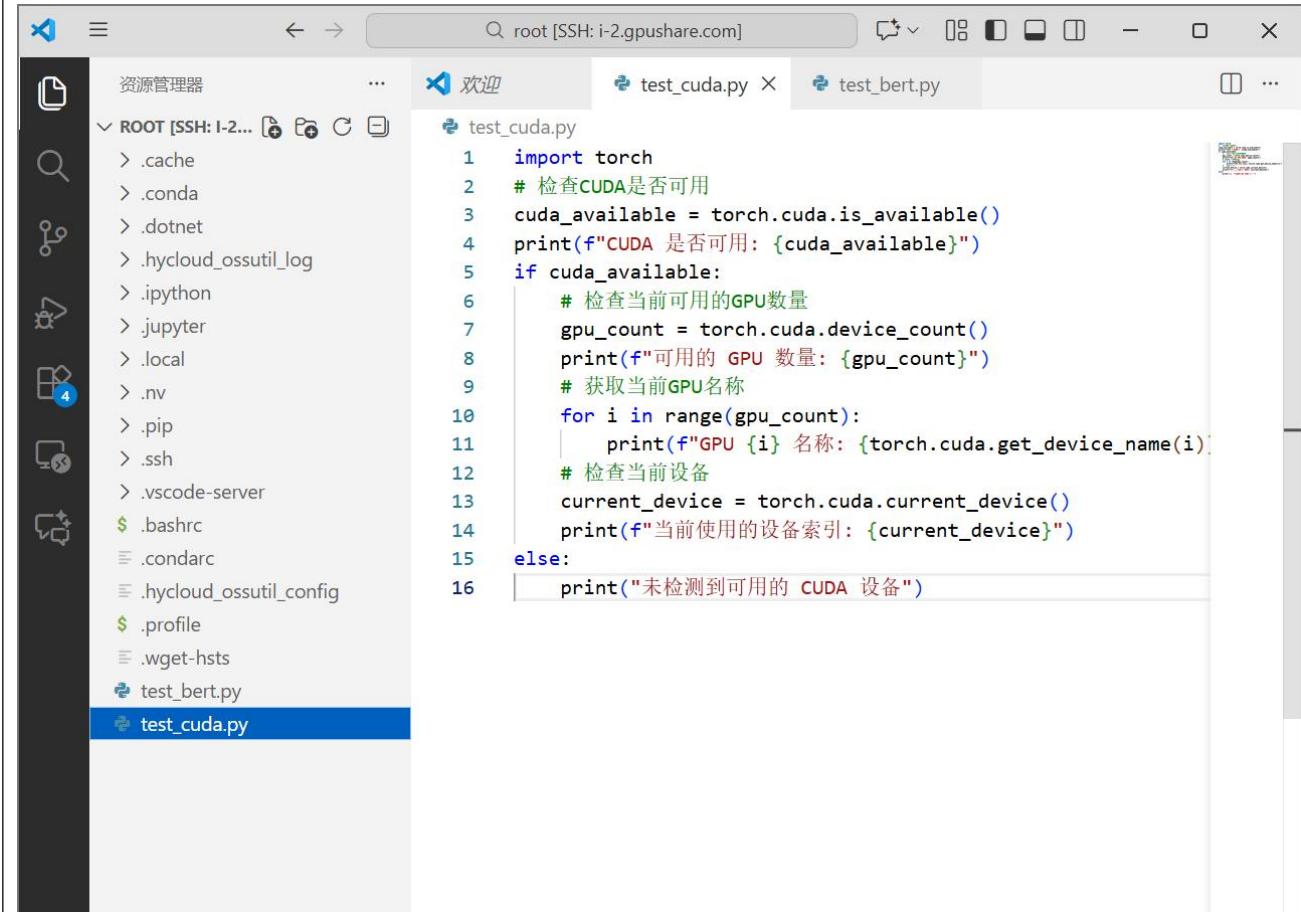


# 山东大学计算机科学与技术学院

## 大数据分析实践课程实验报告

学号: 202300130113	姓名: 丁正旸	班级: 23 数据		
实验题目: BERT 环境配置	实验日期: 2025. 11			
实验学时: 2	实验日期: 2025. 11			
<b>实验目标:</b> 配置可运行 BERT 的远程服务器				
<b>实验过程与内容:</b> <b>创建服务器实例:</b> BERT 是基于 PyTorch 构建的，所以我们在创建服务器时就可直接选择。 <b>镜像实例选择:</b> PyTorch==2.0.0/Cuda==11.7.0/Python==3.8				
				
<b>vscode 远程连接:</b> <b>采用 vscode 插件 remote SSH 远程连接</b>				
				
<b>环境测试:</b> 因为在创建服务器时已经预装好了 PyTorch，那么直接运行代码 test_cuda.py 测试 GPU 环境：				

```
● (base) root@I2595b5652a00801617:~# python test_cuda.py  
CUDA 是否可用: True  
可用的 GPU 数量: 1  
GPU 0 名称: NVIDIA GeForce RTX 3090  
当前使用的设备索引: 0
```

#### BERT 测试:

BERT 是基于 Transformer 架构的，想要运行还要 pip install transformers  
安装好后运行 test\_bert.py 代码测试：

```
(base) root@I2595b5652a00801617:~# python test_bert.py  
pytorch_model.bin: 100%|██████████| 412M/412M [00:07<00:00, 57.2MB/s]  
/usr/local/miniconda3/lib/python3.8/site-packages/torch/_utils.py:776: UserWarning:  
  arning: TypedStorage is deprecated. It will be removed in the future and UntypedStorage  
  will be the only storage class. This should only matter to you if you are using storages  
  directly. To access UntypedStorage directly, use tensor.untyped_storage()  
  instead of tensor.storage()  
  return self.fget.__get__(instance, owner())  
Some weights of BertForSequenceClassification were not initialized from the  
model checkpoint at hfl/chinese-roberta-wwm-ext and are newly initialized:  
['classifier.bias', 'classifier.weight']  
You should probably TRAIN this model on a down-stream task to be able to use  
it for predictions and inference.  
文本：今天的天气真好，我感到非常开心！  
预测的情感：积极  
概率：[[0.366321 0.633679]]
```

#### 结论与体会：

本次实验成功搭建了适配中文情感分析任务的运行环境，解决了 Hugging Face 模型的网络访问限制、Tokenizer 与模型的类型匹配等问题，调用 hfl/chinese-roberta-wwm-ext 模型完成了中文文本的情感预测测试，体会到模型资源的公开性、Tokenizer 与模型的类型匹配度是中文 NLP 环境配置的关键细节，国内镜像源的适配需结合模型的仓库同步情况灵活调整；远程服务器的 GPU 资源搭配本地代码调试的模式，既利用了高性能硬件提升推理效率，又降低了环境配置的操作门槛。