Cong Xie

Staff Research Scientist – Bytedance Inc.

+1 2178196957 • xcgoner1108@gmail.com • congxie1108.github.io
Google Scholar Link

Research Interests

Large-scale and Distributed Machine Learning, System-algorithm Co-design for Machine Learning, Efficient Machine Learning, Non-convex Optimization

Work Experience

Bytedance Inc. Research Scientist 2021-

- Accelerate distributed ML systems.
- Train foundational large language models.
- Establish solutions with system-algorithm co-designs for faster and more efficient machine learning.
- Build ML systems with both promising performance in practice, and rigorous guarantees in theory.

Education

Ph.D. of Computer Science, University of Illinois at Urbana Champaign 2016–2021
GPA: 3.96/4.0,
Advisor: Indranil Gupta and Oluwasanmi Koyejo

M.S. of Computer Science, Shanghai Jiao Tong University 2013–2016
GPA: 2.59/3.3,
Advisor: Wu-Jun Li and Zhihua Zhang

B.S. of Computer Science, Shanghai Jiao Tong University 2009–2013
GPA: 3.52/4.3, 85.38/100

Pre-graduation Industrial Experience

Amazon Inc. Applied Scientist Intern 2018, 2019, 2020

- Accelerate distributed SGD with message compression and infrequent synchronization.
- Accelerate distributed training of BERT language model via local SGD.
- Accelerate tensor reduction/averaging on multiple GPUs.

Honors & Awards

J.P. Morgan 2020 AI Research PhD Fellowship Awards *Website link* 2020
Chinese National Scholarship *Top 2%* 2015
SJTU Academic Excellence Scholarship Class-B *Top 10%* 2012
SJTU Academic Excellence Scholarship Class-C *2 Times, Top 20%* 2009–2011

Publication

1. **SDP4Bit: Toward 4-bit Communication Quantization in Sharded Data Parallelism for LLM Training**
   Jinda Jia, **Cong Xie**, Hanlin Lu, Daoce Wang, Hao Feng, Chengming Zhang, Baixi Sun, Haibin Lin, Zhi Zhang, Xin Liu, Dingwen Tao.
   Advances in Neural Information Processing Systems (*NeurIPS*), *2024.* (*Acceptance rate: 25%*).
2. **MegaScale: Scaling Large Language Model Training to More Than 10,000 GPUs**
   Ziheng Jiang, Haibin Lin, Yinmin Zhong, Qi Huang, Yangrui Chen, Zhi Zhang, Yanghua Peng, Xiang Li, **Cong Xie**, Shibiao Nong, Yulu Jia, Sun He, Hongmin Chen, Zhihao Bai, Qi Hou, Shipeng Yan, Ding Zhou, Yiyao Sheng, Zhuo Jiang, Haohan Xu, Haoran Wei, Zhang Zhang, Pengfei Nie, Leqi Zou, Sida Zhao, Liang Xiang, Zherui Liu, Zhe Li, Xiaoying Jia, Jianxi Ye, Xin Jin, Xin Liu.
   21st USENIX Symposium on Networked Systems Design and Implementation (*NSDI*), *2024.* (*Acceptance rate: 19%*).
3. **SAPipe: Staleness-Aware Pipeline for Data Parallel DNN Training**
   Yangrui Chen, **Cong Xie**, Meng Ma, Juncheng Gu, Yanghua Peng, Haibin Lin, Chuan Wu, and Yibo Zhu.
   Advances in Neural Information Processing Systems (*NeurIPS*), *2022.* (*Acceptance rate: 25%*).
4. **ZenoPS: A Distributed Learning System Integrating Communication Efficiency and Security**
   Cong Xie, Sanmi Koyejo, Indranil Gupta.
   MDPI Algorithms 15.7 (*2022*).
5. **CSER: Communication-efficient SGD with Error Reset**
   Cong Xie, Shuai Zheng, Sanmi Koyejo, Indranil Gupta, Mu Li, and Haibin Lin.
   Advances in Neural Information Processing Systems (*NeurIPS*), *2020.* (*Acceptance rate: 20%*).
6. **Zeno++: Robust Fully Asynchronous SGD**
   Cong Xie, Sanmi Koyejo, and Indranil Gupta.
   International Conference on Machine Learning (*ICML*), *2020.* (*Acceptance rate: 22%*)
7. **Local AdaAlter: Communication-Efficient Stochastic Gradient Descent with Adaptive Learning Rates**
   Cong Xie, Sanmi Koyejo, Indranil Gupta, and Haibin Lin.
   NeurIPS workshop on Optimization for Machine Learning (*OPT2020*). *Preprint: https://arxiv.org/abs/1911.09030*
8. **Asynchronous Federated Optimization**
   Cong Xie, Sanmi Koyejo, and Indranil Gupta.
   NeurIPS workshop on Optimization for Machine Learning (*OPT2020*). *Preprint: https://arxiv.org/abs/1903.03934*
9. **Baechi: Fast Device Placement of Machine Learning Graphs**
   Beomyeol Jeon, Linda Cai, Pallavi Srivastava, Jintao Jiang, Xiaolan Ke, Yitao Meng, **Cong Xie**, Indranil Gupta.
   Proc. ACM Symposium on Cloud Computing (*ACM SoCC*), *2020.*
10. **SLSGD: Secure and Efficient Distributed On-device Machine Learning**
    Cong Xie, Sanmi Koyejo, and Indranil Gupta.
    European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (*ECML PKDD*), *2019.* (*Acceptance rate: 18%*)
11. **Fall of Empires: Breaking Byzantine-tolerant SGD by Inner Product Manipulation**
    Cong Xie, Sanmi Koyejo, and Indranil Gupta.
    Uncertainty in Artificial Intelligence (*UAI*), *2019.* (*Acceptance rate: 26%*)
12. **Zeno: Distributed Stochastic Gradient Descent with Suspicion-based Fault-tolerance**
    Cong Xie, Sanmi Koyejo, and Indranil Gupta.

