Using the Transformer

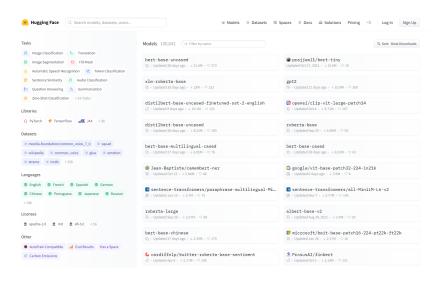
Add-On: Which model to use?



Learning goals

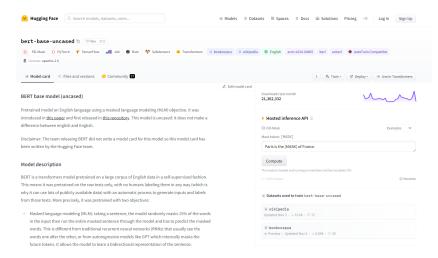
- Huggingface model hub
- Grasp the difficulty of model choice

THE HUGGINGFACE HUB



A SINGLE MODEL

Example: https://huggingface.co/bert-base-uncased



MODEL RECYCLING

A variety of models:

- Everyone can push his/her fine-tuned model to the hub
- Result: A multitude of different models available for each architecture
- → Burning Question: Which one should I use??
 - The "original" pre-trained version
 - · A fine-tuned model on a similar task
 - An ensemble?
 - An averaged model?
 - Something else?

MODEL RECYCLING



Project website: https://ibm.github.io/model-recycling/ Paper: https://arxiv.org/pdf/2211.00107.pdf

MODEL RECYCLING

Best models per architectures

Pretrained	Best model	Avg.	Pretrained Avg.	Ranking
roberta-base	ibm/ColD-Fusion-itr13-seed2	78.72	76.22	link
bert-base- uncased	ffgcc/InfoCSE-bert-base	74.28	72.20	link
bert-base-cased	Dylan1999/bert-finetuned- squad-accelerate	74.07	72.43	link
t5-base	adit94/nlpcharade	78.23	75.45	link
google/t5-v1_1- base	anshoomehra/t5-v1-base-s2- auto-qgen	74.27	68.82	link