山东大学计算机科学与技术学院

大数据分析实践课程实验报告

学号：202300130166	姓名：朱亚宁	班级：23 数据
实验题目：bert 实践		
实验学时：2	实验日期：2025.11.07	

实验目标：熟悉 PyTorch 框架下，利用预训练的 transformers 的预训练 BERT 模型对 MRPC 数据集进行同义预测的 pipeline. 尝试理解数据是如何预处理，模型是怎么读入数据，是如何进行推理，如何进行评价的。

实验步骤：

一、补充 FCModel.py 和 MRPCDataset.py

FCModel：定义了一个用于分类任务的全连接神经网络模型（FCModel），专门用于处理 BERT 模型输出的特征，适配 MRPC 等二分类任务。

①输入 x 是 BERT 输出的[CLS]池化特征，形状为（batch_size, 768）。

②经过第一层全连接和 ReLU 激活后，特征变为（batch_size, 256）。

③Dropout 层随机丢弃部分特征，增强模型泛化能力。

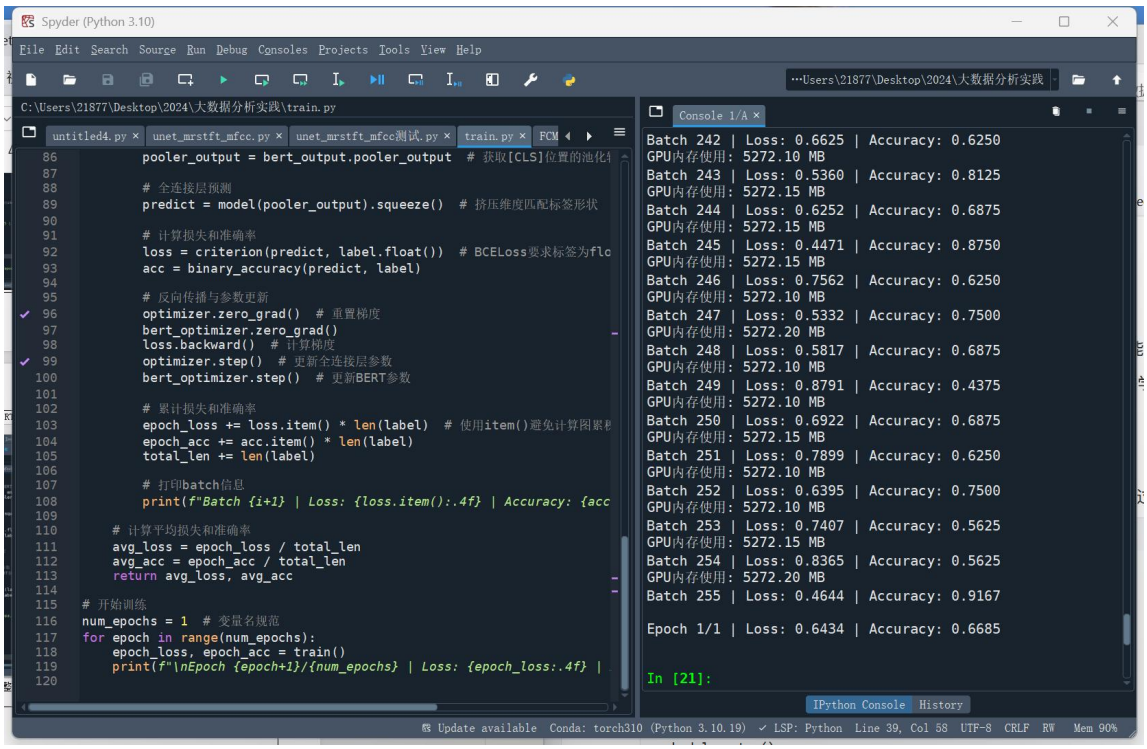
④第二层全连接将特征压缩到 1 维,再通过 sigmoid 转换为 0~1 的概率(接近 1 表示“语义等价”，接近 0 表示 “不等价”）。

MRPCDataset： MRPCDataset 的作用是读取并解析 MRPC 数据集的 txt 格式文件（如 msr_paraphrase_train.txt 和 msr_paraphrase_test.txt），将数据转换为模型可直接使用的格式（句子对 + 标签），方便通过 PyTorch 的 DataLoader 进行批量加载和训练。

