



2025年本科生进组宣讲

赵鹏

zhaop@lamda.nju.edu.cn

2025.02.20

个人信息

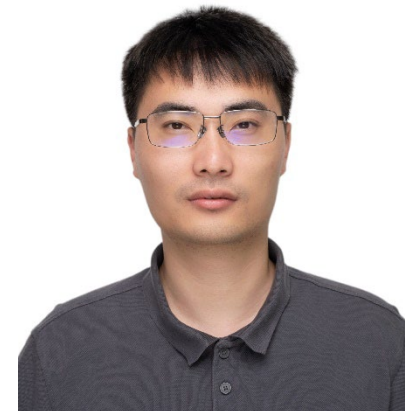
赵鹏

就职：准聘副教授

LAMDA研究所，人工智能学院

主页：<https://www.pengzhao-ml.com/>

邮箱：zhaop@lamda.nju.edu.cn



研究方向

- 机器学习理论、在线学习、强化学习
- 大模型优化、随机优化、大规模优化等

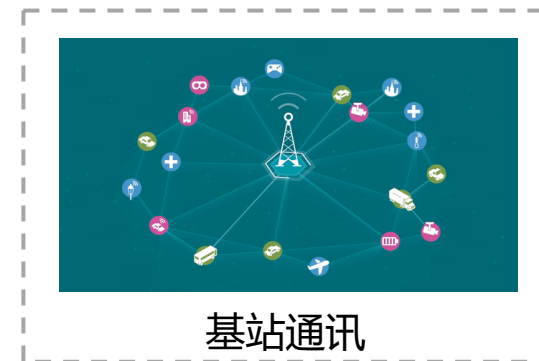
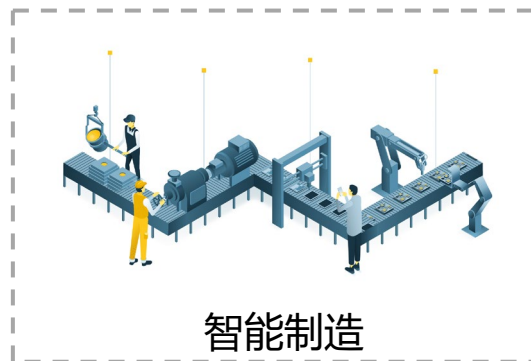
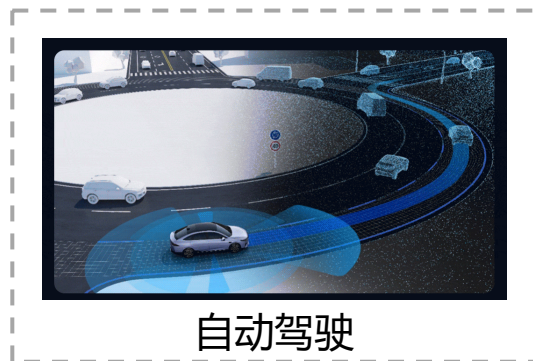


研究方向

- 机器学习是这一轮人工智能热潮的核心推动技术



- 越来越多应用中，环境呈现“开放动态”，数据呈现“流式积聚”



