Mingyu Xu

+86 18801200516
1999.05-
Apply for a deep learning researcher
xumingyu16@gmail.com

Education

Institute of Automation, Chinese Academy of Sciences Sep. 2021 – Jul. 2024
Master of Engineering, Pattern Recognition and Intelligent Systems

School of mathematical sciences, Peking University Sep. 2017 – Jul. 2021
Bachelor of Science, Computational Mathematics

Experience

Pretrained engineer Nov.2023 – Mar.2025
Baichuan Inc.

- Various hybrid structures to reduce the use of KV cache without significantly reducing performance.
- Study the relationship between position encoding and context length.
- Lightly increase the model depth of the transformer to enhance its capabilities.

AGI Researcher Mar.2025 – Present
ByteDance Seed Edge

- Propose Deltaformer that breaks through the TC^0 expressive power limitation of Transformers.

Selected Publication

Before ChatGPT came out, I conducted research in machine learning and published some papers on out of distribution detection[7] and weakly supervised learning[5,6] and win in some workshop[8,9]. After ChatGPT came out, I discovered the dawn of general artificial intelligence and began to try to understand the backbone of large language models and improve it[3,4]. At first, I made limited improvements to the depth of the transformer[2]. Recently, I have been trying to gain a more thorough understanding and improvement of transformers from the theoretical complexity and propose Deltaformer[1].

[1] **[Arxiv 2025]** *Understanding Transformer from the Perspective of Associative Memory.* Shu Zhong*, **Mingyu Xu***, Tenglong Ao* et al.

[2] **[ICML2025]** *KV Shifting Attention Enhances Language Modeling.* **Mingyu Xu** et al.

[3] **[ACL2025 Findings]** *ShortGPT: Layers in Large Language Models are More Redundant Than You Expect.* Xin Men*, **Mingyu Xu***, Qingyu Zhang*, Qianhao Yuan* et al.

[4] **[NeurIPS2024]** *Base of RoPE Bounds Context Length.* **Mingyu Xu***, Xin Men*, Bingning Wang* et al.

[5] **[ICASSP2024]** *Pseudo Labels Regularization for Imbalanced Partial-Label Learning.* **Mingyu Xu** et al.

[6] **[NeurIPS2023]** *ALIM: adjusting label importance mechanism for noisy partial label learning.* **Mingyu Xu***, Zheng Lian* et al.

[7] **[NeurIPS2023]** *VRA: variational rectified activation for out-of-distribution detection.* **Mingyu Xu** et al.

[8] **[ACMMM2023workshop]** *Humor detection system for MuSE 2023: contextual modeling, pesudo labelling, and post-smoothing* **Mingyu Xu** et al.

[9] **[ACMMM2021workshop]** *Multimodal Emotion Recognition and Sentiment Analysis via Attention Enhanced Recurrent Model.* Licai Sun*, **Mingyu Xu*** et al.

Awards

Gold medal, Chinese Mathematical Olympia 2016 year

1st, ACMMM2021 Workshop MuSe 2021 year

1st, ACMMM2024 Workshop MuSe 2024 year

Other Things about me.

Python is my main programming language, I can also write a some **Triton** to efficiently implement Deltaformer which I am working on. If you are interested in Deltaformer, you can listen to my talk on the **Asap Seminar** website. I am an **INTP** according to MBTI. From a longer-term perspective, I hope that something can make all humanity live better and I realized that artificial intelligence has this potential. The emergence of ChatGPT makes me feel that AGI can be achieved in a near future. Every choice we make is about what kind of AGI can ultimately become.