Bolin Ding

Researcher
Institute for Intelligent Computing
Alibaba Group, USA

Mobile: (+1) 217-299-6145
E-mail: bolin.ding@alibaba-inc.com
https://bolinding.github.io/

EXPERIENCE

4/23 – present	Researcher (Senior Director), Data Analytics and Intelligence Lab, Alibaba Group
4/18 – 4/23	Senior Staff Engineer, Data Analytics and Intelligence Lab, Alibaba Group
8/17 – 4/18	Principal Researcher, Data Management, Exploration and Mining Group, Microsoft Research
8/15 – 8/17	Lead Researcher, Data Management, Exploration and Mining Group, Microsoft Research
8/12 – 8/15	Senior Researcher, Data Management, Exploration and Mining Group, Microsoft Research

EDUCATION

University of Illinois at Urbana-Champaign, Urbana, US
- Ph.D., Department of Computer Science, 2012
- Advisor: Jiawei Han
- Thesis: Privacy-Preserving Data Publishing and Analytics using Data Cubes

The Chinese University of Hong Kong, Hong Kong, China
- M.Phil., Department of Systems Engineering and Engineering Management, 2007
- Advisor: Jeffery Xu Yu
- Thesis: Advanced Query Processing in Databases

Renmin University of China, Beijing, China
- B.S., Department of Math and Applied Mathematics, 2005
- Advisor: Qing Zhu and Shan Wang
- Final-year Project: Keyword Search in Databases

RESEARCH INTERESTS AND PROJECTS

My recent research focuses on:
- **Agent systems and programming models for large language models**
- **Data processing systems for training and evaluating large-scale machine learning models**
- **Data privacy** (including definitions, algorithms, and systems)
- **Federated data management, analytics, and learning** (e.g., developing federated matrix factorization algorithms with privacy guarantees and building a federated learning system, FederatedScope)
- **Making systems intelligent and efficient with machine learning and optimization techniques** (including physical computing systems such as databases and machine learning systems, and social systems)

In particular, I enjoy developing algorithms and building systems in the following projects:

DPaaS (Data Privacy as a Service): Developing a series of privacy-preserving data collection, sharing, analysis, and learning techniques, for example, multi-dimensional and multi-source data sharing and OLAP under local differential privacy and/or MPC, and federated learning with formal privacy guarantees for vertically collaborative learning and device-server collaborative learning. Building a system with which these techniques can be deployed on various data platforms.

FederatedScope (a federated learning platform): We have open-sourced an easy-to-use federated learning package, FederatedScope. It provides comprehensive functionalities including privacy protection, personalization, auto-tuning of federated machine learning models, based on a programming framework with which one can conveniently develop and deploy her/his own federated models in various settings about how data and models are distributed, e.g., vertically/horizontally collaborative learning, and cross-silo/cross-device federated learning

services. Our team is actively contributing to the package and enriching its functionality for different types algorithms and scenarios, e.g., FederatedScope-GNN for federated learning of graph neural networks and FederatedScope-Real for large-scale cross-device federated learning.

AgentScope (an agent-based programming framework for developing LLM applications): We develop a programming framework called AgentScope which abstracts each LLM instance with auxiliary components and functionalities such as memory and RAG as an LLM agent. AgentScope allows developers to easily assemble multiple LLM agents into a traditional or AI-driven application. From the view of programming languages, AgentScope aims to offer agent-oriented programming (AOP) as a new programming model to organize the design and implementation of next-generation LLM-empowered applications. AgentScope also offers necessary native services for agents, including communication between agents (locally or remotely), web search, and code execution.

System4AI: Optimizing the infrastructure and pipeline for training machine learning (large language) models, starting from the data pre-processors to model evaluators. We built Data-Juicer, a one-stop multimodal data processing system to make data higher-quality, juicier, and more digestible for training and tuning large language models. At an earlier time, we developed a series of automated machine learning techniques (AutoML) in order to enable developers and data scientists with limited machine learning expertise and resources to train high-quality models, including auto tuning of hyperparameters, auto feature selection, and neural architecture search for machine learning models. Some of these techniques have been deployed into Alibaba's cloud AutoML products.

AI4System: Is it possible for models to learn to be a statistician, a database administrator, an index, a query processor, or a query optimizer? We developed a series of "learning-to-be" techniques for different database components, as well as a middleware system, called PilotScope, to deploy these learned components into databases. PilotScope offers a programming framework to bridge the gaps between AI4DB (Artificial Intelligence for Databases) algorithms and database systems, with which researchers can develop and train their own learned components, deploy them without modifying the code in database engines, and compare them fairly with alternatives in actual database systems.

AI4SocialScience: We conduct interdisciplinary research between Social Science (starting from, e.g., Economics) and Machine Learning. Besides traditional theoretical study, we build a multi-agent reinforcement learning platform to conduct simulation study and to investigate scenarios which are too complicated to be formulated as clean and solvable theoretical/math problems. For example, in recommendation systems, we study how users would react to different data privacy policies, and how the utility of platforms and the utility of individual users are affected.

SELECTED PUBLICATIONS

Refereed Conference Publications

1. Weirui Kuang, Bingchen Qian, Zitao Li, Daoyuan Chen, Dawei Gao, Xuchen Pan, Yuexiang Xie, Yaliang Li, Bolin Ding, and Jingren Zhou: FederatedScope-LLM: A Comprehensive Package for Fine-tuning Large Language Models in Federated Learning, in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'24*), 2024.

2. Fangyuan Zhao, Zitao Li, Xuebin Ren, Bolin Ding, Shusen Yang, and Yaliang Li: VertiMRF: Differentially Private Vertical Federated Data Synthesis, in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'24*), 2024.

3. Feijie Wu, Zitao Li, Yaliang Li, Bolin Ding, and Jing Gao: FedBiOT: LLM Local Fine-tuning in Federated Learning without Full Model, in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'24*), 2024.

4. Daoyuan Chen, Yaliang Li, and Bolin Ding: Multi-modal Data Processing for Foundation Models: Practical Guidances and Use Cases, in *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'24*), 2024.

