

朱裕昌



- 性别：男
- 研究方向：多模态大模型、大语言模型、图学习公平性
- 籍贯：广东梅州
- 邮箱：zhuych27@mail2.sysu.edu.cn
- 出生年月：1997.4
- 电话：13416486396

教育背景

- 2022.09 – 2026.6 (预计) 中山大学 计算机技术 博士 导师：陈亮
论文发表：一作发表 ICML、KDD 等会议/期刊论文 5 篇（详细列表见第 2 页）；获奖：国家奖学金、校长奖学金
- 2019.09 - 2021.06 华南农业大学 车辆工程 硕士 导师：赵祚喜（罗锡文院士团队）
论文发表：学生一作发表 SCI 论文 2 篇；获奖：校优秀硕士毕业论文、研究生一等奖学金
- 2015.09 - 2019.06 华南农业大学 包装工程 本科 平均绩点 4.05 (排名 4/59)
获奖：优秀本科毕业生、优秀本科毕业论文、广东省优秀共青团员、国家励志奖学金、校一等奖学金

实习经历

- 2024.03 – 2025.5 Tencent AI Lab 机器智能组 科研实习生
研究方向及产出：实习期间开展大语言模型合成数据、多模态大模型相关的研究，产出相关学术论文 1 篇、专利 3 项。
1) 在大模型合成数据方面，围绕合成数据质量评估问题，提出从分类角度衡量大语言模型合成数据的多样性，计算速度提升 15% 左右；围绕合成数据误用、滥用问题，研究鲁棒的大语言模型生成内容水印方法，提出基于语义哈希的生成内容水印方法。
2) 在多模态大模型方面，围绕现有多模态生成受限于 Tokenizer 的问题，提出使用局部敏感哈希的方式实现 lookup-free 向量量化，进而构建聚焦于语义信息的 Tokenizer，实现了高特征维度下 100% 的码本使用率。

研究课题

- 多模态大模型相关组件及应用研究
针对多模态大模型理解与生成问题，研究多模态大模型相关组件（Tokenizer）以及多模态大模型的下游任务应用。
1) 在多模态生成方面，围绕多模态生成中非文本的生成受限于 Tokenizer 的问题，研究兼顾生成与重建任务的 Tokenizer，提出使用局部敏感哈希的方式实现 lookup-free 向量量化并基于此实现聚焦于语义信息的 Tokenizer，目前已实现高特征维度下 100% 的码本使用率，同时保证 SOTA 相当的重建性能（进行中）。
2) 在多模态理解方面，围绕多模态服务数据监管复杂的问题，研究多模态服务数据合规检测、异常检测、质量监测，引入检索增强生成、有监督微调、思维链等技术提升模型在复杂推理中的性能。
- 大语言模型合成数据研究
针对大语言模型生成合成数据后可能产生的不良影响，研究大语言模型合成数据的评估、影响、管理。
1) 在合成数据评估方面，围绕合成数据的多样性缺乏合理评估方案，提出从分类角度衡量大语言模型合成数据的多样性，保证评估准确性的同时计算速度提升约 15%。
2) 在合成数据影响方面，围绕合成数据的多样性对下游任务模型影响不明，实验验证了合成数据多样性的正向作用，但是需要避免分布偏移的影响。
3) 在合成数据管理方面，围绕合成数据可能存在误用、滥用的问题，提出基于语义哈希的鲁棒生成内容水印方法。
- 图学习公平性研究
针对图学习算法（图神经网络）公平性问题，研究从数据、模型、训练等方面提升图神经网络决策的公平性。
1) 在数据方面，围绕对于图结构以及消息传递机制放大训练数据中的偏见问题，提出通过夏普利值来衡量边对公平的贡献，以此实现数据中边权重的预处理，实现公平的消息传递。
2) 在模型方面，围绕公平性与下游任务性能权衡的问题，通过敏感属性解离避免任务相关信息的删除，同时消除敏感属性对模型决策的影响，在下游任务性能无下降的情况下实现超越 SOTA 的公平性能。
3) 在训练方面，围绕敏感属性未知、多敏感属性场景下的公平问题，提出基于知识蒸馏、不变性学习的公平图神经网络训练框架，在敏感属性未知、多敏感属性场景下实现超越 SOTA 的公平性能。

个人总结

- 具有顶会论文发表经验，熟悉团队合作的方式开展科研，参与过多模态大模型、大语言模型相关的项目
- 熟悉 Python 语言、Pytorch 深度学习框架，具有多学科交叉知识背景
- 熟悉 Transformers、TRL、PEFT 等大语言模型相关的库

❖ 已发表

[1]

Measuring Diversity in Synthetic Datasets. **Yuchang Zhu**, Huizhe Zhang, Bingzhe Wu, Jintang Li, Zibin Zheng, Peilin Zhao, Liang Chen, Yatao Bian. 2025

ICML

[2]

One Fits All: Learning Fair Graph Neural Networks for Various Sensitive Attributes. **Yuchang Zhu**, Jintang Li, Yatao Bian, Zibin Zheng, Liang Chen. 2024

KDD, oral

[3]

Fair Graph Representation Learning via Sensitive Attribute Disentanglement. **Yuchang Zhu**, Jintang Li, Zibin Zheng, Liang Chen. 2024

WWW, oral

[4]

The Devil is in the Data: Learning Fair Graph Neural Networks via Partial Knowledge Distillation. **Yuchang Zhu**, Jintang Li, Liang Chen, Zibin Zheng. 2024

WSDM

[5]

FairAGG: Towards Fair Graph Neural Networks via Fair Aggregation. **Yuchang Zhu**, Jintang Li, Liang Chen, Zibin Zheng. 2024

TCSS 期刊

[6]

A Smartphone-Based Six-DOF Measurement Method With Marker Detector. **Yuchang Zhu**, Yuan Huang, Yuanhong Li, Zhi Qiu, Zuoxi Zhao. 2022

TIM 期刊

❖ Preprint

[1]

What Matters in LLM-generated Data: Diversity and Its Effect on Model Fine-Tuning. **Yuchang Zhu**, Huazhen Zhong, Qunshu Lin, Haotong Wei, Xiaolong Sun, Zixuan Yu, Minghao Liu, Zibin Zheng, Liang Chen. 2025

arXiv

[2]

SaGIF: Improving Individual Fairness in Graph Neural Networks via Similarity Encoding. **Yuchang Zhu**, Jintang Li, Zibin Zheng, Liang Chen. 2025

arXiv