

Text_Classifier_Pytorch文档



背景：基于Pytorch的文本分类框架。

源码：https://github.com/wzzzd/text_classifier_pytorch

说明：基于T4卡实际测试与官方readme部分细节有所不同，同时会出现一些bug，本文档记录如何使用nativa environment跑通框架。

一、环境

1. 硬件

GPU	每秒浮点运算次数	显卡内存
NVIDIA Tesla T4	7.95 TFLOPS	15 GB
GPU带宽	通道数	PCIE带宽
320.06 GB/s	16	15.75 GB/s
可使用CPU数	可使用内存	硬盘
12	64 GB	LOGICAL VOLUME/LOGICAL VOLUME
硬盘带宽	可用空间	CPU型号
385.72 MB/s	200 GB	Intel(R) Xeon(R) Gold 6271 CPU @ 2.60GHz
主板型号		
YZMB-01130-10C		

2. pip包

```
1  numpy==1.19.2
2  pandas==1.1.5
3  scikit_learn==0.24.2
4  torch==1.7.1
5  tqdm==4.62.3
6  transformers==4.15.0
7  apex==0.1
8  datasets==1.4.0
```

3. Bug

a. `from datasets import Dataset` cannot import name 'Dataset' - Google 搜索

替换datasets版本

b. `RuntimeError: CUDA error: CUBLAS_STATUS_INTERNAL_ERROR when calling
`cublasCreate(handle)`` - Google 搜索

替换torch版本

c. `AttributeError: module 'torch.distributed' has no attribute '_all_gather_base'`

重新安装apex

```
1 $ pip uninstall apex
2 $ rm -rf apex
3
4 $ git clone https://github.com/ptrblck/apex.git
5 $ cd apex
6 $ git checkout apex_no_distributed
7 $ pip install -v --no-cache-dir ./
```

二、使用说明

1. 修改Config.py参数配置
2. `python main.py`