Twitter Thread by Yoshitomo Matsubara



Yoshitomo Matsubara
@yoshitomo_cs

As promised in paper, torchdistill now supports ■ @huggingface transformers, accelerate & datasets packages for deep learning & knowledge distillation experiments with ■■ LOW coding cost■

https://t.co/4cWQIL8x1Z

Paper, new results, trained models and Google Colab are ■

1/n

Paper: https://t.co/GX9JaDRW2r
Preprint: https://t.co/ttiihRjFmG

This work contains the key concepts of torchdistill and reproduced ImageNet results with KD methods presented at CVPR, ICLR, ECCV and NeurIPS

Code, training logs, configs, trained models are all available■

2/n

With the latest torchdistill, I attempted to reproduce TEST results of BERT and apply knowledge distillation to BERT-B/L (student/teacher) to improve BERT-B for GLUE benchmark

BERT-L (FT): 80.2 (80.5)
BERT-B (FT): 77.9 (78.3)
BERT-B (KD): 78.9 ■

https://t.co/RAjZPcfGyR

3/n

The pretrained model weights are available @huggingface model hub■
https://t.co/mYapfFGoxH

For these experiments, I used Google Colab as computing resource■■

So, you should be able to try similar experiments based on the following examples! ■
https://t.co/U9lVqmKj5d

4/n