Xuezhou Zhang

Boston University
Center for Computing & Data Sciences
665 Commonwealth Ave, Boston, MA 02215

Phone: (310) 717-9440
Email: xuezhouz@bu.edu
Homepage: https://zhangxz1123.github.io/

Employment

- **Boston University**, Boston, MA   2023 - present
  - Tenure-track Assistant Professor, Faculty of Computing & Data Sciences
- **Princeton University**, Princeton, NJ   2021 - 2023
  - Postdoctoral Research Associate, hosted by Professor Mengdi Wang.
- **Alibaba Group**, Redmond, WA   Summer 2019
  - Research Intern, hosted by Shaojian He. RL-based recommendation system.
- **Microsoft Research**, Redmond, WA   Summer 2018
  - Research Intern, hosted by Rich Caruana. Explainable Machine Learning.

Education

- Ph.D. Computer Sciences, University of Wisconsin-Madison, 2016 - 2021.
  - Thesis: Adversarial Learning in Sequential Decision Making.
  - Advisor: Professor Xiaojin (Jerry) Zhu
- M.S. Computer Sciences, University of Wisconsin-Madison, 2016 - 2018.
- B.S. Applied Mathematics, *Summa Cum Laude* (highest Latin Honor), University of California, Los Angeles, 2012 - 2016.

Publication

Refereed Publications

1. Jiacheng Guo*, Zihao Li, Huazheng Wang, Mengdi Wang, Zhuoran Yang, **Xuezhou Zhang**. Provably Efficient Representation Learning with Tractable Planning in Low-Rank POMDP. In *The 40th International Conference on Machine Learning (ICML)*. 2023.

2. Chengzhuo Ni*, Yuda Song*, **Xuezhou Zhang**, Zihan Ding, Chi Jin, Mengdi Wang. Representation Learning for Low-rank General-sum Markov Games. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

3. Yiding Chen*, **Xuezhou Zhang***, Kaiqing Zhang, Mengdi Wang, Xiaojin Zhu. Byzantine-Robust Online and Offline Distributed Reinforcement Learning. In *The 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

4. Alekh Agarwal*, Yuda Song*, Wen Sun*, Kaiwen Wang*, Mengdi Wang*, **Xuezhou Zhang***. Provable Benefits of Representational Transfer in Reinforcement Learning. In *The 36th Annual Conference on Learning Theory (COLT 2023)*, 2023.

5. Shubham Kumar Bharti*, **Xuezhou Zhang**, Adish Singla, Jerry Zhu. Provable Defense against Backdoor Policies in Reinforcement Learning. In *The Thirty-Sixth Conference on Neural Information Processing Systems(NeurIPS)*, 2022.

6. Shuoguang Yang*, **Xuezhou Zhang**, Mengdi Wang. Decentralized Gossip-Based Stochastic Bilevel Optimization over Communication Networks. In *The Thirty-Sixth Conference on Neural Information Processing Systems(NeurIPS)*, 2022.

7. Hui Yuan*, Chengzhuo Ni, Huazheng Wang, **Xuezhou Zhang**, Le Cong, Csaba Szepesvári, Mengdi Wang. Bandit Theory and Thompson Sampling-Guided Directed Evolution for Sequence Optimization. In *The Thirty-Sixth Conference on Neural Information Processing Systems(NeurIPS)*, 2022.

8. **Xuezhou Zhang***, Yuda Song, Masatoshi Uehara, Mengdi Wang, Alekh Agarwal, Wen Sun. Efficient reinforcement learning in block mdps: A model-free representation learning approach. *In The 39th International Conference on Machine Learning (ICML)*. 2022.

9. Chengzhuo Ni*, Ruiqi Zhang, Xiang Ji, **Xuezhou Zhang**, Mengdi Wang. Optimal Estimation of Policy Gradient via Double Fitted Iteration. *In The 39th International Conference on Machine Learning (ICML)*. 2022.

10. Ruiqi Zhang*, **Xuezhou Zhang**, Chengzhuo Ni, Mengdi Wang. Off-Policy Fitted Q-Evaluation with Differentiable Function Approximators: Z-Estimation and Inference Theory. *In The 39th International Conference on Machine Learning (ICML)*. 2022.

11. Masatoshi Uehara*, **Xuezhou Zhang**, Wen Sun. Representation Learning for Online and Offline RL in Low-rank MDPs. *The Tenth International Conference on Learning Representations (ICLR)*. 2022.

12. **Xuezhou Zhang***, Yiding Chen, Xiaojin Zhu, Wen Sun. Corruption-Robust Offline Reinforcement Learning. In *The 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2022.

13. **Xuezhou Zhang***, Yiding Chen, Xiaojin Zhu, Wen Sun. Robust Policy Gradient against Strong Data Corruption. *In The 38th International Conference on Machine Learning (ICML)*. 2021.

14. **Xuezhou Zhang***, Shubham Kumar Bharti, Yuzhe Ma, Adish Singla, Xiaojin Zhu. The Sample Complexity of Teaching-by-Reinforcement on Q-Learning. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

15. Huajie Shao*, Jun Wang, Haohong Lin, **Xuezhou Zhang**, Aston Zhang, Heng Ji and Tarek Abdelzaher. Controllable and Diverse Text Generation in E-commerce. In *WWW*, 2021.

