ckkissane first commit 5bde1d5 · 4 months ago History

Code Blame 49 lines (44 loc) · 1.01 KB Raw

```python
# %%
from utils import *
from trainer import Trainer
# %%
device = 'cuda:0'

base_model = HookedTransformer.from_pretrained(
    "gemma-2-2b",
    device=device,
)

chat_model = HookedTransformer.from_pretrained(
    "gemma-2-2b-it",
    device=device,
)

# %%
all_tokens = load_pile_lmsys_mixed_tokens()

# %%
default_cfg = {
    "seed": 49,
    "batch_size": 4096,
    "buffer_mult": 128,
    "lr": 5e-5,
```

Symbols

Find definitions and references for functions and other symbols in this file by clicking a symbol below or in the code.

Filter symbols r

const device
const base_model
const chat_model
const all_tokens
const default_cfg
const cfg
const trainer

```
        "num_tokens": 400_000_000,
        "l1_coeff": 2,
        "beta1": 0.9,
        "beta2": 0.999,
        "d_in": base_model.cfg.d_model,
        "dict_size": 2**14,
        "seq_len": 1024,
        "enc_dtype": "fp32",
        "model_name": "gemma-2-2b",
        "site": "resid_pre",
        "device": "cuda:0",
        "model_batch_size": 4,
        "log_every": 100,
        "save_every": 30000,
        "dec_init_norm": 0.08,
        "hook_point": "blocks.14.hook_resid_pre",
        "wandb_project": "YOUR_WANDB_PROJECT",
        "wandb_entity": "YOUR_WANDB_ENTITY",
    }
    cfg = arg_parse_update_cfg(default_cfg)

    trainer = Trainer(cfg, base_model, chat_model, all_tokens)
    trainer.train()
    # %%
```