

Caltech



signSGD with Majority Vote is Communication Efficient and Byzantine Fault Tolerant

Jeremy Bernstein*

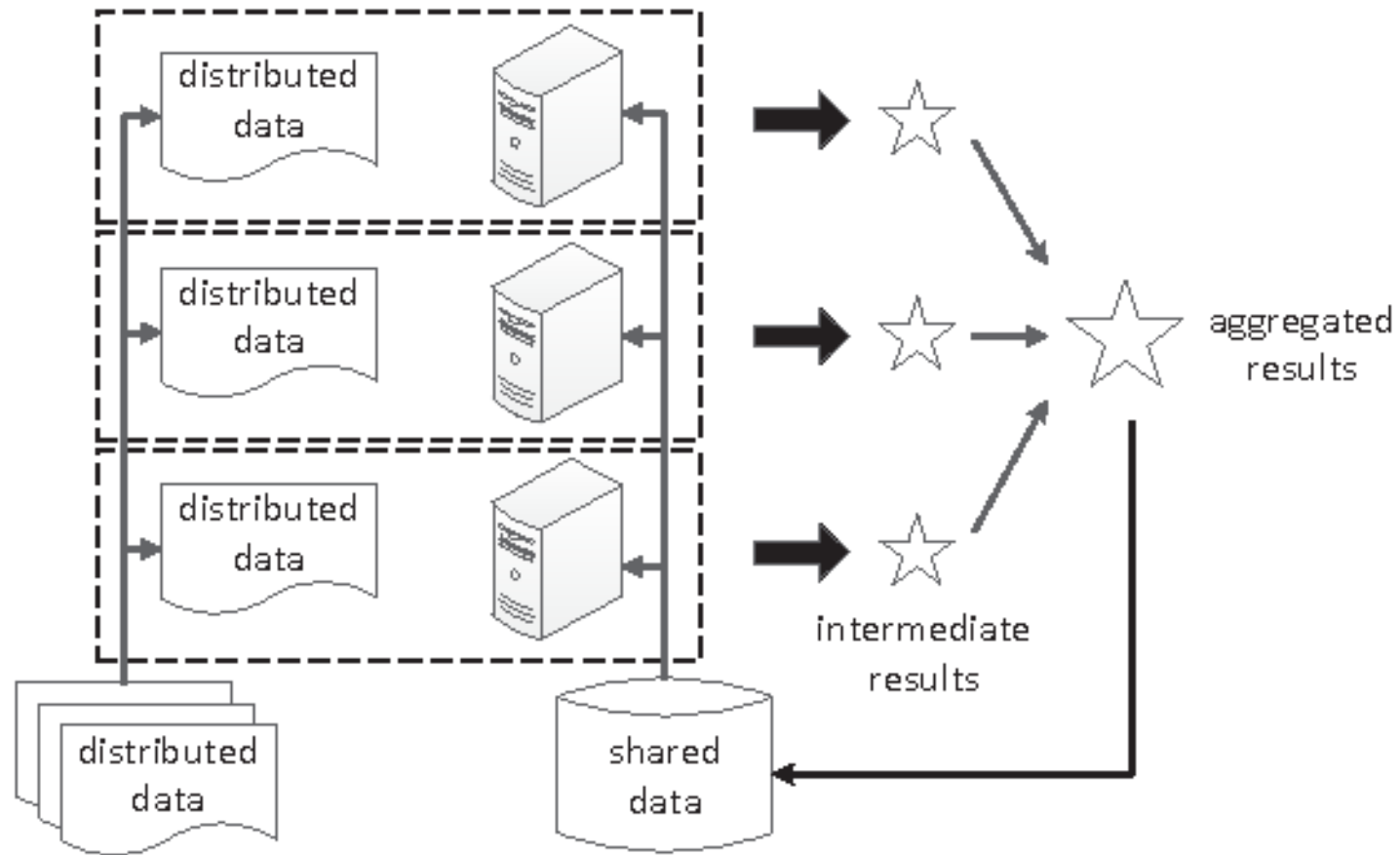
Jiawei Zhao*

Kamyar Azizzadenesheli

Anima Anandkumar



Standard Distributed Machine Learning



Problem

- time cost of communicating gradients
- network faults for large machine counts.

