nanoT5: A PyTorch Framework for Pre-training and Fine-tuning T5-style Models with Limited Resources

THE UNIVERSITY of EDINBURGH

Piotr Nawrot
University of Edinburgh

PiotrNawrot/nanoT5

I have this interesting idea to improve Transformer architecture but how should I test it? **It is impossible!**

Pre-training requires enormous compute, model is written in Jax, I don't know how to process the data. I don't even know which dataset to choose because there are millions of them.

I feel you but have you seen **nanoT5**? It is exactly **what you need**.

We reproduce T5-base pre-training and fine-tuning in 16 hours on a single A100 GPU *

Pre-training

We strictly follow the T5-v.1.1-base[1] (248M) pre-training configuration. We pre-train the model using original T5-objective on 8.5B tokens from C4 dataset. It takes 16 hours on 1xA100 GPU.

Written in PyTorch 2.0

We support multi-GPU training. We use *torch.compile, mixed precision,* and memory opts to make it work on older GPUs.

Reproducible

We include detailed instructions how to reproduce our results including environment, configs, training curves, checkpoints, etc

Fine-tuning

We reproduce fine-tuning pipeline of Tk-Instruct[2] on Super Natural-Instructions meta-dataset. We strictly follow the recipe proposed by Tk-instruct and use it to evaluate models from HF Hub and those trained in nanoT5.

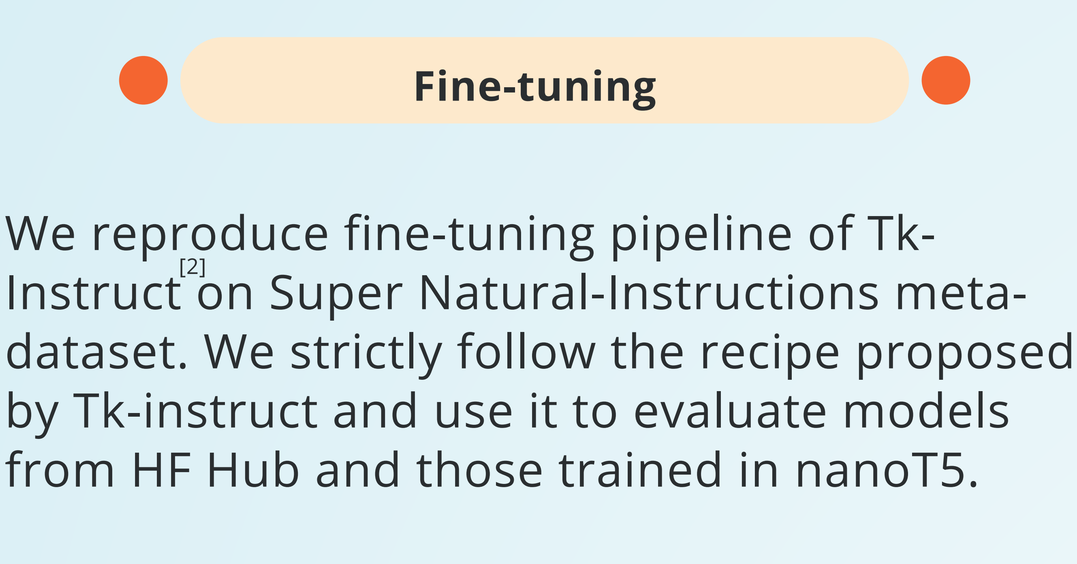
Minimalistic

We expose the training loop and provide optimised T5 implementation written in native PyTorch.

Modular

Our implementation makes it very easy to swap any part of the process such as model, dataset or optimiser.

Results



Optimised

Training starts within minutes after downloading the repo. Pre-training and Fine-tuning of 250M model takes 17 hours.

Robust

850+ stars
50+ forks
0 Issues

References:
[1] Noam Shazeer "GLU Variants Improve Transformer" https://arxiv.org/abs/2002.05202
[2] Wang et al. "Super-NaturalInstructions: Generalization via Declarative Instructions on 1600+ NLP Tasks" https://arxiv.org/abs/2204.07705

"nanoT5 is the ideal template for your LLM research"
~Happy User