Iman Tabrizian

(647) 551-9065 | iman.tabrizian@mail.utoronto.ca | linkedin.com/in/tabrizian | github.com/tabrizian

Education

University of Toronto — Toronto, Canada
Master of Applied Science in Electrical and Computer Engineering — *January 2019 – December 2020*

- **GPA: 3.94 / 4.0**

Amirkabir University of Technology — Tehran, Iran
Bachelor of Science in Computer Engineering — *September 2014 – August 2018*

- **GPA: 17.97 / 20**

Experience

Deep Learning System Software Intern — July 2020 – Present
NVIDIA — *Toronto, Canada*

- Developed native Python backend to allow running Python models in Triton Inference Server
- Contributed to the Model Analyzer project which is used for benchmarking of deep learning inference models before being served using Triton

Research Assistant — January 2020 – Present
Vector Institute — *Toronto, Ontario*

- Research on the quantization algorithms for data-parallel SGD
- Designing robust gradient aggregation algorithms for byzantine tolerant federated learning

Publications

Adaptive Gradient Quantization for Data-Parallel SGD — To appear in NeurIPS 2020
Fartash Faghri, **Iman Tabrizian***, *Ilia Markov, Dan Alistarh, Daniel M. Roy, Ali Ramezani-Kebrya*

Adaptive Quantization for Data-Parallel SGD — ML4HPC Workshop @ ASPLOS 2020
Ali Ramezani-Kebrya, ***Iman Tabrizian****, Fartash Faghri, Daniel M. Roy*

Representation of Federated Learning via Worst-Case Robust Optimization Theory
Saeedeh Parsaeefard, ***Iman Tabrizian****, Alberto Leon-Garcia* — NewInML Workshop @ NeurIPS 2019

Projects

Parallel Radix Sort on GPU | *CUDA*

- Implemented radix sort using CUDA on GPU as a part of Udacity CS344 course.
- Used Hillis-Steele algorithm for the parallel scan section of the sorting.

Profiling of Co-located Deep Learning Training Jobs | *Nsight Compute, Nsight Systems, Python*

- Developed a benchmarking framework to study the interference patterns between the jobs placed on the same GPU
- Developed an algorithm for the scheduling of DL jobs in GPU clusters using co-location

Honors and Awards

ACM ASPLOS '20 Travel Grant Recipient valued $1600 USD
Electrical and Computer Engineering Edward S. Rogers Sr. Admission Scholarships

Technical Skills

Languages: Python, C/C++, CUDA, Java, JavaScript, HTML/CSS
Frameworks: PyTorch, Kubernetes
Developer Tools: Git, Docker, Google Cloud Platform, VS Code
Libraries: pandas, NumPy, Matplotlib, Scipy

*Equal Contribution