

리뷰 - ImageNet in 15 min

Share



임도형

Jul 04, 2018

개요

원제 : Extremely Large Minibatch SGD: Training ResNet-50 on ImageNet in 15 Minutes

[리뷰 - ImageNet in 1 Hour](#)의 방법을 따라 해 본것. 유의미하다는 실증 예.

8K에 1시간 이었는데 4배 키운 32K로 15분(=1시간/4)의 시간이 소요됨.

고찰

성능저하를 방지하려면 이러저러한 방법들이 시도되어야 한다.

방법

이전 논문에서는 mini batch 사이즈가 8K보다 커지면 성능 저하가 있었다.

이 논문에서는 성능저하를 방지하기 위해 다음을 사용.

- RMSprop Warm-up
- Slow-Start Learning Rate Schedule
- Batch Normalization without Moving Averages

min batch 사이즈 32K(=32 x 1024), GPU 1024개 사용

RMSprop Warm-up

초기에는 RMSprop를 사용. 이후 SGD를 사용.

자세한 내용은 논문 부록에 기술되어 있음.

학습율 스케줄

초기 상태를 좀더 오래 유지하고, 초기 학습율을 더 작게 했다.

$$\eta_{base} = 0.1 \cdot \frac{n b_{local}}{256}, n = \text{워커수 } 1024, b_{local} = \text{워커의 mini batch 사이즈 } 32.$$

그래서 기반 학습율은 12.8

초기 40 epoch : 0.5 x 기반 학습율

다음 30 epoch : 0.075 x 기반 학습율

다음 15 epoch : 0.01 x 기반 학습율

다음 5 epoch : 0.001 x 기반 학습율

이동 평균 없이 Batch Normalization

잘 이해 안됨.

SW

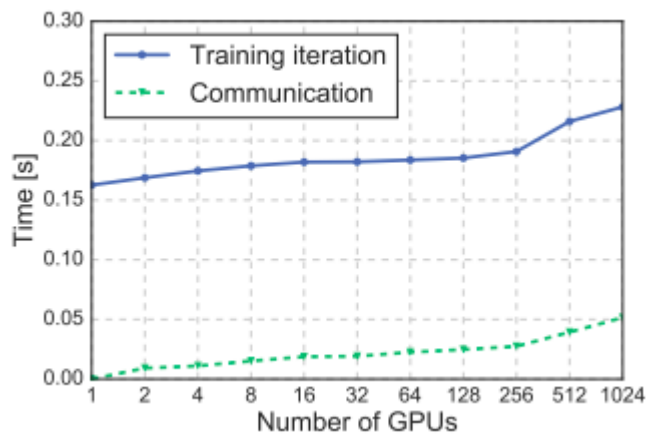
- Chainer 3.0.0rc1 : deep learning framework
- ChainerMN 1.0.0 : multi-node distribute deep learnig 애드온 패키지
- NCCL 2.0.5 : NVIDIA Collective Communication Library
- Open MPI 1.10.2

실험 결과

Team	Hardware	Software	Minibatch size	Time	Accuracy
He <i>et al.</i> [5]	Tesla P100 × 8	Caffe	256	29 hr	75.3 %
Goyal <i>et al.</i> [4]	Tesla P100 × 256	Caffe2	8,192	1 hr	76.3 %
Codreanu <i>et al.</i> [3]	KNL 7250 × 720	Intel Caffe	11,520	62 min	75.0 %
You <i>et al.</i> [10]	Xeon 8160 × 1600	Intel Caffe	16,000	31 min	75.3 %
This work	Tesla P100 × 1024	Chainer	32,768	15 min	74.9 %

성능은 떨어지지 않는다.

mini batch 사이즈와 실행 시간은 정확히 반비례이다.



CPU 갯수가 늘수록 communication 시간이 커짐.

Reference

- paper : https://www.preferred-networks.jp/docs/imagenet_in_15min.pdf

👍 좋아요 처음으로 좋아하는 사람이 되볼까요?

레이블 없음 ✎