Qin Shi

shi622@purdue.edu • shqii1j.github.io

Education

Purdue University — **Indiana, IN**
Doctor of Philosophy in Statistics — *2023–present*
GPA – 4.0/4.0

National University of Singapore — **Singapore**
Master of Statistics — *2022–2023*
GPA – 4.67/5.0

Shanghai University of Finance and Economics — **Shanghai, China**
Bachelor of Science in Data Science — *2018–2022*
GPA – 3.45/4.0, Ranking – 9/41

Publications

[1] Jiawei Du, **Qin Shi**, and Joey Tianyi Zhou. Sequential subset matching for dataset distillation. *Advances in Neural Information Processing Systems* (*NeurIPS*), 36, 2023.

Research & Work Experience

Purdue University — Department of Statistics — **West Lafayette, IN**
Research Assistant — *July 2024 - Present*
Mentor: Qifan Song, Joey Tianyi Zhou, Jiawei Du

- Applied data distillation to vision-language dataset and developed an efficient algorithm to improve Vision-Language tasks with better understanding.
- Researched understanding of LLMs from a statistical perspective and their security vulnerabilities.

A*STAR's Institute of High-Performance Computing — **Singapore**
Research Intern — *Sept. 2022 - June. 2023*
Mentor: Joey Tianyi Zhou, Jiawei Du

- Proposed a novel and generalized training scheme of data distillation (DD) to improve the performance of the gradients-matching-based DD method.
- Addressed the coupling issue within the synthetic dataset, which previously limited its ability in effectively condense additional high-level features.
- The proposed training scheme is verified to improve several gradient-matching-based DD methods and achieve state-of-the-art in various datasets.

YumChina — **Shanghai, China**
Data Scientist Intern — *Apr. 2022 - June. 2022*

- Built multi-objective recommendation system in a large-scale dataset and successfully applied it in practice. The model's off-line test set AUC reached 0.73.
- Built deep structured semantic model to generate combination-meal template based on user's food purchasing behavior.

MarcPoint — **Shanghai, China**
NLP Algorithm Engineer Intern — *Sep. 2021 – Jan. 2022*

- Proposed an algorithm to remove meaningless comments from social media. The algorithm could achieve 80% filter rate in a human-labeled test set and 4 min/million pieces of speed, which has been put into use.
- Built ACNN-CRF model to implement named entity recognition (NER) on social media data. Compared to Bert-CRF model, the running speed of this model was improved by 11%.
- Participated in the project of "MEMO: Rethinking Named Entity Recognition Evaluation Metrics, Datasets and Models". Proposed a new method of NER dataset partition and a new evaluation measure method for MER models.

Academic Services

Conference Reviewer: European Conference on Computer Vision (ECCV)

Other Skills

Technical Skills: Pytorch, Python, R-Programming, SQL-Server, C Language