

# 杨珍

手机：(+86) 13070106218 · 邮箱：yangz21@mails.tsinghua.edu.cn



## 研究兴趣

- 预训练模型，多模态预训练，大模型数学推理
- 图表示学习，对比学习，稠密检索

## 教育背景

|                    |                   |
|--------------------|-------------------|
| 清华大学，计算机科学与技术系，博士  | 2021.09 - 至今      |
| • 导师：唐杰            |                   |
| 清华大学，集成电路工程，硕士（保送） | 2016.09 - 2019.07 |
| • 导师：任天令           |                   |
| 西安电子科技大学，电子封装技术，本科 | 2012.09 - 2016.07 |

## 科研经历

大模型推理方向 – **Mathematical Reasoning** 2023.1 – 至今

### *VisScience*: 多模态科学数据基准

- 多模态大语言模型 (MLLM) 通过集成文本和视觉信息以实现复杂场景中的视觉理解，在各种任务中展示了有前景的能力。尽管有几个基准旨在评估从视觉问答到复杂问题解决等任务中的 MLLM，但大多数基准主要集中在数学或一般视觉理解任务上。这揭示了当前基准中的一个关键差距，这些基准常常忽视了物理和化学等其他关键科学学科的纳入。
- 为了弥补这一差距，我们精心构建了一个名为 *VisScience* 的综合基准，用于评估跨数学、物理和化学三个学科的多模态科学推理。该基准包含源自 K12 教育的 3,000 个问题，均匀分布在每个学科 1,000 个问题。*VisScience* 中的问题涵盖 21 个不同的主题，并分为五个难度级别，在每个学科内提供广泛的主题。
- 通过 *VisScience*，我们对 25 个具有代表性的 MLLM 在科学推理方面的表现进行了详细评估。实验结果表明，闭源 MLLM 通常优于开源模型。观察到的最佳性能包括 Claude3.5-Sonnet 的数学准确率 53.4%、GPT-4o 的物理准确率 38.2%、Gemini-1.5-Pro 的化学准确率 47.0%。这些结果强调了 MLLM 的优势和局限性，提出了未来改进的领域，并强调了开发能够有效处理多模态科学推理的不同需求的模型的重要性。

### *MathGLM-Vision*: 多模态数学大模型

- 大语言模型 (LLM) 在数学推理方面表现出了显着的能力，特别是在基于文本的数学问题上。然而，当前的多模态大语言模型 (MLLM)，尤其是那些专门研究数学的模型，往往主要集中于解决几何问题，而忽略了其他数学领域中可用的视觉信息的多样性。此外，这些专门的数学 MLLM 的几何信息源自多个公共数据集，这些数据集的多样性和复杂性通常受到限制。
- 我们专注于构建一个名为 *MathVL* 的微调数据集，基于此数据集，在 CogVLM-32B 和 CogVLM2 上进行监督微调 (SFT) 来开发一系列称为 *MathGLM-Vision* 的数学 MLLM。实验表明在 *MathVista-GPS* 上实现了 40 65% 性能提升。此外，在 *MathVL-test* 数据集上评估了 *MathGLM-Vision* 的有效性。结果表明，该模型不仅在专门的数学测试中超越了它们的骨干模型，而且还保留了一般视觉语言理解领域的泛化能力。

### *MathGLM2*: 数据增强的数学大模型

- 尽管目前最强的大语言模型 GPT-4 在英文数学问题 GSM8K 上取得了 92% 的准确率，但其在 K9 的中文数学问题的准确率仍然只有 67%，远未达到人类解决数学问题的能力。而其他的大语言模型解决中文数学问题的能力则远低于 GPT-4，如何提升现有大语言模型的数学推理能力，实现接近甚至超越 GPT-4?

