



22.02.2022

# The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

Jonathan Frankle, Michael Carbin, 2019

Кирилл Сизов



Снижение количество параметров в нейросетях может оптимизировать требования к вычислительным мощностям, хранению без потери качества моделей.



# Гипотеза лотерейного билета

Абстрактно

*Гипотеза лотерейного билета* - это гипотеза, которая говорит нам, что случайно инициализированная нейронная сеть имеет внутри себя подсеть, которая при изолированном обучении может соответствовать качеству исходной сети после обучения не более чем за то же количество итераций.



## Гипотеза лотерейного билета

Формально

- Пусть  $f(x, \theta)$  нейронная сеть с начальными параметрами  $\theta = \theta_0 \sim \mathcal{D}_\theta$ .
- Пусть при обучении с SGD на тренировочной выборке,  $f$  достигает минимальный loss  $L$  на  $j$  итерации с тестовым ассигасу  $\alpha$ .
- Рассмотрим обучение  $f(x, m \odot \theta)$ , где  $m \in \{0, 1\}^{|\theta|}$  - маска отобранных параметров с начальной инициализацией  $m \odot \theta_0$
- Пусть при нем достигается минимальный loss  $L'$  на  $j'$  итерации с тестовым ассигасу  $\alpha'$
- Гипотеза лотерейного билета утверждает, что  $\exists m$ , что  $j' \leq j, \alpha' \geq \alpha, \|m\|_0 \ll |\theta|$



## Выигрышные билеты

Подсети  $f(x, m \odot \theta)$ , которые при изолированном обучении имеют качество не хуже при меньшем количестве итераций называют *выигрышными билетами*.



## Поиск выигрышных билетов

One-shot pruning

1. Случайно инициализируем нейронную сеть  $f(x, \theta_0)$ ,  $\theta_0 \sim \mathcal{D}_\theta$ :
2. Обучим сеть за  $j$  итераций, получая параметры  $\theta_j$
3. Сократим  $p\%$  наименьших весов  $\theta_j$ , создавая маску  $m$
4. Сбросить параметры до  $\theta_0$ , создавая выигрышный билет  $f(x, m \odot \theta_0)$



## Поиск выигрышных билетов

Iterative pruning

- Идея: делать обрезку не однократно, а в течение  $n$  раундов.
- В каждом раунде происходит обучение, сокращение и сброс параметров.
- Во время сокращения отбрасываются  $p^{\frac{1}{n}}\%$  из неотсеянных параметров.



## Зачем

1. Повысить качество прогнозов и времени работы моделей.
2. Создавать более продвинутые архитектуры.
3. Улучшить теоретическое понимание нейронных сетей.





