Daize Dong

Personal Page | Google Scholar | dzdong2019@gmail.com

Summary

I am a graduate in Computer Science & Mathematics and Applied Mathematics from UESTC, now spending my gap year in Shanghai AI Lab and Westlake University. I am dedicated to uncovering the intrinsic properties of neural networks with theoretical guarantees, aiming to develop robust and effective AI systems. My research interests primarily revolve around:

1. Understanding the representation abilities (e.g. interpretability and robustness) and inherent statistical features (e.g. sparsity and scaling property) of neural networks.
2. Investigating neural networks' fundamental architectures to enhance efficiency (e.g. Mixture of Experts) and mathematical completeness (e.g. Graph Neural Networks).
3. Leveraging artificial intelligence to promote the scientific researches (e.g. AI for physics, biology, psychology).

Education

University of Electronic Science and Technology of China *Sep. 2019 – Jul. 2023*

Bachelor of Computer Science & Mathematics and Applied Mathematics *GPA: 3.91/4.00*

Research Experience

OpenGVLab, Shanghai Artificial Intelligence Laboratory *Jul. 2023 – Jul. 2024*

Research Assistant

Instructor: Dr. Xiaoye Qu. Supervisor: Prof. Yu Cheng

Mixture of Experts, Efficient Methods, Large Language Models

- Got a comprehensive understanding of the pipeline for pre-training and fine-tuning Large Language Models (LLMs).
- Explored efficient methods like Mixture of Experts (MoE) for training effective LLMs with low cost.
- Conducted research on enhancing the representation power of conditional networks like MoE.

Center for Artificial Intelligence Research and Innovation, Westlake University *Apr. 2023 – Present*

Research Assistant

Collaborator: Zhangyang Gao. Supervisor: Prof. Stan Z. Li

Molecular Generation, AI for Drug Discovery and Development

- Explored the strategies for 2D and 3D molecular representation learning and generation.
- Conducted research on the unified molecular modelling framework using pure transformers.

Data Intelligence Group, University of Electronic Science and Technology of China *Jul. 2022 – Mar. 2023*

Research Intern

Instructor: Prof. Wen Li

Domain Adaptation, Transfer Learning, Computer Vision

- Explored the theories and algorithms for unsupervised and self-supervised learning.
- Conducted research on knowledge transfer strategies for Multi-Target Domain Adaptation (MTDA).

Publications

1. **A Graph is Worth K Words: Euclideanizing Graph using Pure Transformer.** [Paper]
   Zhangyang Gao*, **Daize Dong***, Cheng Tan, Jun Xia, Bozhen Hu, Stan Z. Li.
   The 41st International Conference on Machine Learning (ICML 2024).
2. **iDAT: inverse Distillation Adapter-Tuning.** [Paper]
   Jiacheng Ruan, Jingsheng Gao, Mingye Xie, **Daize Dong**, Suncheng Xiang, Ting Liu, Yuzhuo Fu
   2024 IEEE International Conference on Multimedia and Expo (ICME 2024). ***(Oral)***
3. **PAD-Net: An Efficient Framework for Dynamic Networks.** [Paper]
   Shwai He, Liang Ding, **Daize Dong**, Boan Liu, Fuqiang Yu, Dacheng Tao.
   Proceedings of The 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023).
4. **SparseAdapter: An Easy Approach for Improving the Parameter-Efficiency of Adapters.** [Paper]
   Shwai He, Liang Ding, **Daize Dong**, Miao Zhang, Dacheng Tao.
   Findings of The 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).
5. **SD-Conv: Towards the Parameter-Efficiency of Dynamic Convolution.** [Paper]
   Shwai He, Chenbo Jiang, **Daize Dong**, Liang Ding.
   IEEE/CVF Winter Conference on Applications of Computer Vision, 2023 (WACV 2023)

* Equal Contribution

Preprints

1. **Demystifying the Compression of Mixture-of-Experts Through a Unified Framework.** [Paper] [Code]
   Shwai He*, **Daize Dong***, Liang Ding, Ang Li.
   Under Review by the Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024).
2. **ExFusion: Efficient Transformer Training via Multi-Experts Fusion.**
   Jiacheng Ruan, **Daize Dong**, Xiaoye Qu, Tong Zhu, Ting Liu, Yuzhuo Fu, Yu Cheng.
   Under Review by the Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024).
3. **DLO: Dynamic Layer Operation for Efficient Vertical Scaling of LLMs.**
   Zhen Tan*, **Daize Dong***, Xinyu Zhao, Jie Peng, Yu Cheng, Tianlong Chen.
   Under Review by The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).
4. **Dynamic Data Mixing Maximizes Instruction Tuning for Mixture-of-Experts.** [Paper] [Code]
   Tong Zhu, **Daize Dong**, Xiaoye Qu, Jiacheng Ruan, Wenliang Chen, Yu Cheng.
   Under Review by The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).
5. **Blending and Aggregating the Target for Blended-Target Domain Adaptation.**
   Tong Chu, **Daize Dong**, Jinhong Deng, Lixin Duan, Wen Li.
   Under Review by IEEE Transactions on Image Processing (IEEE-TIP).

Projects

LLaMA-MoE: Building Mixture-of-Experts from LLaMA with Continual Pre-training. [Paper] [Code] *Jul. 2023 – Dec. 2023*
Under Review by The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).

- Conducted research on the framework to integrate the mixture-of-experts (MoE) structure into existing LLMs.
- Explored multiple methods to initialize the converted MoE model using pretrained parameters from the LLM.
- Proposed a novel random split strategy with output-scaling to recover model performance.

Technical Skills

Natural Languages: English (TOEFL 100), Mandarian.
Programming Languages: Python, C/C++, Java, Matlab, etc.
Deep Learning Tools: PyTorch, Hugging-Face Transformers, Torch-Lightning, DeepSpeed, etc.

Relevant Courses

Deep Learning: Machine Learning, Artificial Intelligence, Deep Learning for Computer Vision, Deep Learning for Natural Language Processing, Knowledge Representation and Reasoning, Data Mining and Big Data Analysis.
Optimization Algorithm: Optimization Theory and Methods, Introduction to Algorithms.
Mathematics: Differential Calculus, Linear Algebra, Probability Theory, Stochastic Process, Discrete Mathematics, Graph Theory, Multivariate Statistical Analysis, Causal Inference.
Computer Science: Computer Organization and Architecture, Compiler Principles, Computer Operating Systems, Database Principles and Applications, Information Retrieval, Software Engineering.