- 我们专注于借助 CoT 的解题过程及自训的打分模型，构造了一批高质量数学数据，并在 MathGLM2-6B 上实现了微超 GPT-4 的结果。
- 然而，我们发现 MathGLM2-6B 的泛化能力并不突出，更倾向于在训练过的数据集上表现出更佳的性能。而 GPT-4 则具备高泛化性，其在各种类型的数学题目上均表现出最强性能。

### MathGLM: GPT 无需计算器即可解决数学问题

- 一个常见的误解便是大语言模型并不能精确执行高位数的算数运算，尤其是涉及超过 8 位数字乘法的运算以及涉及小数、分数的运算。此外，目前大语言模型对中文数学问题的解决能力尚未得到充分探索，一个关键性的问题便是大语言模型能解决什么难度的中文数学问题？
- 此项工作旨在挑战这种误解。在训练数据充足的情况下，20 亿参数的 MathGLM 模型能够准确地执行多位算术运算，准确率几乎可以达到 100%，其结果显著超越最强大语言模型 GPT-4 在相同测试数据上 18.84% 的准确率。
- 以 GLM-10B 为基座模型训练得到的 MathGLM-10B 模型在 5000 条中文数学的测试数据上实现了接近 GPT-4 的性能。

### 稠密检索方向

2022.02 – 2023.01

#### TriSampler: 一种更好的用于密集检索的负采样原理

AAAI'24

- 负采样是密集检索中的关键技术，对于检索模型的有效训练至关重要，并对检索性能产生重大影响。虽然现有的负采样方法通过利用 hard 负例取得了值得称赞的进展，但该领域尚未建立全面解决负例选构建和负采样分布设计的指导原则。
- 此项工作着手于密集检索中负采样的理论分析。这一探索揭示了准三角原理，阐明了查询 (query)、正文档 (positive document) 和负文档 (negative document) 之间的三角形关系。
- 在这一指导原则的推动下，提出了 TriSampler，一种简单而高效的负采样方法。TriSampler 的关键在于它能够在规定的约束区域内选择性地采样更多信息的负样本。实验评估表明，TriSampler 在各种代表性检索模型中始终获得卓越的检索性能。

### 对比学习方向

2022.02 – 2023.01

#### BatchSampler: 视觉、语言、图上的对比学习中的小批量采样方法

KDD'23

- 批量对比学习是一种最先进的自监督方法，它可以使语义相似的实例接近，同时在小批量内将不同的实例分开，其成功的关键是负共享策略。然而，现有的工作主要集中在加载每个小批量实例后设计更好的负采样技术，如 HCL 和 DCL。
- 此项工作建议从输入数据中全局采样实例以形成小批量。目标是让小批量自然地包含尽可能多的 hard 负例，这些负例拥有和正样本不同的标签，是真负例，但与正样本有相似表示。基于以上的研究动机，提出了 BatchSampler，一种简单且通用的小批量采样技术，以提供 hard 的真实例的小批量样本，主要包括两个技术：邻近图构建和邻近图采样。
- 实验表明，BatchSampler 可以持续提高对比模型的性能，在视觉领域 ImageNet-100 上的 SimCLR；语言领域 STS 上的 SimCSE；图领域上的 GraphCL 和 MVGRL 均取得不同程度的性能提升。

### 多模态预训练方向

2022.02 – 2023.01

#### ViLTA: 通过文本增强增强视觉语言预训练

ICCV'23

- 视觉语言预训练 (VLP) 方法的关键目标是通过基于 Transformer 的架构联合学习视觉和文本特征。现有技术通常关注如何对齐视觉和文本特征，但对提高模型鲁棒性和加速模型收敛的策略探索不足。
- 此项工作提出了一种新方法 ViLTA，由两个组件组成，以进一步促进模型学习图像文本对之间的细粒度表示。对于掩码语言建模 (MLM)，提出了一种交叉蒸馏方法来生成软标签以增强模型的鲁棒性，从而缓解了将掩码词的同义词视为单热标签中的负样本的问题。对于图像文本匹配 (ITM)，利用当前的语言编码器根据语言输入的上下文合成 hard 负样本，以此来鼓励模型学习高质量的表示。
- 实验表明，ViLTA 可以在各种视觉语言任务上取得更好的性能。对基准数据集的大量实验证明了 ViLTA 的有效性及其在视觉语言预训练方面的巨大潜力。

