Agniva Chowdhury

Personal URL | LinkedIn | Google Scholar
2111 NE 25th Ave, Hillsboro, OR 97124, USA
agniva.chowdhury@intel.com | (765)-775-0297

Summary

AI Research Scientist with a strong foundation in machine learning, statistics, and computational mathematics. Specialized in designing and implementing fast and scalable randomized algorithms for large scale data analysis. Proficient in Python, R, and C++. Demonstrated research productivity, including publications and presentations in top-tier AI/ML venues such as NeurIPS, ICML, CVPR, AAAI, UAI, JMLR, etc.

Education

Purdue University — West Lafayette, IN, USA
Ph.D. in Statistics — Aug 2015 - Dec 2021

Indian Institute of Technology Kanpur — Kanpur, UP, India
M.Sc. in Statistics — Jul 2009 - May 2011

University of Calcutta — Kolkata, WB, India
B.Sc. in Statistics — Jul 2006 - May 2009

Experience

Intel Corporation — Hillsboro, OR, USA
AI Research Scientist — Jul 2024- Present

As an AI Research Scientist at Intel, I develop and refine ML models for cyber-attack detection, collaborate with teams, and apply advanced statistical and mathematical tools to improve AI technologies.

Oak Ridge National Laboratory — Oak Ridge, TN, USA
Postdoctoral Research Associate — Jan 2022- Jun 2024

At ORNL, I advanced my doctoral research by applying randomized methods to large linear systems for scientific applications like medical imaging, focusing on accelerating classification of rare events and tackling both over-constrained and under-constrained problems for more precise predictions. My work also explored scientific machine learning for physical systems and the impact of regularization, as detailed in my publications.

Purdue University — West Lafayette, IN, USA
Research Assistant — Jan 2017- Dec 2021
Teaching Assistant — Aug 2015- Dec 2016

During my PhD, I specialized in designing randomized algorithms for large-scale matrix computations to solve statistical and optimization problems, utilizing RandNLA and subspace embeddings to achieve fast, approximate solutions. My research concentrated on randomized matrix sketching and preconditioning, with applications across statistics, machine learning, and optimization, including ridge regression and matrix factorization.

HSBC — Kolkata, WB, India
Analyst - Decision Science — Nov 2012- Mar 2015

I built and monitored predictive models to improve marketing strategies for various products, offering key advice to support decisions focused on customer needs. My projects ranged from segmenting customers by value to creating models that helped sell foreign exchange services and education insurance more effectively.

EXL Service — Gurugram, Haryana, India
Senior Programmer Analyst — Jul 2011- Nov 2012

I have collaborated with various clients across the globe, delivering insights via thorough data analysis, including a project where I developed an Early-Claim Propensity model for a leading Indian life insurance firm. This model, which became a benchmark for best practices, was designed to predict early policy claims based on customers' previous insurance behaviors.