Proposal

Worker → Server:

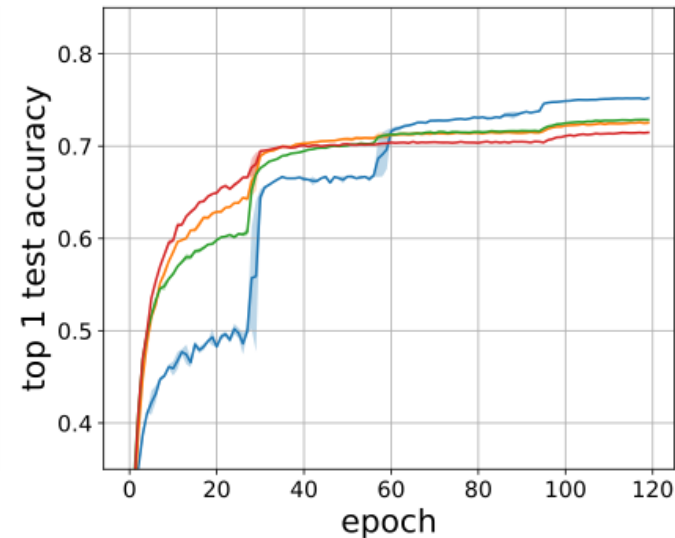
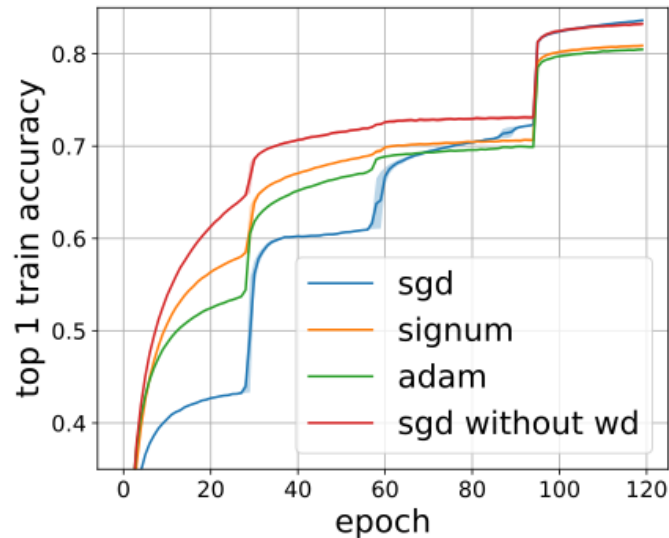
sign(g)

Server → Workers:

sign[\sum sign(g)]

Motivation from Sign

- signSGD: Compressed Optimization for Non-Convex Problems
ICML 2018 by Jeremy Bernstein, etc.
- Both theoretical and experimental provement



Framework

- Nvidia all-reduce, for intra-communication
- Signum with Majority Vote, for inter-communication (NV links)