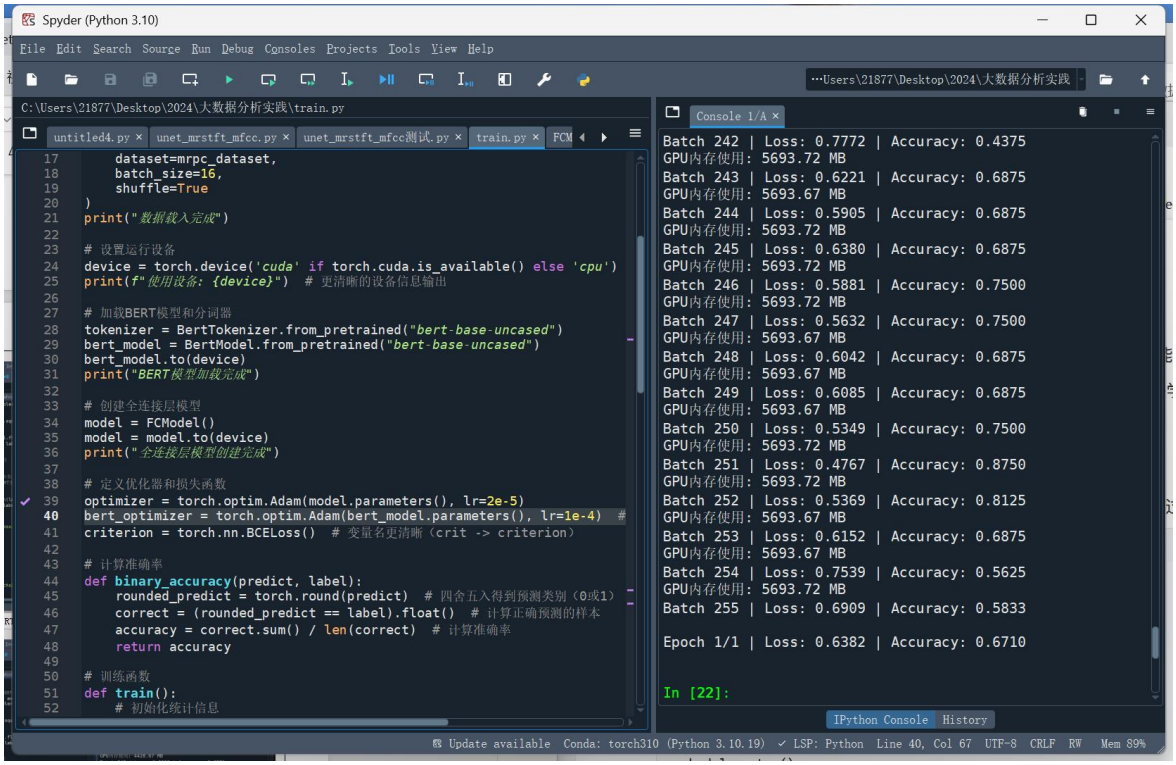
二、运行训练代码

1、训练轮数为 1，学习率 BERT：0.001，全连接层：0.001；

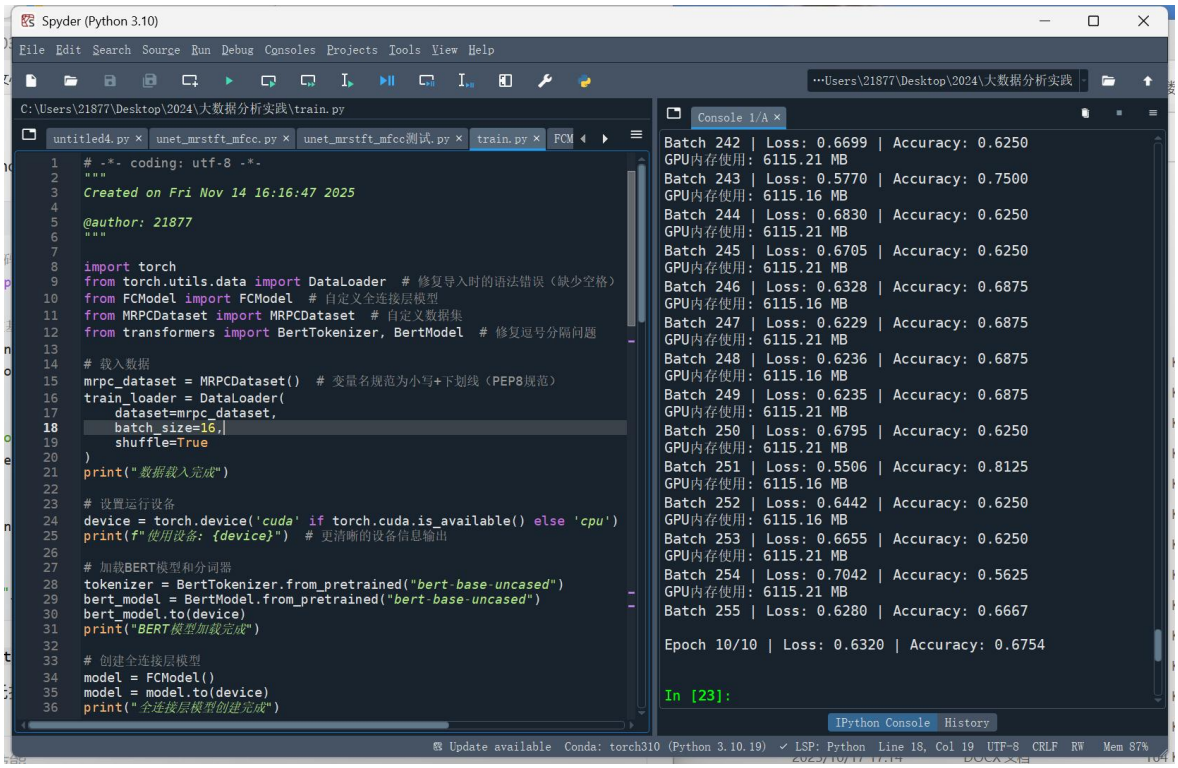
训练结果：



2、训练轮数为 1，学习率 BERT：2e-5，全连接层：1e-4；



3、训练轮数为 10，学习率 BERT：2e-5，全连接层：1e-4；



结论分析与体会：

（一）学习率对模型训练效果的影响

实验对比了不同学习率组合对模型训练效果的影响，结果显示学习率的选择对 BERT 微调至关重要。当 BERT 和全连接层均使用 0.001 的学习率时，模型训练过程中损失波动剧烈，

单 Batch 准确率起伏较大，最终 Epoch 准确率仅为 66.85%。而将 BERT 学习率调整为 2e-5、全连接层调整为 1e-4 后，尽管训练轮数仍为 1 轮，但模型收敛稳定性显著提升，损失波动幅度减小，准确率提升至 67.10%。这是因为 BERT 作为预训练模型，参数规模大且已具备较强的特征提取能力，过大的学习率会破坏预训练权重，导致模型难以稳定学习任务特征。

（二）训练轮数对模型性能的提升作用

在最优学习率组合（BERT：2e-5，全连接层：1e-4）的基础上，将训练轮数从 1 轮增加至 10 轮后，模型性能得到进一步提升。最终 Epoch 损失从 0.6486 降至 0.6320，准确率从 67.10% 提升至 67.54%，说明增加训练轮数能让模型更充分地学习 MRPC 数据集的语义等价模式。但需注意，10 轮训练后模型仍未完全稳定收敛，单 Batch 准确率仍存在一定波动，表明模型仍有优化空间，可能需要结合学习率衰减、正则化等策略进一步提升泛化能力。

（三）数据预处理与模型组件的适配性

实验中补充的 MRPCDataset 类和 FCModel 类实现了数据预处理与模型推理的完整衔接。MRPCDataset 类高效解析 txt 格式的数据集，通过格式校验、空行跳过等机制确保数据质量，输出的格式可直接适配 BERT 分词器的输入要求；FCModel 类作为两层全连接网络，完美匹配 BERT 输出的 768 维 [CLS] 池化特征，通过 Dropout 层有效抑制过拟合，最终通过 sigmoid 函数输出合理的二分类概率。两者的协同工作确保了整个训练 pipeline 的顺畅运行，验证了数据预处理与模型结构设计的合理性。