Zhanyu Wang

Address: 2400 W El Camino Apt 502, Mountain View, CA, 94040 **Phone**: (765) 409-6928
Email: wzy1993xt@gmail.com **Website**: https://www.stat.purdue.edu/~wang4094/

Education

Purdue University, West Lafayette, IN, U.S. 08/2018 - 12/2023
- Ph.D. in Statistics, Department of Statistics. Advisors: Dr. Jordan Awan and Dr. Guang Cheng.
- Research Area: Differential Privacy. Optimization in Deep Learning. Meta-learning.

Chinese Academy of Sciences, Beijing, China 09/2015 - 07/2018
- M.S. in Statistics, Academy of Mathematics and Systems Science. Advisor: Dr. Lei Li.

Peking University, Beijing, China 09/2011 - 07/2015
- B.S. in Statistics, School of Mathematics Sciences.

On-going projects

- **Wang, Z.**, Awan, J. "De-Biased Parametric Bootstrap Inference on Privatized Data." *TPDP (Theory and Practice of Differential Privacy)*, 2023.

Submitted papers

- Awan, J., **Wang, Z.** "Simulation-based Confidence Intervals and Hypothesis Tests for Privatized Data." *Journal of the American Statistical Association.* (Major Revision.)
- **Wang, Z.**, Cheng, G., Awan, J. "Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies." *Journal of Machine Learning Research.* (Submitted.)

Published papers

- Wang, C., **Wang, Z.**, Sun, W., Cheng, G. "Online Regularization for High-Dimensional Dynamic Pricing Algorithms." *Journal of the American Statistical Association.* 2023.
- **Wang, Z.**, Zhang, X., Yun, H., Teo, C., Chilimbi, T. "MICO: Selective Search with Mutual Information Co-training." *COLING*, 2022 (Oral presentation).
- **Wang, Z.**, Honorio, J. "The Sample Complexity of Meta Sparse Regression." *AISTATS*, 2021.
- Chao, S., **Wang, Z.**, Xing, Y., Cheng, G. "Directional Pruning of Deep Neural Networks." *NeurIPS*, 2020.
- Li, W., **Wang, Z.**, Zhang, Y., Cheng, G. "Variance Reduction on Adaptive Stochastic Mirror Descent." *Machine Learning*, 2022. (Conference version in *NeurIPS workshop, OPT* 2020.)
- Wang, A., **Wang, Z.**, Li, Z., Li, L. "BAUM: Improving genome assembly by adaptive unique mapping and local overlap-layout-consensus approach." *Bioinformatics*, 34.12 (2018), 2019-2028.

Talks and posters

Theory and Practice of Differential Privacy (TPDP 2023), Boston, U.S. 09/2023
- (Poster) De-Biased Parametric Bootstrap Inference on Privatized Data

Joint Statistical Meetings (JSM 2023), Toronto, Canada 08/2023
- (Talk) De-Biased Parametric Bootstrap Inference on Privatized Data

Statistics and Optimization in Data Science Workshop, Purdue University 05/2023
- (Poster) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Midwest Machine Learning Symposium (MMLS 2023), Chicago, U.S. 05/2023
- (Poster) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Eighth Bayesian, Fiducial, and Frequentist conference (BFF8), Cincinnati, U.S. 05/2023
- (Poster) Simulation-based Confidence Intervals and Hypothesis Tests for Privatized Data

Future Leaders Summit, Ann Arbor, Michigan, U.S. 04/2023
- (Talk) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Machine Learning for Business (MGMT 47500), Purdue University 02/2023
- (Talk) MICO: Selective Search with Mutual Information Co-training

Joint Statistical Meetings (JSM 2022), Washington D.C., U.S. 08/2022
- (Poster) Differentially Private Bootstrap: New Privacy Analysis and Inference Strategies

Honors

- Honorable mention of the Student Paper Competition in Joint Statistical Meetings (JSM) 2024
- Top (10%) reviewers of NeurIPS 2023
- Bilsland Dissertation Fellowship. Purdue University 2023
- Second Place Winner Poster Award. Statistics and Optimization in Data Science Workshop. 2023
- Outstanding Poster Award (5/141). Midwest Machine Learning Symposium (MMLS). 2023
- Travel Grant. Purdue Graduate Student Government. 2023
- Top (10%) reviewers of NeurIPS 2022
- Outstanding (10%) reviewers of ICML 2022
- Ross Fellowship. Purdue University 2018
- China National Scholarship (Graduate) 2017
- Silver Medal, National Final, Chinese Mathematics Olympiad. Chinese national training team for IMO. 2011

Teaching and Consulting

Purdue University, *Teaching Assistant*, West Lafayette, US 08/2021 - 05/2022

- STAT 303 Coordinator and Head TA. Teaching probability and statistics to undergraduate students.