International Conference on Machine Learning (*ICML*), *2019.* (*Acceptance rate: 23%*)

13. **Distributed Power-law Graph Computing: Theoretical and Empirical Analysis**
    Cong Xie, Ling Yan, Wu-Jun Li, and Zhihua Zhang.
    Advances in Neural Information Processing Systems (*NeurIPS*), *2014.*
14. **A Scalable and Extensible Framework for Superposition-Structured Models**
    Shenjian Zhao, **Cong Xie**, and Zhihua Zhang.
    The Thirtieth Conference on Artificial Intelligence (*AAAI-16*), *2015.*
15. **Wishart Mechanism for Differentially Private Principle Components Analysis**
    Wuxuan Jiang, **Cong Xie**, and Zhihua Zhang.
    The Thirtieth Conference on Artificial Intelligence (*AAAI-16*), *2015.*
16. **Feature Extraction and Ensemble Decision Tree Classifier in Plant Failure Detection**
    Cong Xie, Donglin Yang, Yixiang Huang, and Donglai Sun.
    Annual Conference of the Prognostics and Health Management Society (*IEEE PHM2015 Data Challenge Winner Paper*), *2015.*

Preprints

1. **Compressed Communication for Distributed Training: Adaptive Methods and System**
   Yuchen Zhong, **Cong Xie**, Shuai Zheng, Haibin Lin.
   https://arxiv.org/abs/2105.07829
2. **Phocas: dimensional Byzantine-resilient stochastic gradient descent**
   Cong Xie, Sanmi Koyejo, and Indranil Gupta.
   https://arxiv.org/abs/1805.09682
3. **Distributed Power-law Graph Computing: Theoretical and Empirical Analysis**
   Cong Xie, Ling Yan, Wu-Jun Li, and Zhihua Zhang.
   (*Long version of a conference paper*)
4. **S-PowerGraph: Streaming Graph Partitioning for Natural Graphs by Vertex-Cut**
   Cong Xie, Wu-Jun Li, and Zhihua Zhang.
   http://arxiv.org/abs/1511.02586
5. **A New Relaxation Approach to Normalized Hypergraph Cut**
   Cong Xie, Wu-Jun Li, and Zhihua Zhang.
   http://arxiv.org/abs/1511.02595

Academic Service

Journal Reviewer:

- Journal of Machine Learning Research (JMLR) 2015
- ACM Transactions on Autonomous and Adaptive Systems (TAAS) 2016
- IEEE Transactions on Signal Processing 2019, 2020
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS) 2020
- Journal of Computer Science and Technology (JCST) 2021
- Transactions on Signal and Information Processing over Networks (SIPN) 2021
- Transactions on Network Science and Engineering (TNSE) 2020,2021
- Frontiers in Artificial Intelligence 2021
- Journal of Systems Architecture 2021
- Transactions on Knowledge Discovery from Data (TKDD) 2022

- Distributed Computing 2021, 2022
- IEEE Transactions on Computers (TC) 2022
- IEEE Access 2024
- TMLR 2024

Conference Reviewer:

- MLSys, ICML, NeurIPS, AISTATS, ICLR, UAI, DISC, AAAI, IJCAI 2018-

Teaching Experience

Teaching Assistant, *CS446, Machine Learning* 2018
Department of Computer Sciences, University of Illinois at Urbana-Champaign

Teaching Assistant, *CS435, Introduction to Machine Learning* 2014
Department of Computer Sciences, Shanghai Jiao Tong University, China