

Uai-train for Deecamp 使用手册

这是一份给 Deecamp2018 T2I 小组的 Uai-train 使用说明文档。 版本号: v0.1

--修改版, 供其他小组使用

--添加其他小组需要修改的位置

--传到 GitHub 中了

Created by team16(T2I) fahai ,2018.07.31

一、创立账户

1. 创立 Ucloud 账户, 该步骤交由组长完成
2. 创立云主机账户, 该步骤交由任意具有管理员权限账户的队员完成, 目前具有管理员账户的组员有: [REDACTED] (注: 请管理员慎用管理员权限, 可以使用普通权限解决绝不动用管理员权限)
3. 云主机 ip: [REDACTED]

Run code:

```
sudo adduser {username} ##添加用户
```

```
sudo gpasswd -a {username} docker ##添加加入 docker 组
```

二、配置

1. GitHub 克隆

```
git clone https://github.com/zzuzayu/uai.git
```
2. 初始化配置
 - a) `cd uai`
 - b) `./getdocker.sh` # 已经配好 docker 环境的无需执行此步
 - c) `source ./path.sh` # 添加环境变量
3. 在 config.txt 中修改常用配置文件

```
1 basic_image="uhub.service.ucloud.cn/uashare/gpu_uaiservice_ubuntu-16.04_python-2.7_tensorflow-1.8.0:v1.0" # -b 设置基础镜像
2 uhub_name="uai_test" # -u 设置镜像库名称
3 add_software="" # -a 需要添加的环境 如 RUN pip install torch torchvision 即可添加pytorch,如果需要多行, 请使用"\n"分隔
4 bucket=t2i #ufile bucket
```

4. 添加 docker 用户组

```
sudo gpasswd -a {username} docker
```
5. Docker 登陆 uhub

```
docker login uhub.service.ucloud.cn
#username: ucloud 账户
```

#passwd: ucloud 密码

6. 写好 config.cfg

详见第三章 3.a) ii

7. 测试

```
uai.sh -n setaname -c codepath
```

成功示例：在 ucloud 的 uhub 中可以找到刚刚建立上传的镜像

```
uaihub: digest: sha256:8f7d1282c04b6f20142681084bbf5cfd4747c486d891442e4d6f38785e898e size: 9776
*****
run code:
docker build -t uhub.service.ucloud.cn/uai_test/zfh_test3:uaihubtrain -f /data/t2i/zfh_tmp.Dockerfile /data/t2i
docker push uhub.service.ucloud.cn/uai_test/zfh_test3
```

失败举例：未登陆 uhub

```
denied: requested access to the resource is denied
*****
run code:
docker build -t uhub.service.ucloud.cn/uai_test/cgy_test1:uaihubtrain -f /data/t2i/cgy_tmp.Dockerfile /data/t2i
docker push uhub.service.ucloud.cn/uai_test/cgy_test1
```

其他失败情况请联系钟发海同学。

三、使用

0. Uai-train 简介

可以认为 uai-train 有三个部分构成：code、data、output。其中 code 在镜像打包过程中提交到 uhub 中，data 为 ufile 中的数据输入源，output 为指定 ufile 输出文件夹。

1. 代码准备

```
from uaitrain.arch.tensorflow import uflag #引入 UAI Train 相关参数
FLAGS = tf.app.flags.FLAGS
```

uaihubtrain 提供三个变量：FLAGS.data_dir, FLAGS.log_dir, FLAGS.output_dir
代码中对数据所在文件夹（FLAGS.data_dir），输出文件夹（FLAGS.output_dir），
log 保存文件夹（FLAGS.log_dir）做相应的替换即可

请将所有代码放入统一文件夹下

官方示例文档：<https://docs.ucloud.cn/ai/uai-train/tutorial/tf-mnist/coding>

2. 代码写入镜像并上传

```
uai.sh -n set_a_name -c code_path
```

其他可选参数详见 Usage：

```
"-b image, -- basic_image" 选择基础镜像
"-a cmd, ---- add_software" 添加其他环境
"-h uhub name,uhub_name" #设置自己的 uhub 镜像库名称
"Please modify common variables in config.txt"也可以修改 config.txt
```

默认基础环境 tf1.8、py2.7

如果有其他基础环境需求，请联系钟发海同学，请使用 -b xximage 选择基础镜像或者修改 config.txt 中的 basic_image 的值

3. Uai-train 平台训练

a) 使用 ufile 上传数据到 ufile 平台（使用 ucloud 网页上传亦可）

```
i. /data/filemgr-linux64.elf/filemgr-linux64 --action mput --
bucket t2i --dir /where/path/your_data/ --trimpath /where/
```

