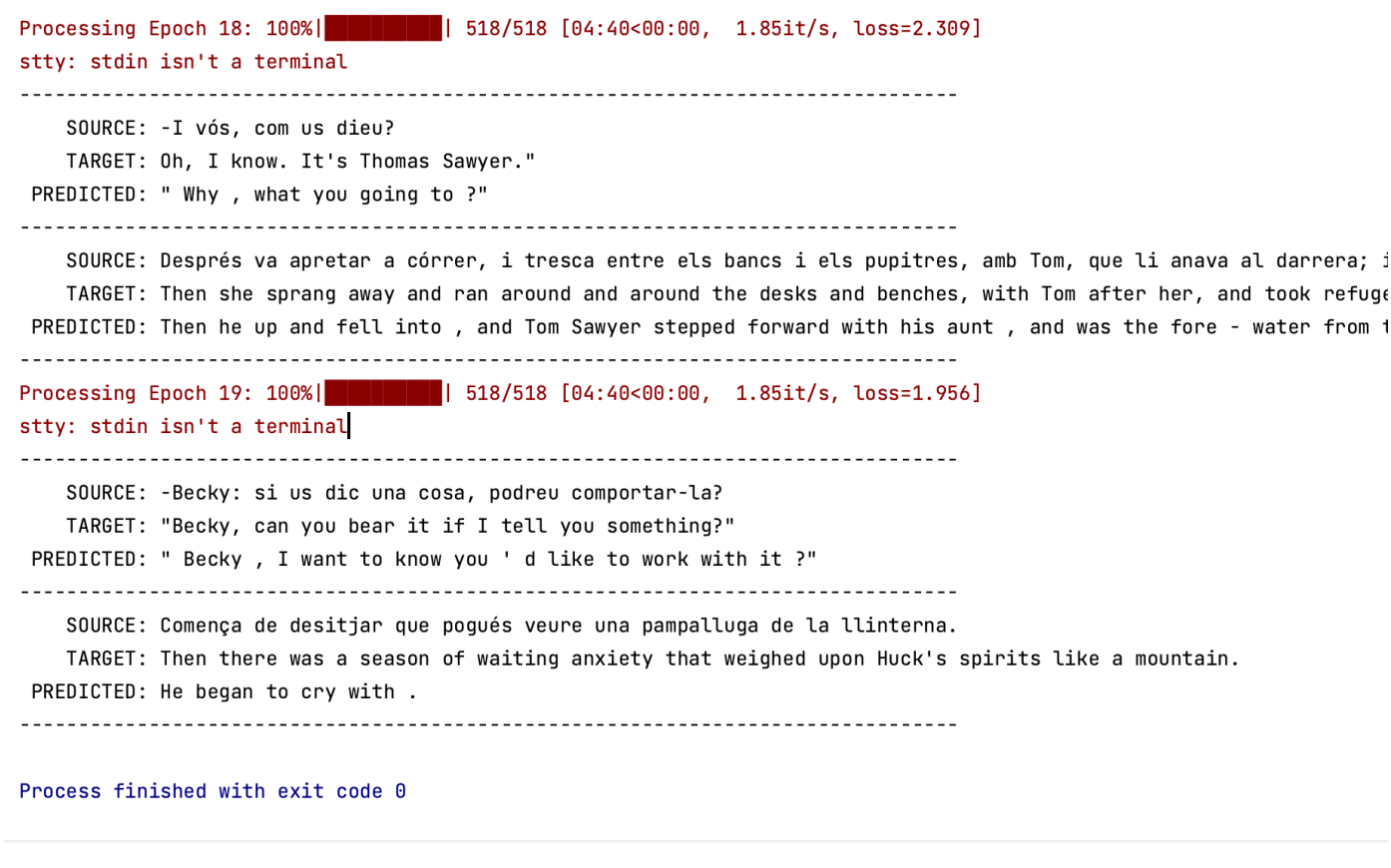


   - Helsinki-NLP文件夹：存放训练模型（由于模型太大，无法上传到github，因此放到百度网盘中，链接: https://pan.baidu.com/s/1J1E12hlJQcIdP0WxDtDCNw 提取码: 4n4w ）
   - config.py：训练设置
   - dataset.py：处理数据集
   - model.py：定义transformer模型
   - test.py：测试模型函数
   - train.py：训练模型函数
   - tokenizer.json：数据集

