S-Data Science Lab

Hongteng Xu

09/06/2023

Overview

- **Our Mission**
  - New AI Paradigms
    - Model Architectures, Optimization Methods, etc.
  - AI-Empowered Scientific Discovery
    - AI4Science, ***AI4Health, AI4Math***
- **Tech Questions**
  - Structured Data (e.g., Graphs and Sequences) Inference, Analysis, Generation, and ***Control***
- **Applications**
  - Drug design
  - Network analysis
  - **3D Graphics**
  - **Mathematical Problem Solver**

Your Missions (Our Goals)

- **Strong Mathematical Modeling Skill**
  - Abstract Ability: Describe Real-World Problems in Math
  - Association Ability: Connect Knowledge/Tech in Different Fields
  - Problem Solver: Learn Optimization Methods and Theory
- **Outstanding Communication Skill**
  - Listening, Writing, Presentation, English Skill
- **Terrific Coding Skill**
  - Fluent on Python, Build and Maintain Toolboxes and Libs (Your Research Fruits)
  - If you would like, teach me other things (e.g., CUDA)
- **A Honest Person THE MOST IMPORTANT!**
  - Fight to Academic Misconduct
- **If our goals are inconsistent, we need to talk in person...**

Target Publications

- **Conferences**
  - **CCF-A:** ICML, NeurIPS, ICLR, COLT, KDD, WWW, AAAI, IJCAI, CVPR, ICCV, SIGIR, ACMMM
  - **CCF-B:** ECCV, WSDM, CIKM, UAI, AISTATS, ICASSP, SDM, ICDM, ICME...
- **Journals**
  - JMLR, TPAMI, TIT, SIAM Optimization, Nature Machine Intelligence, ...
  - TMLR, TSP, TNNLS, TKDE, TSIPN, ...
- **1. Let me know when you have other targets.**
- **2. Show me your COMPLETE draft two weeks before ddl.**
- **3. Don't abuse coauthorship.**
- **4. Be careful, make your work systematic.**

Workload and Payment

- **Payment Strategy**
  - Time: <10 months
  - Amount: Depend on funding status
- **Internship and visiting scholarship**
  - For summer internships, no restrictions.
    - Be responsible for yourselves.
  - For the internships in other time periods
    - Let me know in advance. Correlated with your research.
  - Open to international exchange and visiting projects.
    - Let me know in advance. Correlated with your research.
- **Flexibility of Workload**
  - Do not absent group/project meetings without notices
  - No more rules

Project Management

- **https://github.com/SDS-Lab**
  - Release our code through this organization and fork to yourself.
  - Maintain your works on your own.
- Generally, each work is associated with **papers, codes, and patents.**

What We Did Last Year?

- **Accepted Works (Let us cheer up:))**
  - My work: TPAMI + 2, ACMMM 2023, IJCAI 2023 + 2
  - Team work: AAAI 2023 + 2, CIKM 2023 + 1
  - Every PhD has at least one submission/accepted paper.
- **Projects:**
  - CAAI-Huawei Mindspore （9w, Nov 2023）
  - NSFC重大研究计划培育项目（80w，Dec 2025）
  - MOOC Funding（2w，Done）
  - NSFC青基（30w，Dec 2024）
  - RUC Startup Funding (15w，Dec 2024)
  - NSFC海外优青（300w，Dec 2025）

What Will We Do This Year?

- **Projects:**
  - 中央引导地方人工智能专项（1800w，申请中）
  - 302 Hospital Project (推进中)
  - CAAI-Huawei Mindspore (9w, Nov. 2023)
  - **Prepare for writing and applying other proposals**
- **Research work:**
  - 1st-year PhD: at least submit one paper.
  - 2nd-year PhD: one accepted work + at least submit one more paper.
  - 3rd-year PhD: at least submit one paper **based on your own idea.**
  - **Sharpen your reading, talking and writing skills.**

What Will We Do This Year?

- **Research**
  - **More reading workloads.** Improve your reading and investigation efficiency.
    - Try to compress your preparation time for reading groups
    - Push me with papers, rather than let me push you
  - **More informative group meeting report.**
    - It is a good chance to share your research and interact with others.
  - After reading sufficiently, more brain storms
    - Try to propose research ideas by yourselves
  - **Coding and math**
    - **Be familiar with basic concepts, knowledge and tools in your field**

What Will We Do This Year?

- **AI4Science**
  - **New Model Architectures for AlphaFold and Others (Shen)**
  - **LLM for Molecular Retrosynthesis/Synthesis (Shen, Fanmeng)**
  - **Shape-based Molecular Generation (Fanmeng)**
  - **Toxicology Prediction (Fanmeng)**
  - **OT-based GNN Learning (Minjie)**
  - **Quaternion-based Conformation Analysis (Angxiao)**
  - **Self-supervised Learning for Single-cell Analysis (Fengjiao, Angxiao)**

What Will We Do This Year?

- **Point Process for Healthcare**
  - **EOT-based TPP Learning (Qingmei)**
  - **NTPP Clustering (Qingmei)**
  - **LLM for TPP (Yuxin)**
  - **Granger Causality for NTPP (Ke, Yuxin)**
- **Toolboxes and Benchmarks (Dec. 2023)**
  - **PoPPy2 (Yuxin, Qingmei)**
  - **OT Package, developing POT API + Mindspore version (Minjie, Angxiao)**
- **Write ML Book (Hongteng, Jan. 2024)**

Thanks & QA!