Bin Wu

wubeen@outlook.com · bin-wu

Education

Hong Kong University of Science and Technology, Hong Kong, China 09/2012 – 10/2018

Doctor of Philosoply in Computer Science

Fudan University, Shanghai, China 09/2008 – 07/2012

Bachelor of Science in Computer Science

Work Experience

DAMO Academy & Alibaba Cloud, *Alibaba Group* Hangzhou, China, 12/2018 – Present

Research Scientist Drove innovation from **DATA+AI** related research to production, publishing novel methods in top-tier conferences while deploying them at scale to solve key business challenges.

Noah's Ark Lab, *Huawei* Hong Kong, 08/2017 – 10/2018

Researcher Pioneered research in meta-learning and few-shot learning and validated these approaches by successfully applying them to solve real-world problems.

Projects

Agent DevOps Platform on **DMS** 09/2024 – present

Team Leader

- Directed the end-to-end development of an Agent DevOps platform on Alibaba Cloud DMS, successfully onboarding and supporting over 100 enterprise customers. Engineered a production-grade version of the open-source Dify framework on Kubernetes, re-architecting its core for high availability and boosting API performance to sustain over 1,000 QPS.
- Architected and built a flexible model serving and finetuning engine, enabling clients to securely deploy and manage a variety of large language models (including Qwen, DeepSeek, and Kimi) in their private environments.

Artificial Intelligence for IT Operations (AIOps) Program 04/2021 – 08/2024

Technical Leader

- We introduced **ShapleyIQ**, a novel diagnosis service that quantifies the true impact of component anomalies in complex microservice systems. It precisely identifies performance root causes. The solution is deployed on a 2,000-machine cluster at the internal system of Alibaba Database Product team, where it has cut system failures by 83% and achieved 97.3% accuracy, far surpassing all baselines.
- We proposed **OneShotSTL**, an accurate online decomposition algorithm with an O(1) update time complexity. Extensive experiments show that our method is up to 1,000 times faster than state-of-the-art alternatives on downstream tasks like anomaly detection and forecasting, while maintaining comparable or even better accuracy. This method had been implemented in Lindorm, whch is a cloud-native, multi-model database service offered by Alibaba Cloud.
- We developed **Andromeda**, a novel framework that significantly improves automatic database debugging by pairing a Large Language Model with a sophisticated retrieval-augmented generation (**RAG**) system. This RAG strategy is uniquely designed to synthesize context from heterogeneous sources—including past troubleshooting tickets, manuals, and telemetry data—to produce expert-level solutions that outperform existing methods.

Natural language to SQL (NL2SQL) 06/2019 – 06/2021

Technical Leader We introduced a novel framework that combines a versatile, template-guided generation method called the CatSQL sketch with a Semantics Correction post-processing routine. This dual approach leverages database domain knowledge to fix errors, achieving state-of-the-art accuracy and up to 63x higher throughput on the challenging Spider benchmark. This method topped the Yale Spider benchmark leaderboard with a score of 78.

Vector Database (AnalyticDB-V, Alibaba Cloud) 01/2019 – 01/2022

Algorithm Technical Leader

- We developed AnalyticDB-V (ADBV), a hybrid analytic engine from Alibaba that extends SQL to support queries on high-dimensional vectors. ADBV integrates a novel Approximate Nearest Neighbor Search (ANNS) algorithm and an accuracy-aware cost-based optimizer, delivering superior performance in real-world applications on Alibaba Cloud.
- We proposed LEQAT, a query-aware optimizer that generates a unique, efficient search plan for each individual query. By using a machine learning model to predict the distribution of a query's nearest neighbors, LEQAT dynamically selects the best partitions to probe, reducing latency by up to 58% and improving throughput by up to 3.9 times.

Meta Learning 08/2017 – 10/2018

Technical Leader Developed a novel Deep Meta-Learning framework that established a new state-of-the-art on public few-shot image benchmarks. Engineered a separate solution for few-shot user intent detection, which substantially outperformed established baselines like SVM and Siamese Networks by leveraging existing meta-learning algorithms.

Wander Join and *XDB* 10/2014 – 03/2017

Core contributor and researcher Accompolished standalone implementation (C++) of *Wander Join* which could resolve the online aggregation problem $>$ 10X faster than existing methods. Implemented the *wander join* algorithm in PostgreSQL: introduced new keywords to parser part, added new components into optimizer part and executor part. For TPCH queries, our method could return $<$ 1% error result within seconds for large datasets.

Counting triangles in sublinear time 08/2013 – 09/2014

Researcher Proposed and implemented three approximate triangle counting algorithms on very large graphs that could output an approximate answer within seconds.

