

# 预训练模型学习情况周报9

汇报人：闵德海

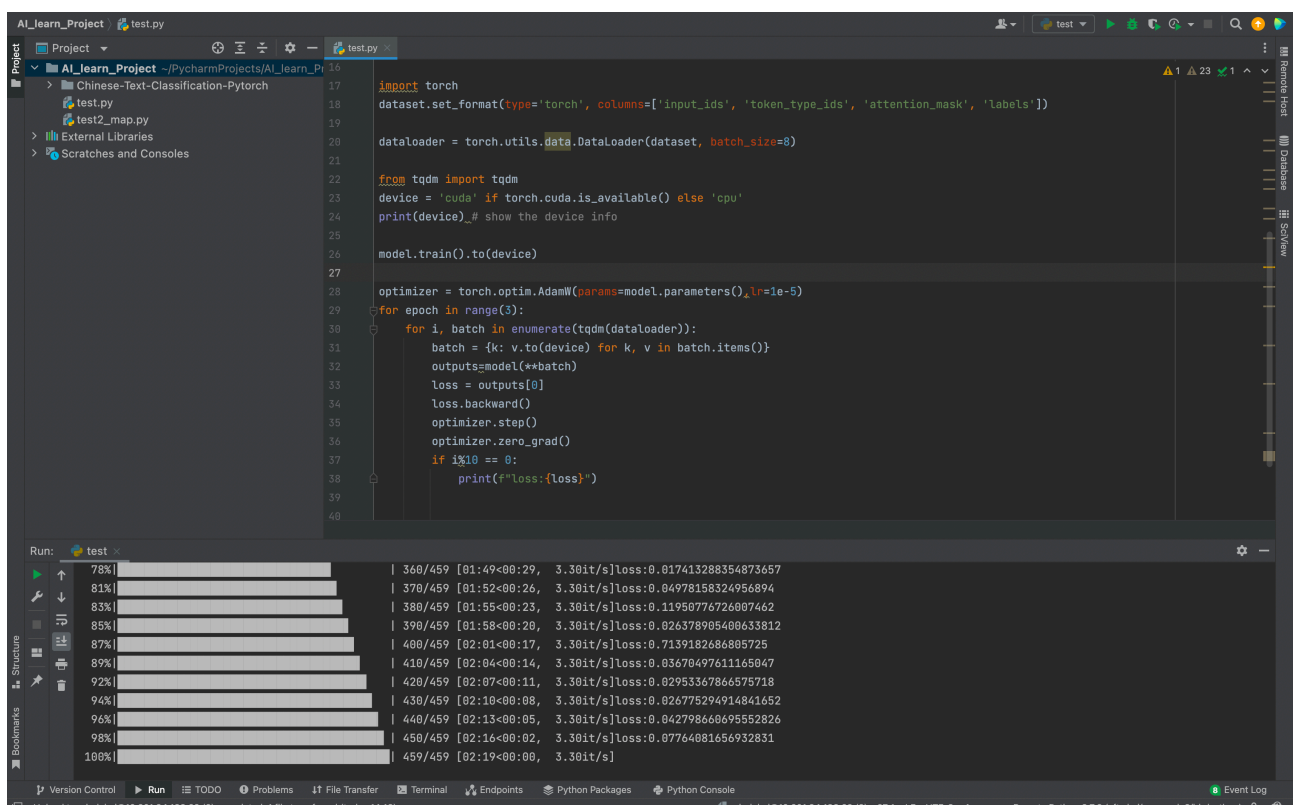
## 本周主要学习内容：

本周重点学习了Bert的模型架构和原理，阅读了提出该模型的论文《BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding》。

观看了论文的讲解视频，对Bert的模型架构有了比较全面的理解，理解了通过Masked Language Model 和 Next sequence prediction 两个任务训练模型对word的document-level级理解的核心思想。

并整理了该小节的学习笔记在个人博客上：[Bert的学习笔记](#)。

同时阅读了huggingface平台上Transformer的tutorial和Datasets的tutorial，并学习了使用Bert模型和Glue数据集去做一些下游任务的Fine-tune操作。



```
16
17 import torch
18 dataset.set_format(type='torch', columns=['input_ids', 'token_type_ids', 'attention_mask', 'labels'])
19
20 dataloader = torch.utils.data.DataLoader(dataset, batch_size=8)
21
22 from tqdm import tqdm
23 device = 'cuda' if torch.cuda.is_available() else 'cpu'
24 print(device)_# show the device info
25
26 model.train().to(device)
27
28 optimizer = torch.optim.AdamW(params=model.parameters(), lr=1e-5)
29 for epoch in range(3):
30     for i, batch in enumerate(tqdm(dataloader)):
31         batch = {k: v.to(device) for k, v in batch.items()}
32         outputs=model(**batch)
33         loss = outputs[0]
34         loss.backward()
35         optimizer.step()
36         optimizer.zero_grad()
37         if i%10 == 0:
38             print(f"loss:{loss}")
39
40
```

Run: test

Progress	Time	Loss
78%	360/459 [01:49<00:29, 3.30it/s]	loss:0.017413288354873657
81%	370/459 [01:52<00:26, 3.30it/s]	loss:0.04978158324956894
83%	380/459 [01:55<00:23, 3.30it/s]	loss:0.11950776726807462
85%	390/459 [01:58<00:20, 3.30it/s]	loss:0.026378985408633812
87%	400/459 [02:01<00:17, 3.30it/s]	loss:0.7139182686805725
89%	410/459 [02:04<00:14, 3.30it/s]	loss:0.03670497611165847
92%	420/459 [02:07<00:11, 3.30it/s]	loss:0.02953367866575718
94%	430/459 [02:10<00:08, 3.30it/s]	loss:0.026775294914841652
96%	440/459 [02:13<00:05, 3.30it/s]	loss:0.04279866869552826
98%	450/459 [02:16<00:02, 3.30it/s]	loss:0.07764081656932831
100%	459/459 [02:19<00:00, 3.30it/s]	

Upload to mindedhai@10.201.94.183:22 (2) completed: 1 file transferred (today 14:16)

目前进度如下，红色对勾为当前完成的任务，圆圈代表正在进行中的任务。

