

大数据分析实践实验四

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

组名: Give a five to big data	姓名: 杨笑语 赵汉哲 康海洋 李姿含 李嘉欣	班级: 数据 23
实验题目: Bert 环境配置		
实验学时: 2		实验日期: 2025/10/31
实验目标:		
对动手实践利用机器学习方法分析大规模数据有进一步了解, 并学习如何利用远程环境进行工程代码的调试		
实验步骤:		
租借服务器版		
1. 租借配套的云服务器		

腾讯云

实例 上海 2 其它地域

请您对当前页面进行打分 ★ ★ ★ ★ ★ ×

① 腾讯云禁止客户利用云服务从事虚拟货币相关业务, [查看声明](#)。如您有虚拟货币行为或感染挖矿木马,请参考[自助清理手册](#)进行处理。

新建实例 ··· 开机 关机 重启 续费 销毁/退还 更多操作

多个关键字只支持精准查询,用竖线“|”分隔,多个过滤标签用回车键分隔

ID/名称	状态	监控	可用区	实例类型	操作系统	实例配置
ins-5m17rcg5 未命名	运行中	已监控	上海二区	GPU计算型...	Ubuntu 16.04 LTS (2018-04-11)	4核 20GB 5M 系统盘:通用型SSD云硬盘
ins-ctrqhzzh 未命名	已关机 ① 关机不计费...	未监控	上海五区	标准型S5	Ubuntu 16.04 LTS (2018-04-11)	2核 4GB 5Mb 系统盘:通用

共 2 条

2. 使用 `conda create` 命令创建 base 环境以外的虚拟环境 并且 激活新环境

- 创建好的新环境用命令查看,名字是 `my_project_env`

```
(base) ubuntu@VM-16-4-ubuntu:~$ conda env list

# conda environments:
#
# * -> active
# + -> frozen
base          * /home/ubuntu/miniconda3
my_project_env           /home/ubuntu/miniconda3/envs/my_project_env
```

- 激活新环境

```
(base) ubuntu@VM-16-4-ubuntu:~$ conda activate my_project_env
(my_project_env) ubuntu@VM-16-4-ubuntu:~$
```

3. 安装所需要的包

- 我们指定了 PyTorch 及其配套库 `torchvision` 和 `torchaudio` 的版本，并加入了 `+cu110` 标签，确保安装的是 CUDA 11.0 编译的 GPU 版本。

```
# 安装 PyTorch 1.7.1 + cu110 版本
pip install torch==1.7.1+cu110 torchvision==0.8.2+cu110 torchaudio==0.7.2 -f https://download.pytorch.org/whl/torch_stable.html
Looking in indexes: http://mirrors.tencentyun.com/pypi/simple
Looking in links: https://download.pytorch.org/whl/torch_stable.html
Collecting torch==1.7.1+cu110
  Downloading https://download.pytorch.org/whl/cu110/torch-1.7.1%2Bcu110-cp39-cp39-linux_x86_64.whl (115
6.7 MB)
    ━━━━━━━━━━━━━━━━ 0.3/1.2 GB 11.7 MB/s eta 0:01:11
```

- 接着安装 Transformers

```
COLLECTING JUJU110 (from transformers==4.10.0)
  Downloading http://mirrors.tencentyun.com/pypi/packages/1e/e8/685f47e0d754320684db4425a0967f7d3fa70126
bffd76110b7009a0090f/joblib-1.5.2-py3-none-any.whl (308 kB)
Installing collected packages: tokenizers, urllib3, tqdm, regex, pyyaml, packaging, joblib, idna, hf-xet,
, fsspec, filelock, click, charset_normalizer, certifi, sacremoses, requests, huggingface-hub, transform
ers
Successfully installed certifi-2025.10.5 charset_normalizer-3.4.4 click-8.1.8 filelock-3.19.1 fsspec-202
5.10.0 hf-xet-1.2.0 huggingface-hub-0.36.0 idna-3.11 joblib-1.5.2 packaging-25.0 pyyaml-6.0.3 regex-2025
.10.23 requests-2.32.5 sacremoses-0.1.1 tokenizers-0.12.1 tqdm-4.67.1 transformers-4.18.0 urllib3-2.5.0
(my_project_env) ubuntu@VM-16-4-ubuntu:~$
```

学校服务器版

1. 使用学校的服务器

使用 FinaShell 测试服务器是否连接成功, 文件是否同步上传....

```
连接主机...
连接主机成功
Welcome to Ubuntu 20.04.6 LTS (GNU/Linux 5.15.0-89-generic x86_64)

 * Documentation: https://help.ubuntu.com
 * Management: https://landscape.canonical.com
 * Support: https://ubuntu.com/advantage

Expanded Security Maintenance for Infrastructure is not enabled.

410 updates can be applied immediately.
276 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable

126 additional security updates can be applied with ESM Infra.
Learn more about enabling ESM Infra service for Ubuntu 20.04 at
https://ubuntu.com/20-04

New release '22.04.5 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Your Hardware Enablement Stack (HWE) is supported until April 2025.
Last login: Thu Nov  6 15:48:07 2025 from 172.25.179.203
(base) bioinfo202200101012@rtx2080t:~$
```

