CHENLU YE

University of Illinois Urbana-Champaign
chenluy3@illinois.edu | https://chenluye99.github.io/

RESEARCH INTERESTS

- Reinforcement learning from human feedback (RLHF) for aligning large language model;
- Reinforcement Learning Theory;
- Statistical Machine Learning.

EDUCATION

University of Illinois Urbana-Champaign *Urbana, USA*
Ph.D. student, *Computer Science* *2024.8 - present*
Advisor: Prof. Tong Zhang

The Hong Kong University of Science and Technology *Hong Kong, China*
MPhil, *Artificial Intelligence* *2021.9 - 2024.8*
Advisor: Prof. Tong Zhang and Prof. Yuan Yao

University of Science and Technology of China *Hefei, China*
Bachelor of Science, *Statistics* *2017.9 - 2021.6*

RESEARCH

Self-rewarding correction for mathematical reasoning *2024.12 - 2025.2*
University of Illinois Urbana-Champaign
Advisor: Prof. Tong Zhang from UIUC

- We post-trained a single large language model to have the capabilities of self-rewarding and self-correcting simultaneously.
- We formulated the problem as multi-turn mathematical reasoning and combined supervised fine-tuning (SFT) and direct preference learning (DPO) with the self-generated dataset.
- Our approach outperformed intrinsic self-correction and achieved performance comparable to systems with external reward models.

Sharper Rate and Logarithmic Regret Bound for KL-regularized RL *2024.08 - 2025.2*
University of Illinois Urbana-Champaign
Advisor: Prof. Tong Zhang from UIUC, Prof. Quanquan Gu from UCLA

- We studies the role of KL-regularization and data coverage for KL-regularized contextual bandits and MDPs.
- We found that the KL constraint leads to a sharper sample complexity, and the sample complexity also enjoys a better coverage dependence on the data-collector policy under an on-policy framework.
- Furthermore, we obtain **logarithmic** bounds with optimistic reward estimation.

Algorithm Designs in Reinforcement Learning from Human Feedback (RLHF) *2023.11 - 2024.8*
The Hong Kong University of Science and Technology
Advisor: Prof. Tong Zhang, Prof. Nan Jiang from UIUC

- We formulated the real-world RLHF process as a reverse-KL regularized contextual bandits for preference satisfying Bradley-Terry (BT) model and a the reverse-KL regularized minimax game under general preference, respectively. We studied its theoretical property by proposing statistically efficient algorithms with finite-sample theoretical guarantee.
- We connected our theoretical findings with practical algorithms (e.g. DPO, RSO, iterative IPO), offering new tools and insights for the algorithmic design of alignment algorithms.

Corruption-Robust Reinforcement Learning with General Function Approximation *2022.9 - 2024.2*
University of California, Los Angeles, and The Hong Kong University of Science and Technology
Advisor: Prof. Tong Zhang, Prof. Quanquan Gu from UCLA

- We developed a series of corruption-robust algorithms based on uncertainty weighting for online and offline, value-based and model-based settings.
- We provided theoretical analysis for each algorithm, which enjoy an optimal regret dependence on the corruption level. We implemented the offline algorithm practically under various data-corruption scenarios, which outperforms the state-of-the-art.

Optimal Sample Selection Through Uncertainty Estimation and Its Application in Deep Learning *2020.3 - 2021.8*
HKUST
Advisor: Prof. Tong Zhang, and Prof. Yuan Yao from HKUST

- We proposed a theoretically optimal and computationally efficient sample selection approach.
- We effectively applied it to deep learning and is robust to misspecification (by down-weighting highly uncertain samples).

Provably Efficient Learning in High-Dimensional Batched Bandits *2020.3-2021.7*
USTC
Advisor: Prof. Zhaoran Wang from Northwestern University

- We designed an efficient algorithm for high-dimensional linear contextual bandits with batched feedback.
- Our algorithm enjoyed nearly the same regret order as the sequential case.
- We extended the algorithm to low-rank matrix bandits.

EXPERIENCE

University of California, Los Angeles:
Visiting Reseach Scholar advised by Prof. Quanquan Gu *2023.8 - 2023.12*

The Hong Kong University of Science and Technology:
Teaching Assistant: EMIA 2020 - Cross-disciplinary Design Thinking *Fall 2022*

University of Science and Technology of China:
Teaching Assistant: Mathematical Statistics *Fall 2020*
Teaching Assistant: Linear Algebra *Spring 2020*

HONORS AND AWARDS

Gold Prize for Outstanding Student Scholarship (1/40) *2020.9*

Bronze Prize for Outstanding Student Scholarship *2019.9*

Bronze Prize for Outstanding Student Scholarship *2018.9*

SELECTED PUBLICATIONS AND PREPRINTS

(* denotes alphabetical order or equal contribution)

[1] Heyang Zhao* Chenlu Ye*, Wei Xiong, Quanquan Gu, Tong Zhang, "Logarithmic Regret for Online KL-Regularized Reinforcement Learning", [Preprint].

[2] Chenlu Ye, Yujia Jin, Alekh Agarwal, Tong Zhang, "Catoni Contextual Bandits are Robust to Heavy-tailed Rewards", [Preprint].

[3] Heyang Zhao Chenlu Ye, Quanquan Gu, Tong Zhang, "Sharp Analysis for KL-Regularized Contextual Bandits and RLHF", [Preprint].

[4] Chenlu Ye*, Wei Xiong*, Yuheng Zhang*, Hanze Dong*, Nan Jiang, Tong Zhang, "Online iterative reinforcement learning from human feedback with general preference model", [NeurIPS 2024].

[5] Wei Xiong*, Hanze Dong*, Chenlu Ye*, Han Zhong, Nan Jiang, Tong Zhang, "Iterative preference learning from human feedback: Bridging theory and practice for rlhf under kl-constraint", [ICML 2024]

[6] Chenlu Ye*, Jiafan He*, Quanquan Gu, Tong Zhang, "Towards robust model-based reinforcement learning against adversarial corruption", [ICML 2024].

[7] Chenlu Ye*, Rui Yang*, Quanquan Gu and Tong Zhang, "Corruption-Robust Offline Reinforcement Learning with General Function Approximation", [NeurIPS 2023].

[8] Yong Lin*, Chen Liu*, Chenlu Ye*, Qing Lian, Yuan Yao and Tong Zhang,"Optimal Sample Selection Through Uncertainty Estimation and Its Application in Deep Learning", [Preprint].

[9] Chenlu Ye, Wei Xiong, Quanquan Gu, and Tong Zhang, "Corruption-Robust Algorithms with Uncertainty Weighting for Nonlinear Contextual Bandits and Markov Decision Processes", [ICML 2023].

[10] Jianqing Fan*, Zhaoran Wang*, Zhuoran Yang*, Chenlu Ye*, "Provably Efficient High-Dimensional Bandit Learning with Batched Feedbacks", [Preprint].

PROFESSIONAL ACTIVITY

Conference Reviewer: ICML, NeurIPS, ICLR, AISTAT.

Journal Reviewer: Machine Learning, Artificial Intelligence.