Guanduo Chen

gdchen22@m.fudan.edu.cn | woooooow-pro.github.io | wooooooow-pro | Guanduo Chen

Education

Master of Science in Computer Science and Technology, Fudan University September 2022 – June 2025 (Expected)

- Research Interests: LLMSys, AI4DB

B.Eng in Computer Science and Technology, Fudan University September 2018 – July 2022

Work Experience

Meituan, Ltd. January 2025 –Present
Large Multi-modality Model Architecture Intern Shanghai, China

- **Vision Encoder Long Context Training Support:** Designed and developed a context-parallelism mechanism for the vision encoder of a large multi-modality model.

ByteDance, Ltd. December 2023 – April 2024
Database System R&D Intern, ByteHouse Runtime Team Shanghai, China

- **Geographic Data Aggregation Query Benchmark:** Developed and benchmarked geospatial aggregation queries on the NYC Taxi dataset, identifying performance disparities across major database platforms (StarRocks, ClickHouse, PostGIS, DuckDB, and ByteHouse-CE). This analysis led to actionable insights for optimizing geospatial query execution.
- **Geospatial Data Support for CNCH:** Integrated geometry data types into ByteHouse-CNCH (Cloud Native ClickHouse) via geos library, expanding its capabilities and improving data representation efficiency.
- **Geospatial Index Design:** Spearheaded the design and implementation of multi-level indexing (disk and memory cache) for geospatial data, cutting query latency by nearly 50% compared to the base ClickHouse implementation.

Research Experience

Computation Efficient LoRA [Paper] March 2024 – January 2025
Advisor: Prof. Binhang Yuan HKUST Relaxed System Lab

- **Algorithm Development:** Developed CE-LoRA, a high-efficiency algorithm for parameter-efficient fine-tuning (PEFT), which significantly reduced backpropagation costs in large language model training. By leveraging structured sparsity and low-rank approximation techniques, the model achieved a $3.39\times$ improvement in training efficiency without sacrificing accuracy.
- **Theoretical Analysis:** Provided rigorous convergence analysis, proving that CE-LoRA maintains the same convergence rate as LoRA, but with reduced computational overhead.

Learned Range Filter [Paper, Code] February 2023 – December 2023
Advisor: Prof. Siqiang Luo NTU Data Management Research Group

- **Oasis:** Developed Oasis, a learned range filter that segments the key space into non-overlapping intervals and maps data into a bitmap using a linear model-simulated CDF as the hash function. The filter utilizes block-based Elias-Fano compression to reduce space overhead without compromising query efficiency.
- **Oasis+:** Created Oasis+, a hybrid range filter that combines learning-based and hash-based methods to enhance filter applicability and robustness across various workloads.
- **Integration into RocksDB:** Integrated Oasis and Oasis+ into RocksDB and tested their performance, achieving up to $6.2\times$ improvement in query response times.

Text2SQL [Paper, Code] July 2021 – February 2022
Advisor: Prof. X. Sean Wang Fudan DASLab

- **Text2SQL Framework Development** Developed the GAR framework for Text2SQL translation, using a unique "Generate-and-Rank" approach that leverages parsing, generation, and ranking strategies for high-accuracy SQL generation from natural language queries.
- **Benchmarking** Built and tested a complex benchmark with self-joins, analyzing GAR's performance against other end-to-end models, providing crucial insights into its strengths in complex query generation.

Publication

1. CE-LoRA: Computation-Efficient LoRA Fine-Tuning for Language Models **[preprint]**
   Guanduo Chen*, Yutong He*, Yipeng Hu, KunYuan, Binhang Yuan.
2. Oasis: An Optimal Disjoint Segmented Learned Range Filter **[VLDB 2024]**
   Guanduo Chen, Zhenying He, Meng Li, Siqiang Luo.
3. GAR: A Generate-and-Rank Approach for Natural Language to SQL Translation **[ICDE 23]**
   Y Fan, Z He, T Ren, D Guo, L Chen, R Zhu, **G Chen**, Y Jing, K Zhang, XS Wang