HAO SUN

Available for internship: Mar. 2024 - Aug. 2025

Pembroke College, Cambridge, UK, CB2 1RF

+44 07821-628660 ⋄ hs789@cam.ac.uk ⋄ https://holarissun.github.io/ ⋄ Google Scholar

RESEARCH KEYWORDS

Reinforcement Learning	Large Language Models	Alignment and RLHF
RL for Healthcare and Robotics	Interpretable Machine Learning	Time-Series Modeling

EDUCATION

University of Cambridge (Expected) Aug. 2025
D.Phil. in Applied Mathematics and Theoretical Physics
Thesis: Toward Reality-Centric Reinforcement Learning: Healthcare, Robotics, and Large Language Models
Advisor: Prof. Mihaela van der Schaar.

Chinese University of Hong Kong Sep. 2021
M.Phil. in Information Engineering
Advisor: Prof. Bolei Zhou and Prof. Dahua Lin.

Peking University Sep. 2018
B.Sc. in Physics. Yuanpei Honored Class.
Advisor: Prof. Zhouchen Lin.

INDUSTRIAL EXPERIENCES

Tencent Robotics X. *Research Scientist Intern.* Shenzhen, China. Jun. - Sep. 2021
Amazon AWS Redshift. *Applied Scientist Intern.* Palo Alto, US. (Remote) Jun. - Sep. 2020
Peng Cheng Lab. *Research Scientist Intern.* Shenzhen, China. Jun. - Sep. 2019

SELECTED WORKS

17. Query-Dependent Prompt Evaluation and Optimization with Offline Inverse RL ICLR 2024
Hao Sun, Alihan Hüyük, Mihaela van der Schaar

- Key Words: *Inverse-RL; RLHF; Alignment; Off-Policy Evaluation;*
- Insight: We propose Prompt-OIRL, showing that Inverse RL can be used for offline query-dependent prompt evaluation and optimization. It does not require interactions with the LLMs during learning yet achieves superior performance on arithmetic reasoning tasks.

16. Accountability in Offline RL: Explaining Decisions with a Corpus of Examples NeurIPS 2023
Hao Sun, Alihan Hüyük, Daniel Jarrett, Mihaela van der Schaar

- Key Words: *Explanable RL; Offline-RL;*
- Insight: We introduce an effective algorithm to enhance interpretability and accountability in offline RL. This research is critical for responsibility-sensitive applications like finance and healthcare.

15. Exploit Reward Shifting in Value-Based Deep Reinforcement Learning NeurIPS 2022
Hao Sun, Lei Han, Rui Yang, Xiaoteng Ma, Bolei Zhou

- Key Words: *Value-Based DRL; Offline RL; Exploration; Exploitation;*
- Insight: A positive reward shifting leads to conservative exploitation, while a negative reward shifting leads to curiosity-driven exploration.

14. Policy Continuation with Hindsight Inverse Dynamics (Spotlight) NeurIPS 2019
Hao Sun, Zhizhong Li, Dahua Lin, Bolei Zhou

- Key Words: *Self-Imitate RL; Supervised Learning for RL*
- Insight: For the first time in the field, we show supervised learning can be applied to improve sample efficiency and stability of goal-conditioned RL tasks.

RECENT PREPRINTS

13. Dense Reward for Free in Reinforcement Learning from Human Feedback 2024
Alex J Chan, **Hao Sun**, Samuel Holt, Mihaela van der Schaar

- Key Words: *RLHF; Credit Assignment;*
- Insight: the attention weights of reward models in RLHF can guide credit assignment to accelerate and stabilize the learning process.

12. Retrieval-Augmented Thought Process as Sequential Decision Making 2024
Thomas Pouplin, **Hao Sun**, Samuel Holt, Mihaela Van der Schaar

- Key Words: *Thought Processes; Information Retrieval; Monte-Carlo Tree Search;*
- Insight: we model the reasoning process of language models as a sequential decision-making problem, and apply MCTS as an efficient planner for the task.

11. Reinforcement Learning in the Era of LLMs: What is Essential? What is Needed? 2023
Hao Sun

- Key Words: *RLHF; Prompting; Tutorial on RL;*
- Insight: (1) RLHF is online IRL rather than offline RL. (2) RLHF is better than SFT because imitation learning alleviates the compounding error problem. (3) Insight of RM can be generalized to other LLM applications except alignment. (4) RLHF is more challenging than conventional IRL due to action space dimensionality and reward sparsity. (5) The superiority of PPO in RLHF may originate from its stability.

10. DataCOPE: Rethinking Off-Policy Evaluation Problems from a Data-Centric Perspective 2023
Hao Sun, Alex Chan, Nabeel Seedat, Alihan Hüyük, Mihaela van der Schaar

- Key Words: *Off-Policy Evaluation; Uncertainty Quantification; Data-Centric AI*
- Insight: We demonstrate the importance of the data-centric perspective of Off-Policy Evaluation. OPE is not only a challenge for learning algorithms, but also a challenge for the quality of data.

