张宇昂

Zhang Yu'ang

2002.04
18902012189
zya1412@mail.ustc.edu.cn
实验室：KDELab@USTC
联系邮箱：zya1412@mail.ustc.edu.cn

教育经历

24.9 – **硕士**, *中国科学技术大学*, 计算机应用技术
GPA: 3.90/4.30　加权平均分: 90.7/100
核心课程：高级数据库系统 (98) 计算机应用数学 (94) 计算机系统 (92) 高级人工智能 (90)
研究方向：缓存优化 / KVCache 压缩 / 向量检索加速

20.9 – 24.6 **本科**, *中国科学技术大学*, 计算机科学与技术
GPA: 3.55/4.30　加权平均分: 86.23/100
核心课程：编译原理与技术 (96) 数据结构 (95) 计算机体系结构 (92)

项目经历

25.10 – 26.1 **大模型 RAG 场景 KV Cache 复用优化 (HiKV)**, *队长 & 核心算法设计*
针对大模型 RAG 长文本推理中 KV Cache 复用精度低、首字延迟 (TTFT) 高的问题，设计层级筛选式文档重算算法。
- **核心算法设计**：提出 HiKV 算法，解决跨文档 Cross-Attention 缺失导致的幻觉问题。设计双维度优化策略：按文档相关性动态分配重算比例，按 Transformer 层级设计金字塔式预算，严格控制重算比例 ≤ 30%。
- **工程化落地**：深入理解并修改 **vLLM (v0.7.3)** 源码，在 Qwen2-7B / DeepSeek-R1-Distill 等模型上完成适配。
- **性能收益**：在 Dureader/MusiQue 数据集上验证，实现了精度损失与 TTFT 的最优权衡，显著优于现有静态 Cache 策略，入围最终的决赛并获得优胜奖。

23.9 – 24.10 **数据库自适应缓存置换算法研究 (iCache)**, *技术负责人*, 华为-CCF 胡杨林基金
解决传统数据库缓存算法无法动态适配业务负载变化、命中率低的核心问题。
- **架构设计**：设计 iCache 核心架构，基于 **Lock-free 4-bit** 统计实现低开销热度计量；结合 **RankNet** 模型对冷数据队列进行未来访问预测，实现精确驱逐。
- **内核级开发**：在 **openGauss** 数据库内核中完成算法落地，搭建 TPC-C 动态负载测试环境。
- **量化成果**：相较原生 Clock-Sweep 算法，缓存缺失率降低 **16.5% – 21.9%**，超额达成项目验收指标。

22.9 – 23.1 **编译器前端与后端实现 (C-Minus-F)**, 独立实现者
从零构建面向 C 语言子集的编译器，实现从源码解析到中间代码优化的完整链路。
- **前端构建**：基于 **Flex/Bison** 完成词法与语法分析，构建 AST 实现源码的结构化解析与基础类型检查。
- **中间代码**：实现从 AST 到线性 IR 的端到端转换，精准处理复杂控制流 (Control Flow) 跳转与函数调用规约。实现了面向内存的 SSA (Static Single Assignment) 形式或标准 IR 格式生成。
- **中端优化**：独立设计并实现 **GVN (全局值编号)** 优化，分析消除冗余计算，显著减少 IR 指令数。

论文发表

DASFAA 2026 **Selectively Learned Cache Eviction Algorithm**, *CCF-B 类会议*, Full Paper
第一作者 (First Author)
- 提出 **SL-Cache** 算法，构建三层缓存架构与双指标热度计量体系。
- 创新预计算效用评估策略，对单次访问对象跳过复杂模型预测，大幅降低推理开销。
- 实验表明：相较 LRU 缺失率降低 **11.9%**，吞吐性能是 SOTA 算法 (3L-Cache) 的 **1.1-2 倍**。

专业技能

大模型推理 熟悉 **vLLM** 框架 KV Cache 相关推理算法源码并具备调试经验；熟悉 KV Cache 选择压缩算法；熟悉 CPU/GPU 多层存储架构下的 KV Cache 优化算法。

数据库内核 熟悉 **openGauss** 缓存管理模块及相关代码，具备内核级调试经验；熟悉学习型/非学习型缓存替换算法及多种缓存优化技术。

向量检索 熟悉 **DiskANN**, HNSW 等向量索引算法及其工程优化技术；了解向量索引调试技术。

编程语言 了解并掌握 **C/C++** (系统级编程), Python (PyTorch/科学计算)。

英语能力 CET-6 (577), TOEFL (101)；具备流利的英文顶会论文阅读与写作能力。