研究方向

- 探究机器学习基础理论 (**Foundations of Machine Learning**), 指导算法设计及实践应用

机器学习与优化：理论与应用

分布漂移

类别未知

特征新增

类别新增

开放环境机器学习

流式机器学习场景

在线学习及优化

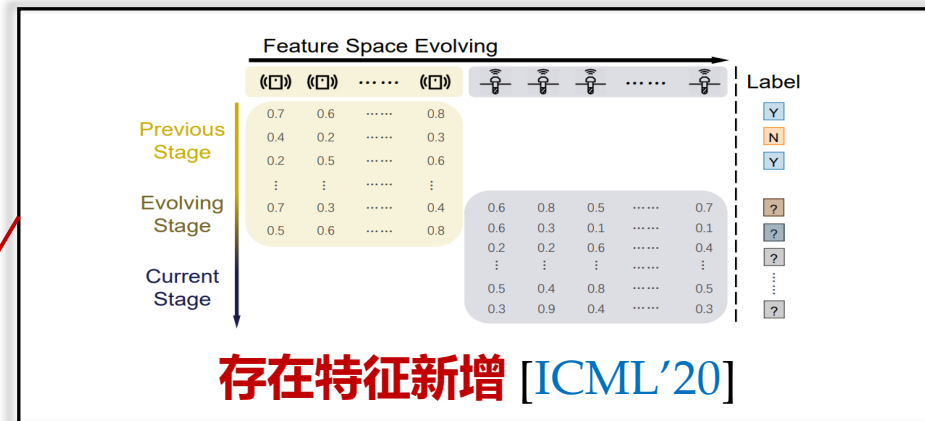
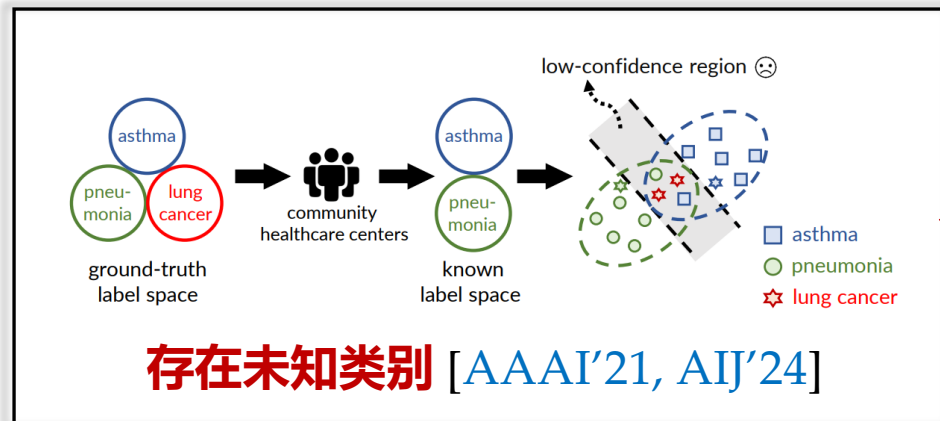
在线预测

强化学习

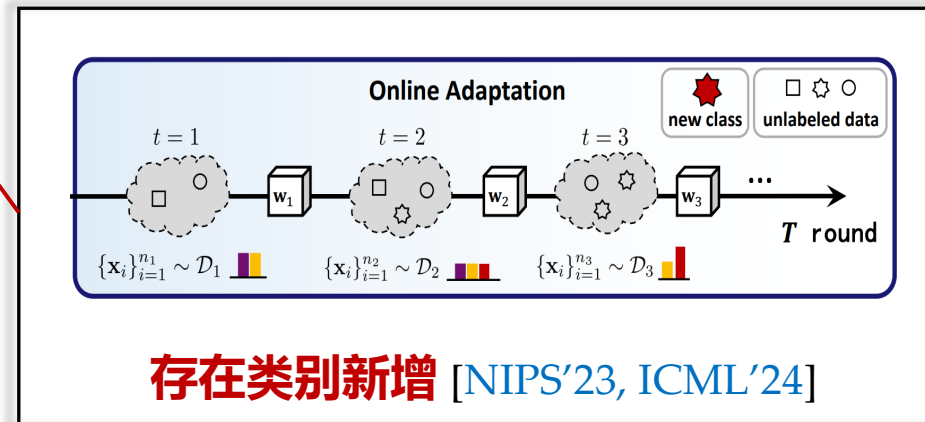
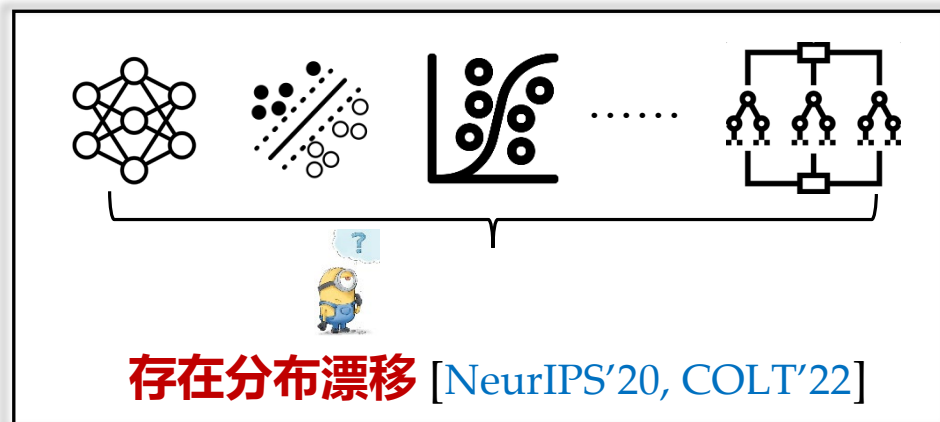
控制理论

