the Lottery Ticket Hypothesis

Практик-исследователь (Пётр Григорьев, БПМИ181)

1. История публикаций
   - ICLR | 2019, Oral, Best Paper
   - 5 версий статьи
     - опубликована только последняя версия
     - дважды менялось название
     - убрали пример с xor
     - добавили аппендикс, в него включили конкретику прунинга, спуска

2. Авторы
   - MIT, первое соавторство
   - Первый автор занимался в основном recognition и privacy
   - Второй автор занимался в том числе оптимизацией сетей (с 2017)
   - Продолжали работу по теме

3. Наиболее повлиявшие работы
   - Learning both weights and connections for efficient neural network - конкретный метод прунинга
     - In [4] and follow-up work, network compression takes place in three iterative steps. First, a large network is trained. Second, weights or units are pruned according to a heuristic. Third, the network is further trained using the already-trained weights. Han et al. find that, without this third retraining step, network performance drops off much earlier in the pruning process. Han et al. also caution that the pruned network should not be re-initialized after training, but do not consider reusing the values to which the surviving weights were initialized in the original network as we do.
   - Understanding dropout. In Advances in neural information processing systems - градиентный спуск на подсетях
     - Follow-up work on dropout [1] has characterized training with dropout as “perform[ing] gradient descent...with respect to...the ensemble of all possible subnetworks” and inference with dropout as approximately computing the average over this ensemble. In the terminology of dropout, our experiment aims to discover a single, particularly successful member of this ensemble of subnetworks.
   - Конкуренты: DEEP-R
     - DEEP R, that enables us to train directly a sparsely connected neural network. DEEP R automatically rewires the network during supervised training so that connections are there where they are most needed for the task, while its total number is all the time strictly bounded

4. Продолжения и цитирования
   - Продолжения:
     - Stabilizing the Lottery Ticket Hypothesis / The Lottery Ticket Hypothesis at Scale
     - Linear Mode Connectivity and the Lottery Ticket Hypothesis
     - The Lottery Ticket Hypothesis for Pre-Trained BERT Networks

    - и др.
- ~1500 цитирований. Мета-мета обзоры, работы про прунинг
    - Explaining explanations: An overview of interpretability of machine learning
    - Rethinking the value of network pruning
    - Importance estimation for neural network pruning
    - и др.

5. Идеи исследований
    - Выявить выигрышные билеты без обучения
    - Аналогичный анализ для сетей другого рода (например, авторы уже делали для bert)

6. Применения
    - Прямое