

Edit New issue



## Week 2 Discussion #2

Open

IMOsbo opened 3 days ago

Trying to get the code running on the HPC cluster - it's going ok so far... It actually installed correctly first try, which in my experience is rare!

Running `train.py` is still giving me issues - looks like a pytorch error (ugh)

```
(grok) [imo2d@c1 ~/DATA6330/DATA6550-Reproducibility/Code/Osborne/grok (main)] $ ./scripts/train.py
Namespace(random_seed=-1, gpu=0, max_epochs=None, max_steps=100000, batchsize=0, n_layers=2, n_heads=4, d_model=128, dropout=0.0,
weight_noise=0.0, non_linearity='relu', max_context_len=50, math_operator='+', operand_length=None, train_data_pct=5, warmup_steps=10,
anneal_lr_steps=100000, anneal_lr=False, max_lr=0.001, weight_decay=0, weight_decay_kind='to_zero', noise_factor=0, save_activations=False,
save_outputs=False, logdir='/home/imo2d/DATA6330/DATA6550-Reproducibility/Code/Osborne/grok', datadir='/home/imo2d/DATA6330/DATA6550-
Reproducibility/Code/Osborne/grok/data')
Traceback (most recent call last):
  File "/home/imo2d/DATA6330/DATA6550-Reproducibility/Code/Osborne/grok/./scripts/train.py", line 14, in <module>
    print(grok.training.train(hparams))
  File "/home/imo2d/DATA6330/DATA6550-Reproducibility/Code/Osborne/grok/grok/training.py", line 703, in train
    model = TrainableTransformer(hparams).float()
  File "/home/imo2d/DATA6330/DATA6550-Reproducibility/Code/Osborne/grok/grok/training.py", line 50, in __init__
    self.hparams = hparams # type: ignore
  File "/projects/imo2d/grok/lib/python3.9/site-packages/torch/nn/modules/module.py", line 2029, in __setattr__
    super().__setattr__(name, value)
AttributeError: can't set attribute
```

It looks like its caused by code for an older version of Pytorch: see here: <https://discuss.pytorch.org/t/pytorch-lightning-module-cant-set-attribute-error/121125>

I fixed that error, but it looks like there'll be more... will keep y'all updated.

Create sub-issue
👍 1

richardhoeohn 3 days ago
Collaborator

Greta job!!!

😊

IMOsbo 3 days ago
Owner Author

Ok, I've (hopefully?) gotten it working!

I found a pull request that adds some package versions and changes the hparam issue: [openai/grok#41](#)

Steps to reproduce:

```
# clone the repo
git clone https://github.com/openai/grok

# pull the pull request to a local branch
git fetch origin pull/41/head:dependency_fix

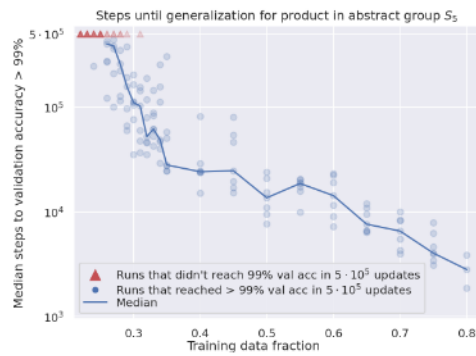
# checkout that branch
git checkout dependency_fix

# create a virtual environment
python -m venv /path/to/your/environment
# source the virtual environment
source /path/to/your/environment/bin/activate

# run the suggested install instructions
```

```
pip install -e .
python scripts/train.py
```

It's currently running on the cluster right now - I'm getting about 400 epochs/minute on a research GPU node, which means it will finish in roughly... 1.73 days. Ugh. I also tried using the a100 node, but got roughly the same results.



I was looking at some of the other graphs in the paper - it looks like we could give it more training data and less validation. Right now it's running off 50% of the dataset - bumping that up to 80% or 90% should be able to give us results a lot quicker - we can crank through 100,000 iterations in about 4 hours. Gonna have to dig through the code to figure out how to change this - probably going to be something with the hparam setup?

I'll also look and see if we can parallelize this any?



richardhoehn 3 days ago

Collaborator ...

I think this is good info @IMOsbo I am working on the paper to add details on cloning of the repository using the correct GIT commands.

Would you like to chat via soom for a minute to make sure we are on the same page?

<https://mtsu.zoom.us/j/86827727198>



IMOsbo 3 days ago

Owner Author ...

Yeah, I'd be able to meet... when works best for you?



richardhoehn 3 days ago

Collaborator ...