博弈理论

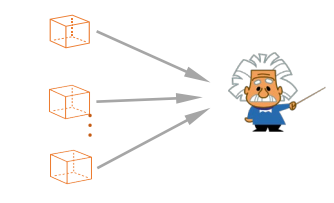
一、开放环境机器学习理论



开放环境
学习



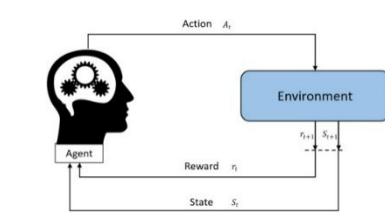
二、在线学习及优化



非稳态学习

自适应环境的
优化算法

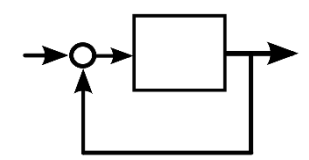
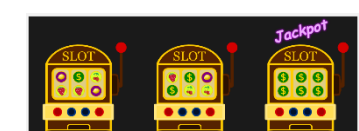
[NIPS' 24, JMLR'24,
ICML'24, NIPS'23,
ICML'23, COLT'22]



强化学习/ Bandits理论

平衡探索于利用
设计理论保证算法

[NIPS' 24, AAAI'24,
NIPS'23, ICML'22,
COLT'22, JMLR'22]



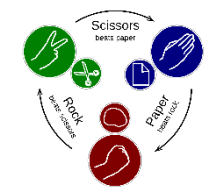
在线控制

设计一般场景下
在线控制策略

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + B\mathbf{u}_t + \xi_t$$
$$\mathbf{y}_t = C\mathbf{x}_t + \mathbf{e}_t$$

[ICML'24, JMLR'23,
NIPS'22, AISTATS'22]

在线学习/ 优化



博弈理论

设计快速求解
博弈均衡算法

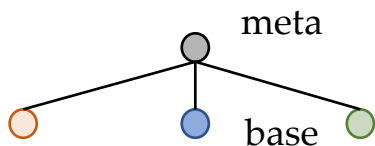
[ICML'22, ICML'23,
NIPS'24, NIPS'23]

	Rock	Scissors	Paper
Rock	(0,0)	(1,-1)	(-1,1)
Scissors	(-1,1)	(0,0)	(-1,1)
Paper	(1,-1)	(-1,1)	(0,0)

二、在线学习及优化

在线优化理论

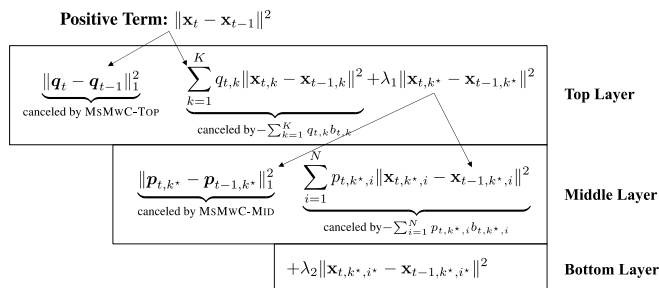
通过**集成**处理环境中不确定性



自适应：合理利用环境信息

$$\sum_{t=1}^T \langle r_{t,i^*} - m_{t,i^*} \rangle^2 \leq \begin{cases} \sum_{t=1}^T \langle \nabla f_t(\mathbf{x}_t), \mathbf{x}_t - \mathbf{x}_{t,i^*} \rangle^2, & (\text{exp-concave \& strongly convex}) \\ V_T + \sum_{t=2}^T \|\mathbf{x}_{t,i^*} - \mathbf{x}_{t-1,i^*}\|^2 + \sum_{t=2}^T \|\mathbf{x}_t - \mathbf{x}_{t-1}\|^2. & (\text{convex}) \end{cases}$$