5. Zitao Li, Bolin Ding, Liuyi Yao, Yaliang Li, Xiaokui Xiao, and Jingren Zhou: Performance-Based Pricing for Federated Learning via Auction, in *Proceedings of the 50th International Conference on Very Large Data Bases* (*VLDB'24*), 2024.
   (also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 17(6): 1269-1282, 2024)

6. Rong Zhu, Lianggui Weng, Wenqing Wei, Di Wu, Jiazhen Peng, Yifan Wang, Bolin Ding, Defu Lian, Bolong Zheng, and Jingren Zhou: PilotScope: Steering Databases with Machine Learning Drivers, in *Proceedings of the 50th International Conference on Very Large Data Bases* (*VLDB'24*), 2024.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 17(5): 980-993, 2024)

7. Dawei Gao, Haibin Wang, Yaliang Li, Xiuyu Sun, Yichen Qian, Bolin Ding, and Jingren Zhou: Text-to-SQL Empowered by Large Language Models: A Benchmark Evaluation, in *Proceedings of the 50th International Conference on Very Large Data Bases* (*VLDB'24*), 2024.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 17(5): 1132-1145, 2024)

8. Lianggui Weng, Rong Zhu, Di Wu, Bolin Ding, Bolong Zheng, and Jingren Zhou: Eraser: Eliminating Performance Regression on Learned Query Optimizer, in *Proceedings of the 50th International Conference on Very Large Data Bases* (*VLDB'24*), 2024.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 17(5): 926-938, 2024)

9. Pengfei Li, Wenqing Wei, Rong Zhu, Bolin Ding, Jingren Zhou, and Hua Lu: ALECE: An Attention-based Learned Cardinality Estimator for SPJ Queries on Dynamic Workloads, in *Proceedings of the 50th International Conference on Very Large Data Bases* (*VLDB'24*), 2024.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 17(2): 197-210, 2024)

10. Fei Wei, Ergute Bao, Xiaokui Xiao, Yin Yang, and Bolin Ding: AAA: An Adaptive Mechanism for Locally Differential Private Mean Estimation, in *Proceedings of the 50th International Conference on Very Large Data Bases* (*VLDB'24*), 2024.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 17(8): 1843-1855, 2024)

11. Daoyuan Chen, Yilun Huang, Zhijian Ma, Hesen Chen, Xuchen Pan, Ce Ge, Dawei Gao, Yuexiang Xie, Zhaoyang Liu, Jinyang Gao, Yaliang Li, Bolin Ding, and Jingren Zhou: Data-Juicer: A One-Stop Data Processing System for Large Language Models, in *Companion of the 2024 International Conference on Management of Data* (*SIGMOD'24*), 2024.

12. Rong Zhu, Lianggui Weng, Bolin Ding, and Jingren Zhou: Learned Query Optimizer: What is New and What is Next, in *Companion of the 2024 International Conference on Management of Data* (*SIGMOD'24*), 2024.

13. Jiajun Li, Runlin Lei, Sibo Wang, Zhewei Wei, and Bolin Ding: Learning-based Property Estimation with Polynomials, in *Companion of the 2024 International Conference on Management of Data* (*SIGMOD'24*), 2024.
(also appeared in the Journal " *Proceedings of the ACM on Management of Data*", 2(3): 1-27, 2024)

14. Yanxi Chen, Xuchen Pan, Yaliang Li, Bolin Ding, and Jingren Zhou: EE-LLM: Large-Scale Training and Inference of Early-Exit Large Language Models with 3D Parallelism, in *Proceedings of the 41st International Conference on Machine Learning* (*ICML'24*), 2024.

15. Kexin Huang, Ziqian Chen, Xue Wang, Chongming Gao, Jinyang Gao, Bolin Ding, and Xiang Wang: Auctionformer: A Unified Deep Learning Algorithm for Solving Equilibrium Strategies in Auction Games, in *Proceedings of the 41st International Conference on Machine Learning* (*ICML'24*), 2024.

16. Zhen Qin, Daoyuan Chen, Bingchen Qian, Bolin Ding, Yaliang Li, and Shuiguang Deng: Federated Full-Parameter Tuning of Billion-Sized Language Models with Communication Cost under 18 Kilobytes, in *Proceedings of the 41st International Conference on Machine Learning* (*ICML'24*), 2024.

17. Yichen Qian, Yongyi He, Rong Zhu, Jintao Huang, Zhijian Ma, Haibin Wang, Yaohua Wang, Xiuyu Sun, Defu Lian, Bolin Ding, and Jingren Zhou: UniDM: A Unified Framework for Data Manipulation with Large Language Models, in *Proceedings of the 7th Annual Conference on Machine Learning and Systems* (*MLSys'24*), 2024.

18. Yin Lin, Bolin Ding, H. V. Jagadish, and Jingren Zhou: SMARTFEAT: Efficient Feature Construction through Feature-Level Foundation Model Interactions, in *Proceedings of the 14th Conference on Innovative Data Systems Research* (*CIDR'24*), 2024.

19. Youbang Sun, Zitao Li, Yaliang Li, and Bolin Ding: Improving LoRA in Privacy-preserving Federated Learning, in *Proceedings of the 12th International Conference on Learning Representations* (*ICLR'24*), 2024.

20. Xue Wang, Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, and Rong Jin: CARD: Channel Aligned Robust Blend Transformer for Time Series Forecasting, in *Proceedings of the 12th International Conference on Learning Representations* (*ICLR'24*), 2024.

21. Zhen Wang, Yaliang Li, Bolin Ding, Yule Li, and Zhewei Wei: Exploring Neural Scaling Law and Data Pruning Methods For Node Classification on Large-scale Graphs, in *Proceedings of the ACM Web Conference* (*WWW'24*), 2024.

22. Yuzheng Hu, Fan Wu, Qinbin Li, Yunhui Long, Gonzalo Munilla Garrido, Chang Ge, Bolin Ding, David A. Forsyth, Bo Li, and Dawn Song: SoK: Privacy-Preserving Data Synthesis, in *Proceedings of 2024 IEEE Symposium on Security and Privacy* (*SP*), 2024.

23. Dawei Gao, Daoyuan Chen, Zitao Li, Yuexiang Xie, Xuchen Pan, Yaliang Li, Bolin Ding, and Jingren Zhou: FS-Real: A Real-World Cross-Device Federated Learning Platform, in *Proceedings of the 49th International Conference on Very Large Data Bases* (*VLDB'23*), 2023.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 16(12): 4046-4049, 2023)

24. Pengfei Li, Hua Lu, Rong Zhu, Bolin Ding, Long Yang, and Gang Pan: DILI: A Distribution-Driven Learned Index, in *Proceedings of the 49th International Conference on Very Large Data Bases* (*VLDB'23*), 2023.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 16(9): 2212-2224, 2023)

25. Xu Chen, Zhen Wang, Shuncheng Liu, Yaliang Li, Kai Zeng, Bolin Ding, Jingren Zhou, Han Su, and Kai Zheng: BASE: Bridging the Gap between Cost and Latency for Query Optimization, in *Proceedings of the 49th International Conference on Very Large Data Bases* (*VLDB'23*), 2023.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 16(8): 1958-1966, 2023)

26. Rong Zhu, Wei Chen, Bolin Ding, Xingguang Chen, Andreas Pfadler, Ziniu Wu, and Jingren Zhou: Lero: A Learning-to-Rank Query Optimizer, in *Proceedings of the 49th International Conference on Very Large Data Bases* (*VLDB'23*), 2023.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 16(6): 1466-1479, 2023)

27. Yuexiang Xie, Zhen Wang, Dawei Gao, Daoyuan Chen, Liuyi Yao, Weirui Kuang, Yaliang Li, Bolin Ding, and Jingren Zhou: FederatedScope: A Flexible Federated Learning Platform for Heterogeneity, in *Proceedings of the 49th International Conference on Very Large Data Bases* (*VLDB'23*), 2023.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 16(5): 1059-1072, 2023)

28. Bolin Ding, Yiding Feng, Chien-Ju Ho, Wei Tang, and Haifeng Xu: Competitive Information Design for Pandora's Box, in *Proceedings of ACM-SIAM Symposium on Discrete Algorithms* (*SODA'23*), 2023.

