

庞可涵的个人简历

+86 132 6451 8388 pangkehan@buaa.edu.cn 北京市

教育经历

北京航空航天大学 计算机学院 软件工程 博士研究生 2022.09 – 至今

- 导师: Wenfei Fan 院士、陆平 副教授
- 研究方向: Graph Data Mining, Graph Knowledge Reasoning

北京邮电大学 计算机学院 计算机科学与技术 工学学士 2018.09 – 2022.07

- GPA: 3.7/4.0 Rank: 7.6%

工作经历

深圳计算科学研究院 基础研究部 实习研究员 2023.07 – 至今

- 在担任该职位期间, 本人主要研究图数据挖掘、图数据质量、图数据增强、GNNs (图神经网络) 可解释性、图数据计算系统架构设计以及大规模图数据上LLMs (大语言模型) 的低资源场景适配、数据处理应用等。在任职期间, 已成功产出多篇学术论文。

ICDE External Reviewer 2024 Second Round

科研经历

《Making It Tractable to Detect and Correct Errors in Graphs》(已录用待刊出)

- Wenfei Fan, Kehan Pang, Ping Lu and Chao Tian. (TODS) 2023
- 为解决静态图上的关联规则在时序图上语义存在欠缺的问题, 本人在该篇论文提出了一类对时序图上的节点进行实时性排序的模型, 以及将静态图上的关联规则适配至动态图的方法, 显著增强了在数据动态更新场景下图关联规则对过时数据、错误数据检测的性能

《Global and Local Explanations for Negative GNN Predictions》(在投)

- Kehan Pang, DanDan Lin and Wenfei Fan.
- 为以人类可感知方式解释GNNs在电商网络、金融数据中检测到的风险用户、欺诈行为, 并解决传统可解释模型冷启动的问题, 本人在该工作中提出了一种使用具有丰富先验知识的RA-LLM生成解释规则的知识蒸馏框架, 显著增强了对GNNs模型的本地、全局反事实解释能力

《Enhancing and Cleaning Labels for Graph Neural Networks》(在投)

- Wenfei Fan, Kehan Pang and Chao Tian.
- 为解决GNNs难以在具有非结构化与半结构化信息的图数据中高效地学习到关键特征, 以及语义上相较1-WL算法存在欠缺的挑战, 本人在该工作中提出了一种通过LLM在非结构化与半结构化的知识图谱中抽取特征、并根据这些特征使用Revised-1-WL算法生成伪标签的框架, 显著提升了GNNs在下游分类任务的性能

《Efficient Mixture of Experts based on Large Language Models for Low-Resource Data Preprocessing》

- Mengyi Yan, Yaoshu Wang, Kehan Pang, Min Xie and Jianxin Li. (SIGKDD) 2024
- 为解决MoE模型在融合多专家时可能带来潜在的信息覆盖、信息冗余问题, 本人在该篇论文中从理论上证明了数据增强时Meta-path的选取具有Church-Rosser性质, 并负责设计RAG模块的信息瓶颈网络架构, 为本篇论文提供了理论保障, 并显著提升了MoE模型相较单专家在下游任务上的性能

《GPU-Accelerated Graph Cleaning with a Single Machine》(在投)

- Wenchao Bai, Wenfei Fan, Jiahui Jin, Shuhao Liu, Kehan Pang and Xiaoke Zhu.
- 为增强资源受限场景下图计算系统对大规模图数据的处理效率, 本工作提出了一种将多机图计算系统转单机图计算系统的框架, 本人在该工作中主要负责设计规则挖掘的多机转单机调度方法、大图的Blocking方法

《Linking Entities across Relations and Graphs》

- Wenfei Fan, Ping Lu, Kehan Pang, Ruochun Jin and Wenyuan Yu. (TODS) 2022
- 为解决异源、异构数据集下实体对齐的语义不一致问题，本人在该论文中提出了一种实现跨异构图与关系表的实体对齐算法；同时，本人在该论文提出了应对数据量增长的数据分区索引算法、增量实体对齐算法，保证了系统在兆级数据下的运行效率

♥ 获奖情况

北京市大学生数学竞赛(理工类甲组)一等奖	2019第30届、2020第31届
全国大学生数学竞赛(非数学类)一等奖	2019第11届、2020第12届
北京邮电大学大学生创新创业训练计划市级项目奖	2020年9月
美国大学生数学建模竞赛H奖	2020年2月
北京邮电大学本科生二等奖学金	2019年、2020年、2021年
北京航空航天大学计算机学院学业奖学金	2022年、2023年

⚙ 专业技能

- 语言：英语-CET6
- 编程技能
 - 可熟练在Linux、Windows环境下使用Python、C++、Shell编写程序；熟悉Git、Conda、Docker等工具的使用；
 - 可熟练使用Numpy、Pandas等框架对数据进行分析、预处理、清洗；
 - 可熟练使用Pytorch、DGL、PyG等框架进行图神经网络构建；熟悉Graph-structure数据上常用的机器学习、特征工程、表示学习、数据挖掘、可解释学习算法；
 - 可熟练对LLM进行微调、推理；熟悉常用RAG、IR算法，可熟练在各种下游任务以及业务场景上应用LLM；