midGPT: simple hackable repo for FSDP 1B+ base training

Nick Landolfi, Yiding Jiang, Allan Zhou
Stanford University; Supported by TPU Research Cloud

Open source repository for training 1B+ parameter base models

- 500 lines of python; funded by Google's TPU Research Cloud; check out on Github →

- like nanoGPT, but with fully-sharded data parallelism (FSDP)

  - easy to experiment with new initializations, optimizers, architectures

available GPT implementations

midGPT

minGPT nanoGPT
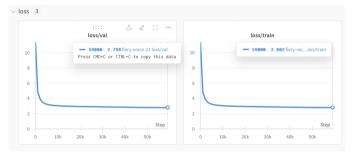
- JAX sharding utilities remove friction like hydrofoils on a boat

Reproduce GPT-2 (124M) in 8 hours using midGPT and 128 TPUs

▶ FSDP saves **time** and **money**

    ▶ TPU v3 128 slice: 8 hours/\$2,250 vs. TPU v3 8 slice: 10 days/\$4,224;

▶ 60,000 steps × 2,048 batch size × 1024 block size ≈ 126B tokens; 124M parameters

Sampling from midGPT trained model: "midGPT is a project to…"

Prompt:

```
midGPT is a project to
```

> […] 3 Items

Samples

∨ [3 Items

  0:
  midGPT is a project to help create new tools that can be used in many situations. If you have an idea for one or more, please send it to me at .

  In summary: The use of data in programming is changing. The future of programming

  1:
  midGPT is a project to study and catalogue solar power and distributed solar technologies.

  As well as being a specialist in the field, the team is working with the likes of solar-powered solar power company SunPower, who is designing the next generation of energy storage systems,

  2:
  midGPT is a project to create a map that helps navigate around the world.

  I contacted them about the idea. We could only find out after a long process that we'd spent hundreds of hours trying to reach a good solution. What they didn't seem to realize is

]

Sampling from midGPT trained model: "innovation endeavors is..."

Prompt:

```
innovation endeavors is
```

› […] *3 Items*

Samples

⌄ [*3 Items*

  0:

innovation endeavors is not limited to building high-performing businesses. In fact, many of the early adopters are more like themselves. Their efforts, while not as disruptive to the economy as they might sound, may be paying off in other ways.

After the

  1:

innovation endeavors is to produce a living, breathing living.

The group of people who will make their living from the work in our lab are the people at Stanford, the people at the Stanford Research Institute, and the people at the MIT Sloan Conference. The group

  2:

innovation endeavors is still afoot, as more details emerge.

"We will release the product in 2014," said David Kennedy, CEO of Motorola Mobility. "Our goal is to bring mobile devices to life and to offer our customers a variety

]

midGPT: a simple hackable repo for 1B+ base parameter models

- check it out on Github!

- **challenge**: debugging on multihost environments can be tough

- **insight**: keeping code as simple and flexible as possible helps