莫大–北理工–深北莫
应用数学联合研究中心
Совместный научно-исследовательский центр прикладной математики
МГУ–ППИ–МГУ–ППИ
MSU–BIT–SMBU Joint Research Center of Applied Mathematics

计算数学与控制系
Факультет вычислительной математики и кибернетики
Faculty of Computational Mathematics and Cybernetics

应用数学讲座

Научный Семинар по Прикладной Математике
Research Seminar on Applied Mathematics

应用数学报告（120）

报告人 / Докладчик / Speaker: PETROSIAN OVANES

题 目 / Название / Title: Multi-agent LLM: important results and future directions 时间 / Время / Time：2025.05.13, 18:00-19:30

地点 / Место / Venue: 1-330

摘要 / Аннотация / Abstract:

Research on multi-agent large language model (LLM) training and inference is transforming AI by enabling collaborative systems that outperform single-agent approaches in scalability, adaptability, and complex problem-solving. For instance, Chain of Agents [1] demonstrates how LLMs can collaborate on long-context tasks by chaining specialized sub-agents, while multi-agent Tree-of-Thought approaches improve reasoning using agents that critique and refine outputs iteratively. Such systems enhance robustness by distributing tasks—seen in reflective multi-agent collaboration [2], where agents self-correct via debate. However, ethical risks like biased collusion or unsafe emergent behaviours persist, necessitating alignment research as discussed in Science Robotics and Nature. As multi-agent LLMs advance—powering applications from healthcare to autonomous systems — their ability to balance collaboration, efficiency, and safety will define the next generation of AI.

Multi-agent Large Language Models training and inference research can be structured in the following directions and challenges that will be discussed during the presentation:

- Large Language Model inference acceleration: computation scheduling and forecasting.
- Multi-agent LLM Chain-of-thought inference time optimization and convergency.

访问学者以及需要做报告的学者请联系 ***张晔*** *教授 / Всем желающим выступить с докладами на семинаре просьба обращаться к* ***Чжану Е*** **ye.zhang@smbu.edu.cn**

- Multi-agent LLM communication and training optimization: automatic organization of Multi-agent LLM communication, improved LLM fine-tuning using Shapley value.

[1] Zhang Y, Sun R, Chen Y, et al. Chain of agents: Large language models collaborating on long-context tasks[J]. Advances in Neural Information Processing Systems, 2024, 37: 132208-132237.

[2] Zangwei Zheng, Xiaozhe Ren, Fuzhao Xue, Yang Luo, Xin Jiang, Yang You, Response Length Perception and Sequence Scheduling: An LLM-Empowered LLM Inference Pipeline, Advances in Neural Information Processing Systems, pp. 65517-65530, volume 36, 2023.

个人简介 （Individual Resume）：

Ovanes Petrosian received the Ph.D. degree in applied mathematics from Saint-Petersburg State University, Saint Petersburg, Russia, in 2017. From 2017 to 2022, he was an Associate Professor at the Saint-Petersburg State University. In 2022 he defended his Doctor of Science Degree. In 2023 he became a Professor in Saint-Petersburg State University. Additionally, he was working as RnD-engineer and main RnD-engineer in Siemens and Huawei Russia correspondingly. In Huawei at 2019 he established an RnD laboratory for optimization in Telecom and at 2022 in Saint-Petersburg State University he held a position of Director of AI-center of Saint-Petersburg State University. His research interests include machine learning, deep learning, reinforcement learning, time series analysis, graph theory, discrete probability models and stochastic processes, game theory, and operations research. In 2023 let the team to become TOP 1 in the world industry AI competition «Pump it Up: Data Mining the Water Table» on the DrivenData platform.

访问学者以及需要做报告的学者请联系 ***张晔*** *教授 / Всем желающим выступить с докладами на семинаре просьба обращаться к* ***Чжану Е*** ye.zhang@smbu.edu.cn