29. Zhen Wang, Weirui Kuang, Ce Zhang, Bolin Ding, and Yaliang Li: FedHPO-Bench: A Benchmark Suite for Federated Hyperparameter Optimization, in *Proceedings of the 40th International Conference on Machine Learning* (*ICML'23*), 2023.

30. Daoyuan Chen, Liuyi Yao, Dawei Gao, Bolin Ding, and Yaliang Li: Efficient Personalized Federated Learning via Sparse Model-Adaptation, in *Proceedings of the 40th International Conference on Machine Learning* (*ICML'23*), 2023.

31. Zeyu Qin, Liuyi Yao, Daoyuan Chen, Yaliang Li, Bolin Ding, and Minhao Cheng: Revisiting Personalized Federated Learning: Robustness Against Backdoor Attacks, in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'23*), 2023.

32. Daoyuan Chen, Dawei Gao, Yuexiang Xie, Xuchen Pan, Zitao Li, Yaliang Li, Bolin Ding, and Jingren Zhou: FS-Real: Towards Real-World Cross-Device Federated Learning, in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'23*), 2023.

33. Daoyuan Chen, Wuchao Li, Yaliang Li, Bolin Ding, Kai Zeng, Defu Lian, and Jingren Zhou: Learned Index with Dynamic Epsilon, in *Proceedings of the 11th International Conference on Learning Representations* (*ICLR'23*), 2023.

34. Liuyi Yao, Yaliang Li, Bolin Ding, Jingren Zhou, Jinduo Liu, Mengdi Huai, and Jing Gao: Path-specific Causal Fair Prediction via Auxiliary Graph Structure Learning, in *Proceedings of the ACM Web Conference* (*WWW'23*), 2023.

35. Zitao Li, Bolin Ding, Ce Zhang, Ninghui Li, and Jingren Zhou: Federated Matrix Factorization with Privacy Guarantee, in *Proceedings of the 48th International Conference on Very Large Data Bases* (*VLDB'22*), 2022. (also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 15(4): 900-913, 2021)

36. Renzhi Wu, Bolin Ding, Xu Chu, Zhewei Wei, Xiening Dai, Tao Guan, and Jingren Zhou: Learning to be a Statistician: Learned Estimator for Number of Distinct Values, in *Proceedings of the 48th International Conference on Very Large Data Bases* (*VLDB'22*), 2022.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 15(2): 272-284, 2021)

37. Jinglin Peng, Bolin Ding, Jiannan Wang, Kai Zeng, and Jingren Zhou: One Size Does Not Fit All: A Bandit-Based Sampler Combination Framework with Theoretical Guarantees, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'22*), 2022.

38. Amrita Roy Chowdhury, Bolin Ding, Somesh Jha, Weiran Liu, and Jingren Zhou: Strengthening Order Preserving Encryption with Differential Privacy, in *Proceedings of the 29th ACM Conference on Computer and Communications Security* (*CCS'22*), 2022.

39. Rong Zhu, Ziniu Wu, Chengliang Chai, Andreas Pfadler, Bolin Ding, Guoliang Li, and Jingren Zhou: Learned Query Optimizer: At the Forefront of AI-driven Databases, in *Proceedings of the 25th International Conference on Extending Database Technology* (*EDBT'22*), 2022.

40. Daoyuan Chen, Dawei Gao, Weirui Kuang, Yaliang Li, and Bolin Ding: pFL-Bench: A Comprehensive Benchmark for Personalized Federated Learning, in *Proceedings of Advances in Neural Information Processing Systems 35* (*NeurIPS'22*), 2022.

41. Runlin Lei, Zhen Wang, Yaliang Li, Bolin Ding, and Zhewei Wei: EvenNet: Ignoring Odd-Hop Neighbors Improves Robustness of Graph Neural Networks, in *Proceedings of Advances in Neural Information Processing Systems 35* (*NeurIPS'22*), 2022.

42. Jiawei Jiang, Lukas Burkhalter, Fangcheng Fu, Bo Li, Bolin Ding, Bo Du, Anwar Hithnawi, and Ce Zhang: VF-PS: How to Select Important Participants in Vertical Federated Learning, Efficiently and Securely, in *Proceedings of Advances in Neural Information Processing Systems 35* (*NeurIPS'22*), 2022.

43. Yaliang Li, Bolin Ding, and Jingren Zhou: A Practical Introduction to Federated Learning, in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'22*), 2022.

44. Zhen Wang, Weirui Kuang, Yuexiang Xie, Liuyi Yao, Yaliang Li, Bolin Ding, and Jingren Zhou: FederatedScope-GNN: Towards a Unified, Comprehensive and Efficient Package for Federated Graph Learning, in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'22*), 2022.
Best Paper Award in ADS

45. Zhen Wang, Zhewei Wei, Yaliang Li, Weirui Kuang, and Bolin Ding: Graph Neural Networks with Node-wise Architecture, in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'22*), 2022.

46. Jiajun Li, Zhewei Wei, Bolin Ding, Xiening Dai, Lu Lu, Jingren Zhou: Sampling-based Estimation of the Number of Distinct Values in Distributed Environment, in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'22*), 2022.

47. Yupeng Hou, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Bolin Ding, and Ji-Rong Wen: Towards Universal Sequence Representation Learning for Recommender Systems, in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'22*), 2022.

48. Dawei Gao, Yuexiang Xie, Zimu Zhou, Zhen Wang, Yaliang Li, and Bolin Ding: Finding Meta Winning Ticket to Train Your MAML, in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'22*), 2022.

49. Yuexiang Xie, Zhen Wang, Yaliang Li, Ce Zhang, Jingren Zhou, and Bolin Ding: iFlood: A Stable and Effective Regularizer, in *Proceedings of the 10th International Conference on Learning Representations* (*ICLR'22*), 2022.

50. Yiqing Xie, Zhen Wang, Carl Yang, Yaliang Li, Bolin Ding, Hongbo Deng, and Jiawei Han: KoMen: Domain Knowledge Guided Interaction Recommendation for Emerging Scenarios, in *Proceedings of the ACM Web Conference* (*WWW'22*), 2022.

51. Chong Chen, Fei Sun, Min Zhang, and Bolin Ding: Recommendation Unlearning, in *Proceedings of the ACM Web Conference* (*WWW'22*), 2022.

52. Shaoyun Shi, Yuexiang Xie, Zhen Wang, Bolin Ding, Yaliang Li, and Min Zhang: Explainable Neural Rule Learning, in *Proceedings of the ACM Web Conference* (*WWW'22*), 2022.

53. Shanlei Mu, Yaliang Li, Wayne Xin Zhao, Jingyuan Wang, Bolin Ding, and Ji-Rong Wen: Alleviating Spurious Correlations in Knowledge-aware Recommendations through Counterfactual Generator, in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR'22*), 2022.