Implementation – Signum with Majority Vote

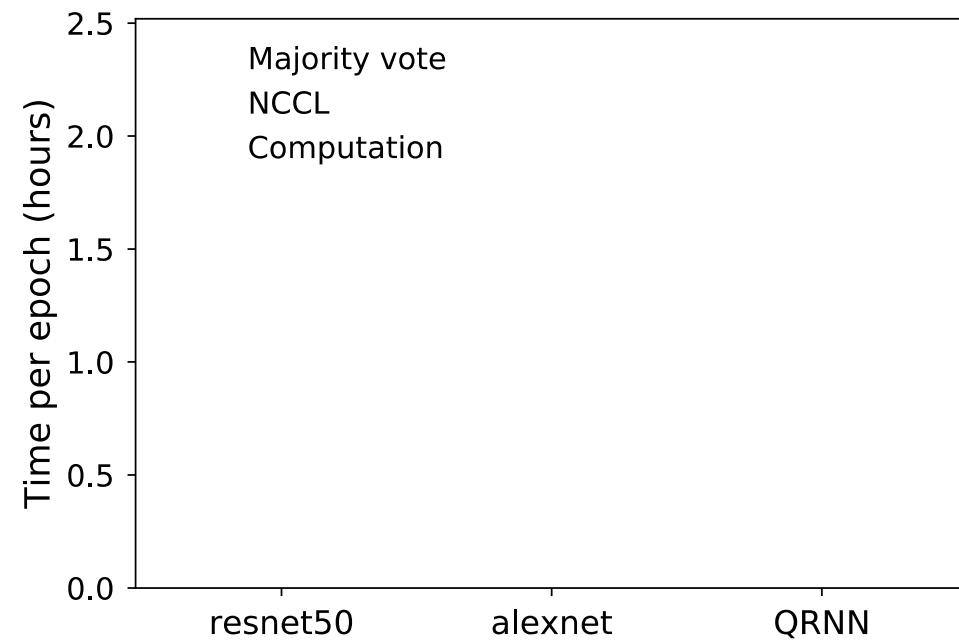
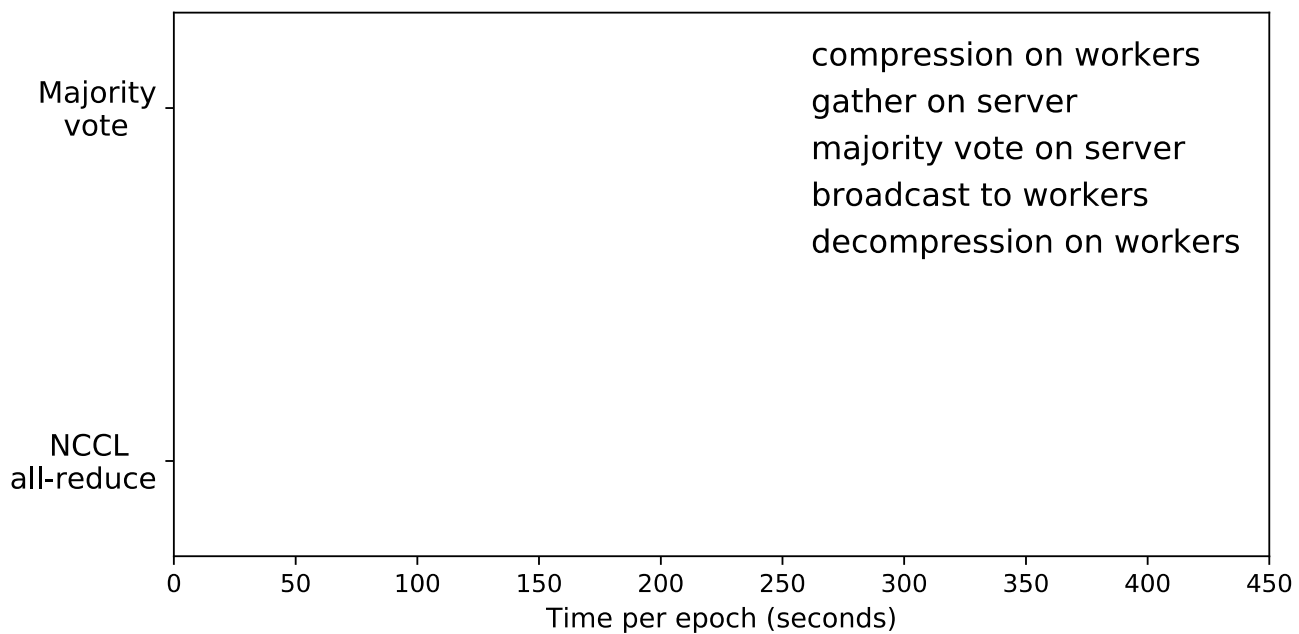
- Backend selection (Like MPI, Gloo, NCCL)
- PyTorch only support 8-bits tensor at least.