I'm on now - If you want to chat a bit and make sure we both use our time wisley. <https://mtsu.zoom.us/j/86827727198>



richardhoehn 3 days ago · edited by richardhoehn

Edits Collaborator ...

```
File Edit View Run Kernel Tabs Settings Help
+
Filter files by name
/
Name Last Modified
data-6550-reproduce 36 minutes ago
ondemand an hour ago

rhoehn@c1:~/data-6550-rex
-rw-r--r-- 1 rhoehn domain users 16 Oct 5 2023 .esd_auth
-rw-r--r-- 1 rhoehn domain users 73 Nov 22 2023 .git-credentials
-rw-r--r-- 1 rhoehn domain users 29 Sep 22 2023 .gitconfig
drwxr-xr-x 2 rhoehn domain users 6 Mar 30 13:34 .ipynb_checkpoints
drwxr-xr-x 2 rhoehn domain users 6 Mar 30 13:33 .ipython
drwxr-xr-x 3 rhoehn domain users 33 Mar 30 14:08 .jupyter
drwxr-xr-x 5 rhoehn domain users 57 Mar 30 13:36 .local
drwxr-xr-x 3 rhoehn domain users 26 Sep 21 2023 .nv
drwxr-xr-x 2 rhoehn domain users 164 Oct 5 2023 .ssh
drwxr-xr-x 5 rhoehn domain users 79 Mar 30 14:02 data-6550-reproduce
drwxr-xr-x 3 rhoehn domain users 18 Mar 30 13:33 ondemand

[rhoehn@c1 ~]$ cd data-6550-reproduce/
[rhoehn@c1 data-6550-reproduce]$ source ~/data-6550-reproduce/environment/bin/activate
(environment) [rhoehn@c1 data-6550-reproduce]$ cd grok/
(environment) [rhoehn@c1 grok]$ python scripts/train.py
Namespace(random_seed=-1, gpu=0, max_epochs=None, max_steps=100000, batchsize=0, n_layers=2, n_heads=4, d_model=128, dropout=0.0, weight_noise=
0.0, non_linearity='relu', max_context_len=50, math_operators='+', operand_length=None, train_data_pct=5, warmup_steps=10, anneal_lr_steps=10000
0, anneal_lr=False, max_lr=0.001, weight_decay=0, weight_decay_kind='to zero', noise_factor=0, save_activations=False, save_outputs=False, logd
ir='/home/rhoehn/data-6550-reproduce/grok', datadir='/home/rhoehn/data-6550-reproduce/grok/data')
/home/rhoehn/data-6550-reproduce/environment/lib/python3.9/site-packages/pytorch_lightning/core/lightning.py:2054: DeprecationWarning: `torch.d
istributed._sharded_tensor` will be deprecated, use `torch.distributed._shard.sharded_tensor` instead
from torch.distributed._sharded_tensor import pre_load_state_dict_hook, state_dict_hook
GPU available: True, used: True
TPU available: False, using: 0 TPU cores
IPU available: False, using: 0 IPU cores
/home/rhoehn/data-6550-reproduce/environment/lib/python3.9/site-packages/pytorch_lightning/trainer/connectors/logger_connector/logger_connector
.py:59: LightningDeprecationWarning: Setting `Trainer(flush_logs_every_n_steps=1000)` is deprecated in v1.5 and will be removed in v1.7. Please
configure flushing in the logger instead.
rank_zero_deprecation(
LOCAL_RANK: 0 - CUDA_VISIBLE_DEVICES: [0]
Set SLURM handle signals.

| Name | Type | Params
0 | transformer | Transformer | 455 K
455 K Trainable params
0 Non-trainable params
455 K Total params
1.822 Total estimated model params size (MB)
Epoch 0: 0% | 0/4 [00:00<?, ?it/s]
/home/rhoehn/data-6550-reproduce/environment/lib/python3.9/site-packages/pytorch_lightning/loops/optimization/closure.py:35: LightningDeprecati
onWarning: One of the returned values {'x_lhs', 'partial_train_loss', 'partial_train_accuracy', 'partial attentions', 'partial_values', 'learni
ng_rate', 'y_hat_rhs'}) has a 'grad_fn'. We will detach it automatically but this behaviour will change in v1.6. Please detach it manually: `ret
urn {'loss': ..., 'something': something.detach()}`
rank_zero_deprecation(
```

10,000

richardhoehn 3 days ago

Collaborator ...

All,

My run finished with 50,000 epochs. I've added the metrics-50000-Epochs/csv file to my /Code folder. Feel free to grab it and see if we can plot the results to resemble those from the paper.

That being said—would you all be open to meeting this week on **Tuesday at 5:00 PM** to review where we are and what we need to complete for submission on Wednesday?

Thanks,  
Richard



Kwoods132 2 days ago

Collaborator ...

5:00 pm this Tuesday works for me.



Hroge19007 2 days ago

Collaborator ...

