
SELF EVALUATION

- **Strong programming skills** (designed and implemented multiple software used in production, 10+ years of programming experiences, and familiar with 10+ programming languages and many technologies).
- **Strong theory background** (8 theoretical papers in top AI conferences such as NIPS and ICML).
- **Fast learner** (learn new technologies very fast in new groups, and develop new algorithms or improve existing systems quickly).

EDUCATION

Ph.D. Computer Science **University of Rochester** **2015-present**

- Advisor: Ji Liu.
- GPA: 4.00/4.00, with exceptional performance (A+) in Advanced Algorithms, Dynamic Lang. & Soft. Dev., Operating Systems, and Computational Complexity.

B.S. Physics **University of Science and Technology of China** **2011-2015**

- Yan Jici Talent Program in Physics.
- GPA: 4.08/4.30. Ranking: 1/63 (Class 003) and 3/264 (School of the Gifted Young).

RESEARCH INTERESTS

- **Optimization for machine learning**
 - Large scale distributed machine learning / deep learning.
 - Asynchronous parallel algorithms.
 - Decentralized parallel algorithms.
 - Communication efficient distributed algorithms.
 - Compositional optimization.
- **Mobile computing**
 - Model compression for efficient computing on mobile devices.
 - Decentralized federate learning.
- **Reinforcement learning**
 - Game AI.
 - Bandit.
- **Natural language processing**
 - Variational autoencoders.

EXPERIENCE

Research Intern **Kwai Inc. Y-Lab** **Oct 2018-Feb 2019**

- Mentor: Wen Zheng.
- **GPU based large scale deep learning training system for sparse features**
 - Design and implement new GPU based large scale distributed deep learning system and infrastructure for Kwai Inc's advertisement group. The system features new training algorithms to handle huge amount of sparse features and neural networks with large embedding layers.
 - The system is used to train on datasets containing billions of samples.
 - The system developed by me improves the efficiency of the existing system by above 10x faster, and uses much less resources. The system replaced the existing offline training platform, and will replace the online training platform and other training systems in the future.
- **Computer vision machine learning task training process acceleration**
 - For the Face Landmarks project frequently used in the product, redesign the I/O implementation with a data loader buffering technology, and implement multi-machine distributed training.

- The accelerated version designed by me is 12x faster than the original implementation without sacrificing the quality of the trained model.
- **Model compression for efficient inference on mobile devices**
 - Design new algorithm based on network slimming to adaptively generate smaller neural networks with sacrificed model quality as small as possible.
 - The algorithm designed by me saves 28% of the computation cost on mobile devices for the Face Landmarks model used in production. The algorithm is used by many other projects now.

Research Summer Intern IBM Thomas J. Watson Research Center May 2018-Aug 2018

- Mentor: Luis Lastras.
- **Improving variational autoencoders**
 - Under the supervision of Dr. Luis Lastras, for variational autoencoders (VAE), without training on the conventional ELBO loss function, we designed an algorithm to train VAE by training directly on the original loss function using importance sampling, which can improve the test performance of the trained model. The method can be used in many tasks such as natural language processing.

Research Intern Tencent AI Lab (Seattle) Aug 2017-Dec 2017

- Advisor: Ji Liu.
- **Game AI for King of Glory** (The King of Glory is the most popular MOBA game in China, with 100 million daily active users.)
 - Design and implement an efficient learning platform called KOG-SGAME which supports both supervised learning and reinforcement learning. The platform is later used in the whole Seattle Game AI team to develop algorithms.
 - Co-design the game arena for different King of Glory game AIs to compete.
 - On the platform designed by me I trained the first AI defeating the internal rule based AI.
 - Before leaving for 1 vs 1 games our AI is about top 10% human performance.

