Jie Liu

392 Noble Dr – Merced – CA, 95348

□ +1 469-910-4827 • ✉ jliu279@ucmerced.edu • ◉ github.com/caterby

RESEARCH INTEREST

o (1) System optimization (especially runtime system) for efficient machine learning in parallel and distributed systems; (3) Database system; (3) On-device artificial intelligence.

EDUCATION

University of California, Merced   Aug. 2018 - Dec. 2022
    Ph.D. in Computer Science   *GPA: 3.9/4.00*
The University of Texas at Dallas, Richardson, TX   Aug. 2016 - May. 2018
    M.S. in Intelligent System   *GPA: 3.62/4.00*
Xi'an Jiaotong University, Shaanxi, China   Aug. 2013 - May. 2016
    M.S. in Software Engineering   *GPA: 3.27/4.00*
Northwestern Polytechnic University, Shaanxi, China   Aug. 2009 - May. 2013
    B.S. in Software Engineering   *GPA: 3.52/4.00*

PROFESSIONAL EXPERIENCE

Research Intern - Futurewei Cloud Lab
Huawei Research Group, Seattle, Washington   *May. 2020 - Aug. 2020*

o Using deep learning models to estimate the cardinality of SQL queries and the estimation confidence, the state of the art estimation accuracy can be achieved.

o Proposed a novel method to encode the SQL queries into more appropriate vectors for the learning models.

o The selectivity estimation algorithm can produce increasingly more accurate answers as it processes more queries.

Graduate Research Assistant - Parallel Architecture, System, and Algorithm Lab (PASA)
University of California, Merced   *Jan. 2019 - Now*

o Introduced a benchmark suite and a tool to study the performance of training deep learning models on a representative resource constraint mobile device (the NVIDIA TX2).

o Used an adaptive neural network-based approximation to accelerate high performance applications.

o Introduced an auto-labeling system which includes a runtime system that efficiently schedules and executes auto-labeling workloads on heterogeneous mobile processors. The system was implemented using C++ with Native Development Kit (NDK) on Android 9.0 and evaluated on Samsung S9 with Snapdragon 845 SoC.

Graduate Research Assistant - Data Management and Analytics Lab
University of Texas at Dallas, Richardson, Texas   *Mar. 2017 - May. 2018*

o Developed efficient techniques for addressing challenges in Data Stream Mining, e.g., Concept Drift (evolving class boundary), emergence of novel classes, scarcity of labeled data, Sampling-bias, Class Imbalance, etc.

RESEARCH PROJECTS

Self-Adaptive Auto-Labeling System for Heterogeneous Mobile Processors
(*Machine Learning, System Design, Java, C++, OpenCL*), CSE Dept, UC Merced   *Nov. 2019 - May. 2020*

o Propose a self-adaptive auto-labeling system to assign labels for data on mobile devices.

o Presented a new labeling algorithm to process dynamically generated data by mobile devices with unknown labels.

o Introduced a high performance runtime system to efficiently leverage heterogeneous mobile processors for labeling.

Characterization of Training Deep Learning Models on Mobile Device
(*Deep Learning, High Performance Computing, Python*), CSE Dept, UC Merced   *Jan. 2019 - May. 2019*

o Performed a variety of experiments on mobile devices to study the performance of training deep learning models.

o Introduce a too to study performance of training deep learning models on mobile devices, from the perspectives of memory consumption, hardware utilization, and power consumption.

Multistream Classification with Relative Density Ratio

(*Data Mining, Python*), *CSE Dept, UT-Dallas* *Mar. 2017 - Mar. 2018*

o Completed a multi-stream classification framework with asynchronous concept drift detection.

o Proposed a gradient-based technique to automatically learn model parameters for relative density ratio estimation.

o Used real-world and synthetic datasets to empirically evaluate our approach, and compared the results with baseline.

