

XINJIE SHEN

EDUCATION

South China University of Technology (SCUT), Guangzhou, China

Sep. 2021 – Present

Undergraduate student in Artificial Intelligence **Homepage:** <https://xinjie-shen.com> **GPA:** 3.76/4.0

Reserach Interests: Graph Learning, Quantitative Finance, Automatous LLM Agent **Chairman** of AIA (AI Association) in SCUT

EXPERIENCE

Automatous R&D LLM Agent and Quantitative Finance Strategy in MSRA

Jan. 2024 – Present

Working with Dr. Xiao Yang, Microsoft Research Asia. Research on Automatic Research and Development (R&D) and Quantitative Finance Strategy. Imagine a world where the R&D process is fully automated, and the LLM can automatic generate analysis results of every proposed ideas and propose new ideas. This is the world we are building, one paper accepted by ICLR 2024 AGI workshop, **external demo for public will be released in 2024.**

Explore Heterophily, Oversmoothing in Graph Mind Group

Dec. 2023 – Present

Advised by Prof. Yujun Yan, Dartmouth College. Explore the heterophily and oversmoothing in graph neural networks, developing principles for graph-based ML models that are both expressive and generalizable. One **very novel paradigm** for graph neural networks. Currently, extending more comprehensive experiments and analysis. Aiming to submit to **NeurIPS 2024.**

Multi-relational Geometric Information Interaction Research Group

Apr. 2023 – Dec. 2023

Advised by Prof. Danyang Wu, Xi'an Jiaotong University. Research on the graph neural network, geometric representation learning and their applications for RNA, protein, and etc. **Two paper accepted, two papers are under review (denoted with *).**

- **(WWW 2024 Co-First author) FinReport: Explainable Stock Earnings Forecasting via News Factor Analyzing Model**
Propose an explainable stock earning framework via news factor analyzing model. Formalizing news into semantic graph and learn embedding of it as a more expressible factor. Stock earning foresting and explainable earning module are built via aggregating and utilizing news factor and numeric stock factor, achieving SOTA performance on A-stock dataset. A detailed report will be generated with the power of LLM.
- **(TNNLS 2024 Co-First author) Cross-view Approximation on Grassmann Manifold for Multiview Clustering**
Propose a novel framework for multi-view clustering that leverages the Grassmann manifold to approximate an orthogonal indicator matrix from multiple graph and feature views. Introduced adaptive weights to handle view inconsistency and devised an efficient algorithm to optimize the objective function. Demonstrated superior performance on four real-world datasets.
- ***(TKDE 2024 First author) NP²L: Negative Pseudo Partial Labels Extraction for Graph Neural Networks**
Propose an unsupervised framework to exact negative relationship between nodes, based on pseudo partial labels, and augment the original graph into a signed graph. Then signed graph neural networks are used for learning node embeddings, achieving SOTA performance on both link prediction and node classification tasks.
- **(WWW 2024 short Second author) Simple Multigraph Convolution Networks** Propose a simple but effective multigraph convolution networks, focusing on exacting edge-level and subgraph-level credible matrix to help cross-view interaction.

CTR/CVR Prediction& Smart Bidding Developing In Tec-Do

Sep. 2022 – Jan. 2024

Data Mining and Machine Learning Engineer in Tec-Do (Future Unicorn), BIC Business Intelligence Center.

Two patents granted, two patents published and two patents under review. Awarded as BEST Project of the year.

- Build CTR (Click-through rate)/CVR (Convert-through rate) predication model though GBDT and GNN for the Tec-Do Ads DSP platform, defeating 3rd party (Cusper) SaaS services. Achieved ROI 1.4x from 0.7x.
- Use DQN Implemented smart bidding module under constraint budget. 30% ROI improvement.
- Apply negative pseudo partial labels to sampling negative samples for recall model on bipartite graph.
- Build workflow generating real-person model picture for e-commerce based on diffusion model.

Multimodal Fitness Intelligent Guidance System

Sep. 2021 – Aug. 2022

Directed by Prof. Lin Shu. **Awarded Second Prize in the 2022 Intel Cup Undergraduate Electronic Design Contest.**

Build Mass Fitness Intelligent Guidance System based on Multimodal Physiological and Behavioral Data Fusion.

- Design high robustness gesture recognition and rating model using in gym environment.
- Design a recommendation algorithm based on community and user profiles with VGG.