♥ Awards and Professional Services

- **Conference Program Committee**: ICDE Research track PC(2025, 2022), ICDE Industrial Track(2025, 2024, 2021), KDD Research Track(2023, 2022).
- **Reviewer**: TKDE, The Journal of VLDB.
- **ACM SIGMOD Research Highlight Award**, 2017
- **ACM SIGMOD Best Paper Award**, 2016
- HKUST travel grant 2016, Studentship from HKUST (2012 – 2017)
- Excellent graduate of Fudan University, Excellent student of Fudan University, Excellent student leader of Fudan University

⚙ Skills

- Programming Languages: Python, C/C++ > Java
- Deep Learning platforms: Pytorch, Tensorflow, MXNet

ℹ Selected Publications

1. Sibei Chen, Ju Fan, **BIN WU**, Nan Tang, Chao Deng, Pengyi Wang, Ye Li, Jian Tan, Feifei Li, Jingren Zhou. "Automatic database configuration debugging using retrieval-augmented language models", SIGMOD'25
2. Han Fu, Chang Liu, **BIN WU**, Feifei Li, Jian Tan, Jianling Sun. "Catsql: Towards real world natural language to sql applications", VLDB'23.
3. Xiao He, Jian Tan, **BIN WU**, Feifei Li, Xinping Zhang, Gaozhong Liang, Jinfeng Xu. "Active Sampling for Sparse Table by Bayesian Optimization with Adaptive Resolution" ICDE'23.
4. Xiao He, Ye Li, Jian Tan, **BIN WU**, Feifei Li. "OneShotSTL: One-shot seasonal-trend decomposition for online time series anomaly detection and forecasting", VLDB'23.
5. Ye Li, Jian Tan, **BIN WU**, Xiao He, Feifei Li. "Shapleyiq: Influence quantification by shapley values for performance debugging of microservices", ASPLOS'23.
6. Liang Lin, Yuhan Li, **BIN WU**, Huijun Mai, Renjie Lou, Jian Tan, Feifei Li. "Anser: Adaptive Information Sharing Framework of AnalyticDB", VLDB'23.

7. Chunhui Shen, Qianyu Ouyang, Feibo Li, Zhipeng Liu, Longcheng Zhu, Yujie Zou, Qing Su, Tianhuan Yu, Yi Yi, Jianhong Hu, **BIN WU**. "Lindorm TSDB: A cloud-native time-series database for large-scale monitoring systems", VLDB'23.
8. Pengcheng Zhang, Bin Yao, Chao Gao, **BIN WU**, Xiao He, Feifei Li, Yuanfei Lu, Chaoqun Zhan, Feilong Tang. "Learning-based query optimization for multi-probe approximate nearest neighbor search", VLDBJ'23.
9. Yuan Qiu, Wei Dong, Ke Yi, **BIN WU**, Feifei Li. "Releasing Private Data for Numerical Queries", KDD'23.
10. Yuan Qiu, Yilei Wang, Ke Yi, Feifei Li, **BIN WU**, Chaoqun Zhan. "Weighted distinct sampling: Cardinality estimation for SPJ queries", SIGMOD'21.
11. Chuangxian Wei, **BIN WU**, Sheng Wang, Renjie Lou, Chaoqun Zhan, Feifei Li, Yuanzhe Cai. "Analyticdb-v: A hybrid analytical engine towards query fusion for structured and unstructured data", VLDB'20.
12. Feifei Li, **BIN WU**, Ke Yi, Zhuoyue Zhao. "Wander join and XDB: online aggregation via random walks", TODS'19.
13. Fengwei Zhou, **BIN WU**, Zhenguo Li. "Deep meta-learning: Learning to learn in the concept space". *arXiv preprint arXiv:1802.03596*, 2018.
14. Feifei Li, **BIN WU**, Ke Yi, Zhuoyue Zhao. "Wander join and XDB: online aggregation via random walks", SIGMOD Record'17.
15. Feifei Li, **BIN WU**, Ke Yi, Zhuoyue Zhao. "Wander join: Online aggregation for joins", SIGMOD'16.
16. Feifei Li, **BIN WU**, Ke Yi, Zhuoyue Zhao. "Wander join: Online aggregation via random walks", SIGMOD'16.
17. **BIN WU**, Ke Yi, Zhenguo Li. "Counting triangles in large graphs by random sampling", TKDE'16.
18. Xiangyang Xue, Wei Zhang, Jie Zhang, **BIN WU**, Jianping Fan, Yao Lu. "Correlative multi-label multi-instance image annotation", ICCV'11.
19. Wei Zhang, Xiangyang Xue, Jianping Fan, Xiaojing Huang, **BIN WU**, Mingjie Liu. "Multi-kernel multi-label learning with max-margin concept network", IJCAI'11.