54. Jiaxiang Liu, Karl Knopf, Yiqing Tan, Bolin Ding, and Xi He: Catch a Blowfish Alive: A Demonstration of Policy-Aware Differential Privacy for Interactive Data Exploration, in *Proceedings of the 47th International Conference on Very Large Data Bases* (*VLDB'21*), 2021.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 14(12): 2859-2862, 2021)

55. Ergute Bao, Yin Yang, Xiaokui Xiao, and Bolin Ding: CGM: An Enhanced Mechanism for Streaming Data Collection with Local Differential Privacy, in *Proceedings of the 47th International Conference on Very Large Data Bases* (*VLDB'21*), 2021.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 14(11): 2258-2270, 2021)

56. Yang Li, Yu Shen, Wentao Zhang, Jiawei Jiang, Yaliang Li, Bolin Ding, Jingren Zhou, Zhi Yang, Wentao Wu, Ce Zhang, and Bin Cui: VolcanoML: Speeding up End-to-End AutoML via Scalable Search Space Decomposition, in *Proceedings of the 47th International Conference on Very Large Data Bases* (*VLDB'21*), 2021.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 14(11): 2167-2176, 2021)

57. Shuyuan Yan, Bolin Ding, Wei Guo, Jingren Zhou, Zhewei Wei, Xiaowei Jiang, and Sheng Xu: FlashP: An Analytical Pipeline for Real-time Forecasting of Time-Series Relational Data, in *Proceedings of the 47th International Conference on Very Large Data Bases* (*VLDB'21*), 2021.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 14(5): 721-729, 2021)

58. Tianhao Wang, Min Xu, Bolin Ding, Jingren Zhou, Cheng Hong, Zhicong Huang, Ninghui Li, and Somesh Jha: Improving Utility and Security of the Shuffler-based Differential Privacy, in *Proceedings of the 47th International Conference on Very Large Data Bases* (*VLDB'21*), 2021.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 13(13): 3545-3558, 2020)

59. Yaliang Li, Zhen Wang, Yuexiang Xie, Bolin Ding, and Ce Zhang: AutoML: A Perspective where Industry Meets Academy, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'21*), 2021.

60. Siqing Li, Liuyi Yao, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Tonglei Guo, Bolin Ding, and Ji-Rong Wen: Debiasing Learning based Cross-domain Recommendation, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'21*), 2021.

61. Yuexiang Xie, Zhen Wang, Yaliang Li, Bolin Ding, Nezihe Merve Gürel, Ce Zhang, Minlie Huang, Wei Lin, and Jingren Zhou: FIVES: Feature Interaction Via Edge Search for Large-Scale Tabular Data, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'21*), 2021.

62. Mingyu Xiao, Sen Huang, Yi Zhou, and Bolin Ding: Efficient Reductions and A Fast Algorithm of Maximum Weighted Independent Set, in *Proceedings of the ACM Web Conference* (*WWW'21*), 2021.

63. Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam: Unified Conversational Recommendation Policy Learning via Graph-based Reinforcement Learning, in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR'21*), 2021.

64. Min Xu, Bolin Ding, Tianhao Wang, and Jingren Zhou: Collecting and Analyzing Data Jointly from Multiple Services under Local Differential Privacy, in *Proceedings of the 46th International Conference on Very Large Data Bases* (*VLDB'20*), 2020.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 13(11): 2760-2772, 2020)

65. Zhuolun Xiang, Bolin Ding, Xi He, and Jingren Zhou: Linear and Range Counting under Metric-based Local Differential Privacy, in *Proceedings of the IEEE International Symposium on Information Theory* (*ISIT'20*), 2020.

66. Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li: Simple and Deep Graph Convolutional Networks, in *Proceedings of the 37th International Conference on Machine Learning* (*ICML'20*), 2020.

67. Ming Chen, Zhewei Wei, Bolin Ding, Yaliang Li, Ye Yuan, Xiaoyong Du, and Ji-Rong Wen: Scalable Graph Neural Networks via Bidirectional Propagation, in *Proceedings of Advances in Neural Information Processing Systems 33* (*NeurIPS'20*), 2020.

68. Zhiqiang Tao, Yaliang Li, Bolin Ding, Ce Zhang, Jingren Zhou, and Yun Fu: Learning to Mutate with Hypergradient Guided Population, in *Proceedings of Advances in Neural Information Processing Systems 33* (*NeurIPS'20*), 2020.

69. Huaxiu Yao, Xian Wu, Zhiqiang Tao, Yaliang Li, Bolin Ding, Ruirui Li, and Zhenhui Li: Automated Relational Meta-learning, in *Proceedings of the 8th International Conference on Learning Representations* (*ICLR'20*), 2020.

70. Hengtong Zhang, Yaliang Li, Bolin Ding, and Jing Gao: Practical Data Poisoning Attack Against Next-Item Recommendation, in *Proceedings of the ACM Web Conference* (*WWW'20*), 2020.

71. Ruiyang Ren, Zhaoyang Liu, Yaliang Li, Wayne Xin Zhao, Hui Wang, Bolin Ding, and Ji-Rong Wen: Sequential Recommendation with Self-attentive Multi-adversarial Network, in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (*SIGIR'20*), 2020.

72. Daoyuan Chen, Yaliang Li, Minghui Qiu, Zhen Wang, Bofang Li, Bolin Ding, Hongbo Deng, Jun Huang, Wei Lin, and Jingren Zhou: AdaBERT: Task-Adaptive BERT Compression with Differentiable Neural Architecture Search, in *Proceedings of the 29th International Joint Conference on Artificial Intelligence* (*IJCAI'20*), 2020.

73. Zhaoyang Liu, Haokun Chen, Fei Sun, Xu Xie, Jinyang Gao, Bolin Ding, and Yanyan Shen: Intent Preference Decoupling for User Representation on Online Recommender System, in *Proceedings of the 29th International Joint Conference on Artificial Intelligence* (*IJCAI'20*), 2020.

74. Yaliang Li, Houping Xiao, Zhan Qin, Chenglin Miao, Lu Su, Jing Gao, Kui Ren, and Bolin Ding: Towards Differentially Private Truth Discovery for Crowd Sensing Systems, in *Proceedings of the 40th IEEE International Conference on Distributed Computing Systems* (*ICDCS'20*), 2020.

75. Min Xu, Tianhao Wang, Bolin Ding, Jingren Zhou, Cheng Hong, and Zhicong Huang: DPSAaS: Multi-Dimensional Data Sharing and Analytics as Services under Local Differential Privacy, in *Proceedings of the 45th International Conference on Very Large Data Bases* (*VLDB'19*), 2019.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 12(12): 1862-1865, 2019)

76. Tianhao Wang, Bolin Ding, Jingren Zhou, Cheng Hong, Zhicong Huang, Ninghui Li, and Somesh Jha: Answering Multi-Dimensional Analytical Queries under Local Differential Privacy, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'19*), 2019.

77. Joshua Allen, Bolin Ding, Janardhan Kulkarni, Harsha Nori, Olga Ohrimenko, and Sergey Yekhanin: An Algorithmic Framework for Differentially Private Data Analysis on Trusted Processors, in *Proceedings of Advances in Neural Information Processing Systems 32* (*NeurIPS'19*), 2019.

