

Shenghe Zheng

+86 13395909577 | shenghez.zheng@gmail.com

EDUCATION

Harbin Institute of Technology
Master Candidate in Computer Science

Aug. 2023 - Jun. 2025 (Expected)
Massive Data Computing Lab

Harbin Institute of Technology, Honors School(Top 5% in HIT)
B.E. in Computer Science

Aug. 2019 - Jun. 2023
GPA: 91.18/100

HONORS & AWARDS

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 (Top 10%)	Sept. 2023
Second Prize at the 14th Undergraduate Academic Forum, Harbin Institute of Technology	Nov. 2021
Honorable Award in American Mathematics Modelling Contest for College Students	Apr. 2021
Outstanding Student of Harbin Institute of Technology (Top 10%)	2019-2022
Renmin Scholarship (Top 10%)	2019-2022

PUBLICATIONS

- **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.
- Mu T, Wang H, **Zheng S**, et al. TSC-AutoML: Meta-learning for Automatic Time Series Classification Algorithm Selection[C]//2023 IEEE 39th International Conference on Data Engineering (ICDE). IEEE, 2023: 1032-1044.
- 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

Graph Data Augmentation Sept. 2023 - Jan. 2024
Supervised by Prof. Hongzhi Wang, Massive Data Computing Lab, HIT Harbin, China

- 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 is under-reviewed.

Neural Architecture Search Based on Self-Supervised Learning Sept. 2022 - Feb. 2023
Supervised by Prof. Hongzhi Wang, Massive Data Computing Lab, HIT Harbin, China

- 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
Supervised by Prof. Hongzhi Wang, Massive Data Computing Lab, HIT Harbin, China

- We propose a novel approach for automatic algorithm selection on the Time Series Classification (TSC) tasks called TSC-AutoML. It marks the first attempt in TSC to extract historical experiences from training data and select algorithms for new datasets autonomously.

- 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.

Automatic Machine Learning on Classification Tasks

Oct. 2020 - Mar. 2021

Supervised by Prof. Hongzhi Wang, Massive Data Computing Lab, HIT

Harbin, China

- Our system, named Assassin, automates the process of Algorithm Selection and Hyperparameter Optimization. Users are only required to upload the classification dataset, and Assassin can determine the optimal algorithm and hyperparameters accordingly.
- Experimental results confirm that Assassin is capable of selecting high-performance algorithms for new datasets.
- Paper published at VLDB 2021.

Recommendation System Based on Automated Machine Learning

Apr. 2021 - Sept. 2021

Supervised by Prof. Hongzhi Wang, Massive Data Computing Lab, HIT

Harbin, China

- We develop an automated machine learning evaluation approach that utilizes a performance predictor learning from training information to accelerate the performance evaluation of sequence recommendation(SR) models. Our objective is to predict the compatibility between SR models and tasks with minimal training epochs.

PROJECTS

Multi-level Automatic Model Ensemble based on Large Language Model(LLM)

Nov. 2023 - Now

Supervised by Prof. Hongzhi Wang

- In complex industrial scenarios, it is challenging for a single model to encompass all pertinent information. Multiple models are required to collaboratively delve into the underlying mechanisms of industrial systems, thereby facilitating informed decision-making processes.
- We propose a Model Ensemble approach based on LLM. Features are extracted from user input requirements by LLM, and AutoML techniques are then employed to make optimal decisions regarding Model Ensemble based on these requirement 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

Massive Data Computing Lab, Harbin Institute of Technology *Research Intern*

Sept. 2020 - Jun. 2023

Supervised by Prof. Hongzhi Wang

- Research on Automatic Machine Learning(AutoML) in various tasks.
- Research on accelerating the evaluation process of Neural Architecture Search(NAS).
- Research 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

- Research on Knowledge Distillation for Natural Language Processing.

State Key Laboratory of Robotics and Systems, HIT *Research Intern*

Mar. 2020 - Jun. 2021

Supervised by Assistant 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)

Developer Tools: Pycharm, VS Code, IntelliJ Idea

Technologies/Frameworks: Pytorch, Scikit-Learn, Linux, GitHub