2. **训练截图**

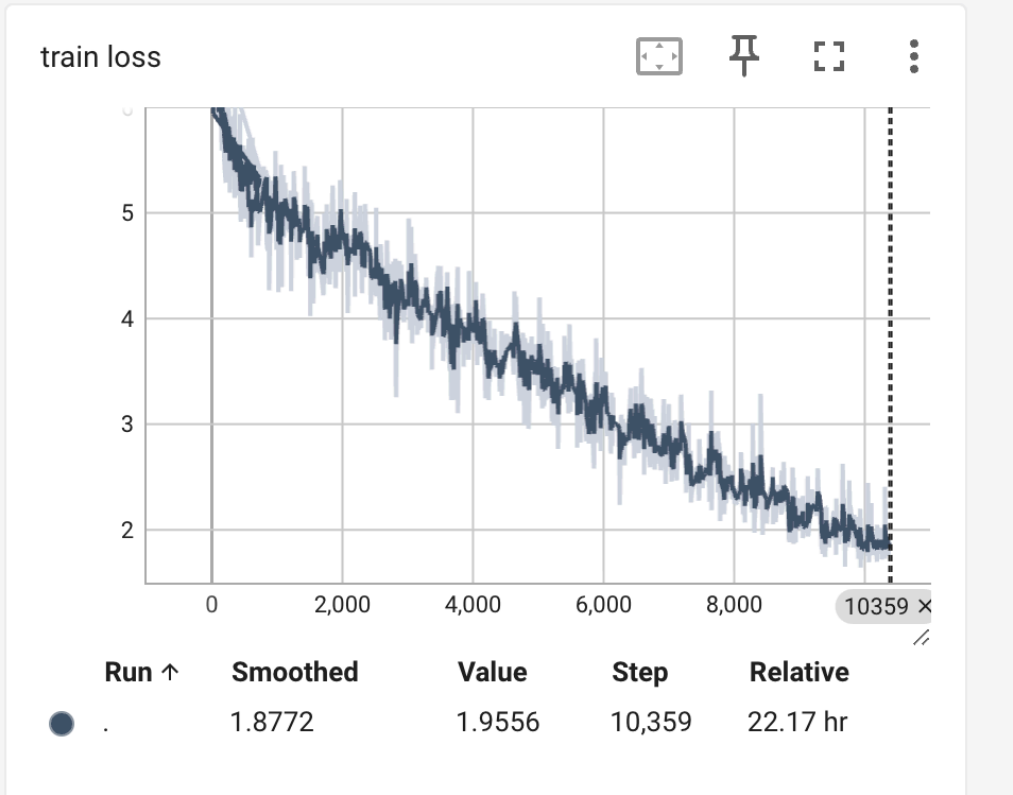
   在本地pycharm跑的，一共训练了20个epoch，没有彻底训练完，loss还在收敛，最后一张为loss图。

```
/Users/tongyaole/anaconda3/envs/deeplearning/bin/python /Users/tongyaole/Downloads/pytorch-transformer-main/train.py
Using device: mps
Device name: <mps>
Max length of source sentence: 220
Max length of target sentence: 210
No model to preload, starting from scratch
Processing Epoch 00: 100%|████████| 518/518 [04:41<00:00,  1.84it/s, loss=5.365]
stty: stdin isn't a terminal
--------------------------------------------------------------------------------
    SOURCE: Ningú entra ni sortí pel carreró; ningú que s'assemblés a l'espanyol entra o sortí per la porta de la taverna.
    TARGET: Nobody entered the alley or left it; nobody resembling the Spaniard entered or left the tavern door.
 PREDICTED: The , and the , and the , and .
--------------------------------------------------------------------------------
    SOURCE: La família Harper, i la tia Polly i els seus, s'endolaven amb molt de sentiment i moltes llagrimes.
    TARGET: The Harpers, and Aunt Polly's family, were being put into mourning, with great grief and many tears.
 PREDICTED: The of the of the of the of the .
--------------------------------------------------------------------------------
Processing Epoch 01: 100%|████████| 518/518 [04:40<00:00,  1.85it/s, loss=5.050]
stty: stdin isn't a terminal
--------------------------------------------------------------------------------
    SOURCE: L'esglai de Tom havia afeblit tots els seus muscles.
    TARGET: Tom's fright weakened every muscle in his body.
 PREDICTED: The boys was a of the .
--------------------------------------------------------------------------------
    SOURCE: Sa tia n'estava tota preocupada; començа de fer-li sofrir tota mena de provatures amb medicines.
    TARGET: His aunt was concerned.
 PREDICTED: The boys was a of the of the of the .
--------------------------------------------------------------------------------
Processing Epoch 02: 100%|████████| 518/518 [04:41<00:00,  1.84it/s, loss=4.432]
stty: stdin isn't a terminal
--------------------------------------------------------------------------------
    SOURCE: Voldria tenir-ne, ara!
    TARGET: I wish I had some now."
 PREDICTED: I ' ll be a - a - !
--------------------------------------------------------------------------------
```



```
Processing Epoch 18: 100%|████████| 518/518 [04:40<00:00,  1.85it/s, loss=2.309]
stty: stdin isn't a terminal
--------------------------------------------------------------------------------
    SOURCE: -I vós, com us dieu?
    TARGET: Oh, I know. It's Thomas Sawyer."
 PREDICTED: " Why , what you going to ?"
--------------------------------------------------------------------------------
    SOURCE: Després va apretar a córrer, i tresca entre els bancs i els pupitres, amb Tom, que li anava al darrera; i
    TARGET: Then she sprang away and ran around and around the desks and benches, with Tom after her, and took refuge
 PREDICTED: Then he up and fell into , and Tom Sawyer stepped forward with his aunt , and was the fore - water from t
--------------------------------------------------------------------------------
Processing Epoch 19: 100%|████████| 518/518 [04:40<00:00,  1.85it/s, loss=1.956]
stty: stdin isn't a terminal
--------------------------------------------------------------------------------
    SOURCE: -Becky: si us dic una cosa, podreu comportar-la?
    TARGET: "Becky, can you bear it if I tell you something?"
 PREDICTED: " Becky , I want to know you ' d like to work with it ?"
--------------------------------------------------------------------------------
    SOURCE: Començа de desitjar que pogués veure una pampalluga de la llinterna.
    TARGET: Then there was a season of waiting anxiety that weighed upon Huck's spirits like a mountain.
 PREDICTED: He began to cry with .
--------------------------------------------------------------------------------

Process finished with exit code 0
```