78. Cedric Renggli, Bojan Karlas, Bolin Ding, Feng Liu, Kevin Schawinski, Wentao Wu, and Ce Zhang: Continuous Integration of Machine Learning Models: A Rigorous Yet Practical Treatment, in *Proceedings of the 2nd SysML Conference* (*SysML'19*), 2019.

79. Silu Huang, Chi Wang, Bolin Ding, and Surajit Chaudhuri: Efficient Identification of Approximate Best Configuration of Training in Large Datasets, in *Proceedings of the 33rd AAAI Conference on Artificial Intelligence* (*AAAI'19*), 2019.

80. Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin: A Distributed Algorithm for Differentially Private Heavy Hitters, in *Privacy Preserving Machine Learning, NeurIPS Workshop* (*PPML'18*), 2018.

81. Joshua Allen, Bolin Ding, Janardhan Kulkarni, Harsha Nori, Olga Ohrimenko, and Sergey Yekhanin: An Algorithmic Framework For Differentially Private Data Analysis on Trusted Processors, in *Privacy Preserving Machine Learning, NeurIPS Workshop* (*PPML'18*), 2018.

82. Azade Nazi, Bolin Ding, Vivek R. Narasayya, and Surajit Chaudhuri: Efficient Estimation of Inclusion Coefficient using HyperLogLog Sketches, in *Proceedings of the 44th International Conference on Very Large Data Bases* (*VLDB'18*), 2018.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 11(10): 1097-1109, 2018)

83. Yaliang Li, Chenglin Miao, Lu Su, Jing Gao, Qi Li, Bolin Ding, Zhan Qin, and Kui Ren: An Efficient Two-Layer Mechanism for Privacy-Preserving Truth Discovery, in *Proceedings of the 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'18*), 2018.

84. Adam Dziedzic, Jingjing Wang, Sudipto Das, Bolin Ding, Vivek R. Narasayya, and Manoj Syamala: Columnstore and B+ tree - Are Hybrid Physical Designs Important?, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'18*), 2018.

85. Bolin Ding, Harsha Nori, Paul Li, and Joshua Allen: Comparing Population Means under Local Differential Privacy: with Significance and Power, in *Proceedings of the 32nd AAAI Conference on Artificial Intelligence* (*AAAI'18*), 2018.

86. Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin: Collecting Telemetry Data Privately, in *Proceedings of Advances in Neural Information Processing Systems 30* (*NIPS'17*), 2017.

87. Yongxin Tong, Libin Wang, Zimu Zhou, Bolin Ding, Lei Chen, Jieping Ye, and Ke Xu: Flexible Online Task Assignment in Real-Time Spatial Data, in *Proceedings of the 43rd International Conference on Very Large Data Bases* (*VLDB'17*), 2017.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 10(11): 1334-1345, 2017)

88. Surajit Chaudhuri, Bolin Ding, and Srikanth Kandula: Approximate Query Processing: No Silver Bullet, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'17*), 2017.
SIGMOD Keynote Talk

89. Dominik Moritz, Danyel Fisher, Bolin Ding, and Chi Wang: Trust, but Verify: Optimistic Visualizations of Approximate Queries for Exploring Big Data, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (*CHI'17*), 2017.

90. Yongxin Tong, Jieying She, Bolin Ding, Lei Chen, Tianyu Wo, Ke Xu: Online Minimum Matching in Real-Time Spatial Data: Experiments and Analysis, in *Proceedings of the 42nd International Conference on Very Large Data Bases* (*VLDB'16*), 2016.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 9(12): 1053-1064, 2016)

91. Samuel Haney, Ashwin Machanavajjhala, and Bolin Ding: Design of Policy-Aware Differentially Private Algorithms, in *Proceedings of the 42nd International Conference on Very Large Data Bases* (*VLDB'16*), 2016.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 9(4): 264-275, 2015)

92. Bolin Ding, Silu Huang, Surajit Chaudhuri, Kaushik Chakrabarti, and Chi Wang: Sample + Seek: Approximating Aggregates with Distribution Precision Guarantee, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'16*), 2016.

93. Kukjin Lee, Arnd Christian Konig, Vivek Narasayya, Bolin Ding, Surajit Chaudhuri, Brent Ellwein, Alexey Eksarevskiy, Manbeen Kohli, Jacob Wyant, Praneeta Prakash, Rimma Nehme, Jiexing Li, and Jeff Naughton: Operator and Query Progress Estimation in Microsoft SQL Server Live Query Statistics, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'16*), 2016.

94. Srikanth Kandula, Anil Shanbhag, Aleksandar Vitorovic, Matthaios Olma, Robert Grandl, Surajit Chaudhuri, and Bolin Ding: Quickr: Lazily Approximating Complex Ad-Hoc Queries in Big Data Clusters, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'16*), 2016.

95. Haiming Jin, Lu Su, Bolin Ding, Klara Nahrstedt, and Nikita Borisov: Enabling Privacy-Preserving Incentives for Mobile Crowd Sensing Systems, in *Proceedings of the 36th IEEE International Conference on Distributed Computing Systems (ICDCS'16)*, 2016.

96. Yongxin Tong, Jieying She, Bolin Ding, Libin Wang, and Lei Chen: Online Mobile Micro-Task Allocation in Spatial Crowdsourcing, in *Proceedings of the 32nd International Conference on Data Engineering* (*ICDE'16*), 2016.

97. Fotis Psallidas, Bolin Ding, Kaushik Chakrabarti, and Surajit Chaudhuri: S4: Top-k Spreadsheet-Style Search for Query Discovery, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'15*), 2015.

98. Mohan Yang, Bolin Ding, Surajit Chaudhuri, and Kaushik Chakrabarti: Finding Patterns in a Knowledge Base using Keywords to Compose Table Answers, in *Proceedings of the 41st International Conference on Very Large Data Bases* (*VLDB'15*), 2015.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 7(14): 1809-1820, 2014)

99. Herodotos Herodotou, Bolin Ding, Shobana Balakrishnan, Geoff Outhred, and Percy Fitter: Scalable Near Real-Time Failure Localization of Data Center Networks, in *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'14*), 2014.

100. Yanyan Shen, Kaushik Chakrabarti, Surajit Chaudhuri, Bolin Ding, and Lev Novik: Discovering Queries based on Example Tuples, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'14*), 2014.

101. Xi He, Ashwin Machanavajjhala, and Bolin Ding: Blowfish Privacy: Tuning Privacy-Utility Trade-offs using Policies, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'14*), 2014.

102. Zhenhui Li, Bolin Ding, Fei Wu, Tobias Kin Hou Lei, Roland Kays, and Margaret C. Crofoot: Attraction and Avoidance Detection from Movements, in *Proceedings of the 39th International Conference on Very Large Data Bases* (*VLDB'13*), 2013.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 7(3): 157-168, 2013)

103. Fangbao Tao, Bolin Ding, and et al.: EventCube: Multi-Dimensional Search and Mining of Structured and Text Data, in *Proceedings of the 19th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'13*), 2013.

104. Bolin Ding, Haixun Wang, Ruoming Jin, Jiawei Han, and Zhongyuan Wang: Optimizing Index for Taxonomy Keyword Search, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'12*), 2012.