Publications/Preprints

1. **A. Chowdhury** and P. Ramuhalli. *A Provably Accurate Randomized Sampling Algorithm for Logistic Regression*. In Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI), 2024.
2. S. Fadnavis, **A. Chowdhury**, J. Batson, P. Drineas, and E. Garyfallidis. *Patch2Self2: Self-supervised Denoising on Coresets via Matrix Sketching*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.
3. A. Bose, M. Burch, **A. Chowdhury**, P. Paschou, and P. Drineas, *Structure-informed clustering for population stratification in association studies*. BMC Bioinformatics 24, p. 411, 2023.
4. F. Liu, **A. Chowdhury**. *Deep Learning with Physics Priors as Generalized Regularizers.* NeurIPS AI for Science Workshop. 2023.
5. **A. Chowdhury**, G. Dexter, P. London, H. Avron, and P. Drineas. *Faster Randomized Interior Point Methods for Tall/Wide Linear Programs*. Journal of Machine Learning Research (JMLR), 23(336), pp.1-48, 2022.
6. G. Dexter, **A. Chowdhury**, H. Avron, and P. Drineas. *On the Convergence of Inexact Predictor-Corrector Methods for Linear Programming*. In Proceedings of the 39th International Conference on Machine Learning (ICML), 2022. **Selected for long presentation**.
7. **A. Chowdhury**, A. Bose, S. Zhou, D. P. Woodruff, and P. Drineas. *A Fast, Provably Accurate Approximation Algorithm for Sparse Principal Component Analysis Reveals Human Genetic Variation Across the World*. In Proceedings of the 26th Annual Conference on Research in Computational Molecular Biology (RECOMB), 2022.
8. **A. Chowdhury**, P. London, H. Avron, and P. Drineas. *Faster Randomized Infeasible Interior Point Methods for Tall/Wide Linear Programs*. In Advances in Neural Information Processing Systems (NeurIPS), 2020.
9. **A. Chowdhury**, P. Drineas, D. P. Woodruff, and S. Zhou. *Approximation Algorithms for Sparse Principal Component Analysis*. arXiv:2006.12748, 2020.
10. A. Bose, M. C. Burch, **A. Chowdhury**, P. Paschou, and P. Drineas. *CluStrat: A Structure Informed Clustering Strategy for Population Stratification*. In Proceedings of the 24th Annual Conference on Research in Computational Molecular Biology (RECOMB), 2020.
11. **A. Chowdhury**, J. Yang, and P. Drineas. *Randomized Iterative Algorithms for Fisher Discriminant Analysis*. In Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI), 2019. **Selected for oral presentation.**
12. **A. Chowdhury**, J. Yang, and P. Drineas. *Structural Conditions for Projection-Cost Preservation via Randomized Matrix Multiplication*. Linear Algebra and its Applications, vol 573, pp. 144-165, 2019.
13. **A. Chowdhury**, J. Yang, and P. Drineas. *An Iterative, Sketching-based Framework for Ridge Regression*. In Proceedings of the 35th International Conference on Machine Learning (ICML), 2018.

Oral Presentations

1. Randomized Numerical Linear Algebra for Scientific Machine Learning. Applied Mathematics Group Seminar. Computational Science Initiative (CSI). Brookhaven National Laboratory, Upton, NY, USA, May 2024.
2. Randomized Linear Algebra for Interior Point Methods. Mathematics in Computation (MiC) Seminar. Oak Ridge National Laboratory, Oak Ridge, TN, USA, Feb 2024.
3. A Provably Accurate Randomized Sampling Algorithm for Logistic Regression. Mathematics and Computer Science (MCS) Seminar. Argonne National Laboratory, Lemont, IL, USA, Jan 2024.
4. Randomized Linear Algebra for Interior Point Methods. Microsoft Research India. Bengaluru, Karnataka, India, Dec 2023.
5. *Randomized Linear Algebra for Interior Point Methods.* SIAM Conference on Optimization (OP23). Seattle, WA, USA, May 2023.
6. *Randomized Numerical Linear Algebra and its Applications.* Flash talk in ORNL's AI Initiative mid-year review. Oak Ridge National Laboratory, TN, USA, Mar 2023.

7. *On the Convergence of Inexact Predictor-Corrector Methods for Linear Programming.* Bi-weekly meeting of Data-Driven Decision Control for Complex Systems Project (DnC2S). Oak Ridge National Laboratory, Oak Ridge, TN, USA, Aug 2022 (virtual).
8. *Faster Matrix Algorithms via Randomized Sketching & Preconditioning.* Bi-weekly meeting of Data-Driven Decision Control for Complex Systems Project (DnC2S). Oak Ridge National Laboratory, Oak Ridge, TN, USA, Mar 2022 (virtual).
9. Speeding-up Linear Programming using Randomized Linear Algebra. Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA, Jun 2021 (virtual).
10. Speeding-up Linear Programming using Randomized Linear Algebra. Michael Mahoney's Research Group, UC Berkeley, Berkeley, CA, USA, Oct 2020 (virtual).
11. Randomized Iterative Algorithms for Fisher Discriminant Analysis. Graduate Students Seminar, Department of Statistics, Purdue University, West Lafayette, IN, USA, Apr 2020 (virtual).
12. Randomized Iterative Algorithms for Fisher Discriminant Analysis. 35th Conference on Uncertainty in Artificial Intelligence (UAI), Tel Aviv, Israel, Jul 2019.
13. *An Iterative, Sketching-based Framework for Ridge Regression.* 35th International Conference on Machine Learning (ICML), Stockholm, Sweden, Jul 2018

