Yihua Zhang

Ph.D. Student in Computer Science

(+1) 517-980-3880 zhan1908@msu.edu www.yihua-zhang.com NormalUhr Yihua Zhang

Personal Information

I am a first-year Ph.D. student in computer science at Michigan State University, where I am advised by **Dr. Sijia Liu**. I am interested in the optimization foundation of **trustworthy and scalable machine learning**, including the optimization theories to improve the robustness, explainability, fairness, and scalability of current machine learning algorithms.

I am actively looking for research intern opportunities.

Education

Doctor of Computer Science 01 2022 — Present
Michigan State University, East Lansing, USA
Advisor: Dr. Sijia Liu

Bachelor of Engineering in Automation and Mechanical Engineering 09 2015 — 06 2019
Huazhong University of Science and Technology

Publications

Pre-print Articles

[1] **Y. Zhang***, Y. Yao*, P. Ram, P. Zhao, T. Chen, M. Hong, Y. Wang, S. Liu, "Advancing Model Pruning via Bi-level Optimization", (*Under Review*), PDF.

[2] G. Zhang*, **Y. Zhang***, Z. Zhang, W. Fan, Q. Li, S. Liu, S. Chang "Fairness Reprogramming", (*Under Review*), PDF.

[3] B. Hou, J. Jia, **Y. Zhang**, G. Zhang, Y. Zhang, S. Liu, S. Chang "TextGrad: Advancing Robustness Evaluation in NLP by Gradient-Driven Optimization", (*Under Review*), PDF.

[4] P. Khanduri, I. Tsaknakis, **Y. Zhang**, J. Liu, S. Liu, J. Zhang, M. Hong "Linearly Constrained Bilevel Optimization: A Smoothed Implicit Gradient Approach", (*Under Review*), PDF.

[5] H. Li, S. Zhang, M. Wang, **Y. Zhang**, P. Chen, S. Liu "Theoretical Characterization of Neural Network Generalization with Group Imbalance", (*Under Review*), PDF.

Conference Papers

[6] G. Zhang*, S. Lu*, **Y. Zhang**, X. Chen, P. Chen, Q. Fan, L. Martie, M. Hong , S. Liu, "Distributed Adversarial Training to Robustify Deep Neural Networks at Scale", 38th Conference on Uncertainty in Artificial Intelligence (*UAI'22 - Oral, acceptance rate 5%*), PDF, code.

[7] **Y. Zhang***, G. Zhang*, P. Khanduri, M. Hong, S. Chang, S. Liu, "Fast-BAT: Revisiting and Advancing Fast Adversarial Training through the Lens of Bi-level Optimization", 39th International Conference on Machine Learning (*ICML'22*), PDF, code.

[8] T. Chen*, Z. Zhang*, **Y. Zhang***, S. Chang , S. Liu , Z. Wang "Quarantine: Sparsity Can Uncover the Trojan Attack Trigger for Free", Computer Vision and Pattern Recognition Conference 2022 (*CVPR'22*), PDF, code.

Research of Interest

Bilevel Optimization in Deep Learning 02 2019 - Present

Bilevel optimization (BLO) is challenging mathematical problem, while many of the deep learning problems can be naturally formulated as a BLO and thus, the effective and efficient algorithms to solve BLO is very cherished by the research community. My research in this direction are as follows:

- Summarize different BLO formulations and corresponding theories/algorithms (current research) in deep learning. Develop a ToolBox for BLO in Python.
- Design effective and efficient BLO algorithms for specific deep learning tasks, such as pruning [1] and adversarial training [4, 7].

- Provide new perspectives to interpret the current deep learning tasks and possible existing algorithms.
- Publications: [1], [4], [7]

Trustworthy Machine Learning 02 2019 - Present

The robustness of the deep learning models have become a research hotspot in the last decade. However, to build a trustworthy machine learning algorithm requires more. My research interest in this topic is summarized as follows:

- Design effective, efficient, and scalable robust training algorithm [3, 6-7] to improve the robustness of the deep learning models against adversarial attacks.
- Improve the fairness of the model through adversarial reprogramming [2].
- Design defense strategy against backdoor attacks [8].
- Publications: [2], [3], [6], [7], [8]

Academic Activities

- **Reviewer**: CVPR'22, ICLR'22, ICML'22, NeurIPS'22
- **TPC** for KDD'22 Workshop 4th Workshop on Adversarial Learning Methods for Machine Learning and Data Mining.
- **Student Chair** for ICML'22 Workshop AdvML: New Frontiers in Adversarial Machine Learning.
- **TPC** for NeurIPS'21 Workshop NFFL: New Frontiers in Federated Learning: Privacy, Fairness, Robustness, Personalization and Data Ownership.

Skills

Programming Languages Python, C++, Java, C
Libraries PyTorch, OpenCV, NumPy, Matplotlib.

Award and Achievements

- National Scholarship, , by Ministry of Education of China (Top2%) 2017
- National Scholarship, , by Ministry of Education of China (Top2%) 2016

References

Prof. **Sijia Liu** Department of Computer Science and Engineering, Michigan State University, East Lansing, USA liusiji5@msu.edu.

Prof. **Shiyu Chang** Department of Computer Science, University of California, Santa Barbara, USA shiyu@ucsb.edu.

Prof. **Mingyi Hong** Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, USA mhong@umn.edu.

Last updated: June 13, 2022.