105. Arnd Christian König, Bolin Ding, Surajit Chaudhuri, and Vivek Narasayya: A Statistical Approach Towards Robust Progress Estimation, in *Proceedings of the 38th International Conference on Very Large Data Bases* (*VLDB'12*), 2012.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 5(4): 382-393, 2011)

106. Yongxin Tong, Lei Chen, and Bolin Ding: Discovering Threshold-based Frequent Closed Itemsets over Probabilistic Data, in *Proceedings of the 28th International Conference on Data Engineering* (*ICDE'12*), 2012.

107. Lu Su, Yong Yang, Bolin Ding, Jing Gao, Tarek F. Abdelzaher, and Jiawei Han: Hierarchical Aggregate Classification with Limited Supervision for Data Reduction in Wireless Sensor Networks, in *Proceedings of the 9th International Conference on Embedded Networked Sensor Systems* (*SenSys'11*), 2011.

108. Ruoming Jin, Lin Liu, Bolin Ding, and Haixun Wang: Distance-Constraint Reachability Computation in Uncertain Graphs, in *Proceedings of the 37th International Conference on Very Large Data Bases* (*VLDB'11*), 2011.
(Also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 4(9): 551-562, 2011)

109. Bolin Ding, and Arnd Christian König: Fast Set Intersection in Memory, in *Proceedings of the 37th International Conference on Very Large Data Bases* (*VLDB'11*), 2011.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 4(4): 255-266, 2011)

110. Zhenhui Li, Cindy Xide Lin, Bolin Ding, and Jiawei Han: Mining Significant Time Intervals for Relationship Detection, in *Proceedings of 2011 International Symposium on Spatial and Temporal Databases* (*SSTD'11*), 2011.

111. Bolin Ding, Marianne Winslett, Jiawei Han, and Zhenhui Li: Differentially Private Data Cubes: Optimizing Noise Sources and Consistency, in *Proceedings of ACM Conference on Management of Data* (*SIGMOD'11*), 2011.

112. Bo Zhao, Jiawei Han, Cindy Xide Lin, and Bolin Ding: TEXplorer: Keyword based Object Ranking and Exploration in Multidimensional Text Databases, in *Proceedings of 2011 ACM International Conference on Information and Knowledge Management* (*CIKM'11*), 2011.

113. David Lo, Bolin Ding, Lucia, and Jiawei Han: Bidirectional Mining of Non-Redundant Recurrent Rules from a Sequence Database, in *Proceedings of the 27th International Conference on Data Engineering* (*ICDE'11*), 2011.

114. Xin Jin, Jiawei Han, Liangliang Cao, Jiebo Luo, Bolin Ding, and Cindy Xide Lin: Visual Cube and On-Line Analytical Processing of Images, in *Proceedings of 2010 ACM International Conference on Information and Knowledge Management* (*CIKM'10*), 2010.

115. Bolin Ding, Yintao Yu, Bo Zhao, Cindy Xide Lin, Jiawei Han, and ChengXiang: Keyword Search in Text Cube: Finding Top-K Aggregated Cell Documents, in *Proceedings of 2010 NASA Conference on Intelligent Data Understanding* (*CIDU'10*), 2010.

116. Zhenhui Li, Bolin Ding, Jiawei Han, and Roland Kays: Swarm: Mining Relaxed Temporal Moving Object Clusters, in *Proceedings of the 36th International Conference on Very Large Data Bases* (*VLDB'10*), 2010.
(also appeared in the Journal "*Proceedings of the VLDB Endowment*" (*PVLDB*), 3(1): 723-734, 2010)

117. Zhenhui Li, Bolin Ding, Jiawei Han, and Roland: Mining Periodic Behaviors for Moving Objects, in *Proceedings of the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (*KDD'10*), 2010.

118. Bolin Ding, Bo Zhao, Cindy Xide Lin, Jiawei Han, and Chengxiang Zhai: TopCells: Keyword-Based Search of Top-k Aggregated Documents in Text Cube, in *Proceedings of the 26th International Conference on Data Engineering* (*ICDE'10*), 2010.

119. Lu Su, Bolin Ding, Yong Yang, Tarek Abdelzaher, and Guohong Cao: oCast: Optimal Multicast Routing Protocol for Wireless Sensor Networks, in *Proceedings of the 17th IEEE International Conference on Network Protocols* (*ICNP'09*), 2009.

120. Bolin Ding, David Lo, Jiawei Han, and Siau-Cheng Khoo: Efficient Mining of Closed Repetitive Gapped Subsequences from a Sequence Database, in *Proceedings of the 25th International Conference on Data Engineering* (*ICDE'09*), 2009.

121. Cindy Xide Lin, Bolin Ding, Jiawei Han, Feida Zhu, and Bo Zhao: Text Cube: Computing IR Measures for Multidimensional Text Database Analysis, in *Proceedings of 2008 International Conference on Data Mining* (*ICDM'08*), 2008.

122. Luiz Mendes, Bolin Ding, and Jiawei Han: Stream Sequential Pattern Mining with Precise Error Bounds, in *Proceedings of 2008 International Conference on Data Mining* (*ICDM'08*), 2008.

123. Lu Qin, Jeffrey Xu Yu, Bolin Ding, and Yoshiharu Ishikawa: Monitoring Aggregate k-NN Objects in Road Networks, in *Proceedings of the 20th International Conference on Scientific and Statistical Database Management* (*SSDBM'08*), 2008.

124. Bolin Ding, Jeffrey Xu Yu, and Lu Qin: Finding Time-Dependent Shortest Paths over Large Graphs, in *Proceedings of the 11th International Conference on Extending Database Technology* (*EDBT'08*), 2008.

125. Jiefeng Cheng, Jeffrey Xu Yu, Bolin Ding, Haixun Wang, and Philip S. Yu: Fast Graph Pattern Matching, in *Proceedings of the 24th International Conference on Data Engineering* (*ICDE'08*), 2008.

126. Lu Qin, Jeffrey Xu Yu, and Bolin Ding: TwigList: Make Twig Pattern Matching Fast, in *Proceedings of the 12th International Conference on Database Systems for Advanced Applications* (*DASFAA'07*), 2007.

127. Jiefeng Cheng, Jeffrey Xu Yu, and Bolin Ding: Cost-Based Query Optimization for Multi Reachability Joins, in *Proceedings of the 12th International Conference on Database Systems for Advanced Applications* (*DASFAA'07*), 2007.

128. Bolin Ding, Jeffrey Xu Yu, Shan Wang, Lu Qin, Xiao Zhang, and Xuemin Lin: Finding Top-k Min-Cost Connected Trees in Databases, in *Proceedings of the 23rd International Conference on Data Engineering* (*ICDE'07*), 2007.
Best Student Paper Award

129. Shan Wang, Zhaohui Peng, Jun Zhang, Lu Qin, Sheng Wang, Jeffrey Xu Yu, Bolin Ding: NUITS: A Novel User Interface for Efficient Keyword Search over Databases, in *Proceedings of the 32nd International Conference on Very Large Data Bases* (*VLDB'06*), 2006.