Poster Presentations

1. *A Provably Accurate Randomized Sampling Algorithm for Logistic Regression.* AAAI 2024. Vancouver, BC, Canada, Feb 2024.
2. *Randomized Linear Algebra for Interior Point Methods.* AI Expo. Oak Ridge National Laboratory, TN, USA, Sep 2023.
3. *Faster Randomized Infeasible Interior Point Methods for Tall/Wide Linear Programs.* NeurIPS 2020.
4. *An Iterative, Sketching-based Framework for Ridge Regression*. TRIPODS Madison Summer School 2018, Madison, USA.
5. *An Iterative, Sketching-based Framework for Ridge Regression*. ICML 2018, Stockholm, Sweden.
6. *An Iterative, Sketching-based Framework for Ridge Regression*. 9th International Purdue Symposium on Statistics, 2018 West Lafayette, USA.
7. An Iterative, Sketching-based Framework for Ridge Regression. Conference on Scientific Computing and Approximation 2018 (in honor of Walter Gautschi), West Lafayette, USA.

Technical Skills

Python, PyTorch, Tensorflow, Git, R, MATLAB, SAS, C++, SQL, LaTeX, Markdown, Excel VBA

Teaching Experience

Lab Instructor: STAT 350: Introduction to Statistics (Fall 2016), STAT 301: Elementary Statistical Methods (Fall 2015).

TA and Grader: CS 590RA: Randomized Algorithms (Fall 2019), STAT 519: Introduction to Probability (Spring 2016, Fall 2016), STAT 512: Applied Regression Analysis (Fall 2015, Spring 2016), STAT 501: Experimental Statistics I (Summer 2016)

Honors and Awards

Travel Awards: NeurIPS 2020 (as complimentary registration); UAI 2019, Tel Aviv, Israel; ICML 2018, Stockholm, Sweden.

Invited to the workshops on

- “Randomized Numerical Linear Algebra, Statistics, and Optimization” organized by Center for Discrete Mathematics and Theoretical Computer Science (DIMACS), at Rutgers University, New Jersey.
- “Randomized Numerical Linear Algebra and Applications” organized by Simons Institute for the Theory of Computing at the University of California, Berkeley.
- “Fundamentals of Data Analysis” (TRIPODS Madison summer school 2018) organized by Institute for Foundations of Data Science (IFDS) at the University of Wisconsin–Madison.

Professional Activities

Membership: Society for Industrial and Applied Mathematics (SIAM), Association for the Advancement of Artificial Intelligence (AAAI).

Journal reviewing: IEEE Transactions on Signal Processing (TSP), ACM Journal of Experimental Algorithmics (JEA), Journal of Machine Learning Research (JMLR), ACM Transactions on Algorithms (TALG), IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Journal of Computational and Graphical Statistics (JCGS), SIAM Journal on Scientific Computing (SISC), IEEE Transactions on Knowledge and Data Engineering (TKDE), SIAM Journal on Matrix Analysis and Applications (SIMAX), Information and Inference: A Journal of the IMA (IMAIAI), Linear Algebra and its Applications (LAA), Applied and Computational Harmonic Analysis (ACHA).

Conference reviewing: AAAI Conference on Artificial Intelligence (AAAI), 2024; International Conference on Artificial Intelligence and Statistics (AISTATS) 2022; International Conference on Machine Learning (ICML) 2020, 2021; Neural Information Processing System (NeurIPS) 2020.

Committee service: Graduate student member of the Diversity and Inclusion Committee 2019-21, Department of Statistics, Purdue University.

Relevant Graduate Coursework

Big Data Theory and Methods, Randomized Algorithms for Big Data Matrices, Computational Statistics, Probability and Stochastic Processes, Linear Models, Regression Techniques, Statistical Inference, Bayesian Statistics.

References

Available upon request.