

乔一明

通信地址: 中国北京市海淀区清华大学, 100084 (邮政编码)
yimingqiao3163@gmail.com / qiaoym21@mails.tsinghua.edu.cn

我是清华大学交叉信息研究院的博士生, 研究方向为数据库管理系统, 尤其是向量化执行、查询优化和数据压缩方面。我最近也在关注检索增强生成与向量数据库。

教育经历

清华大学 交叉信息研究院, 计算机科学与技术专业, 博士研究生在读 GPA: 3.76/4.0, 导师: 张焕晨, 高逸涵	2021 年 9 月 - 2026 年 6 月 中国, 北京
南京邮电大学 贝尔英才学院, 计算机科学与技术专业, 工学学士 GPA: 92.9/100.0, 排名: 1/29; CET-6: 542	2017 年 9 月 - 2021 年 6 月 中国, 南京

社会实践

访问学生 Centrum Wiskunde & Informatica 导师: Peter Boncz	2025 年 2 月 - 至今 阿姆斯特丹, 荷兰
投资实习生 英诺天使基金	2023 年 5 月 - 2023 年 8 月 中国, 北京
软件开发实习生 eBay	2020 年 9 月 - 2020 年 11 月 中国, 上海
软件开发实习生 Oracle	2019 年 11 月 - 2020 年 5 月 中国, 南京
交换生 南京大学	2018 年 9 月 - 2019 年 6 月 中国, 南京

科研项目

DuckDB 的下一代执行引擎 导师: Peter Boncz, 张焕晨	2025 年 2 月 - 至今
--	-----------------

正在为 DuckDB 开发下一代查询执行引擎, 该项目延续了 Robust Predicate Transfer (SIGMOD'25) 的研究成果, 旨在避开数据库中的 cardinality estimation 问题, 提升查询执行效率。

向量化执行中的小数据块问题 导师: 张焕晨	2023 年 8 月 - 2024 年 11 月
---------------------------------	--------------------------

发现了向量化执行中的小数据块问题, 其中 filter 算子和 hash join 算子会减少数据块中的有效元组的数量, 导致性能下降。通过对查询执行过程中出现的小数据块进行合并, 可以实现更高的性能。该方案已在 DuckDB 中实现, 在 Join Order benchmark、TPC-H 以及 TPC-DS 等标准基准测试中实现了平均 10% 的性能提升。

此工作已发表于 SIGMOD'25

面向内存 OLTP 数据库的快速语义压缩 导师: 高逸涵, 张焕晨	2021 年 2 月 - 2024 年 5 月
---	-------------------------

开发了 Blitzcrank, 一种用于 OLTP 数据库的高速压缩方法, 减少了 85% 的内存使用, 并且仅带来 19% 的性能影响。通过引入新的语义模型和一种名为 “Delayed Coding” 的快速编码技术, 解决了低压缩时延与支持数据更新的挑战, 从而提高了 OLTP 场景压缩的速度和压缩率。

此工作已发表于 VLDB'24

基于神经网络的光谱去模糊

2018 年 7 月 - 2019 年 12 月

导师: 朱虎

开发了一个用于红外光谱仪的端到端神经网络框架, 解决了老化设备中频段重叠和噪声的问题。该方法使用 dilated convolutions 和 self-paced learning, 优于传统的偏微分方程方法, 提高了光谱去模糊的性能。

此工作已发表于 *IEEE Transactions on Industrial Informatics*

学术成果

- [1] **Yiming Qiao**, Huanchen Zhang, “Data Chunk Compaction in Vectorized Execution,” In: Proceedings of the ACM on Management of Data (**SIGMOD'25**), 3(1): Article 26, 25 pages.
- [2] **Yiming Qiao**, Yihan Gao, Huanchen Zhang, “Blitzcrank: Fast Semantic Compression for In-memory Online Transaction Processing,” In: *Proceedings of the VLDB Endowment (VLDB'24)* 17, no. 10, pp. 2528 - 2540.
- [3] Hu Zhu*, **Yiming Qiao***, Guoxia Xu, Lizhen Deng, and Yu-Feng Yu. “DSPNet: A Lightweight Dilated Convolution Neural Networks for Spectral Deconvolution with Self-paced Learning.” In: *IEEE Transactions on Industrial Informatics (TII)* 16, no. 12 (2019): 7392-7401. (*Equal Contribution)
- [4] Huihui Wang, Shunmei Meng, **Yiming Qiao**, and Jing Zhang. “Fast Classification Algorithms via Distributed Accelerated Alternating Direction Method of Multipliers.” In: *Proceedings of 2019 IEEE International Conference on Data Mining (ICDM'19)*, Nov. 2019, pp. 1354 - 1359.

专利

- [1] 陈兴国, **乔一明**, 刘威, 朱洁, “一种面向用户的自定义体育赛事解说增强方法,” 中国发明专利 (CN111477244B), 2020 年 4 月申请, 2023 年 9 月授权。
- [2] 邹志强, 李林睿, 常舒予, **乔一明**, “基于天文光谱数据的离群天体分类方法,” 中国发明专利 (CN114282585A), 2020 年 9 月申请, 2024 年 11 月授权。

其他

获奖

- Mitacs 研究实习项目, 由中国留学基金委全额资助, 2020 年, 为期 12 周。
- 贝尔英才学院金质荣誉勋章 (南京邮电大学), 2020 年。
- 斯坦福大学学术与文化交流项目, 由南京邮电大学全额资助, 2019 年, 为期 2 周。

证书

- 私募股权综合能力提升项目证书, 由清华大学研究生院与全球私募股权研究院联合颁发, 2023 年。

教学

- 助教 - 量化投资与金融优化 (清华 80470273) - 2023 年秋季
- 助教 - 数据挖掘 (清华 40470333) - 2021 年秋季