2. **测试以及bleu指标**

在test.ipynb文件中可以直观看到，在这也附上图片

```
--------------------------------------------------------------------------------
    SOURCE: Digué: -Us agraden les rates?
    TARGET: He said: "Do you love rats?"
 PREDICTED: He said : " You don ' t want you done ?"
--------------------------------------------------------------------------------
    SOURCE: -O Tom! Poden veure en la foscor, igual que els gats.
    TARGET: "Oh, Tom, they can see in the dark, same as cats.
 PREDICTED: " Oh , Tom , they can see in the dark , same as cats .
--------------------------------------------------------------------------------
    SOURCE: -I després… i després… Bé, no n'estic cert, pero em sembla com si haguéssiu manat a
    Sid que se n'anes i…i…i… -I que?
    TARGET: "And then--and then--well I won't be certain, but it seems like as if you made Sid go
     and--and--" "Well?
--------------------------------------------------------------------------------
    SOURCE: Els minyons feren allo que havien fet ben sovint altres vegades: anaren a la reixa de
     la cel·la, i donaren a Potter tabac i mistos.
    TARGET: The boys did as they had often done before--went to the cell grating and gave Potter
    some tobacco and matches.
 PREDICTED: The boys did as they had often done before -- went to the grating and gave Potter
 some tobacco and matches .
--------------------------------------------------------------------------------
    SOURCE: I aixo no és res!
    TARGET: And that ain't all.
 PREDICTED: And that ain ' t all .
--------------------------------------------------------------------------------
bleu_total: BLEU = 0.49 2.2/0.6/0.3/0.2 (BP = 1.000 ratio = 18.600 hyp_len = 93 ref_len = 5)
```

3. **总结**

通过测试可以看出模型还存在一些问题，如空格不对以及部分词语翻译不对。但是大部分还是可以正确翻译，如果继续训练至收敛，模型会有不错的效果。因为没有找到gpu训练，所以用的mac训练，训练比较慢而且发热比较严重，就没有继续训练下去。