Shenghe Zheng

+86 13395909577 | shenghez.zheng@gmail.com

Research Interests: Swarm Intelligence, Neural Architecture Search

Education

Harbin Institute of Technology Aug. 2023 - Jan. 2026 (Expected)
Master Candidate in Massive Data Computing Lab, Computer Science *GPA: 93.88/100 Rank: 1/129*

Harbin Institute of Technology, Honors School (Top 5% in HIT) Aug. 2019 - Jun. 2023
B.E. in Computer Science *GPA: 91.18/100*

HONORS & AWARDS

National Scholarship (Top 1%)	Nov. 2024
Tencent Scholarship (Top 3%)	Jun. 2024
Outstanding Student of Harbin Institute of Technology (Top 5%)	May 2024
Outstanding Graduate of Harbin Institute of Technology (Top 5%)	Jun. 2023
Special Graduate Scholarship *2 (Top 10%)	Sept. 2023/2024
Honorable Award in American Mathematics Modeling Contest for College Students	Apr. 2021
Outstanding Student of Harbin Institute of Technology (Top 10%)	2019-2022
Renmin Scholarship (Top 10%)	2019-2022

PUBLICATIONS

- [**NeurIPS 2024**] **Zheng S**, Wang H, Liu X. IntraMix: Intra-Class Mixup Generation for Accurate Labels and Neighbors[J]. Advances in Neural Information Processing Systems, 2024, 37: 75095-75124.
- [**AAAI 2024**] **Zheng S**, Wang H, Mu T. DCLP: Neural Architecture Predictor with Curriculum Contrastive Learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(15): 17051-17059.
- [**ICDE 2023**] Mu T, Wang H, **Zheng S**, et al. TSC-AutoML: Meta-learning for Automatic Time Series Classification Algorithm Selection[C]//2023 IEEE 39th ICDE. IEEE, 2023: 1032-1044.
- [**VLDB 2021**] Mu T, Wang H, **Zheng S**, et al. Assassin: an automatic classification system based on algorithm selection[J]. Proceedings of the VLDB Endowment, 2021, 14(12): 2751-2754.

RESEARCH EXPERIENCES

Physical Reasoning Model Feb. 2025 - Now

- We propose that physical capability is crucial for models to understand the real world. However, our tests reveal that existing models exhibit weak physical abilities. Therefore, we develop a physical reasoning model by optimizing the entire pipeline—from data collection and cleaning to model training—to enhance the model's physical capabilities.
- Two paper is under review, and projects are in progress.

Efficient Knowledge Reuse Based on Model Merging Aug. 2024 - Now

- We propose leveraging existing models to efficiently construct a single model capable of solving multiple tasks through model merging, thereby reducing the storage and inference pressure associated with deploying large models.
- To efficiently align different fine-tuned models for merging, we discover the relationship between parameter frequency-domain information and task performance, enabling a filtering-based approach for efficient model merging.
- Two paper is under review, and other projects are in progress.

Graph Data Augmentation Sept. 2023 - Jan. 2024

- We harness the generation capability of augmentation methods to propose a plug-and-play graph augmentation method. It is designed to mitigate the prevalent challenge of sparse labels within graph datasets.
- To mitigate the noise inherent in the edges of graph data, we propose a simple yet impactful neighbor selection method for nodes. With this method, nodes generated as augmented samples can efficiently identify suitable neighbors, enhancing the information content of the graph.
- Paper was accepted by NeurIPS2024.

Neural Architecture Search Based on Self-Supervised Learning Sept. 2022 - Feb. 2023

- We utilize a self-supervised contrastive learning approach to train a neural network performance predictor using unlabeled data. Our objective is to accelerate the performance evaluation process in Neural Architecture Search (NAS) by directly predicting the performance of searched neural architectures using the neural predictor.

- To alleviate the training difficulty of contrastive learning in NAS, we introduce a novel curriculum-based method to guide contrastive training. This approach enhances the performance of the neural predictor in NAS.
- Paper published at AAAI 2024.

Automatic Algorithm Selection for Time Series Data Jul. 2021 - Mar. 2022

- We propose TSC-AutoML, a novel approach for automatic algorithm selection in Time Series Classification(TSC). It is the first attempt in TSC to leverage historical training data to autonomously select algorithms for new datasets.
- TSC-AutoML employs a reinforcement learning policy to measure the similarity between time series datasets, automatically recommending algorithms for time series datasets based on similarity.
- Paper published at ICDE 2023.

PROJECTS

Multi-level Automatic Model Ensemble based on Large Language Model Nov. 2023 - Now

Supervised by Prof. Hongzhi Wang Harbin, China

- In complex industrial scenarios, it is challenging for a single model to encompass all pertinent information. Multiple models are needed to collaboratively explore underlying mechanisms and enable informed decision-making.
- We propose a Model Ensemble approach using LLM, which extracts features from user input and applies AutoML techniques to optimize decisions based on these features and datasets.
- We propose a multi-level Model Ensemble method, considering the model ensemble as a graph construction process where each node represents a distinct model. By leveraging the distinctions and correlations among models, our objective is to achieve enhanced interpretability and predictive capability.

INTERNSHIP

Shanghai Artificial Intelligence Laboratory *Research Intern* Jan. 2025 - Now

Supervised by Dr. Peng Ye.

- Responsible for research on large model module reuse, multi-agent system, and reasoning model.

The Chinese University of Hong Kong *Research Intern* Apr. 2024 - Sept. 2024

Supervised by Prof. James Chen.

- Responsible for research on efficient deep learning methods for out-of-distribution generalization scenarios.

Massive Data Computing Lab, Harbin Institute of Technology *Research Intern* Sept. 2020 - Jun. 2023

Supervised by Prof. Hongzhi Wang.

- Reasearch on Model Reuse for Efficient AI.
- Reasearch on accelerating the evaluation process of Neural Architecture Search(NAS).
- Reasearch on Graph Data Augmentation for sparse labels on graph.
- Participation in Completion of National Natural Science Foundation of China.

SCIR, Harbin Institute of Technology *Research Intern* Aug. 2020 - Sept. 2021

Supervised by Prof. Wanxiang Che.

- Reasearch on Knowledge Distillation for Natural Language Processing.

State Key Laboratory of Robotics and Systems, HIT *Research Intern* Mar. 2020 - Jun. 2021

Supervised by Prof. Lianzheng Ge

- Research and Simulation of Flexible Robotic Arm Control.
- Research on Design and Control of Robot Cars.

Harbin Institute of Technology Baisi Tang Group *Lecturer* Sept. 2019 - Jun. 2020

- Responsible for delivering course lectures. During my tenure as a lecturer, the organization was recognized as the Top-10 Assistance Volunteer Organization at Harbin Institute of Technology for the 2019 - 2020 academic year.

HIT Intelligent Data Club *Director of the Academic Department* Oct. 2022 - Now

- Organized and prepared diverse technical sharing sessions and academic discussions, fostering opportunities for student research and providing a platform for learning and exchange.

TECHNICAL SKILLS

Program Languages: Python, Java, C/C++, SQL
Language: CET-6: 525, Mandarin (mother tongue)