~~--threads 2~~

- ii. 请写好 config.cfg, https://docs.ucloud.cn/storage_cdn/ufile/tools

使用说明

该工具用于在命令行做上传/下载/删除等针对文件的操作, 如果有需要动态上传文件的场景, 请使用API进行内容管理操作。上传超过100MB大小的文件, 网络条件较差, 和UFile的服务器之间的连接经常断开的场景请使用工具中的分片上传 (MPUT)。使用前请修改当前目录下的配置文件config.cfg, 把API密钥加入到配置项:

```
{
    "public_key" : "paste your public key here",
    "private_key" : "paste your private key here",
    "proxy_host" : "www.cn-bj.ufileos.com",
    "api_host" : "api.spark.ucloud.cn"
}
```

不同地域的proxy_host是不一样的, 具体如下:

```
北京
外网: www.cn-bj.ufileos.com
B机房内网: www.ufile.cn-north-02.ucloud.cn
C机房内网: www.ufile.cn-north-03.ucloud.cn
D机房内网: www.ufile.cn-north-04.ucloud.cn
上海二
外网: www.cn-sh2.ufileos.com
内网: www.internal-cn-sh2-01.ufileos.com
香港
```

云主机一般是北京 D 机房内网

- iii. ufileup.sh **datapath** **ufilename** (bucketname)
iv. **datapath** 为数据路径, **ufilename** 为上传到 ufile 平台后文件夹名称。i)
ii) 可直接使用 iii) 替代, 上传成功后在 ufile 平台中将可以看到有
username_data_ufilename/file 样式文件, **bucketname** 设定你们的ufile,
是一个可选参数, 因为你同样可以在 config.txt 中做修改

b) 创建训练任务

- i. 公钥和私钥获取: 产品与服务→操作管理→API 密钥 UAPI
ii. 业务组选 t2i, 代码镜像路径选择刚刚生成上传的镜像
iii. 数据输入源、输出源:
http://t2i.ufile.ucloud.com.cn/username_data_ufilename/
[http://t2i.ufile.ucloud.com.cn/username_data_ufilename/outp
ut/](http://t2i.ufile.ucloud.com.cn/username_data_ufilename/output/)

- iv. 训练启动命令: `python /data/your_python_file.py ...`

官方文档: <https://docs.ucloud.cn/ai/uai-train/tutorial/tf-mnist/train>

4. ~~Jupyter 服务~~

- a) ~~使用云主机账户密码登陆~~
b) ~~使用 Jupyter 可以在线编辑代码, 开 terminal 等~~
c) 云主机的话建议开 jupyterhub

四、其他团队使用注意事项

1. 如何获得脚本及脚本使用方法请到 Github <https://github.com/zzuzayu/uai> （欢迎 fork 和 star ）
2. 基础环境
 - a) 在 uhub 中的分享镜像库中有基本环境镜像，可以拉取
 - b) 下例是 pytorch-0.4 基础环境拉取
 - c) Uhub→分享镜像库→搜 uaishare→点进去搜 torch-0.4 就有下面的结果了
 - d) 如果你需要其他环境，可以在 uai.sh 添加 `-a "RUN pip install xxx"` 类似参数，也可以将参数写在 config.txt 的 add_software 变量中

容器镜像库-UHub

用户镜像 UCloud镜像 Docker官方镜像 分享镜像库 镜像分类

uai

| 镜像库名称 | 镜像库地址 | 镜像数量↓ | 操作 |
|--------------------------|---------------------------------|-------|------|
| 搜索框: uai; 搜索结果: 4条. 返回列表 | | | |
| uaishare | uhub.service.ucloud.cn/uaishare | 120 | 镜像列表 |
| uai_demo | uhub.service.ucloud.cn/uai_demo | 67 | 镜像列表 |

torch-0.4

| 镜像名称 | 创建时间 | 下载次数↓ | Pull命令 | 操作 |
|--|---------------------|-------|---|---------|
| 搜索框: torch-0.4; 搜索结果: 3条. 返回列表 | | | | |
| uaishare/cpu_uaitrain_ubuntu-16.04_python-3.6_pytorch-0.4.v1.0 | 2018-08-01 14:02:33 | 5 | docker pull uhub.service.ucloud.cn/uaishare/cpu_uaitrain_ubuntu-16.04_python-3.6_pytorch-0.4.v1.0 | 添加分类 详情 |
| uaishare/gpu_uaitrain_ubuntu-16.04_python-3.6_pytorch-0.4.v1.0 | 2018-08-01 13:22:03 | 2 | docker pull uhub.service.ucloud.cn/uaishare/gpu_uaitrain_ubuntu-16.04_python-3.6_pytorch-0.4.v1.0 | 添加分类 详情 |
| uaishare/cpu_uaiservice_ubuntu-16.04_python-3.6_pytorch-0.4.v1.0 | 2018-08-01 14:00:45 | 0 | docker pull uhub.service.ucloud.cn/uaishare/cpu_uaiservice_ubuntu-16.04_python-3.6_pytorch-0.4.v1.0 | 添加分类 详情 |