Zhen Yang

Tsinghua University, Beijing, PR China

+86 13070106218   yangz21@mails.tsinghua.edu.cn

HomePage Zhen   *Google Scholar* Google Scholar

EDUCATION

2021.09-2024.12   **Ph.D.**, Knowledge Engineering Group (KEG), Department of Computer Science and Technology, Tsinghua University. Supervisors: Prof. Jie Tang.

2016.09-2019.07   **M.Phil.**, School of Integrated Circuit, Tsinghua University. Supervisors: Prof. Tian-Ling Ren.

2012.08-2016.07   **B. Eng.**, Electronic Packaging Technology, Xidian University.

RESEARCH INTERESTS

My researches focus on **Mathematical Language Models** and **Graph Representation Learning**. My research goal is to enhance the reasoning of Large Language Models (LLMs) and Multi-Modal Large Language Models (MLLMs) to solve complex mathematical and scientific problem-solving challenges.

– **Mathematical reasoning** in LLMs and MLLMs [MathGLM][MathGLM2][MathGLM-Vision]

– **Scientific Reasoning Benchmark** in MLLMs [VisScience]

– **Graph Representation Learning** [MCNS] [STAM] [RecNS] [BatchSampler] [TriSampler] [MixGCF]

– **Vision Language Pretraining** [ViLTA]

I have accomplished 12 papers, and have published 7 first/co-first authored papers in top-tier conferences and journals(such as 2 KDD, WWW, ICCV, AAAI, TKDE and TPAMI).

AWARD

• Huawei Scholarship, 2023

• Outstanding Thesis Award, Tsinghua University, 2019 (Top 3%)

• Outstanding Graduate of Beijing, 2019 (Top 3%)

• Candidate of Special Scholarship of Tsinghua University, 2018 (Top 30)

• National Scholarship, 2018 (Top 3%)

• 129 Scholarship of Tsinghua University, 2018 (Top 30)

• Outstanding Graduate of Xidian University, 2016

• National Encouragement Scholarship, 2015

• National Scholarship, 2014

PROFESSIONAL SERVICE

• PC Member for Conferences: WWW2023.

• Reviewer for Journals: TKDE.

• Reviewer for Conferences: ICLR2025, WWW2025, AAAI2025, WWW2024, WWW2023, KDD2023.

RESEARCH EXPERIENCE

VisScience: Multi-Modal Scientific Reasoning Benchmark

- Construct a comprehensive benchmark VisScience to evaluate multi-modal scientific reasoning across three disciplines: mathematics, physics, and chemistry.

- Collect 3,000 questions derived from K12 education, evenly distributed with 1,000 questions per discipline. Questions in VisScience cover 21 different topics and are divided into 5 difficulty levels to provide a wide range of topics within each discipline.

- Present a detailed evaluation of the performance of 25 representative MLLMs on scientific reasoning. These results highlight the strengths and limitations of MLLMs and suggest areas for future improvements.

MathGLM-Vision: Mathematical Multi-Modal Large Language Model

- Construct a fine-tuning dataset MathVL, based on which supervised fine-tuning (SFT) is performed on CogVLM-32B and CogVLM2 to develop a family of math MLLMs called MathGLM-Vision.

- Experiments show that 40% to 65% performance improvements are achieved on MathVista-GPS. In addition, the effectiveness of MathGLM-Vision is evaluated on the MathVL-test dataset.

- MathGLM-Vision not only surpasses their backbone models on specialized math tests, but also retains generalization capabilities in the field of general visual language understanding.

MathGLM2: Data-enhanced Mathematical Large Language Model

- The ability of other large language models to solve Chinese math problems is far lower than GPT-4. How to improve the mathematical reasoning ability of current large language models?

- Focus on using the core idea of Chain-of-Thought and self-trained scoring models to construct high-quality mathematical data, and achieve slightly better results than GPT-4 on MathGLM2-6B.

- However, we found that the generalization ability of MathGLM2-6B is not outstanding, and it tends to show better performance on the trained dataset. GPT-4 has high generalization and shows the strongest performance on various types of math problems.

MathGLM: GPT Can Solve Math Problems Without a Calculator

- A common misconception is that large language models cannot accurately perform high-digit arithmetic operations, especially operations involving multiplication of more than 8 digits and operations involving decimals and fractions.