Therefore, we need a packing/unpacking system

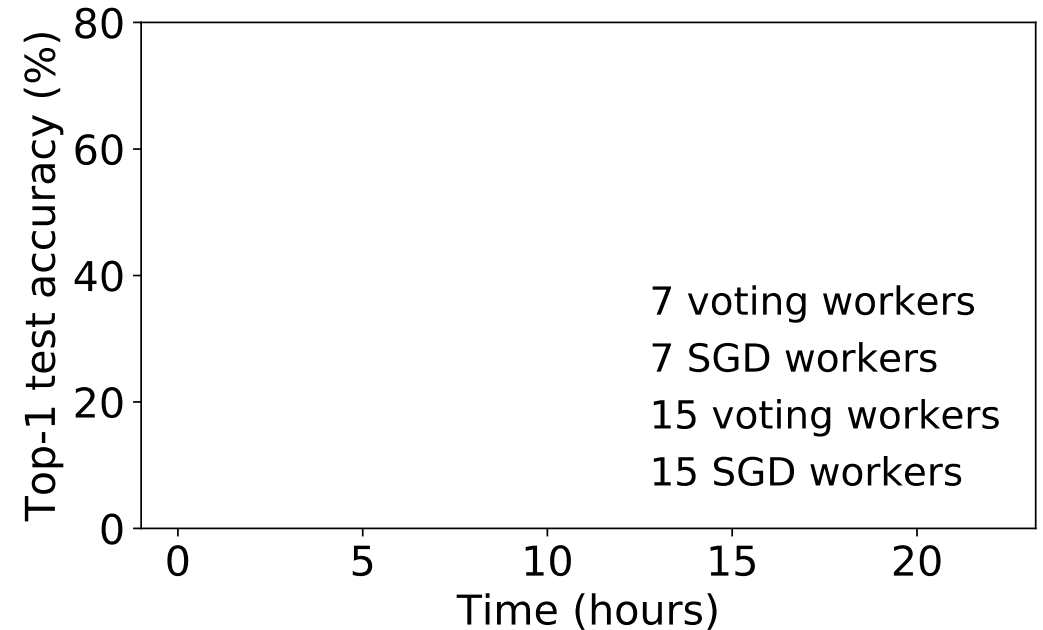
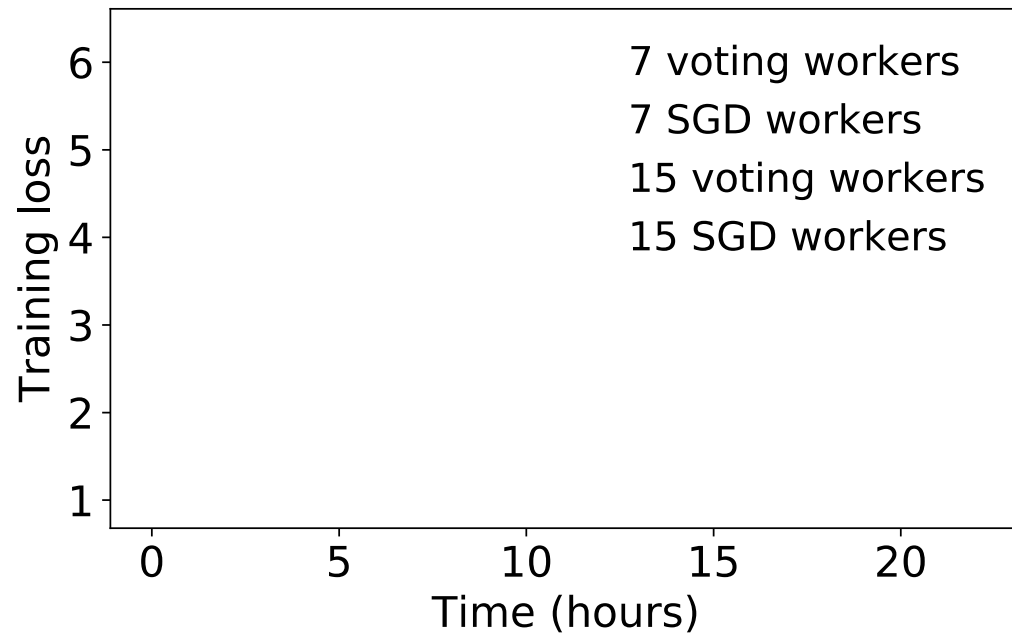
Compression System

- Put 32 32-bits tensor into one 32-bits tensor, just taking the sign.
- Efficiency is a critical problem
- Use cuda-kernel and cost-effective compression policy.