最优性：确保集成误差可控



在线博弈理论

在复杂环境下求解博弈均衡策略

$$\text{Dynamic NE-regret: } \left| \sum_{t=1}^T x_t^\top A_t y_t - \sum_{t=1}^T x_t^{*\top} A_t y_t^* \right|$$

利用算法稳定性进行快速求解

$$\sum_{t=1}^T \langle A_t y_t, x_{t,i} - u_t^x \rangle \lesssim \frac{1 + P_T^x}{\eta_i^x} + \eta_i^x \sum_{t=2}^T \|A_t y_t - A_t y_{t-1}\|^2 - \frac{1}{\eta_i^x} \sum_{t=2}^T \|x_{t,i} - x_{t-1,i}\|^2$$

$$\sum_{t=1}^T \langle -A_t x_t, y_{t,j} - u_t^y \rangle \lesssim \frac{1 + P_T^y}{\eta_j^y} + \eta_j^y \sum_{t=2}^T \|A_t x_t - A_t x_{t-1}\|^2 - \frac{1}{\eta_j^y} \sum_{t=2}^T \|y_{t,j} - y_{t-1,j}\|^2$$

添加修正项处理算法复杂结构

$$\sum_{t=2}^T \|x_t - x_{t-1}\|^2 \lesssim \|p_t - p_{t-1}\|_1^2 + \sum_{i=1}^N p_{t,i} \|x_{t,i} - x_{t-1,i}\|^2$$

$$\sum_{t=1}^T \langle \ell_t, p_t - e_{i^*} \rangle \leq R_T - \sum_{t=1}^T \sum_{i=1}^N p_{t,i} b_{t,i} + \sum_{t=1}^T b_{t,i^*}$$

相消

强化学习理论

考虑非稳态下linear mixture MDP

$$\mathbb{P}_h(s' | s, a) = \phi(s' | s, a)^\top \theta_h^*, \quad \forall (s, a, s')$$

occupancy-based 方法更新策略

$$\hat{q}_{k+1,i} = \arg \max_{q \in \Delta(\mathcal{P}_k, \alpha)} \eta_i \langle q, r_k \rangle - D_\psi(q, \hat{q}_{k,i})$$

$$p_{k+1,i} \propto \exp \left(\varepsilon \sum_{j=1}^k h_{j,i} \right) \text{ with } h_{k,i} = \langle \hat{q}_{k,i}, r_k \rangle$$

policy-based 方法评估策略

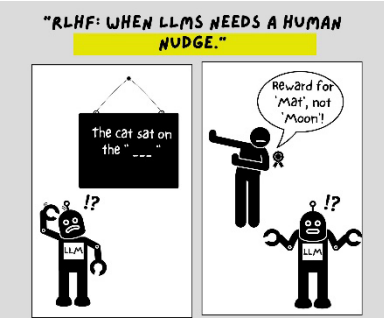
$$\hat{\theta}_{k,h} = \arg \min_{\theta \in \mathbb{R}^d} \sum_{j=1}^{k-1} [\langle \phi_{V_{j,h+1}}(x_{j,h}, a_{j,h}), \theta \rangle - V_{j,h+1}(x_{j,h+1})]^2 / \sigma_{j,h}^2 + \lambda \|\theta\|_2^2$$

$$Q_{k,h}(\cdot, \cdot) = \left[r_{k,h}(\cdot, \cdot) + \min_{\theta \in \mathcal{C}_{k,h}} \langle \theta, \phi_{V_{k,h+1}}(\cdot, \cdot) \rangle \right]_{[0,H]}$$

巧妙结合二者优点，取得最优保障

$$\text{D-Reg}_K \leq \tilde{O} \left(\sqrt{d^2 H^3 K} + \sqrt{H K (H + \bar{P}_K)} \right).$$

三、大模型优化等应用



Online RLHF Pipeline: A Pytorch Implementation

A PyTorch implementation of the paper "Provably Efficient RLHF Pipeline: A Unified View from Contextual Bandits". This repository provides a flexible and modular approach to Reinforcement Learning from Human Feedback (RLHF).

This repository is forked from [OpenRLHF](#) repo, and modified to implement our Online RLHF Pipeline.