Research Summer Intern IBM Thomas J. Watson Research Center May 2017-Aug 2017

- Mentor: Wei Zhang.
- **Large scale decentralized asynchronous parallel training for deep learning**
 - Design a decentralized asynchronous training algorithm, which is very robust in cloud environments with hundreds of GPUs. Prove the convergence and linear speedup property of the new algorithm.
 - The new algorithm is often more than 10x faster than the state of the art distributed training algorithm AllReduce SGD in a cloud environment. The results are published in ICML 2018.

Visiting Student Researcher Stanford University Jul 2014-Aug 2014

- Advisor: Zhi-Xun Shen.
- **Local Electron Spin Resonance (local ESR) apparatus design and building**
 - Design and build a new type of physics apparatus which utilizes the local electron spin resonance effect to scan the local conductivity and permeability of different locations on a surface.

Volunteer Research Assistant Hefei National Laboratory for Physical Science at Microscale 2013-2015

- Advisor: Tao Wu.
- **Data analysis system for NMR applied on high temperature superconductors**
 - Design and implement the data analysis system for Nuclear Magnetic Resonance (NMR) to discover new phenomenon in high temperature superconductors. The system is now used in the lab for collecting and analyzing all data acquired in experiments.

AWARDS & HONORS

- **2018**
 - Intel Student Ambassador
 - ICML Travel Award
- **2017**
 - NIPS Travel Award
- **2015**
 - NIPS Travel Award
 - Outstanding Undergraduate of University of Science and Technology of China
 - Honorary Rank of Academic Achievement of the Grade 2011 Undergraduates
 - Honor of Graduating from Yan Jici Talent Program in Physics
- **2014**
 - National Scholarship
- **2013**
 - National Scholarship
 - 1st Prize of 5th China Undergraduate Mathematical Contest (Anhui division)
 - Grand Prize of USTC Research Oriented Physics Experiment Competition
 - 2nd Prize of USTC Undergraduate Mathematical Contest
- **2012**
 - National Encouragement Scholarship
 - 1st Prize of 4th China Undergraduate Mathematical Contest (Anhui division)
 - 2nd Prize of “USTC Star” Forum LOGO Design Competition
- **2011**
 - USTC Scholarship for Outstanding Freshman

PUBLICATIONS

(* means equal contribution)

Hanlin Tang, **Xiangru Lian**, Chen Yu, Tong Zhang, and Ji Liu. “DoubleSqueeze: Parallel Stochastic Gradient Descent with Double-pass Error-Compensated Compression”. In: *International Conference on Machine Learning*. 2019

Xiangru Lian and Ji Liu. “Revisit Batch Normalization: New Understanding and Refinement via Composition Optimization”. In: *Proceedings of Machine Learning Research*. Vol. 89. Proceedings of Machine Learning Research. PMLR, 2019, pp. 3254–3263

Hanlin Tang, **Xiangru Lian**, Ming Yan, Ce Zhang, and Ji Liu. “ D^2 : Decentralized Training over Decentralized Data”. In: *International Conference on Machine Learning*. 2018

Xiangru Lian*, Wei Zhang*, Ce Zhang, and Ji Liu. “Asynchronous Decentralized Parallel Stochastic Gradient Descent”. In: *International Conference on Machine Learning*. 2018 (**Long talk**)

Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, and Ji Liu. “Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent”. In: *Advances in Neural Information Processing Systems*. 2017 (**Oral paper, rate 1%**)

Xiangru Lian, Mengdi Wang, and Ji Liu. “Finite-sum Composition Optimization via Variance Reduced Gradient Descent”. In: *International Conference on Artificial Intelligence and Statistics*. 2017

Yang You*, **Xiangru Lian***, Ji Liu, Hsiang-Fu Yu, Inderjit Dhillon, James Demmel, and Cho-Jui Hsieh. “Asynchronous Parallel Greedy Coordinate Descent”. In: *Advances in Neural Information Processing Systems*. 2016

Xiangru Lian, Huan Zhang, Cho-Jui Hsieh, Yijun Huang, and Ji Liu. “A Comprehensive Linear Speedup Analysis for Asynchronous Stochastic Parallel Optimization from Zeroth-Order to First-Order”. In: *Advances in Neural Information Processing Systems*. 2016

