DUMP

important: https://arxiv.org/abs/2301.04502
Important: https://arxiv.org/abs/2302.06600

Transformers = 🤖

Unstructured Pruning

LTH

A million paper

Transfer

w/o Learned Sparsity

The Lottery Tickets Hypothesis for Supervised and Self-supervised Pre-training in Computer Vision Models
- Performs iterative magnitude pruning (IMP) on the transfer tasks and find subnetworks

w/ Learned Sparsity

Data-Efficient Double-Win Lottery Tickets from Robust Pre-training (2022)
- Uses some variation of IMP to find safe-robust sparse subnetworks

Sparse Transfer Learning via Winning Lottery Tickets (2019)
- Special LTH pretraining that transfers better

One ticket to win them all: generalizing lottery ticket initializations across datasets and optimizers
- Special LTH pretraining that transfers better

- Source tasks size matters for transfer

How well do sparse model transfer? (2022)
- Studies how different (unstructured) pruning pretraining algorithms transfer to new tasks

Bespoke vs. Prêt-à-Porter Lottery Tickets: Exploiting Mask Similarity for Trainable Sub-Network Finding (2020)
- Lottery tickets obtained from the same original network across several image classification tasks share partial structure and can be combined to obtain an even sparser lottery ticket that works well across tasks (but is faster to obtain)

[2005.05232] On the Transferability of Winning Tickets in Non-Natural Image Datasets
- Mixed results on transferring lottery tickets to medical image tasks in the low data regime

DiSparse: Disentangled Sparsification for Multitask Model Compression
- Multitask pruning

🤖 Learning to Win Lottery Tickets in BERT Transfer via Task-agnostic Mask Training

Structured pruning

LTH

Transfer

w/o Learned sparsity

🤖 SPARSEGPT: MASSIVE LANGUAGE MODELS CAN BE ACCURATELY PRUNED IN ONE-SHOT
- Basically uses some kind of sparse regression technique

USSSSSS

w/ learned sparsity

🤖 SPARSEGPT: MASSIVE LANGUAGE MODELS CAN BE ACCURATELY PRUNED IN ONE-SHOT

- Semi-structured as in Learning N:M fine-grained structured sparse neural networks from scratch.
    - Look Into
- Sparsity is learned on the pretraining data after pretraining, uses a sparse regression approach keep the function intact afer pruning
- Some zero-shot experiments, this is why it's classified as transfer
- Should look into the 2:4 and 4:8 semi-structured pruning approaches

Coarsening the Granularity: Towards Structurally Sparse Lottery Tickets

- LTH training w/ structured pruning

Winning the Lottery Ahead of Time: Efficient Early Network Pruning

- LTH training w/ structured pruning

On Iterative Neural Network Pruning, Reinitialization, and the Similarity of Masks

- Investigates structured vs unstructured pruning and studies how weights evolve
- Shows that unstructured pruning with rewinding gives rise to relatively structured pruning patterns

Conv net specific

Pruning Filters for Efficient ConvNets

- Shows that you can prune ~38% of convolution filters of Resnet filters w/ drop in accuracy
- Not zero shot,
- No transfer

Channel Pruning for Accelerating Very Deep Neural Networks

- Title is good enough summary

Transformer specific

Sample vs compute efficiency:

Scaling Laws for transfer

Training Compute-Optimal Large Language Models

ALSO:

Data-Efficient Double-Win Lottery Tickets from Robust Pre-training

-

Head-to-toe