Experiment Result

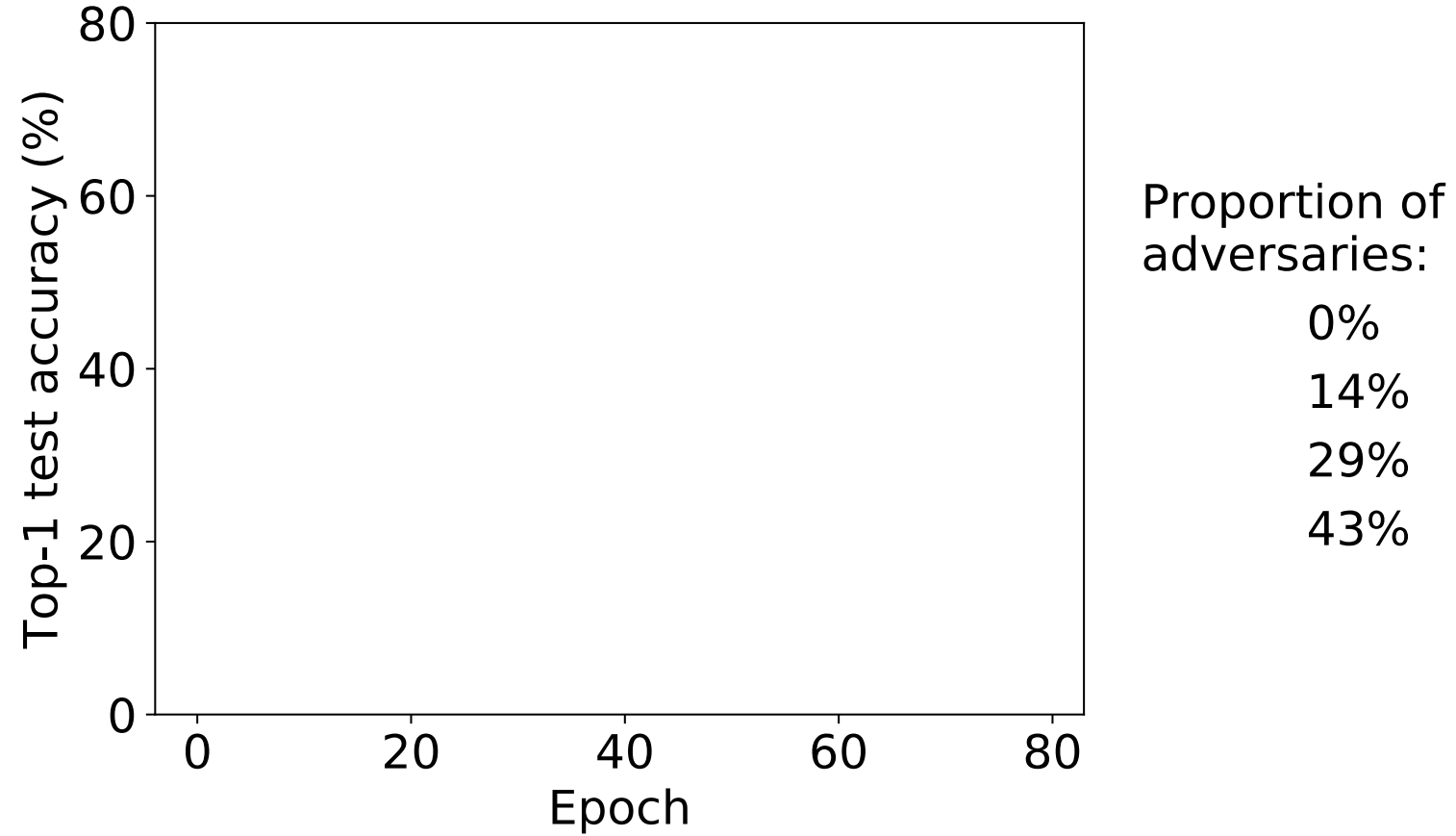


Experiment Result for ImageNet



Cut down 30% time cost with almost the same accuracy!

Experiment Result for robustness experiments



Current Status

- Submitted to ICLR 2019
- Presented at PyTorch Developer Conference 2018, San Francisco

Future Work

- Experiment on QRNN
- Integrating into PyTorch
- Improve efficiency with the help of Nvidia