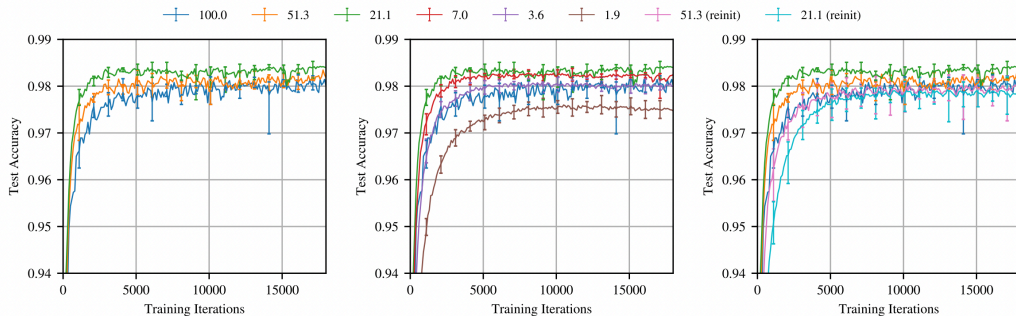
## Результаты

<i>Network</i>	Lenet	Conv-2	Conv-4	Conv-6	Resnet-18	VGG-19
<i>Convolutions</i>		64, 64, pool	64, 64, pool 128, 128, pool	64, 64, pool 128, 128, pool 256, 256, pool	16, 3x[16, 16] 3x[32, 32] 3x[64, 64]	2x64 pool 2x128 pool, 4x256, pool 4x512, pool, 4x512
<i>FC Layers</i>	300, 100, 10	256, 256, 10	256, 256, 10	256, 256, 10	avg-pool, 10	avg-pool, 10
<i>All/Conv Weights</i>	266K	4.3M / 38K	2.4M / 260K	1.7M / 1.1M	274K / 270K	20.0M
<i>Iterations/Batch</i>	50K / 60	20K / 60	25K / 60	30K / 60	30K / 128	112K / 64
<i>Optimizer</i>	Adam 1.2e-3	Adam 2e-4	Adam 3e-4	Adam 3e-4	← SGD 0.1-0.01-0.001 Momentum 0.9 →	
<i>Pruning Rate</i>	fc20%	conv10% fc20%	conv10% fc20%	conv15% fc20%	conv20% fc0%	conv20% fc0%



# Выигрышные билеты в полносвязных сетях

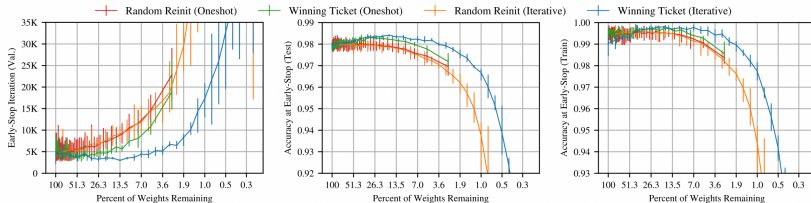
## Пример архитектуры Lenet для MNIST с разными разреженностями маски



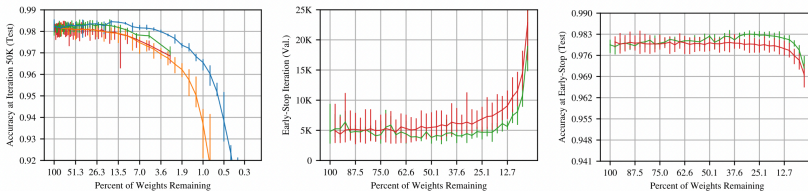


# Выигрышные билеты в полносвязных сетях

## Пример архитектуры Lenet для MNIST с разными разреженностями маски



(a) Early-stopping iteration and accuracy for all pruning methods.



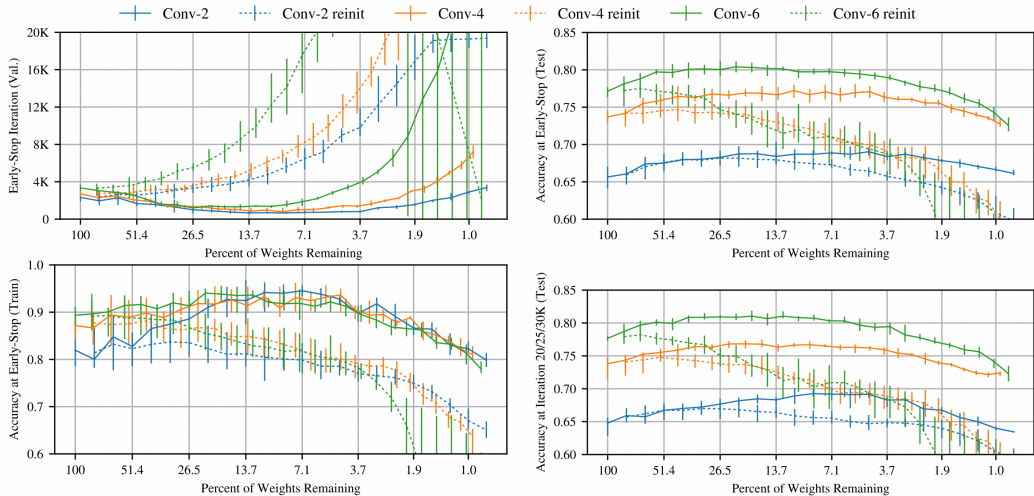
(b) Accuracy at end of training.

