

tensorflow 实验过程可重复

原创

wzg2016 于 2021-10-12 16:55:37 发布 6232 收藏

文章标签: [tensorflow](#) [pytorch](#) [深度学习](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/Strive_For_Future/article/details/120726563

版权

在做科研时使用tensorflow作为算法开发工具。后来有了更好用的pytorch之后, 由于很多基础代码都是用tensorflow编写的, 所以, 就没有转到pytorch上来。

最近在写论文, 比较重视算法的可复现性, 发现在tensorflow中, 相同的参数设置, 相同的种子设置, 相同的输入数据, 相同的优化器, 相同的初始化参数, 在训练过程中会有不同的损失, 且随着迭代次数的增加, 这种“损失差异”的积累最终会导致准确率差异较大的实验结果。同一组参数多组实验, 实验结果浮动较大, 差异之大使自己无法确认自己工作的有效性, 导致在实验分析时无法做控制变量分析。

在网上查找了很多如何固定tensorflow参数, 使实验过程可复现的方法, 但是无一例外都不work, 于是无奈只能转为pytorch。

这里主要记录一下自己在“tensorflow实验过程可重复”上做的调查和测试。

最基本的, 要指定各种随机种子, 然后, 安装tensorflow-determinism,

```
pip install tensorflow-determinism
```

最后, 设置如下:

```
# -*- coding: utf-8 -*-
import os
import random
import numpy as np
import tensorflow as tf
from keras import backend as K
from tfdeterminism import patch

patch()

seed_value = 2021
random.seed(seed)
os.environ['PYTHONHASHSEED'] = str(seed)
np.random.seed(seed)
tf.set_random_seed(seed)
```

但是, 实验结果是然并卵, 完全相同的实验设置, 与数据输入过程, 训练损失仍旧不同。

注:

- tensorflow-determinism 要求指定的tensorflow版本, 如1.14、1.15或2.0。
- 我的tensorflow版本就是1.14
- 然并卵, tfdeterminism并没有帮助, 训练过程仍旧没有复现。

所以，听劝，用pytorch吧：

pytorch的设置如下，实验过程完全可复现。

```
import numpy as np
import torch
seed=args.seed
torch.manual_seed(seed)          # 为CPU设置随机种子
torch.cuda.manual_seed(seed)     # 为当前GPU设置随机种子
torch.cuda.manual_seed_all(seed) # 为所有GPU设置随机种子
from torch.backends import cudnn
cudnn.benchmark = False          # if benchmark=True, deterministic will be False
cudnn.deterministic = True
import random
random.seed(seed)
np.random.seed(seed)
```

参考：[tensorflow-gpu/keras的结果可重复_lzw790222124的博客-CSDN博客](#)