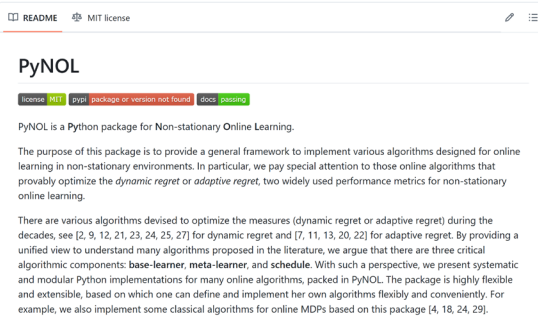
🔥 **Cook Your Own Online-RLHF Recipe!**

Think of Online-RLHF training as cooking - each step is an ingredient that contributes to the final dish. This implementation lets you mix and match different training components to create your perfect RLHF recipe!

📖 **Recipe Format**

```
# A recipe is a list of ingredients (steps):  
# 1. Single ingredient: A string representing one step (e.g., "SFT")  
# 2. Ingredient group: A list of steps that cook together in one iteration
```

大模型的人机对齐 (RLHF)

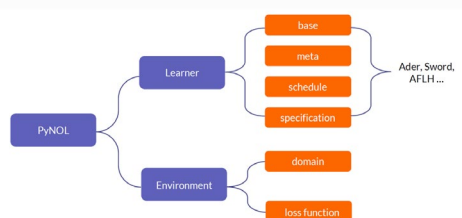


PyNOL

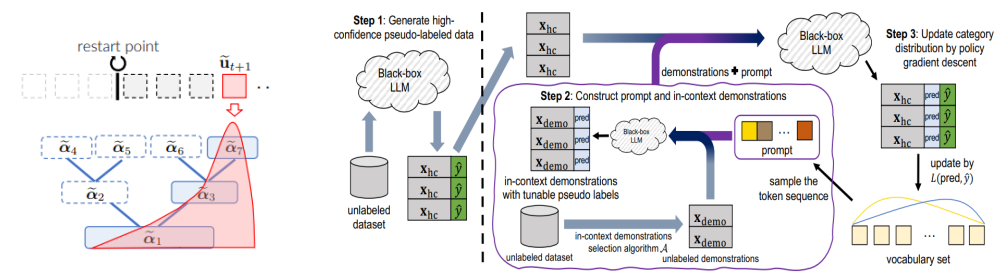
PyNOL is a Python package for Non-stationary Online Learning.

The purpose of this package is to provide a general framework to implement various algorithms designed for online learning in non-stationary environments. In particular, we pay special attention to those online algorithms that provably optimize the *dynamic regret* or *adaptive regret*, two widely used performance metrics for non-stationary online learning.

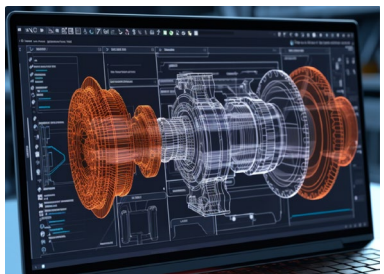
There are various algorithms devised to optimize the measures (dynamic regret or adaptive regret) during the decades, see [2, 9, 12, 21, 23, 24, 25, 27] for dynamic regret and [7, 11, 13, 20, 22] for adaptive regret. By providing a unified view to understand many algorithms proposed in the literature, we argue that there are three critical algorithmic components: **base-learner**, **meta-learner**, and **schedule**. With such a perspective, we present systematic and modular Python implementations for many online algorithms, packed in PyNOL. The package is highly flexible and extensible, based on which one can define and implement her own algorithms flexibly and conveniently. For example, we also implement some classical algorithms for online MDPs based on this package [4, 18, 24, 29].