Purdue University, *Data Science Consultant*, West Lafayette, US 01/2021 - 05/2021

- We propose a method for the defense of various adversarial attacks with meta-learning.

Purdue University, *Teaching Assistant*, West Lafayette, US 08/2019 - 12/2020

- STAT 225 Instructor. Teaching probability to undergraduate students.

Research Experiences

Differentially Private Bootstrap

- Existing work focused on differential private (DP) point estimates of a parameter, but not general-purpose methods to quantify the uncertainty of a DP procedure to perform reliable inference.
- We develop and analyze a DP bootstrap procedure that releases multiple DP bootstrap estimates and constructs confidence intervals based on the released estimates while preserving the privacy of individuals.
- Our method is evaluated and compared to existing methods through simulations and real-world experiments using 2016 Canada Census data. Our procedure is better in terms of statistical accuracy.

Sparse Neural Network Optimizer

- Existing sparsifying methods are mostly based on the intuition of pruning the least important weights in the model, which still degrades the performance of neural networks significantly and requires further fine-tuning.
- We propose a new optimizer, Generalized Regularized Dual Averaging (GRDA), as an improvement of SGD to solve the ℓ_1-norm regularization for promoting sparsity. We prove that GRDA prunes the weights within the null space of the Hessian matrix of the dense model trained by SGD and leaves the performance unaffected.
- We implement GRDA in both PyTorch and TensorFlow and use it to train ResNet50 on ImageNet using GPU and TPU. GRDA only requires resources similar to SGD and still achieves state-of-the-art of sparse models.

Dynamic Pricing

- A key challenge in dynamic pricing policies is their limited adaptability to online uncertainties in learned statistical models, which is crucial to enhancing the effectiveness and responsiveness of pricing strategies.
- We propose the Optimistic Online Regularized Maximum Likelihood Pricing (OORMLP) policy that combines market noise knowledge and consistently valid decision-making for more robust and efficient dynamic pricing.
- We thoroughly evaluate the performance of OORMLP in synthetic and real-world pricing scenarios, which demonstrates improved utilization of high-dimensional model sparsity and maintained a logarithmic regret.

Academic services

Reviewer

- Conferences: ICLR (2023, 2024), ICML (2022, 2023), NeurIPS (2021, 2022, 2023)
- Journals: Scientific reports. Statistical analysis and data mining. Journal of Data Science.

Other services

Graduate Student Organization (Statistics), *Officer*, Purdue University 2019, 2021, 2022

- Served as the Seminar Coordinator in 2021 and 2022, and the Social Chair in 2019.

Prof. Guang Cheng's group meetings, *Coordinator*, Purdue University 2021

Professional careers

Meta, *Research Scientist*, Menlo Park, US 12/2023 - Now

Meta, *Software Engineer Intern (Machine Learning Engineer track)*, Menlo Park, US 05/2022 - 08/2022

- We build a learning-based system for Ads Pacing using primal-dual algorithms. Our next-generation pacer model can adapt to market change promptly and also has strong explainability in terms of marginal cost.
- We implement the pacer using C++ and test it in the production environment of Meta; By experiments on 5% real-world traffic of the company, we have observed a significant increase (0.4%) in Ads value.

Amazon, *Applied Scientist Intern*, Seattle, US 05/2021 - 08/2021

- We propose a selective search framework based on Mutual Information Co-training (MICO) (clustering documents to groups by their similarity, and searching each query only within its most relevant groups). We reduce the search cost to 5% with achieving 99% accuracy compared to searching on all documents.
- MICO is an end-to-end learning model. Its objective function is the mutual information between the two group indices of a query and its related document, both of which are outputs of trainable neural networks.
- In my implementation, the model takes BERT representation of a sentence (query or document title) as input, and it can be efficiently trained on huge dataset (hundreds of GB) with BERT also being finetuned.
- The paper of MICO is accepted by COLING 2022 as Oral Presentation (10%).

Skills

- **Programming Languages**: experienced in Python, R, and C++, familiar with C, Perl, Shell, SQL, etc.
- **Softwares & Systems**: PyTorch, TensorFlow (GPU & TPU), Git, Docker, Linux, etc.