Dong Hu

Lally 09, Rensselaer Polytechnic Institute, Troy, NY
(206)890-1181 · hud3@rpi.edu · https://jurohd.github.io

JOB INTEREST

Research Scientist, Machine Learning Researcher/Scientist, Applied Data Scientist.

EDUCATION

Rensselaer Polytechnic Institute, Troy, NY Fall 2019 – Spring 2024 (expected)
Ph.D. in Computer Science, GPA: 3.91/4.0 Advisor: *Prof.* Alex Gittens
IBM Artificial Intelligence Research Collaboration(AIRC) fellowship
Thesis Proposal: Scalable Cost-Efficient Techniques for Machine Learning

Rensselaer Polytechnic Institute, Troy, NY Spring 2016 – Spring 2019
B.S. in Mathematics Advisor: *Prof.* Jeffery Banks
B.S. in Computer Science Advisor: *Prof.* Heng Ji
Dean's Honor List, GPA: 3.86/4.0

Relevant courses: Machine Learning(ML) from data, ML and Optimization, Convex Optimization, Stochastic Optimization, Randomized Algorithms, Computational Linear Algebra, Information Theory, Security & Privacy for ML

SKILLS

Programming (Proficient)Python, C++, Matlab, LaTeX; (Familiar) C, R, SQL
Frameworks Pytorch, Tensorflow, Keras, CUDA, Scikit-learn

PROFESSIONAL EXPERIENCE

IBM, Yorktown Heights *AIRC scholar*

- Reduced label complexity for non-linear machine learning(*Python*) Summer 2023 – present
  - Aimed at significantly reducing the amount of labeled data required without compromising the model's performance and accuracy.
  - Adapted and applied an advanced sampling algorithm, originally designed for the linear regression context, to the realm of neural networks(*Pytorch*) with non-linear activation function.
  - Conducted comprehensive validations using real-world datasets, ensuring that our approach consistently delivers reliable and tight approximations.
- Sketching for low-rank Tucker decomposition(*Matlab&Tensor Toolbox*) Spring 2022 – present
  - Aimed at breaking down complex data structures(**Tensors**) into simpler parts more quickly, especially when dealing with large volumes of information.
  - Developed and rigorously tested a novel algorithm(*sketched-Tucker-ALS*), designed to streamline the decomposition process while ensuring rapid and reliable convergence(**∼6X faster in runtime** compared with state-of-the-art decomposition algorithms) to accurate results.
  - Introduced an adaptive heuristic practically to speed-up the decomposition process, resulting in even faster convergence(**∼1.5X faster in runtime** compared with *sketched-Tucker-ALS*), and a more efficient(**18% less memory**) use of computational resources.
- Sparse graph based sketching(*Python*) Summer 2020 – Spring 2021
  - Aimed at developing efficient data compressing methods, focused on significantly reduce data size while keeping essential information, specifically for handling **large-scale** and **sparse** datasets.
  - Theoretically defined key parameters to ensure optimal performance, achieving impressive results in preserving data quality and accelerating computational processes.
  - Conducted extensive experiments, verifying our approach's superiority(generated **sparser** sketching matrices than existing sparse sketching benchmarks but meanwhile achieving comparable performance as dense sketching benchmarks) over existing methods, particularly for **matrix approximation** applications. Results and Toolkit is available at *Sparse-Graph-Sketching Toolkit(Python)*.

RESEARCH EXPERIENCE

Rensselaer Polytechnic Institute *Graduate Researcher*

- NoisyCUR Algorithm for matrix completion(*Python*) Fall 2019 – Summer 2020
  - Aimed at efficiently completing data matrices with incomplete observations, while quantifying cost-efficiency and accuracy trade-offs.
  - Proposed the *NoisyCUR Algorithm*, a novel approach tailored for scenarios under **limited budgets**, ensuring that **quality recovery** of missing data are still achievable.
  - Conducted numerical experiments using a variety of datasets, showcasing the exceptional performance of our algorithm when compared to state-of-the-art matrix completion methods, particularly in situations where budget constraints limited us to observing **less than 20%** of the data entries(**improved the reconstruction error by 40%**).

Rensselaer Polytechnic Institute *Undergraduate Researcher*

- Multi-modal Data for Eye-gaze tracking(*Python*) Spring 2019
  - Aimed at enhancing the precision of a Convolutional Neural Network model dedicated to eye-tracking.
  - Applied a decision fusion model, integrating outputs from multiple pre-trained models to **improve the overall prediction accuracy by ∼5%**.
  - **Reduced the overall prediction error by 19%** on datasets featuring multiple people, while addressing the challenges posed by varying angles, distances, and lighting conditions.
- Multi-modal Data for Cognitive Analysis(*Python*) Summer 2018-Spring 2019
  - Fused and analyzed the real-time **multi-modal** data gathered from immersive environment.
  - Applied Speech-to-text technology from **IBM Bluemix** and punctuation restoring algorithm to provide a real-time transcription, ensuring seamless **Human-computer communication and interaction** within the virtual space.
  - Implemented the *gazing object* calculation model, accurately determining the user's focus and interest within the virtual environment.
- Convex Optimization Research (*Matlab&C*) Summer 2018
  - Under the supervision of *Prof.* Yangyang Xu, engaged in an in-depth exploration of **efficient** numerical strategies tailored for solving a variety of **convex programs**.
  - Conducted extensive experiments, evaluating and contrasting the performance of the *iALM (inexact augmented Lagrangian method)* and *ADMM (alternating direction method of multipliers)*, two prominent optimization algorithms.
  - Applied these advanced optimization techniques to a range of practical applications, including compressed sensing, image recovery and second-order cone problems.

PUBLICATIONS

Dong Hu, Alex Gittens, and Malik Magdon-Ismail, "Label complexity reduction for non-linear machine learning", in preparation for ICML 2024 submission

Alex Gittens, ***Dong Hu***, Shashanka Ubaru, and Lior Horesh, "Provable fast and convergent low-rank structured Tucker Decomposition via sketching", in preparation for TMLR 2023 submission

Dong Hu, Shashanka Ubaru, Alex Gittens, Ken Clarkson, Lior Horesh, and Vassilis Kalantzis, "Sparse graph based sketching for fast numerical linear algebra." in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 2021.

Dong Hu, Alex Gittens, and Malik Magdon-Ismail, "NoisyCUR: An algorithm for two-cost budgeted matrix completion," in *Machine Learning and Knowledge Discovery in Databases - European Conference(ECML-PKDD),* 2020

TEACHING EXPERIENCE

Teaching Assistant of ML and Optimization, RPI, Fall 2023
Teaching Assistant of Computer Organizations, RPI, Summer 2023
Undergraduate Mentor of Foundation of Computer Science, RPI Spring 2018