2. 下载实验所需要的环境

- `torch==1.7.0`

• transformers==4.18.0

并且将这些环境装在一个 conda 环境中, 可以不污染其他环境
使用命令查看创建好的环境, 并且切换到此环境中去

```
(base) bioinfo202200101012@rtx2080t:~$ conda env list

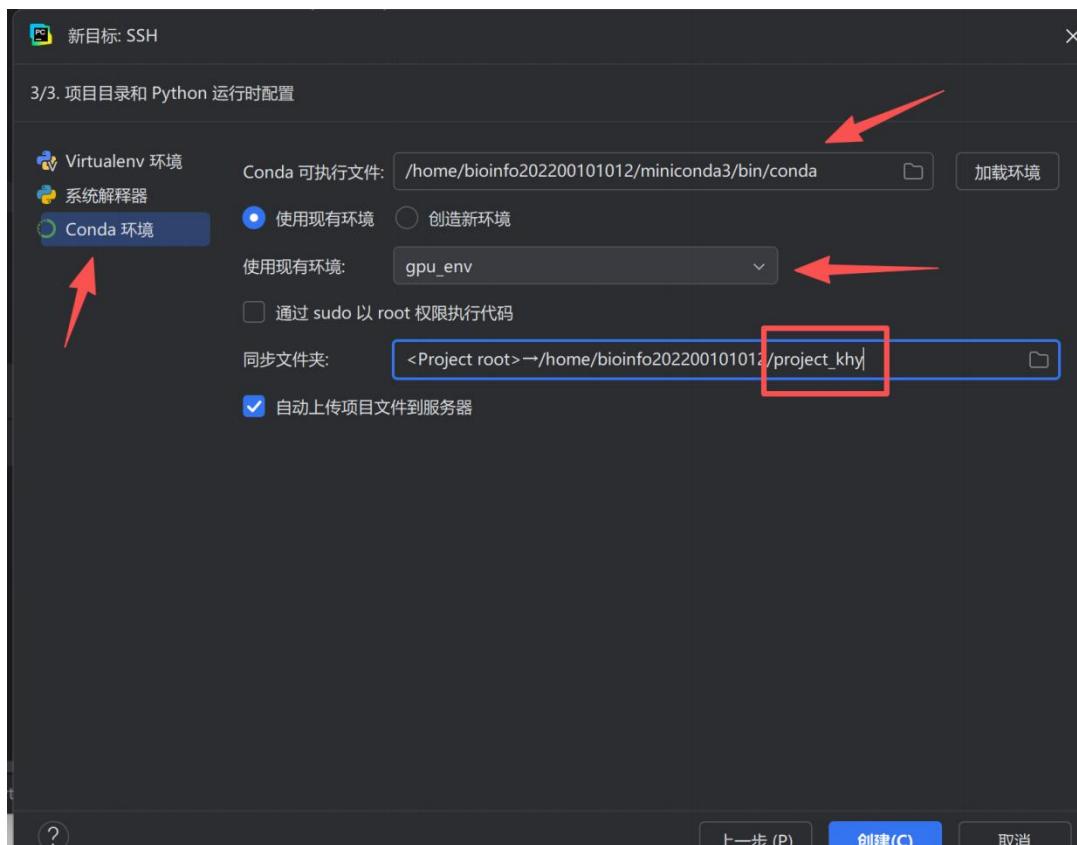
# conda environments:
#
# * -> active
# + -> frozen
base          * /home/bioinfo202200101012/miniconda3
gpu_env           /home/bioinfo202200101012/miniconda3/envs/gpu_env

(base) bioinfo202200101012@rtx2080t:~$ conda activate gpu_env
```

3. Pycharm 配置

主要是文件映射与解释器的选择

文件选择好自己的文件夹, 解释器选择上 conda, 环境选上刚刚创建的



结果图片:

环境配置成功

```
COLLECTING joblib (from version sacremoses==0.1.1)
  Downloading http://mirrors.tencentyun.com/pypi/packages/1e/e8/685f47e0d754320684db4425a0967f7d3fa70126
  bffd76110b7009a0090f/joblib-1.5.2-py3-none-any.whl (308 kB)
Installing collected packages: tokenizers, urllib3, tqdm, regex, pyyaml, packaging, joblib, idna, hf-xet,
, fsspec, filelock, click, charset_normalizer, certifi, sacremoses, requests, huggingface-hub, transformers
Successfully installed certifi-2025.10.5 charset_normalizer-3.4.4 click-8.1.8 filelock-3.19.1 fsspec-2025
.10.0 hf-xet-1.2.0 huggingface-hub-0.36.0 idna-3.11 joblib-1.5.2 packaging-25.0 pyyaml-6.0.3 regex-2025
.10.23 requests-2.32.5 sacremoses-0.1.1 tokenizers-0.12.1 tqdm-4.67.1 transformers-4.18.0 urllib3-2.5.0
(my_project_env) ubuntu@VM-16-4-ubuntu:~$
```