Dong Hu

(206)890-1181 · hud3@rpi.edu · https://jurohd.github.io

RESEARCH INTERESTS

Tensor Decomposition, Matrix Completion, Sketching, Low-rank Approximation, Label Complexity Reduction, High-dimensional Statistics, Randomized Linear Algebra.

EDUCATION

Rensselaer Polytechnic Institute Fall 2019 – Spring 2024 (expected)
Ph.D. in Computer Science, GPA: 3.91/4.0 Advisor: *Prof.* Alex Gittens

Rensselaer Polytechnic Institute Spring 2016 – Spring 2019
B.S. in Mathematics and Computer Science, GPA: 3.86/4.0

RESEARCH EXPERIENCE

Rensselaer Polytechnic Institute *Graduate Research Assistant*

- Reduced label complexity for non-linear machine learning Summer 2023 – Present
  - Addressed the label complexity reduction problem: For a datset, aiming to minimize the size of labeled data subset while ensuring bounded in-sample error on the remaining dataset.
  - Adapted and implemented the polynomial reject-sampling algorithm from ℓ_2-regression scenario for multi-layer neural networks. Provided empirical validation of maintaining tight approximations on various real datasets.
- *NoisyCUR Algorithm* for matrix completion Fall 2019 – Summer 2020
  - Addressed the matrix completion problem, investigated its critical cost-efficiency challenges.
  - Proposed and tested the Noisy-CUR Algorithm, a regression-based matrix completion approach optimized for low budget scenarios.
  - Conducted empirical results on both synthetic and real datasets, demonstrating superior performance over state-of-the-art matrix completion algorithms, especially in low budget settings.

IBM, Yorktown Heights *Artificial Intelligence Research Collaboration(AIRC) scholar*

- Sketching for low-rank Tucker decomposition Spring 2022 – present
  - Focused on speeding up Tucker decomposition problem via sketching in the low-rank setting.
  - Proposed a proximal regularized sketched alternating least squares (Tucker-ALS) algorithm, validated its sub-linear rate of convergence.
  - Demonstrated a heuristic for optimizing sketching rates across iterations using bandit sampling, leading to improved performance and convergence.
- Sparse graph based sketching Summer 2020 – Spring 2021
  - Investigated sparse sketching matrices derived from bipartite graphs for sparse inputs.
  - Theoretically established the minimal right vertices required (sketching size) for these graphs to provide a $(1 \pm \varepsilon)$ ℓ_2-subspace embedding.
  - Empirically validated low distortion and reconstruction error compared to state-of-the-art sketching methods for low-rank matrix approximation problems, with results and code published at the *Sparse-Graph-Sketching Toolkit(Python)*.

PUBLICATIONS

D. Hu, S. Ubaru, A. Gittens, K. Clarkson, L. Horesh, and V. Kalantzis. "Sparse graph based sketching for fast numerical linear algebra." in *International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 2021.

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