BOSTON UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

Thesis

SIMULTANEOUS MULTI-PARTY LEARNING: HYPERGRAPH STOCHASTIC GRADIENT DESCENT

by

JUSTIN CHEN

B.S., Boston University, 2014

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2018

Approved by

First Reader

Andrei Lapets, PhD
Associate Professor of Computer Science

Second Reader

Kate Saenko, PhD
Assistant Professor of Computer Science

Third Reader

Adam Smith, PhD
Professor of Computer Science

To be a warrior is not a simple matter of wishing to be one. It is rather an endless struggle that will go on to the very last moment of our lives. No one is born a warrior, in exactly the same way that no one is born an average person. We make ourselves into one or the other.
Natsume Sōseki

Acknowledgments

Here go all your acknowledgments. You know, your advisor, funding agency, lab mates, etc., and of course your family.

As for me, I would like to thank Jonathan Polimeni for cleaning up old LaTeX style files and templates so that Engineering students would not have to suffer typesetting dissertations in MS Word. Also, I would like to thank IDS/ISS group (ECE) and CV/CNS lab graduates for their contributions and tweaks to this scheme over the years (after many frustrations when preparing their final document for BU library). In particular, I would like to thank Limor Martin who has helped with the transition to PDF-only dissertation format (no more printing hardcopies – hooray !!!)

The stylistic and aesthetic conventions implemented in this LaTeX thesis/dissertation format would not have been possible without the help from Brendan McDermot of Mugar library and Martha Wellman of CAS.

Finally, credit is due to Stephen Gildea for the MIT style file off which this current version is based, and Paolo Gaudiano for porting the MIT style to one compatible with BU requirements.

Justin Chen

Master's Student

CS Department

SIMULTANEOUS MULTI-PARTY LEARNING: HYPERGRAPH STOCHASTIC GRADIENT DESCENT

JUSTIN CHEN

ABSTRACT

.

Contents

List of Tables

List of Figures

List of Abbreviations

ANN	....	Artificial Neural Network
ASGD	....	Asynchronous Stochastic Gradient Descent
BP	....	Backpropagation
C-PSGD	....	Centralized Parallel Stochastic Gradient Descent
D-PSGD	....	Decentralized Parallel Stochastic Gradient Descent
DCG	....	Dynamic Computation Graph
DNN	....	Deep Neural Network
GD	....	Gradient Descent
HgSGD	....	Hypergraph Stochastic Gradient Descent
L-BFGS	....	Limited-memory Broyden-Fletcher-Goldfarb-Shanno
ML	....	Machine Learning
MPC	....	Multi-Party Computation
NEAT	....	Neuroevolution of Augmenting Topologies
NN	....	Neural Network
P2P	....	Peer-to-peer
SGD	....	Stochastic Gradient Descent
SMPL	....	Simultaneous Multi-Party Learning
SSGD	....	Synchronous Stochastic Gradient Descent

Chapter 1

Introduction

1.1 Evolutionary Algorithms

1.2 Democratizing Machine Intelligence

¡Blurb on importance of democracy ¿

Chapter 2

Background

2.1 Deep Learning

2.1.1 Supervised Learning

2.1.2 Unsupervised Learning

Generative Adversarial Networks

We digress to give a concise background on the seminal work by (Goodfellow et al., 2014) on Generative Adversarial Networks.

2.1.3 Semi-Supervised Learning

Semi-Supervised Learning with GANs

(Salimans et al., 2016)

2.2 Gradient-based Learning

2.2.1 Model Parallelism

2.2.2 Data Parallelism

2.2.3 Problems with Distributed Training

Stale Gradients

Parameter Synchronization

2.2.4 Distributed Training Optimizations

(Dean et al., 2012)

Quantization

Sparsification

(Lin et al., 2017)

2.3 Related Work

2.3.1 Centralized Parameter Server

Synchronous Data Parallelism

Asynchronous Data Parallelism

2.3.2 Decentralized Parameter Servers

Chapter 3

P2P Distributed Training

3.1 Hypergraph Stochastic Gradient Descent

Decentralized Paradigm

Cliques and Hypergraphs

Chapter 4

Experiments

4.1 Baseline

4.1.1 Wallclock Performance

4.1.2 Gradient Analysis

4.1.3 Convergence Analysis

4.2 Hypergraph Stochastic Gradient Descent

4.2.1 Wallclock Performance

Random Validation Accuracy

Best Validation Accuracy

Worst Validation Accuracy

Weighted Validation Accuracy

Weighted by Number of Peers

Weighted by Number of Cliques

4.2.2 Gradient Analysis

Random Validation Accuracy

Best Validation Accuracy

Worst Validation Accuracy

Weighted Validation Accuracy

Weighted by Number of Peers

Weighted by Number of Cliques

4.2.3 Convergence Analysis

Random Validation Accuracy

Best Validation Accuracy

Worst Validation Accuracy

Weighted Validation Accuracy

Weighted by Number of Peers

Weighted by Number of Cliques

Chapter 5

Conclusions

5.1 Summary

Time to get philosophical and wordy.

IMPORTANT: In the references at the end of thesis, all journal names must be spelled out in full, except for standard abbreviations like IEEE, ACM, SPIE,https://preview.overleaf INFOCOM, ...

5.2 Future Work

5.2.1 Information Diffusion

Contagion Theory

Percolation Theory

5.2.2 Alternative Optimizations

Second-Order Methods

(Ba et al., 2016)

Deep Neuroevolution

Parameter synchronization Initializing with the same seed and passing seeds used for pseudonumber generators to derive parameter updates instead of passing entire set of parameter matrices as with distributed gradient optimization. (Such et al., 2017)

5.2.3 Anonymous P2P Communication

5.2.4 Optimizations

Max-Flow Algorithms

Routing for Optimal Clique Formation

Appendix A

Proof of Intractability of Model Inversion Against Generative Adversarial Nets

This is the appendix.

References

Ba, J., Grosse, R., and Martens, J. (2016). Distributed second-order optimization using kronecker-factored approximations.

Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., Senior, A., Tucker, P., Yang, K., Le, Q. V., et al. (2012). Large scale distributed deep networks. In *Advances in neural information processing systems*, pages 1223–1231.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Lin, Y., Han, S., Mao, H., Wang, Y., and Dally, W. J. (2017). Deep gradient compression: Reducing the communication bandwidth for distributed training. *arXiv preprint arXiv:1712.01887*.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pages 2234–2242.

Such, F. P., Madhavan, V., Conti, E., Lehman, J., Stanley, K. O., and Clune, J. (2017). Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. *arXiv preprint arXiv:1712.06567*.

CURRICULUM VITAE

Justin Chen

Basically, this needs to be worked out by each individual, however the same format, margins, typeface, and type size must be used as in the rest of the dissertation.