

Yutong Dai

☎ 217.721.9646 · ✉ yud319@lehigh.edu · 🏠 212 Orchard Dr., Whitehall, PA 18052 · 🔗 ydopt.ml

EDUCATION

Lehigh University , PA, USA <i>Ph.D.</i> in Industrial and Systems Engineering	Sept.2019 – Dec.2023
University of Illinois at Urbana-Champaign , IL, USA <i>M.S.</i> in Statistics	Sept.2017 – May.2019
Sichuan University , Chengdu, China <i>B.S.</i> in Mathematics with honors (concentration in Statistics)	Sept.2013 – Jun.2017

SELECTED PUBLICATIONS

A Variance-Reduced and Stabilized Proximal Stochastic Gradient Method with Support Identification Guarantees for Structured Optimization [\[Link\]](#) 2023

Collaborators: Guanyi Wang, Franke E. Curtis, and Daniel P. Robinson Accepted by *AISTATS'23*

- Proposed a stochastic algorithm for minimizing a class of composite functions with the structured sparse solutions by utilizing variance reduction and support identification; the novel variance reduction technique does not need either any full gradient evaluation or storages of past stochastic gradients
- Derived the complexity bound on identifying the solution support
- Conducted extensive experiments to suggest the theoretical complexity bound predicts the algorithm's performance well

Tackling Data Heterogeneity in Federated Learning with Class Prototypes [\[Link\]](#) 2023

Collaborators: Zeyuan Chen, Junnan Li, Shelby Heinecke, Lichao Sun, and Ran Xu Accepted by *AAAI'23*

- Proposed a novel method to tackle data heterogeneity with the class imbalance in personalized Federated Learning by combining the uniformity and semantics of class prototypes
- Developed a unified and extensible framework to test the state-of-the-art Federated Learning algorithms

Inexact Proximal Gradient Methods with the Certified Support Identification Property [\[Link\]](#) 2022

Collaborators: Daniel P. Robinson Submitted to *Optimization Methods and Software*

- Proposed adaptive termination conditions for inexact proximal gradient subproblems by monitoring reduction in the function value and optimality measure
- Proved the algorithm's certified ability to find a stationary point with structured sparsity patterns
- Conducted extensive experiments to validate algorithms' support identification property and its superior performance over the state-of-the-art inexact proximal gradient algorithms

A Subspace Acceleration Method for Minimization Involving a Group Sparsity-Inducing Norm [\[Link\]](#) 2022

Collaborators: Daniel P. Robinson and Frank E. Curtis Accepted by *SIAM Journal on Optimization*

- Proposed a new optimization paradigm (GroupFaRSA) for minimizing a class of composite functions with the structured sparse solutions by utilizing support identification, domain decomposition, and subspace acceleration techniques
- Proved GroupFaRSA's super-linear local convergence rate
- Conducted extensive numerical experiments on solving large scale group- ℓ_1 regularized logistic regression and linear regression problems to validate GroupFaRSA's superior performance over four state-of-the-art methods

INTERNSHIP

Salesforce Research May.2022 – Aug.2022

Position: Research Intern *Manager: Zeyuan Chen & Ran Xu* Palo Alto, CA

- Conducted research in personalized Federated Learning to tackle data heterogeneity with the class imbalance
- Developed a unified and extensible framework to test the state-of-the-art Federated Learning algorithms
- The work is accepted by AAAI 2023 conference and is filed as a non-provisional patent

Anheuser-Busch InBev Jan.2018 – May.2019

Position: Data Scientist Intern *Supervisor: Shang-Jen Yang & Hector Hernandez* Urbana, IL

- Provided analytics and benchmarks of farmer production performance for global agronomist and procurement teams to improve barley productivity
- Revised machine learning algorithms with agronomists' on field knowledge to formulate a global barley production environment model that accounts for complex weather and soil systems
- Developed predictive models to suggest optimal management packages that help growers to hit highest barley yield
- Designed Smart Barley UI/UX prototype in Rshiny to dynamically visualize analytic results, like growers' production performance and highest yield management packages, and delivered it to agronomist teams