- This work aims to challenge this misconception. Leveraging the step-by-step strategy, the 2 billion-parameter MathGLM model can accurately perform multi-digit arithmetic operations with an accuracy of almost 100%.

- The MathGLM-10B model trained with the step-by-step strategy can achieve performance close to GPT-4 on 5,000 Chinese mathematics test data.

ViLTA: Enhancing Vision-Language Pre-training through Textual Augmentation

- For masked language modeling (MLM), a cross-distillation method is proposed to generate soft labels to enhance the robustness of the model, thereby alleviating the problem of treating synonyms of masked words as negative samples in one-hot labels.

- For image-text matching (ITM), the current language encoder is utilized to synthesize hard negative samples according to the context of the language input to encourage the model to learn high-quality representations.

- ViLTA can achieve better performance on various visual language tasks, enhancing model robustness and accelerating model convergence.

PUBLICATION

Authored 12 research papers, with first-author credit on 7 CCF-A ranked publications; Accumulated over 2500 citations on Google Scholar.

1. **Zhen Yang***, Ming Ding*, Chang Zhou, Hongxia Yang, Jingren Zhou, Jie Tang. Understanding Negative Sampling in Graph Representation Learning. **KDD'20. (CCF A)**

2. **Zhen Yang**, Ming Ding, Xu Zou, Jie Tang, Bin Xu, Chang Zhou, Hongxia Yang. Region or Global? A Principle for Negative Sampling in Graph-based Recommendation. **TKDE'22. (CCF A)**

3. **Zhen Yang**, Ming Ding, Bin Xu, Hongxia Yang, Jie Tang. STAM: A Spatiotemporal Aggregation Method for Graph Neural Network-based Recommendation. **WWW'22. (CCF A)**

4. **Zhen Yang**, Tinglin Huang, Ming Ding, Yuxiao Dong, Zhitao Ying, Yukuo Cen, Yangliao Geng, Jie Tang. BatchSampler: Sampling Mini-Batches for Contrastive Learning in Vision, Language, and Graphs. **KDD'23. (CCF A)**

5. **Zhen Yang**, Shao Zhou, Yuxiao Dong, Jie Tang. TriSampler: A Better Negative Sampling Principle for Dense Retrieval. **AAAI'24. (CCF A)**

6. **Zhen Yang**, Ming Ding, Tinglin Huang, Yukuo Cen, Junshuai Song, Bin Xu, Yuxiao Dong, Jie Tang. Does Negative Sampling Matter? A Review with Insights into its Theory and Applications. **TPAMI'24 (CCF A)**.

7. Weihan Wang*, **Zhen Yang***, Bin Xu, Juanzi Li, Yankui Sun. ViLTA: Enhancing Vision-Language Pre-training through Textual Augmentation. **ICCV'23. (CCF A)**

8. Tinglin Huang, Yuxiao Dong, Ming Ding, **Zhen Yang**, Wenzheng Feng, Xinyu Wang, Jie Tang. Mixgcf: An improved training method for graph neural network-based recommender systems. **KDD'21. (CCF A)**

9. **Zhen Yang**, et al. GPT Can Solve Mathematical Problems Without a Calculator. (arXiv)

10. **Zhen Yang**, et al. MathGLM-Vision: Solving Mathematical Problems with Multi-Modal Large Language Model. (arXiv)

11. Zhihuan Jiang*, **Zhen Yang***, et al. VisScience: An Extensive Benchmark for Evaluating K12 Educational Multi-modal Scientific Reasoning. (arXiv)

12. GLM Team, Chatglm: A family of large language models from glm-130b to glm-4 all tools. (arXiv)

PERSONAL SKILLS

Languages & Frameworks: Python, C/C++; Pytorch, Tensorflow

Distributed Training and Inference Frameworks: Megatron, DeepSpeed; Slurm Job Scheduling System

REFERENCES

Prof. Jie Tang (PhD supervisor).

Department of Computer Science and Technology, Tsinghua University, China.

jietang@tsinghua.edu.cn.

Prof. Tian-Ling Ren (PhD supervisor).

School of Integrated Circuit, Tsinghua University, China.

RenTL@tsinghua.edu.cn.