非稳态在线学习工具包PyNOL



大模型的持续微调 (continual learning)



通讯基站调优、工业设计软件等应用

研究成果及国际影响

- 在机器学习顶级学术期刊论文发文50余篇，包括
 - 期刊：JMLR 5 篇、AIJ 1 篇、IEEE TKDE 3篇 等
 - 会议：COLT 2 篇、NeurIPS 13 篇、ICML 8 篇、AITSTATS 7 篇 等

研究成果得到包括麻省理工、哈佛、斯坦福等**国际著名单位研究组**关注



研究论文得到**国际知名学者、领域同行**的引用，并得到积极正面评价



Shankar Sastry

IEEE/IFAC Fellow
美国国家工程院院士
美国艺术与科学学院院士
美国加州大学伯克利分校教授



David Woodruff

国际理论计算机科学顶会
SODA 2024程序主席
国际数据挖掘顶会KDD
2014最佳论文获得者
美国卡耐基梅隆大学教授



John Lygeros

国际控制领域著名会议
HSCC 2014程序主席
国际控制领域著名会议
ECC 2013程序主席
苏黎世联邦理工大学教授



Elad Hazan

国际计算学习理论顶会
COLT 2015程序主席
在线凸优化领域权威
美国普林斯顿大学教授



Kilian Q. Weinberger

AAAI 2011程序主席
ICML 2016程序主席
ICML 2023 PC
康奈尔大学教授

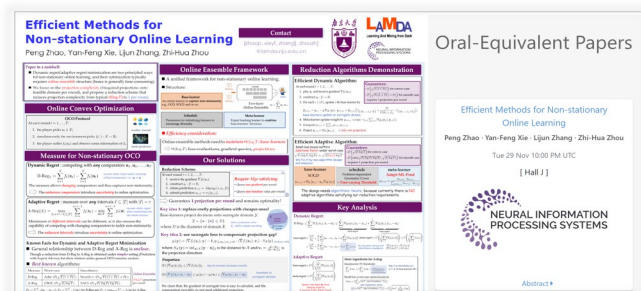


Tuomas Sandholm

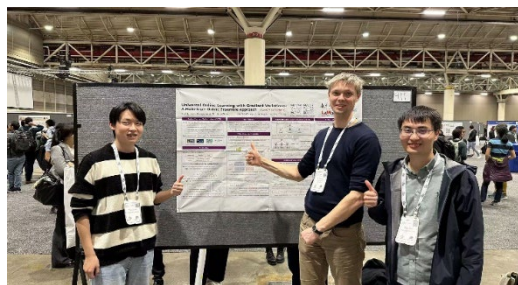
ACM/AAAI/AAAS Fellow
德州扑克AI之父
IJCAI Minsky Medal,
IJCAI McCarthy Award得主
卡耐基梅隆大学教授

研究成果及国际影响

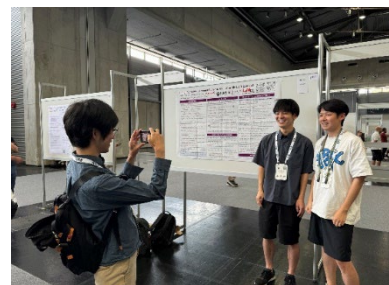
- 研究工作连续三年入选机器学习顶会ICML/NeurIPS的**Oral/Spotlight展示**



NeurIPS'22 Oral
(录用率1.76%)

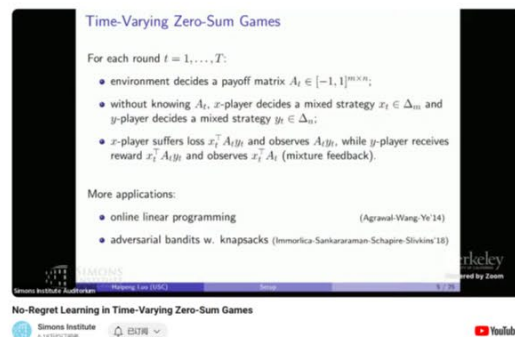
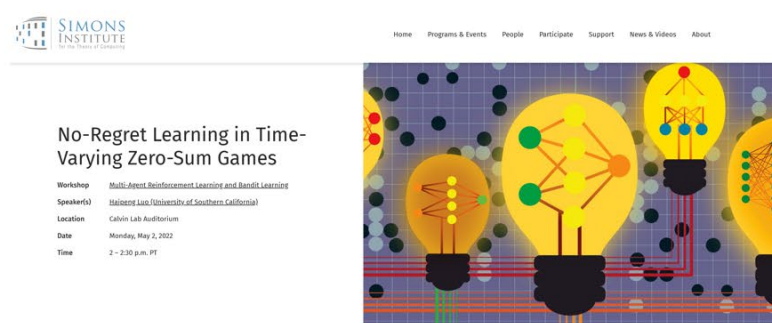