SKILLS

- Programming Languages: Python, C++, SQL, JavaScript (TypeScript)
- Maths: Calculus, Linear Algebra, Complex Functions, Probability, Statistics, Convex Optimization, Graph Theory, Group Theory
- Platform&Tools: Windows, Linux; VSCode, Git, Docker, Nginx, L^AT_EX, Markdown
- Frameworks: Pytorch (Geometric), DGL, Stable Baselines3, Gym, Pandas (Polars), GBDTs (Catboost etc.), Visualization (Matplotlib, Seaborn, Plotly etc.), Scikit-learn, Numpy, Scipy, OpenCV, Flask, Pybind, Cython

1. Advertisement click prediction method based on random gradient attack, **CN116757748B**
2. Multi-task combined prediction method for few-sample advertisement, **CN117035873B**
3. Neural network system based on multidimensional data exploration, **CN117314531A**
4. Electronic commerce platform system based on artificial intelligence, **CN117196788A**
5. A method for patterned clothing fusion based on detail reconstruction, (*Under Review*)
6. An advertisement bidding system based on interval exploration, (*Under Review*)
7. An Ad Delay Conversion Prediction Method Based on Dual Window and Tree Modeling, (*Under Review*)

SELECTED AWARDS

<i>National Prize</i>	Highest scholarship awarded by the Chinese governments (< 0.1%)	2022
<i>TaiHu Innovation Prize</i>	Highest scholarship awarded by the WuXi city governments (1%)	2022
<i>1st Prize</i>	Baidu Paddle Paddle Cup (1%)	2021
<i>2nd Prize</i>	Intel Cup Undergraduate Electronic Design Contest (8%)	2022
<i>1st Prize</i>	First price scholarship awarded by SCUT (1%)	2022
<i>1st Prize</i>	First price scholarship awarded by EXCELLENCE GROUP(SCUT's Alumni) (1%)	2022
<i>Star Volunteer</i>	Volunteer title awarded by SCUT Youth Volunteer Guidance Center	2023
<i>Outstanding Student</i>	Title awarded by SCUT for Work-Study program	2023
<i>Outstanding Student</i>	Title awarded by Baidu Songguo Talent Training Elite Class	2023
<i>Project of the Year</i>	Awarded by Tec-Do for the DSP Platform	2023

RESEARCH INTEREST EXPLANATION

- **Graph Learning:** I am interested in the graph neural networks for various types' graph in deep learning filed, problems such as heterophily, invariant&equivalent in geometric, multi-relational, and etc. I am also interested in the graph learning, which includes graph signal processing, clustering on graph and graph unrolling in traditional machine learning or optimization filed. I love the expressive ability of graph and wild applications in real-world problems.
- **Quantitative Finance:** I am interested in the quantitative finance, which includes quantitative factors mining and analysis, stock prediction, explainable news factor analysis, and etc.
- **Automatous LLM Agent:** I am interested in the astomatous LLM agent, which includes the R&D process, real-world applications in non-toy problems such as fiance and etc. I care the most about LLM's ability in real-world complex scenarios.

ONGOING PROJECTS

- A novel Multimodal Graph Neural Networks for Singe Cells. SOTA performance achieved; Wonderful explainable ability in gene expression level. Currently, writing paper and preparing for submission to Nature Communications.
- A very novel paradigm for graph neural networks. Currently, extending more comprehensive experiments and analysis.
- A novel open problem and its benchmark for LLM. A novel evolving paradigm for this problem. Currently, enriching the benchmark and extending more comprehensive experiments and analysis.

CORE COURSES

- C++ Programming (4.0/4.0), Python Programming (4.0/4.0), Advanced Language Programming (4.0/4.0), Machine Learning (4.0/4.0), Deep Learning and Computer Vision (4.0/4.0), Big Data and Data Mining (4.0/4.0), Optimization Algorithms (4.0/4.0), Metaverse (4.0/4.0), Large Language Model (4.0/4.0), Design Thinking (4.0/4.0).

MISCELLANEOUS

- English Proficiency: CET4 Grade 581 CET6 Grade 576
- Hobby: Anime, Literature (Novels and Poem), Badminton, Running
- In my spare time, I have keen interests in visualization communication design and VI system.
- I am extremely enthusiastic about teaching with vivid visualization and frequent language interaction.