Yanchao Sun

Email : ycs@umd.edu
Website : https://ycsun2017.github.io

Education

- **University of Maryland, College Park** — Maryland, U.S.A.
  Ph.D. in Computer Science. GPA: 3.966/4. Won a Best Paper Award. — *Sep 2018 – Present*
- **Sichuan University** — Chengdu, China
  B.S. in Computer Science and Technology. GPA: 3.9/4 (95/100). Rank: 1/380 — *Sep 2014 – Jun 2018*

Research Interests

- *Reinforcement Learning:* sample efficiency, robustness, and adaptability in sequential decision making
- *Trustworthy Machine Learning:* trustworthiness and adversarial robustness of deep neural networks
- *Representation Learning:* building efficient representation from high-dimensional inputs
- *Transfer Learning:* enabling agents to learn by analogy and master various tasks

Publications

13. **Yanchao Sun**, Shuang Ma, Ratnesh Madaan, Rogerio Bonatti, Furong Huang, and Ashish Kapoor. "SMART: Self-supervised Multi-task pretrAining with contRol Transformers". ICLR 2023 (***Spotlight***).

12. **Yanchao Sun**, Ruijie Zheng, Parisa Hassanzadeh, Yongyuan Liang, Soheil Feizi, Sumitra Ganesh and Furong Huang. "Certifiably Robust Multi-Agent Reinforcement Learning against Adversarial Communication". ICLR 2023.

11. Yuancheng Xu, **Yanchao Sun**, and Furong Huang. "Exploring and Exploiting Decision Boundary Dynamics for Adversarial Robustness". ICLR 2023.

10. Yongyuan Liang*, **Yanchao Sun***, Ruijie Zheng, and Furong Huang. "Efficient Adversarial Training without Attacking: Worst-Case-Aware Robust Reinforcement Learning". (*Equal Contribution.) NeurIPS 2022.

9. Jifeng Hu, **Yanchao Sun**, Hechang Chen, Sili Huang, Haiyin Piao, Yi Chang, and Lichao Sun. "Distributional Reward Estimation for Effective Multi-agent Deep Reinforcement Learning". NeurIPS 2022.

8. Kaiwen Yang, **Yanchao Sun**, Jiahao Su, Fengxiang He, Xinmei Tian, Furong Huang, Tianyi Zhou, and Dacheng Tao. "Adversarial Auto-Augment with Label Preservation: A Representation Learning Principle Guided Approach". NeurIPS 2022 (***Spotlight***).

7. **Yanchao Sun**, Ruijie Zheng, Yongyuan Liang, and Furong Huang. "Who Is the Strongest Enemy? Towards Optimal and Efficient Evasion Attacks in Deep RL". ICLR 2022. (***Best Paper Award*** at the NeurIPS 2021 SafeRL Workshop.)

6. **Yanchao Sun**, Ruijie Zheng, Xiyao Wang, Andrew Cohen, and Furong Huang. "Transfer RL across Observation Feature Spaces via Model-Based Regularization" ICLR 2022.

5. **Yanchao Sun**, Da Huo, and Furong Huang. "Vulnerability-Aware Poisoning Mechanism for Online RL with Unknown Dynamics". ICLR 2021.

4. **Yanchao Sun**, Xiangyu Yin, and Furong Huang. "TempLe: Learning Template of Transitions for Sample Efficient Multi-task RL". AAAI 2021.

3. **Yanchao Sun** and Furong Huang. "Can Agents Learn by Analogy? An Inferable Model for PAC Reinforcement Learning". AAMAS 2020.

2. Jingling Li, **Yanchao Sun**, Jiahao Su, Taiji Suzuki and Furong Huang. "Understanding Generalization in Deep Learning via Tensor Methods". AISTATS 2020.

1. **Yanchao Sun**, Cong Qian, Ning Yang and Philip S. Yu. "Collaborative Inference of Coexisting Information Diffusions". ICDM 2017.

Research Experience

- **Research Assistant** — University of Maryland, College Park
  Advisor: Dr. Furong Huang — *Jun 2019 – Present*
  - **Adversarial Robustness of Deep Reinforcement Learning**
    established a systematical understanding of the robustness of RL agents against adversarial attacks, including both training-time attacks and test-time attacks;
    proposed effective and efficient algorithms for evaluating and improving the robustness of RL agents.
  - **Sample Efficient Multi-task Reinforcement Learning**
    proposed the first PAC-MDP method for multi-task reinforcement learning that could be applied to tasks with varying state/action space.

- **Research Intern** — Microsoft Research, Redmond
  Supervisor: Dr. Shuang Ma — *Jun 2022 – Aug 2022*
  - **Pretraining Representation for Reinforcement Learning Tasks.**
    proposed a self-supervised pretraining framework that works for various downstream control tasks, based on a transformer backbone.

- **AI Research Summer Associate** — JPMorgan Chase & Co., New York
  Supervisor: Dr. Sumitra Ganesh — *Jun 2021 – Aug 2021*
  - **Robustifying Agents in a Communicative Multi-agent System.**
    studied the emergence of adversarial communication in a multi-agent system and how to make agents robust against adversarial communication.

- **Machine Learning Research Intern** — Unity Technologies, San Francisco
  Supervisor: Dr. Andrew Cohen — *May 2020 – Aug 2020*
  - **Cross-domain Transfer RL with Model Regularizers.**
    designed an algorithm that utilizes model-based regularizers to transfer a learned policy to a new task with different observation space, contributed to the ML-Agents toolkit.

- **Research Assistant Intern** — Sichuan University, China
  Advisor: Prof. Ning Yang — *Apr 2016 – Jun 2018*
  - **Collaborative Inference of Coexisting Information Diffusions.**
    built a model that accurately recovers and predicts information diffusion trails in coexisting information diffusion networks (e.g. on social networks), using context-aware tensor decomposition.

Honors and Awards

- **Outstanding Research Assistant Award** (top 2%), University of Maryland, College Park, 2022
- **Best Paper Award** at the NeurIPS SafeRL Workshop, 2021
- Dean's Fellowship, University of Maryland, College Park, 2018
- Special Award of Wang Wen Guo Scholarship, Wuyuzhang Honors College, 2016
- Excellent Student Cadre of Sichuan University, 2016
- National Endeavor Scholarship, China, 2016
- The **1st Prize** of Blue Bridge Cup National C/C++ Programming Contest, Sichuan Province, 2016
- National Scholarship, China, 2015
- Excellent Student of Sichuan University, 2015
- The **1st Prize** of The Seventh Chinese Mathematics Competitions, Sichuan Province, 2015

Professional Services

- Co-organizer of the 1st Reincarnating RL Workshop at ICLR 2023
- Program Committee of NeurIPS 2022 Deep RL Workshop
- Reviewer of International Conference on Learning Representations (ICLR), 2021, 2022,2023
- Reviewer of Advances in Neural Information Processing Systems (NeurIPS), 2021, 2022
- Reviewer of International Conference on Machine Learning (ICML), 2020, 2021, 2022