

ATEC TensorFlowGPU组件使用说明

简介

PAI TensorFlowGPU组件（在组件列表中搜索TensorFlow即可）提供了下列功能：

- 单机TensorFlow的训练环境（已预装大部分主流Python包）
- 独占GPU卡，16核CPU，以及56G内存
- 打通上下游数据，已默认加载输入桩（所引用表）的全部数据到内存中
- 提供NAS路径存储模型

使用说明

环境

python 3.6.2

tensorflow-gpu 1.5.0

已提供的python包与版本列表：

```
absl-py (0.2.0)
backports.ssl-match-hostname (3.4.0.2)
backports.weakref (1.0.post1)
bleach (1.5.0)
boto (2.48.0)
boto3 (1.7.10)
botocore (1.10.10)
bz2file (0.98)
certifi (2018.4.16)
chardet (3.0.4)
docutils (0.14)
enum34 (1.1.6)
funcsigs (1.0.2)
gensim (3.4.0)
h5py (2.7.1)
html5lib (0.9999999)
idna (2.6)
iniparse (0.4)
jieba (0.39)
jmespath (0.9.3)
Keras (2.1.6)
kitchen (1.2.5)
lightgbm (2.1.1)
Markdown (2.6.9)
mkl-fft (1.0.0)
mkl-random (1.0.1)
mock (2.0.0)
numpy (1.14.3)
pandas (0.22.0)
pbr (4.0.2)
Pillow (5.1.0)
pip (9.0.1)
protobuf (3.5.2.post1)
pyodps (0.7.17)
python-dateutil (2.7.2)
pytz (2018.4)
PyYAML (3.12)
requests (2.18.4)
s3transfer (0.1.13)
scikit-learn (0.19.1)
scipy (1.0.1)
setuptools (36.4.0)
six (1.10.0)
sklearn (0.0)
smart-open (1.5.7)
tensorflow (1.5.0)
tensorflow-tensorboard (1.5.1)
torch (0.4.0)
torchvision (0.2.1)
urllib3 (1.22)
Werkzeug (0.14.1)
wheel (0.29.0)
xgboost (0.72)
```

资源

目前单个任务独占整台机器，机器资源如下：
GPU（单机单卡）：

+-----+												
NVIDIA-SMI 390.46				Driver Version: 390.46								
+-----+												
GPU Name		Persistence-M		Bus-Id		Disp.A		Volatile Uncorr. ECC				
Fan	Temp	Perf	Pwr:Usage/Cap		Memory-Usage		GPU-Util		Compute M.			
+-----+												
0	Tesla P4		On		00000000:00:08.0 Off				0			
N/A	34C	P8	7W / 75W		0MiB / 7611MiB				0%		Default	
+-----+												
+-----+												
+-----+												
Processes:								GPU Memory				
GPU		PID	Type	Process name				Usage				
+-----+												
No running processes found												
+-----+												

CPU：无限制，可以独占使用ECS 16核
内存：56GB

输入输出

TensorFlowGPU组件提供4个输入桩与4个输出桩，其中输入桩在代码中以df1(1-4)的形式可以直接引用。例如df1中包含的是第一个输入表的全部数据，类型为pandas.core.frame.DataFrame。
df使用示例：

```
dataset = tf.data.Dataset.from_tensor_slices((df1['x1'],df1['x2'],df1['x3'],df1['y'])).shuffle(buffer_size=1000).batch(300).repeat(10)
iterator = dataset.make_one_shot_iterator()
```

topai(1, df1)则将df1内容写入第一个输出桩，即第一个输出表。topai的第一个参数为1-4，第二个参数类型需要是pandas.core.frame.DataFrame。

保存模型

TensorFlowGPU组件提供了支持POSIX接口的分布式存储系统NAS，用于保存TensorFlow模型。对每个用户都默认提供model_dir作为HOME目录，可以直接使用。

使用方法示例：

```
saver = tf.train.Saver()
ckp_path = model_dir + 'model'
saver.save(sess, ckp_path)
```

注意事项

- 1. 限制每个用户正在运行的+pending 的任务加起来不超过两个(<=2)，超过数量则无法继续提交
- 2. 为避免任务长期占据机器资源，每个任务限制最大时长120分钟，超时任务会被KILL。