16. **Xuezhou Zhang***, Yun-Shiuan Chuang*, Yuzhe Ma, Mark Ho, Joe Austerweil, Xiaojin Zhu. Using Machine Teaching to Investigate Human Assumptions when Teaching Reinforcement Learners. *In The 41st Annual Conference of the Cognitive Science Society (CogSci). 2021.*

17. Rishabh Agarwal*, Nicholas Frosst, **Xuezhou Zhang**, Rich Caruana, Geoffrey Hinton. Neural Additive Models: Interpretable Machine Learning with Neural Networks. In *The Thirty-Fifth Conference on Neural Information Processing Systems(NeurIPS)*, 2021. **Spotlight Presentation.**

18. **Xuezhou Zhang***, Yuzhe Ma, Adish Singla. Task-agnostic Exploration in Reinforcement Learning. In *The Thirty-Fourth Conference on Neural Information Processing Systems(NeurIPS)*, 2020.

19. **Xuezhou Zhang***, Yuzhe Ma, Adish Singla, and Xiaojin Zhu. Adaptive reward-poisoning attacks against reinforcement learning. *In The 37th International Conference on Machine Learning (ICML).* 2020.

20. **Xuezhou Zhang***, Xiaojin Zhu, Laurent Lessard. Online Data Poisoning Attack. *In Learning for Dynamics and Control (L4DC). 2020.* **Oral Presentation.**

21. Yuzhe Ma*, **Xuezhou Zhang**, Wen Sun, and Xiaojin Zhu. Policy poisoning in batch reinforcement learning and control. In *The Thirty-Third Conference on Neural Information Processing Systems(NeurIPS)*, 2019.

22. **Xuezhou Zhang***, Sarah Tan, Paul Koch, Urszula Chajewska, and Rich Caruana. Axiomatic Interpretability for Multiclass Additive Models. In *The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, Research Track, 2019. **Oral Presentation.**

23. Laurent Lessard*, **Xuezhou Zhang***, and Xiaojin Zhu*. An Optimal Control Approach to Sequential Machine Teaching. In *The 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

24. **Xuezhou Zhang***, Xiaojin Zhu, and Stephen J. Wright. Training Set Debugging using Trusted Items. In *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 2018. **Oral presentation.**

25. Yuzhe Ma*, Philippe Rigollet, **Xuezhou Zhang**, Robert D Nowak, and Xiaojin Zhu. Teacher Improves Learning by Selecting a Training Subset. In *The 21st International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2018.

26. Ayon Sen*, Scott Alfeld, **Xuezhou Zhang**, Ara Vartanian, Yuzhe Ma, and Xiaojin Zhu. Training set camouflage. In *Conference on Decision and Game Theory for Security (GameSec)*, 2018. **Oral presentation.**

Pre-Prints

1. Kaixuan Huang*, Yu Wu, **Xuezhou Zhang**, Shenyinying Tu, Qingyun Wu, Mengdi Wang, Huazheng Wang. Provably Efficient Reinforcement Learning for Online Adaptive Influence Maximization. *ArXiv Preprint*, 2022.

2. Amin Rakhsha*, **Xuezhou Zhang***, Xiaojin Zhu, Adish Singla. Reward Poisoning in Reinforcement Learning: Attacks Against Unknown Learners in Unknown Environments. *ArXiv Preprint*, 2021.

Teaching

- **Guest Lecturer**, Department of Computer Sciences, UW Madison

  - Lecture on *Introduction to Theoretical Reinforcement Learning*, in CS 861, 2020 Fall
  - Lecture on *Introduction to Reinforcement Learning*, in CS 760, 2019 Fall

- **Teaching Assistant**, Department of Computer Sciences, UW Madison

  - CS 761 - Advanced Machine Learning, 2017 Spring
  - Math 240 - Discrete Mathematics, 2016 Fall

- **Teaching Assistant**, Department of Mathematics, UCLA

  - Math 32A - Multivariable Calculus, 2015 Fall
  - Los Angeles Math Circle, Year 11-12 group, 2013 - 2014

Professional Service

- **Conference Co-Chair** for

  - Annual Conference on Information Sciences and Systems (CISS), 2022

- **Conference Area Chair** for

  - International Conference on Learning Representations (ICLR), 2023-2024

- **Conference Reviewer** for

  - Conference on Neural Information Processing Systems (Neurips), 2019-2023
  - International Conference on Machine Learning (ICML), 2021-2023
  - International Conference on Learning Representations (ICLR), 2021-2023
  - International Conference on Artificial Intelligence and Statistics (AISTATS), 2020-2023
  - AAAI Conference on Artificial Intelligence (AAAI), 2019-2021

- **Journal Reviewer** for

  - Journal of Machine Learning Research (JMLR)

- Machine Learning Journal (MLJ)
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- IEEE Transactions on Neural Networks and Learning Systems
- Mathematics of Operations Research (MOR)

Invited Talks

1. Provable Benefits of Representational Transfer in Reinforcement Learning.

   - [Dec. 2022] The 3rd Offline RL Workshop at Neurips 2022.

2. Representation Learning in General-Sum Markov Games.

   - [Nov. 2022] Symposium on Frontiers of Machine Learning and Artificial Intelligence, University of Southern California
   - [Nov. 2022] ECE Seminar, Northwestern University.

3. Decentralized Gossip-Based Stochastic Bilevel Optimization over Communication Networks.

   - [Nov. 2022] Multistage Optimization Theory and Methodology, INFORMS 2022.

4. Efficient Rich-observation Reinforcement Learning: A Representation Learning Approach.

   - [May 2022] RL Theory Virtual Seminars, 2022
   - [Mar. 2022] Frontiers of Reinforcement Learning Theory, CISS 2022
   - [Mar. 2022] University of California, Berkeley, 2022

Last updated: August 31, 2023