TAO YU

Gates Hall 325, Cornell University, Ithaca, NY
Homepage: http://www.cs.cornell.edu/∼tyu/
Email: tyu@cs.cornell.edu

EDUCATION

Cornell University, Ithaca, NY, United States *Sep. 2019 - June 2024 (anticipated)*
Ph.D. in Computer Science Dept. of Computer Science
Shanghai Jiao Tong University, Shanghai, China *Sep. 2015 - June 2019*
B.S. in Mathematics and Applied Mathematics (Honors) ZhiYuan college

RESEARCH INTEREST

I am intrigued by the prospect of integrating data geometry into machine learning and NLP, as it helps capture diverse properties exhibited by data across various tasks. I'm also dedicated to developing efficient algorithmic and library solutions to ensure robust numerical computations of low-precision ML models. Additionally, my interests extend to LLMs, machine learning privacy and robustness, along with a curiosity for emerging cognitive learning paradigms such as vector symbolic architectures.

PUBLICATIONS

"Collage: Light-Weight Low-Precision Strategy for LLM Training"
Tao Yu, Gaurav Gupta, Karthick Gopalswamy, Amith R Mamidala, Hao Zhou, Jeffrey Huynh, Youngsuk Park, Ron Diamant, Anoop Deoras, Luke Huan (Under Review)

"Momentum Approximation in Asynchronous Private Federated Learning"
Tao Yu, Congzheng Song, Jianyu Wang, Mona Chitnis (Under Review)

"Shadow Cones: Unveiling Partial Orders in Hyperbolic Space"
Tao Yu, Toni J.B. Liu, Albert Tseng, Christopher De Sa
— In 12th International Conference on Learning Representations (ICLR 2024)

"Coneheads: Hierarchy Aware Attention"
Albert Tseng, **Tao Yu**, Toni J.B. Liu, Christopher De Sa
— In 37th Conference on Neural Information Processing Systems (NeurIPS 2023).

"HyLa: Hyperbolic Laplacian Features For Graph Learning"
Tao Yu, Christopher De Sa
— In 11th International Conference on Learning Representations (ICLR 2023)

"Understanding Hyperdimensional Computing for Parallel Single-Pass Learning"
Tao Yu, Yichi Zhang, Zhiru Zhang, Christopher De Sa
— In 36th Conference on Neural Information Processing Systems (NeurIPS 2022)

"MCTensor: A High-Precision Deep Learning Library with Multi-Component Floating-Point"
Tao Yu, Wentao Guo, Jianan Canal Li, Tiancheng Yuan, Christopher De Sa
— In 39th International Conference on Machine Learning (ICML 2022), Workshop on Hardware Aware Efficient Training (HAET 2022)

"Salvaging Federated Learning by Local Adaptation"
Tao Yu, Eugene Bagdasaryan, Vitaly Shmatikov

"Representing Hyperbolic Space Accurately using Multi-Component Floats"
Tao Yu, Christopher De Sa
— In 35th Conference on Neural Information Processing Systems (NeurIPS 2021)

"Numerically Accurate Hyperbolic Embeddings Using Tiling-Based Models"
Tao Yu, Christopher De Sa.
— In 33rd Conference on Neural Information Processing Systems (NeurIPS 2019 **Spotlight**)

"A New Defense Against Adversarial Images: Turning a Weakness into a Strength"
Tao Yu, Shengyuan Hu*, Chuan Guo, Weilun Chao, Kilian Q. Weinberger
— In 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)

"Simplifying Graph Convolutional Networks"
Felix Wu, Tianyi Zhang, Amauri Holanda de Souza Jr., Christopher Fifty, v, Kilian Q. Weinberger
— In 36th International Conference on Machine Learning (ICML 2019)

"Curvature-based Comparison of Two Neural Networks"
Tao Yu, Huan long, John Hopcroft
— In 24th International Conference on Pattern Recognition (ICPR 2018)

EMPLOYMENT

Applied Research Intern, Amazon *Aug. 2023 - March 2024*
Manager: Luke Huan; Mentor: Gaurav Gupta AWS AI
— Studied the impact of precision strategies on LLM training, by making the following contributions:
1) Designed a metric to track lost information during training and differentiate precision strategies
2) Proposed precision strategies with best trade-off between memory (23% less), throughput (3.7× speed up) and (similar/better) accuracy

Research Intern, Apple *2020, 2021, 2022, 2023 summer*
Manager: Ulfar Erlingsson, Vojta Jina, Mona Chitnis PriML Team
Mentor: Martin Pelikan, Congzheng Song MLPT
— Developed solutions for asynchronous private federated learning (FL) to improve the system latency (1.15–4× speed up) and performance of existing adaptive optimizers, with a minor communication and storage cost
— Proposed personalized FL with cohort adaptation for language models with better performances
— Examined the privacy vulnerabilities in FL, evaluated and attacked different privacy mechanisms (e.g. DP, SeparatedDP) with practical inference and reconstruction attacks
— Constructed data poisoning attacks to theoretically measure lower bounds of privacy leakages in FL

Research Intern, Cornell University *July. 2018 - Dec. 2018*
Supervisor: Kilian Q. Weinberger, Christopher De Sa Dept. of Computer Science
— Proposed algorithms to successfully detect white-box adversarial examples efficiently and accurately based on boundary information
— Simplified graph convolution networks (GCN) to linear models, which significantly speed up training with comparable performances to GCN
— Constructed accurate representations of hyperbolic space, which provably represent points with bounded error for the first time and outperforms state-of-the-art representations

Research Intern, MSAR *March. 2019*
Supervisor: Jifeng Dai Visual Computing
— Studied and proposed efficient spatial attention mechanisms in graph attention networks

PROFESSIONAL ACTICITY

Talks:

- **NeurIPS 2019**, "Numerically Accurate Hyperbolic Embeddings Using Tiling-Based Models"
- **VSAONLINE 2022**, "Understanding hyperdimensional computing for parallel single-pass learning." Invited talk in Vector Symbolic Architectures and Hyperdimensional Computing Workshop
- **PPML 2024**, "Efficient Asynchronous Private Federated Learning." Invited talk in Workshop on Privacy Preserving Machine Learning 2024

PC/Reviewer: NeurIPS, ICML, AISTATS, ICLR, KDD, SDM
Teaching: TA for CS1110, Intro to Computing with Python, Aug. 2019 - May. 2020