Refereed Journal Publications

130. Ziqian Chen, Fei Sun, Yifan Tang, Haokun Chen, Jinyang Gao, and Bolin Ding: Studying the Impact of Data Disclosure Mechanism in Recommender Systems via Simulation, *ACM Transactions on Information Systems* (*TOIS*), 41(3): 60:1-60:26, 2023.

131. Ergute Bao, Xiaokui Xiao, Jun Zhao, Dongping Zhang, and Bolin Ding: Synthetic Data Generation with Differential Privacy via Bayesian Networks, *Journal of Privacy and Confidentiality*, 11(3), 2023.

132. Yongxin Tong, Yuxiang Zeng, Bolin Ding, Libin Wang, and Lei Chen: Two-Sided Online Micro-Task Assignment in Spatial Crowdsourcing, *IEEE Transactions on Knowledge and Data Engineering* (*TKDE*), 33(5): 2295-2309, 2021.

133. Zhenhui Li, Jiawei Han, Bolin Ding, and Roland Kays: Mining Periodic Behaviors of Object Movements for Animal and Biological Sustainability Studies, *Data Mining and Knowledge Discovery*, 24(2): 355-386, 2012.

134. Bolin Ding, Bo Zhao, Cindy Xide Lin, Jiawei Han, Chengxiang Zhai, Ashok Srivastava, and Nikunj C. Oza: Efficient Keyword-Based Search for Top-K Cells in Text Cube, *IEEE Transactions on Knowledge and Data Engineering* (*TKDE*) (Special Issue: Keyword Search on Structured Data), 23(12): 1795-1810, 2011.

135. Zhenhui Li, Jiawei Han, Ming Ji, Lu-An Tang, Yintao Yu, Bolin Ding, Jae-Gil Lee, and Roland Kays: MoveMine: Mining Moving Object Data for Discovery of Animal Movement Patterns, *ACM Transactions on Intelligent Systems and Technology* (*TIST*), 2(4): 37, 2011.

136. Wenfei Fan, Jeffrey Xu Yu, Jianzhong Li, Bolin Ding, and Lu Qin: Query Translation from XPath to SQL in the Presence of Recursive DTDs, *VLDB Journal*, 18(4): 857-883, 2009.

137. Cheqing Jin, Bolin Ding, and Jeffrey Xu Yu: Making Filters Smart in Distributed Data Stream Environments, *Information Sciences*, 179(9): 1348-1361, 2009.

138. Jing Gao, Bolin Ding, Wei Fan, Jiawei Han, and Philip S. Yu: Classifying Data Streams with Skewed Class Distribution and Concept Drifts, *IEEE Internet Computing* (Special Issue: Data Stream Management), 12(6): 37-49, 2008.

139. Qing Zhu, Shan Wang, Bolin Ding, Xiao Zhang, Hongyan Cai, and Jiali Yao: Service-Oriented Search Algorithm on Data Grid, *Chinese Journal of Computers*, 29(7): 1234-1240, 2006 (in Chinese).

Miscellaneous

140. Jiawei Han and Bolin Ding, "Stream Mining", in Ling Liu and M. Tamer Özsu (Eds.), *Encyclopedia of Database Systems*, Springer, 2009.

SOME RESEARCH PRESENTATIONS

1. "Learned Query Optimizer: What is New and What is Next" (a conference tutorial), ACM Conference on Management of Data (*SIGMOD'24*), Santiago, Chile, June 2024.

2. "AI for Systems" (a panelist in the conference panel discussion), ACM Conference on Management of Data (*SIGMOD'24*), Santiago, Chile, June 2024.

3. "Answering Multi-Dimensional Analytical Queries under Local Differential", The 7th Annual Conference on Machine Learning and Systems (*MLSys'24*), Santa Clara, USA, May 2024.

4. "Lero: A Learning-to-Rank Query Optimizer", the 49th International Conference on Very Large Data Bases (*VLDB'23*), Vancouver, Canada, August 2023.

5. "A Practical Introduction to Federated Learning", the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (*KDD'22*), Washington DC, USA, August 2022.

6. "AutoML: A Perspective where Industry Meets Academy", the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (*KDD'21*), Singapore (virtual conference), August 2021.

7. "FlashP: An Analytical Pipeline for Real-time Forecasting of Time-Series Relational Data", the 47th International Conference on Very Large Data Bases (*VLDB'21*), Copenhagen, Denmark (virtual conference), August 2021.

8. "Answering Multi-Dimensional Analytical Queries under Local Differential", ACM Conference on Management of Data (*SIGMOD'19*), Amsterdam, The Netherlands, June 2019.

9. "Efficient Estimation of Inclusion Coefficient using HyperLogLog Sketches", the 44th International Conference on Very Large Data Bases (*VLDB'18*), Rio de Janeiro, Brazil, August 2018.

10. "Sample + Seek: Approximating Aggregates with Distribution Precision Guarantee", ACM Conference on Management of Data (*SIGMOD'16*), San Francisco, USA, June 2016.

11. "Scalable Near Real-Time Failure Localization of Data Center Networks", the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (*KDD'14*), New York, USA, August 2014.

12. "Supporting Efficient and Effective Keyword Search: From Set Intersection to Taxonomies and Text Cubes", Microsoft Research Asia, Beijing, China, March 2012.

13. "Supporting Efficient and Effective Keyword Search: From Set Intersection to Taxonomies and Text Cubes", Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University, Beijing, China, March 2012.

14. "Supporting Efficient and Effective Keyword Search: From Set Intersection to Taxonomies and Text Cubes", Microsoft Research, Redmond, Washington, USA, February 2012.

15. "Privacy-Preserving Data Publishing and Analytics using Data Cubes", NEC Laboratories America, Princeton, New Jersey, USA, February 2012.

16. "Fast Set Intersection in Memory", Google Research, Mountain View, California, USA, February 2012.

17. "Keyword Search in Structural Data: Models, Algorithms, and Operators", The Yahoo!-DAIS Seminar, Department of Computer Science, University of Illinois at Urbana-Champaign, January 2012.

18. "Fast Set Intersection in Memory", the 37th International Conference on Very Large Data Bases (*VLDB'11*), Seattle, Washington, USA, August 2011.

19. "Differentially Private Data Cubes: Optimizing Noise Sources and Consistency", ACM Conference on Management of Data (*SIGMOD'11*), Athens, Greece, June 2011.

20. "Differentially Private Data Cubes: Optimizing Noise Sources and Consistency", The Yahoo!-DAIS Seminar, Department of Computer Science, University of Illinois at Urbana-Champaign, April 2011.

21. "TopCells: Keyword-Based Search of Top-k Aggregated Documents in Text Cube", the 26th International Conference on Data Engineering (*ICDE'10*), Long Beach, California, USA, March 2010.

22. "The State of the Art in Anonymized Data Publishing: Algorithms, Complexity, and Deficiency", Trust and Security Seminar, Information Trust Institute, University of Illinois at Urbana-Champaign, April 2009.

23. "Efficient Mining of Closed Repetitive Gapped Subsequences from a Sequence Database", The Yahoo!-DAIS Seminar, Department of Computer Science, University of Illinois at Urbana-Champaign, October 2008.