Wei Zhang, Suyog Gupta, **Xiangru Lian**, and Ji Liu. “Staleness-aware Async-SGD for Distributed Deep Learning”. In: *International Joint Conference on Artificial Intelligence*. 2016

Xiangru Lian, Yijun Huang, Yuncheng Li, and Ji Liu. “Asynchronous parallel stochastic gradient for non-convex optimization”. In: *Advances in Neural Information Processing Systems*. 2015, pp. 2719–2727 (**Spotlight paper, rate 4%**)

Yongping Wu, Dan Zhao, **Xiangru Lian**, Xiufang Lu, Naizhou Wang, Xigang Luo, Xianhui Chen, and Tao Wu. “NMR evidence for field-induced ferromagnetism in $(\text{Li}_{0.8}\text{Fe}_{0.2})\text{OHFeSe}$ superconductor”. In: *Physical Review B* 91.12 (2015), p. 125107

PROFESSIONAL SERVICES

Reviewer/Program Committee

- **2019:** Neural Information Processing Systems (NIPS), International Conference on Machine Learning (ICML), IEEE Transactions on Neural Networks and Learning Systems (TNNLS), Machine Learning, Neural Networks, Scalable Deep Learning over Parallel and Distributed Infrastructure (ScaDL), Data Mining and Knowledge Discovery (DAMI), Pattern Recognition, Optimization Methods and Software, Journal of Parallel and Distributed Computing (JPDC).
- **2018:** IEEE Transactions on Network Science and Engineering (TNSE), Neurocomputing, IEEE Transactions on Knowledge and Data Engineering (TKDE), AAAI, European Journal of Operational Research (EJOR).
- **2017:** BIT Numerical Mathematics, IEEE Transactions on Knowledge and Data Engineering (TKDE), IEEE Transactions on Signal Processing (TSP).
- **2016:** Neural Information Processing Systems (NIPS).

Invited Talks

- **2019:**
 - Kwai Inc.: *On model compression and distributed training of the Face Landmarks model*.
- **2018:**
 - INFORMS 2018 International Meeting: *Asynchronous Parallel Empirical Variance Guided Algorithms for the Thresholding Bandit Problem*.
 - ICML 2018: *Asynchronous Decentralized Parallel Stochastic Gradient Descent*.
 - ICML 2018: *D2: Decentralized Training over Decentralized Data*.
- **2017:**
 - IBM T.J. Watson Research Center, Optimization for AI: *Accelerating Deep Learning via Decentralized Parallel Optimization*
- **2015:**
 - NIPS 3min spotlight talk: *Asynchronous Parallel Stochastic Gradient for Nonconvex Optimization*

Teaching Assistant

- **2018:** CSC446 – Machine Learning.
- **2017:** CSC484 – Advanced Algorithms.
- **2016:** CSC282 – Design and Analysis of Efficient Algorithms.

TECHNOLOGIES

I have learned many programming languages, for example F#, Clojure, ClojureScript, Java, TypeScript, Hy, Python, C, Racket, Emacs Lisp, Groovy, Haskell, SQL, AQL, Julia, Matlab, Mathematica, C++, C#, Lua, Ruby, \LaTeX , ConTeXt, Bash, Zsh, fish, LabVIEW, HTML, CSS, Sass, and PHP. I learned them mostly from using them in practice when hacking or developing different software.

For technologies, I am familiar with but not limited to

Deep learning frameworks PyTorch (together with MPI, NCCL, 0MQ etc. for distributed communication).

Big data frameworks I am familiar with Hadoop (HDFS, YARN etc.), Hive, and Spark. I mainly learned them in Kwai Inc. to process big data for distributed neural network training.

Containers Docker, and systemd-nspawn.

Linux distributions Arch Linux, Debian/Ubuntu, Gentoo, and NixOS.

Web I am familiar with Clojure and F# web stack, from frontend (with ReactJS) to backend.