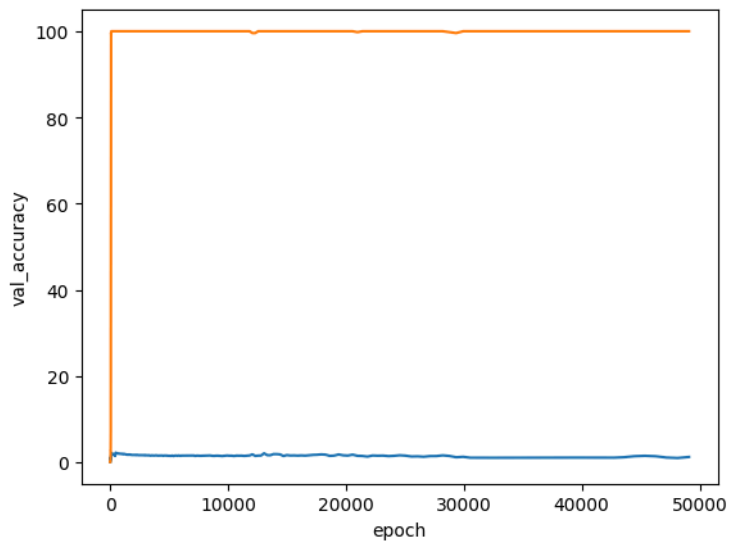
5:00 pm this Tuesday also works with me.



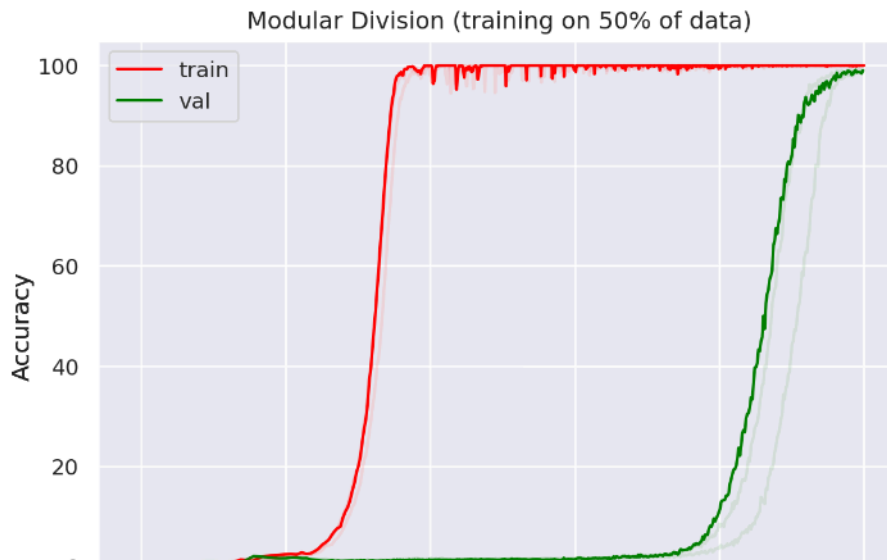
IMOsbo 2 days ago · edited by IMOsbo

Edits Owner Author ...

Ok, so I plotted the results and they're... not great (blue is validation accuracy, orange is training accuracy)



(The paper figure for reference)



richardhoehn yesterday

Collaborator ...

@IMOsbo It took until the next morning... It seemed that as I increased the epoch count the time to process each epoch took longer and longer.

### Zoom Link for this evening:

Richard Hoehn is inviting you to a scheduled Zoom meeting.

Topic: DATA-6550 | Week-B Meeting

Time: Apr 1, 2025 17:00 Central Time (US and Canada)

Join Zoom Meeting

<https://mtsu.zoom.us/j/6153192171?omn=83889836202>

Meeting ID: 615 319 2171

One tap mobile

+13126266799,,6153192171# US (Chicago)

+16465588656,,6153192171# US (New York)



IMOsbo yesterday

Owner Author ...

It doesn't even scale up linearly? That's *really* disappointing (but I think a really important point for our paper)

I tried creating a smaller model that we could overfit on the Titanic dataset to see if it performed any better, but it struggled with the same issues - takes forever. It also tended to diverge after too much training...



Kwoods132 6 hours ago

Collaborator ...

[@richardhoehn](#) and [@IMOsbo](#), do you guys remember how long it took to run the code?



IMOsbo 6 hours ago · edited by IMOsbo

Edits ▾ Owner Author ...

I would hit about 400 epochs / a minute using the cluster, but I never trained it for a large number of epochs either... I only ever ran it for about ~30 minutes - nothing more than 10000 epochs.



richardhoehn now

Collaborator ...

Hi All - I am about to close this down and submit our work.

**I think this might be the best paper so far this semester - Great Job All Around!!!**

Thanks & Havea. good night!



Add a comment

H B I | ≡ <> | 🔗 | ☰ ☷ ☰ | @ 📎 ↶ 📎

Write Preview

Use Markdown to format your comment

📎 Paste, drop, or click to add files

🕒 Close issue ▾

Comment

Assignees

No one - [Assign yourself](#)



Labels

No labels



Projects

No projects



Milestone

No milestone



Relationships

None yet




Development

[Create a branch](#) for this issue or link a pull request.



Notifications

Customize


 Unsubscribe


You're receiving notifications because you're subscribed to this thread.

Participants



→ Transfer issue

 Lock conversation

 Pin issue