24. "Finding Frequent Items in Data Streams", Theory Seminar, Department of Computer Science, University of Illinois at Urbana-Champaign, October 2007.

25. "Finding Top-k Min-Cost Connected Trees in Databases", the 23rd International Conference on Data Engineering (*ICDE'07*), Istanbul, Turkey, April 2007.

SELECTED PATENTS

- Firecommendation Method, Device, and System for Distributed Privacy-preserving Learning, US20230161905A1
  filed by Alibaba, 2022

- Item Recommendation Method and Device for Protecting User Privacy and Learning System, US20230162262A1
  filed by Alibaba, 2022

- Model Processing Method, Apparatus, Storage Medium, and Processor, US20210357752A1
  filed by Alibaba, 2021

- Feature Interaction Via Edge Search, WO2022011553A1
  filed by Alibaba, 2020

- Hyperparameter Recommendation for Machine Learning Method, WO2021000244A1
  filed by Alibaba, 2019

- Data Sharing and Data Analytics Implementing Local Differential Privacy, WO2020248149A1
  filed by Alibaba, 2019

- Method and System for Answering Multi-dimensional Analytical Queries under Local Differential Privacy, WO2020248150A1
  filed by Alibaba, 2019

- Remote Testing Analysis for Software Optimization based on Client-side Local Differential Privacy-based Data, US10902149B2
  filed by Microsoft, 2018

- Database Management using Hyperloglog Sketches, US20190384830A1
  filed by Microsoft, 2018

- Efficient Configuration Selection for Automated Machine Learning, US20200065712A1
  filed by Microsoft, 2018

- Hardware Protection for Differential Privacy, US20190147188A1
  filed by Microsoft, 2017

- Collection of Sensitive Data---such as software usage data or other telemetry data---over Repeated Collection Cycles in Satisfaction of Privacy Guarantees, US10776242B2
  filed as MS #362082.01 by Microsoft, 2017

- Database Queries for Histograms, US10949438B2
  filed as MS #361817.01 by Microsoft, 2017

- Difference Visualization between Data Sets, US10453228B2
  filed as MS #361820.01 by Microsoft, 2017

- Fast Approximate Results and Slow Precise Results, US10552435B2
  filed as MS #361819.01 by Microsoft, 2017

- Workload-driven Recommendations for Columnstore and Rowstore in RDBs, US20180096006A1
  filed as MS #360797.01 by Microsoft, 2017

- Aggregate-query Database System and Processing, US10740328B2
  filed as MS #359953.01 by Microsoft, 2016

- Controlling Approximations of Queries, US10496643B2
  filed as MS #358060.01 by Microsoft, 2016

- Finding Patterns in a Knowledge Base to Compose Table Answers, US20150310073A1
  filed as MS #341279.01 by Microsoft, 2014

- Query Progress Estimation, US20130151504A1
  filed as MS# 334457.01 by Microsoft, 2011

- Fast Set Intersection, US20110314045A1
  filed as MS #329216.01 by Microsoft, 2010

PROFESSIONAL SERVICE

- **NSF Panelist:** 2016

- **Reviewer for Journals:** ACM Transactions on Database Systems, ACM SIGMOD Record, IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on Knowledge Discovery from Data, Theoretical Computer Science, Pattern Recognition, Information Sciences, Knowledge and Information Systems
- **Information Director:** ACM Transactions on Knowledge Discovery from Data, 2009 – 2011
- **Some Program Committee Memberships for Conferences:**
  - ACM International Conference on Management of Data (*SIGMOD*): 2022, 2021, 2020
  - International Conference on Very Large Data Bases (*PVLDB*): 2025, 2024, 2023, 2020, 2019, 2018, 2017
  - IEEE International Conference on Data Engineering (*ICDE*): 2023, 2022, 2021, 2020, 2019
  - Conference on Information and Knowledge Management (CIKM): 2024 (senior PC)
  - ACM Conference on Computer and Communications Security (*CCS*): 2023, 2022
  - ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (*KDD*): 2024 (area chair), 2023 (area chair), 2022, 2021, 2020, 2019, 2018, 2017
  - Neural Information Processing Systems (*NeurIPS*): 2024 (area chair), 2023 (area chair), 2022 (area chair), 2021 (area chair), 2020
  - International Conference on Machine Learning (*ICML*): 2024 (area chair), 2023 (area chair), 2022, 2021
  - International Conference on Learning Representations (*ICLR*): 2024, 2023, 2022, 2021
  - International Conference on Web and Internet Economics (*WINE*): 2022 (senior PC)
  - International Workshop on Privacy-Preserving Data Publication and Analysis (*PrivDB* in *ICDE*): 2013

AWARDS AND HONORS

- Best Paper Award in ADS, KDD'22 — 2022
- AI 2000 Most Influential Scholar Award Honorable Mention in Database — 2022-2024
- 1st place, 2020 NIST Differential Privacy Temporal Map Challenge — 2020
- FY17 Technical Excellence (the only one in Microsoft), Microsoft Privacy, Microsoft — 2017
- Gold, The 2nd Yahoo!-DAIS Research Excellence Award Competition, Yahoo! and UIUC — 2012
- SIGMOD Travel Award (supported by NSF, SAP and the SIGMOD Executive Committee) — 2012
- 1st place, TopCoder Programming Competition College Tour (at University of Illinois) — 2007
- Richard T. Cheng Fellowship Award, University of Illinois at Urbana-Champaign — 2007-08
- Best Student Paper Award, ICDE'07 — 2007
- Outstanding Graduating Senior Award, Renmin University of China — Spring 2005
- Honorable Mention, 2005 ACM-ICPC Programming Contest World Finals — 2005
- 3rd place out of 255 teams, Gold Medal, ACM-ICPC Asia Regional Contest, Shanghai Site — 2004
- 6th place out of 361 teams, Silver Medal, ACM-ICPC Asia Regional Contest, Beijing Site — 2004
- Computerworld Scholarship, China Computerworld Newspaper (2 recipients in RUC) — Fall 2004
- Scholarship for Innovation in Science and Technology, Renmin University (4 recipients) — Fall 2003
- 2nd prize nationwide, China Undergraduate Mathematical Contest in Modeling — 2003
- Scholarship for Academic Excellence, Renmin University of China — 2002,03,04
- 1st place in Yunnan Province, National Olympiad in Informatics, China — 2000
- 8th place in Yunnan Province, National Olympiad in Physics, China — 2000
- 27th place nationwide, National Olympiad in Informatics, China — 1999

TEACHING EXPERIENCE

- *Teaching Assistant*, University of Illinois at Urbana-Champaign
  CS512: Data Mining: Principles and Algorithms, Spring 2012

- *Teaching Assistant*, The Chinese University of Hong Kong
  ERG2018A/B: Advanced Engineering Mathematics, Fall 2005 and Fall 2006
  SEG3550A: Fundamentals in Information System, Spring 2006

- *Team Leader and Student Supervisor*, Renmin University of China
  ACM-ICPC Programming Contest Team, Summer 2003 – Spring 2005