### 图表示学习方向

2019.09 – 2022.02

#### STAM: 一种基于 GNN 推荐的时空聚合方法

WWW'22

- 近年来，基于图神经网络的推荐系统蓬勃发展，其核心是邻居聚合方法。现有技术通常侧重于如何从空间结构信息的角度聚合信息，但对邻居的时间信息探索不足。
- 此项工作提出了一种时空聚合方法 STAM，以有效地将时间信息整合到邻居嵌入学习中。STAM 从空间结构信息和时间信息的角度生成时空邻居嵌入，促进了从空间到时空的聚合方法的发展。STAM

利用 Scaled Dot-Product Attention 来捕获一跳邻居的时间顺序，并利用多头注意力在不同的潜在子空间上执行联合注意力。

- 实验表明，与基于空间的聚合方法相比，STAM 对基于 GNN 的推荐带来了显著的改进，例如，在 MRR@20 方面，MovieLens 为 24%，亚马逊为 8%，淘宝为 13%。

**RecNS: 基于图的推荐中的负采样原则**

TKDE'22

- 基于图的推荐的一个基本挑战是在用户-商品图中只存在观察到的正用户-商品对。负采样是解决一类问题的重要技术，广泛应用于许多推荐方法中。然而，之前的工作只关注负采样分布的设计，而忽略了负采样的采样区域。
- 此项工作提出了三区域原则来指导负采样，原则认为应该在“中间区域”负采样更多的商品，而不应采样相邻和远距离的商品。基于此原则，提出了一种通用的负采样方法 RecNS，设计有两种采样策略：正辅助采样和曝光增强采样。
- 实验表明，RecNS 方法在阿里巴巴数据集上的 Recall@20 显著优于所有负采样基线，例如在 PinSage 上增益为 10.47%，在 NGCF 增益为 6.02%，在 LightGCN 增益为 8.20%。

**MCNS: 图表示学习中负采样的理解与研究**

KDD'20

- 图表示学习为各种网络生成连续的向量表示，但是将高质量的向量表示推向大型节点集的有效性和效率性方面仍具有挑战。采样是实现性能目标的关键点。现有技术通常集中于对正节点进行采样，而负采样策略则没有得到足够的探索。
- 此项工作从目标函数和方差两个角度系统地分析了负采样的作用，从理论上第一次证明了负采样与正采样在确定优化目标和估计方差方面同等重要。在理论的指导下，提出 MCNS，用自对比近似估计正采样分布，用 Metropolis-Hastings 加速负采样过程。
- 在 5 个数据集上评估了 MCNS，这些数据集涵盖了 19 个实验设置，涵盖了广泛的下游图表示学习任务，包括链路预测，节点分类和个性化推荐。实验结果证明了其稳健性和优越性。

## 发表论文

- **Zhen Yang\***, Ming Ding\*, Chang Zhou, Hongxia Yang, Jingren Zhou, Jie Tang. Understanding Negative Sampling in Graph Representation Learning. **KDD'20. (CCF A)**
- **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)**
- **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)**
- **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)**
- **Zhen Yang**, Shao Zhou, Yuxiao Dong, Jie Tang. TriSampler: A Better Negative Sampling Principle for Dense Retrieval. **AAAI'24. (CCF A)**
- **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)**.
- Weihan Wang\*, **Zhen Yang\***, Bin Xu, Juanzi Li, Yankui Sun. ViLTA: Enhancing Vision-Language Pre-training through Textual Augmentation. **ICCV'23. (\* 共同一作). (CCF A)**
- 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)**
- **Zhen Yang**, et al. GPT Can Solve Mathematical Problems Without a Calculator. (arXiv)
- **Zhen Yang**, et al. MathGLM-Vision: Solving Mathematical Problems with Multi-Modal Large Language Model. (arXiv)
- Zhihuan Jiang\*, **Zhen Yang\***, et al. VisScience: An Extensive Benchmark for Evaluating K12 Educational Multi-modal Scientific Reasoning. (arXiv)

## 学术服务

- 会议程序委员、审稿人：WWW2024, WWW2023, KDD2023, TKDE 等

## 个人荣誉

---

- 博士：清华之友-华为奖学金
- 硕士：北京市优秀毕业生、清华大学优秀毕业论文、研究生国家奖学金、清华大学一二九奖学金
- 本科：国家奖学金、国家励志奖学金、校优秀毕业生

## 参考

---

- 唐杰 清华大学 jietang@tsinghua.edu.cn