(c) Early-stopping iteration and accuracy for one-shot pruning.



# Выигрышные билеты в сверточных сетях

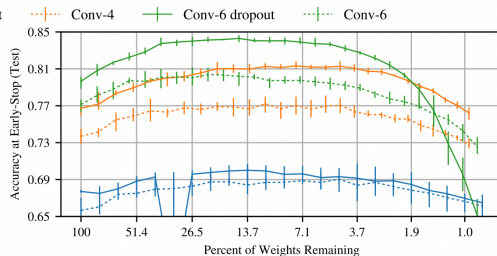
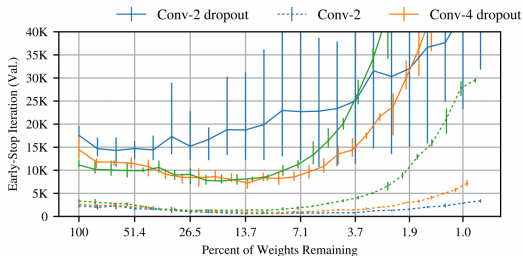
## Архитектура с разным количеством сверточных слоев на CIFAR10





# Выигрышные билеты в сверточных сетях

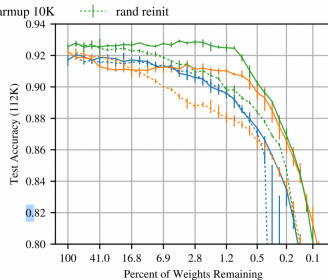
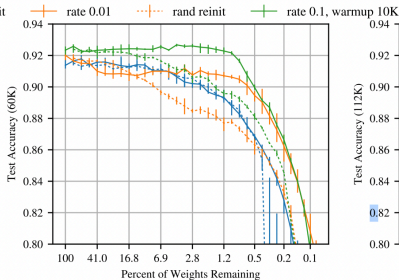
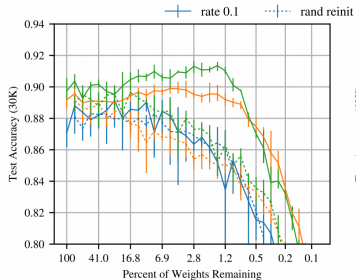
## Связь с dropout





# Практические архитектуры

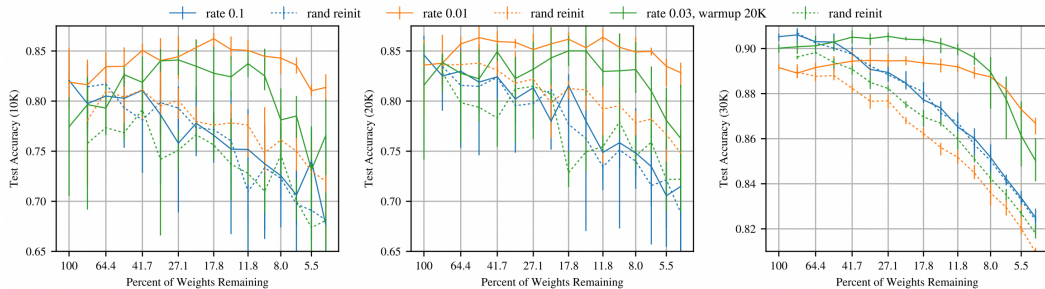
## VGG-19 для CIFAR10





# Практические архитектуры

## Resnet-18 для CIFAR10





## Заключение

На приведенных архитектурах увеличилось качество прогнозов, уменьшилась скорость сходимости и упростилась модель.