Research and Application of Word and Text Similarity Computing Method

(*Text Data Processing, Text Mining , Java*), *XJTU* *Aug. 2015 - Jun. 2016*

o Designed and implemented a novel algorithm for computing the similarity of Chinese words which merges synonyms based on the semantic information provided by a most widely used Chinese Thesaurus.

o Developed algorithms for computing the Chinese text syntactic and semantic similarity via maximum bipartite matching.

o Built a system named Abstract Similarity Comparison through various text mining techniques like text parsing, TF-IDF, Vector Space Model, which can be applied into Text Mining and Plagiarism Detection.

PUBLICATIONS

o **Jie Liu**; Wenqian Dong; Qingqing Zhou; Dong Li, "Deep Ensembles with Uncertainty for Cardinality Estimation". In 47th International Conference on Very Large Data Bases (**VLDB To Appear**), Copenhagen, Denmark, Aug 2021.

o **Jie Liu**; Jiawen Liu; Zhen Xie; Dong Li, "Falme: A Self-Adaptive Auto-Labeling System for Heterogeneous Mobile Processors". (**ICDE Under Submission**), Ottawa, Canada, July 2022.

o Jiawen Liu; Zhen Xie; **Jie Liu**; Dong Li, "RIANN: Real-time Incremental Learning with Approximate Nearest Neighbor on Mobile Devices". (**ICPP Under Submission**), Chicago, Illinois, Aug 2021.

o Zhen Xie; Wenqian Dong; **Jie Liu**; Ivy Peng, Yanbao Ma, Dong Li, "MD-PM: Memoization-based Molecular Dynamics Simulation with Non-Volatile Memory". In 35th ACM International Conference on Supercomputing (**ICS**), Paris, France, June 2021. (Acceptance rate: **24.2%**)

o **Jie Liu**; Jiawen Liu; Zhen Xie; Dong Li, "Falme: A Self-Adaptive Auto-Labeling System for Heterogeneous Mobile Processors". In Proceedings of the 3rd Conference on Machine Learning and Systems (**MLSys**), On-device Intelligence Workshop, Austin, TX, March 2020.

o **Jie Liu**; Jiawen Liu; Wan Du; Dong Li, "Performance Analysis and Characterization of Training Deep Learning Models on Mobile Device", *IEEE 25th International Conference on Parallel and Distributed Systems (**ICPADS**), 2019. (Acceptance rate: **28%**)*

o Wenqian Dong; **Jie Liu**; Zhen Xie; Dong Li, "Adaptive Neural Network-Based Approximation to Accelerate Eulerian Fluid Simulation", *In 31st ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (**SC**), 2019. (Acceptance rate: **22%**)*

o Haque A.; Hemeng T.; Chandra S.; **Jie Liu**; Khan L., "A Framework for Multistream Regression with Direct Density Ratio Estimation", *In 32th AAAI Conference on Artificial Intelligence (**AAAI**), New Orleans, Lousiana, 2018. (Acceptance rate: **17%**)*

o Chandra S.; Haque A.; Hemeng T.; **Jie Liu**; Khan L., "Ensemble Direct Density Ratio Estimation for Multistream Classification", *In 34th IEEE International Conference on Data Engineering (**ICDE**), Paris, France, 2018. (Acceptance rate: **19%**)*

SELECTED ACADEMIC PROJECT

Simulate Tomasulo's Algorithm with Simultaneous Multithreading

o Simulated a modified version of the PowerPc 604 architecture which is a speculative, multi-issue, out of order ALU.

o Added Simultaneous Multithreading capabilities to architecture so that it executes two programs simultaneously.

o Implemented a simplified simulator to evaluate distributed directory cache coherence protocols in CMPs. .

TECHNICAL SKILLS

o **Programming Languages:** Java, Python, C++, C, CUDA, OpenCL, CSS, JavaScript
o **Database & Visualization:** MS SQL, Teradata, Oracle, MySQL, Tableau, PL/SQL
o **Data Analysis:** R, Weka, Scikit-Learn, NLTK, MATLAB, Tensorflow

RELEVANT COURSEWORK

o Computer Architecture	Artificial Intelligence	Design and Analysis of Computer Algorithms
o Machine Learning	Natural Language Processing	Statistical Methods in AI and ML
o Parallel Computing	Operating System	Distributed Computing System