NeurIPS'23 Spotlight
(录用率3.6%)



ICML'24 Spotlight
(录用率3.5%)

- 研究工作受邀到理论计算机科学领域著名的**Simons Institute**报告

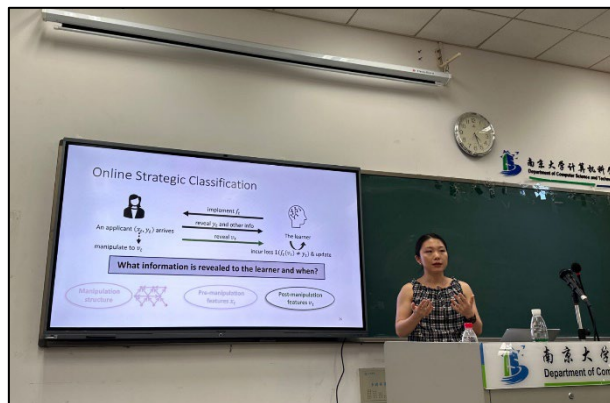


国际学术交流

- 邀请国内外同行前来南大进行报告



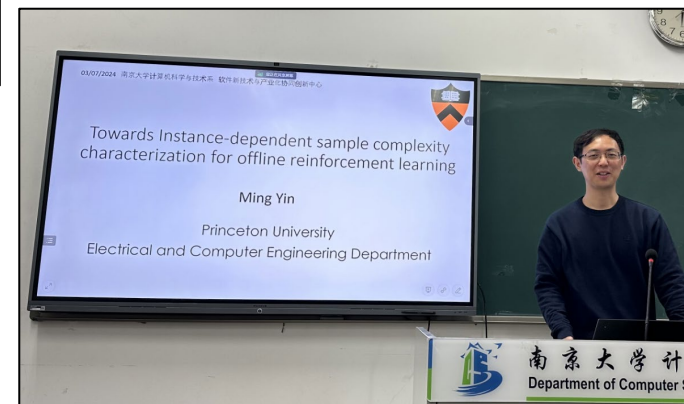
清华特奖得主
MIT博士生 戴言



Harvard Postdoc
马里兰大学助理教授 邵涵



Yale Ph.D.
Weiqiang Zheng



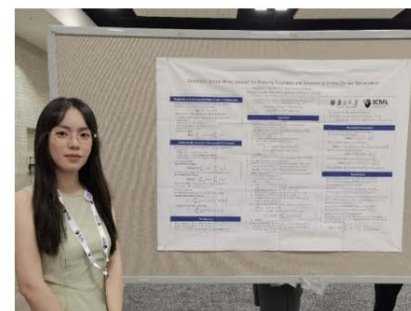
Princeton Postdoc
Ming Yin

本科生培养

- 指导本科生陈思佳同学，研究工作发表于**机器学习顶会ICML'23**，期刊拓展版发表于**顶刊JMLR'24**
- 组内本科生出国访问，推荐支持



匡院大三学生李想，在**加州伯克利大学**交换并参与**Peter Bartlett教授科研组**进行科研实习



人工智能学院大四学生郁航，在**匹兹堡大学**及**卡耐基梅隆大学**进行科研实习并参与课程学习

组内定期团建活动



梅花山春游 (2023.03.02)



皖南秋游 (2023.10.15)



羽毛球乒乓球活动 (weekly)



奥本海默观影 (2023.09.14)

入组培养

- 研究方向
 - 理论方法：在线优化、强化/深度学习理论
 - 算法应用：大模型优化、对齐、持续学习
- 夯实理论基础
 - 周组会：接触前沿学习理论及优化研究
 - 读书班：学习理论、优化方面经典教材
- 工程实践项目
 - 组内mini project训练；大创/竞赛/实践项目

精力有限：~1名理论，~2名算法应用

招生说明：https://www.pengzhao-ml.com/for_student.html



国际视野，尊重学术，氛围和谐

邮箱：zhaop@lamda.nju.edu.cn

或联系

组内博士生谢龔锋 (xieyf@lamda.nju.edu.cn)

谢谢！