9. Meta-RL Solvers Also Solve RL 2023
Hao Sun

- Key Words: *Sample-Efficient RL; Foundation Models for Decision Modeling; Meta-RL*
- Insight: Regarding RL tasks as a generalization over initial state distributions, Meta-RL algorithms can be applied to improve sample efficiency.

SELECTED CONFERENCE AND WORKSHOP PAPERS

8. DAUC: a Density-based Approach for Uncertainty Categorization NeurIPS 2023
Hao Sun, Boris van Breugel, Jonathan Crabbe, Nabeel Seedat, Mihaela van der Schaar

- Key Words: *Uncertainty Quantification; Explainable Machine Learning;*
- Insight: Uncertain examples flagged by various uncertainty quantifications can be categorized into three categories: examples that are similar to misclassifications, examples located at decision boundaries, and OOD.

7. Neural Laplace Control for Continuous-time Delayed Systems AISTATS 2023
Samuel Holt, Alihan Hüyük, Zhaozhi Qian, **Hao Sun**, Mihaela van der Schaar

- Key Words: *Model-Based DRL; Continuous Control; Model Predictive Control;*
- Insight: We study and solve a realistic problem setting in DRL where control signals are continuous in time and systematic delay exists.

6. Supervised Q-Learning can be a Strong Baseline for Continuous Control FMDM@NeurIPS 2022
Hao Sun, Ziping Xu, Yuhang Song, Meng Fang, Bolei Zhou

- Key Words: *Self-Imitate RL; Sample-Efficient RL;*
- Insight: The idea of using supervised policy updates to solve RL problems can be generalized to continuous control tasks.

5. Toward Causal-Aware RL: State-Wise Action-Refined Temporal Difference DRL@NeurIPS 2022
Hao Sun, Taiyi Wang

- Key Words: *Causality-Driven Temporal Difference Learning; Feature Selection;*
- Insight: We introduce two practical algorithms to reduce action space redundancy through causality-aware temporal difference learning.

4. MOPA: a Minimalist Off-Policy Approach to Safe-RL DRL@NeurIPS 2022

Hao Sun, Ziping Xu, Meng Fang, Zhenghao Peng, Bo Dai, Bolei Zhou

- Key Words: *AI Safety; Constrained RL; Sample-Efficient RL;*
- Insight: We introduce a minimalist approach for the Safe-RL challenges by introducing the Early-Terminated MDP. We further propose to use context variables to boost the generalization ability of the RL algorithm under such MDPs.

3. Rethinking Goal-conditioned Supervised Learning and Its Connection to Offline RL ICLR 2022

R. Yang, Y. Lu, W. Li, **H. Sun**, M. Fang, Y. Du, X. Li, L. Han, C. Zhang

- Key Words: *Self-Imitate RL; Offline RL; Goal-Conditioned RL;*
- Insight: A supervised learning approach can also solve the reward of sparse goal-conditioned tasks in offline settings.

2. Adaptive Regularization of Labels AAAI 2021

Qianggang Ding, Sifan Wu, **Hao Sun**, Jiadong Guo, Shu-Tao Xia

- Key Words: *Soft Label Learning; Regularization;*
- Insight: We exploit the informative inherent structure in labels and improve the prediction accuracy of neural networks through regularization.

1. Hierarchical Multi-Scale Gaussian Transformer for Stock Movement Prediction IJCAI 2020

Qianggang Ding, Sifan Wu, **Hao Sun**, Jiadong Guo, Jian Guo

- Key Words: *Time-Series Modeling; Foundation Models;*
- Insight: We improve the forecasting ability of transformers in time-series data and apply it to stock market movement prediction.

TEACHING

Machine Learning Summer School
University of Cambridge. Teaching Assistant. Jun. - Sep. 2022

Deep Reinforcement Learning
Chinese University of Hong Kong. Teaching Assistant. Jan. - Jun. 2020

Final Year Project on Machine Learning
Chinese University of Hong Kong. Teaching Assistant. Aug. 2018 - Jun. 2019

SERVICE

I serve as a reviewer for NeurIPS, ICLR, AISTATS, and AAAI, and a PC member for the CausalML workshop at NeurIPS 2022, RLxLLM workshop at AAAI 2024.

HONOURS

- D.Phil. Scholarship Awarded by ONR Oct. 2021
- M.Phil. Scholarship Awarded by CUHK Aug. 2018
- Outstanding Graduate of Peking University Jul. 2018
- The May-4th Scholarship (The Highest Honor for Undergrad Students in Peking University) Sep. 2017
- The Weiming Scholarship (4 times) Sep. 2014 - 2017
- First Prize in the Big Data Innovation and Entrepreneurship Competition May. 2016
- National Innovation Fund for Undergraduate Research Oct. 2015
- First Prize in China Undergraduate Physics Tournament (CUPT) Aug. 2014

SKILLS

Programming Skills	Mainly work with Python, also write C++, C, HTML
Deep Learning Packages	Mainly work with PyTorch, also use Keras, Tensorflow, Jax
Language	Full proficiency in English. Native Mandarin. A bit of French and Japanese.
Miscellaneous	Climbing, Bouldering, Snowboarding, Ski.