HAO SUN

Pembroke College, Cambridge, UK, CB21RF
+44 07821-628660 ⋄ hs789@cam.ac.uk ⋄ https://holarissun.github.io/ ⋄ Google Scholar

RESEARCH KEYWORDS (2020-2023)

RL via Supervised Learning	Value-Based DRL	Offline RL	Interpretable RL
Diverse Policy Generation	Optimism Exploration	Uncertainty Quantification	Data-Centric OPE
Time-Series Modelling			RL in Language Models

ACADEMIC EXPERIENCES

University of Cambridge Expected 2025
D.Phil. in Applied Mathematics and Theoretical Physics
Advisor: Prof. Mihaela van der Schaar.

Chinese University of Hong Kong 2018 - 2021
M.Phil. in Information Engineering
Advisor: Prof. Bolei Zhou, Prof. Dahua Lin.

Peking University (Yuanpei Honored Class) 2013 - 2018
B.Sc. in Physics, Computer Science, and Management.
Advisor: Prof. Zhouchen Lin.

INDUSTRIAL EXPERIENCES

Tencent Robotics X. Shenzhen, China. Jun. - Sep. 2021
Research Scientist Intern in developing value-based Deep Reinforcement Learning algorithms with Dr. Lei Han, and sample-efficient Reinforcement Learning with Prof. Meng Fang.

Amazon AWS Redshift. Palo Alto, US. (Remote) Jun. - Sep. 2020
Applied Scientist Intern in applying RL in database optimization with Dr. Balakrishnan Narayanaswamy.

Peng Cheng Lab. Shenzhen, China. Jun. - Sep. 2019
Research Scientist Intern in sample-efficient Deep Reinforcement Learning; Apply Machine Learning and Reinforcement Learning in quantitative trading. Advised by Prof. Jian Guo.

SELECTED WORKS

14. Exploit Reward Shifting in Value-Based DRL NeurIPS 2022
Hao Sun, Lei Han, Rui Yang, Xiaoteng Ma, Jian Guo, 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.

13. 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.

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

- Key Words: *Self-Imitate 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

11. Offline Prompt Evaluation and Optimization through Inverse RL 2023
Hao Sun, Alihan Hüyük, Andrew Rashbass, Mihaela van der Schaar

- Key Words: *Off-Policy Evaluation; Inverse-RL; RLHF and RLAIF*
- Insight: We propose Prompt-OIRL, showing that Inverse RL can be used for OFFLINE prompt evaluation and optimization. It does not require interactions with the LLMs during learning.

10. Accountable Batched Control with Decision Corpus 2023
Hao Sun, Alihan Hüyük, Daniel Jarret, Mihaela van der Schaar

- Key Words: *Explanable RL; Offline-RL;*
- Insight: We introduce an effective algorithm for accountable offline RL that is critical for responsibility-sensitive applications like finance and healthcare.

9. 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.

8. Meta-RL Solvers Also Solve RL 2023
Hao Sun, Bolei Zhou

- 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

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. 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.

5. 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.

4. DAUX: a Density-based Approach for Uncertainty eXplanations DFUQ@ICML 2022
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.

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. 2022 - Sep. 2022

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

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

SERVICE

Serve as a reviewer for:

- NeurIPS 2021-2023
- NeurIPS 2022 Causal Machine Learning for Impact Workshop
- AISTATS 2023

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

LEADERSHIP

Central Plains Development Research Association, Vice President Sep. 2016 - Jun. 2017
Organized more than 20 public welfare inspirational mindset-adjust talks in Henan and Qinghai province.
Organized non-profit lectures in Henan, Qinghai, Shanxi, and Shandong provinces to help students prepare for the Independent Enrollment for the College Entrance Exam

Academic Practice Department of Yuanpei College, Vice Minister Sep. 2015 - Jun. 2016
Took charge of preparation activities of Yuanpei College for the Challenge Cup and awarded the highest prize of Wang Xuan Cup.

SKILLS

Programming Skills	Python, C++, C, HTML
Deep Learning Packages	PyTorch, Keras, Tensorflow, Jax
Language	English: TOEFL (106/120)
Miscellaneous	Climbing, Bouldering(v5-6), Snowboard(BASI Lv3), Ski(BASI Lv1).