

EDUCATION

University of Utah <i>Ph.D. in Computer Science (Expected graduation: May 2022)</i>	August 2017- Present
Indian Institute of Technology (IIT) Kanpur <i>B.Tech in Computer Science and Engineering</i>	July 2012- April 2016

SKILLS

Research - Machine Learning, Deep Neural Networks, Topological Data Analysis, Ethics and Fairness in ML
Programming Languages - Python, Javascript, C
Data Science - Tensorflow, Pytorch, Scikit-Learn, Numpy, Pandas, Apache Spark (PySpark), SQL, Jupyter
Frameworks - Vue, Angular, Docker, Terraform, Hadoop, Spark, AWS (ECS, Fargate, API Gateway)

RESEARCH INTEREST

I work on understanding ML models through the lens of Computational Topology and Visualization. I believe that the strong mathematical foundations of topological analysis and interactive capabilities through visualization techniques can help traverse the inherently complex structures of ML models, especially in the context of deep learning. I create frameworks for Topological Data Analysis (TDA) and visualization to analyze and reason about deep learning models, as a step towards explainable and interpretable ML.

WORK EXPERIENCE

OpenSesame, Portland, OR <i>ML Engineer Intern</i>	May 2021 - Aug 2021 <i>Data Science Team</i>
--	---

- Built a data pipeline and NLP based chatbot service to provide course recommendation to learning admins.
- Created the architecture and APIs to deploy trained models in a scalable and secure manner.
- Planned and coordinated integration of the chatbot service with the recommendation system and frontend UI.
- Chatbot prototype identified as one of the key goals for the data science arm of the organization.

VISA Research, Palo Alto, CA <i>Research Intern</i>	May 2019 - Aug 2019 <i>Risk Modelling Team</i>
---	---

- Proposed a method to create auto-encoding models using Recurrent Neural Networks for financial transactions.
- Developed models to create embeddings for entities under weak labels.
- Implemented an end-to-end pipeline in PyTorch to interface with fraud detection and recommendation systems.
- Implemented efficient models capable of handling more than a million data points using pruning and distillation.

Samsung R&D Institute, Bangalore, India <i>Data Engineer</i>	Jun 2016 - July 2017 <i>Data Analytics Team</i>
--	--

- Built and optimized ETL pipelines for processing device logs for Samsung smartphones using Apache Spark.
- Proposed and implemented methods to find recurrent temporal patterns in smartphone usage and app activities.
- Deployed distributed processing pipelines for 500 million daily users with nearly 5 billion data points.
- Integrated the output from above into the intelligence module for Bixby in low-resource high-throughput setting.

PUBLICATIONS

VERB: Visualizing and Interpreting Bias Mitigation Techniques for Word Representations. A. Rathore, S. Dev, J.M. Philips, S. Srikumar, *et al.* Tutorial at [AAAI 2021](#), [KDD 2021](#), demo at NeurIPS 2021. Under review at TVCG. [[code](#)]

TopoAct: Exploring the Shape of Activations in Deep Learning. A. Rathore, N. Chalapathi, S. Palande, B. Wang. Computer Graphics Forum, 2020. [[pdf](#) | [demo](#) | [code](#) | [video](#)]

Autism Classification Using Topological Features and Deep Learning: A Cautionary Tale. A. Rathore, S. Palande, J. S. Anderson, *et al.* 22nd International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2019. [[pdf](#) | [code](#)]

Mapper Interactive: A Scalable, Extendable, and Interactive Toolbox for the Visual Exploration of High Dimensional Data. Y. Zhou, N. Chalapathi, A. Rathore, Y. Zhao and B. Wang, 2021 IEEE 14th Pacific Visualization Symposium (PacificVis), 2021. [[code](#)]

KEY PROJECTS

VERB: Visualizing and Interpreting Bias Mitigation Techniques for Word Representations

- Decomposed multiple debiasing algorithms into atomic operations for interpretability and interactivity.
- Built an interactive system to allow exploration of societal biases present in word vector embeddings and tweak parameters for debiasing algorithms in real-time.
- Chosen as part of the systems presented at the ‘Discover Engineering’ outreach program for students interested in data science across 1000 high-schools in Utah.

TopoAct: Exploring the Shape of Activations in Deep Learning

- Proposed a method to extract topological structures in the activation space of object detection and language modeling neural networks.
- Implemented deep learning models and data processing in Python and PyTorch.
- Built visualization tools using JavaScript, HTML and D3js which allows layer-wise analysis of activation vectors.
- Presented exploration scenarios that provide valuable insights about the learned representations of the deep neural network.

Autism Classification Using Topological Features and Deep Learning

- Developed a novel multi-branch neural network that combines vanilla features with topological information obtained from functional MRI scans to predict autistic vs control patients.
- Used TDA (persistent homology) in conjunction with deep learning to improve classification scores.
- Performed rigorous statistical testing to establish the significance of the accuracy scores.

Scalable Mapper Algorithm for Big Data

- Optimized the Mapper algorithm, a central tool in TDA, to efficiently parallelize the clustering step.
- Achieved speedup of upto 10x on dataset sizes of more than 1 million points.

RELEVANT COURSEWORK

- Machine Learning
- Computational Geometry
- Vis for Data Science
- Natural Language Processing
- Probabilistic Modeling
- Data Mining
- Advanced Algorithms
- Database Systems
- Computational Geometry

TEACHING AND TALKS

- Teaching Assistant for Discrete Mathematics at University of Utah, Spring 2020.
- Teaching Assistant for Theory of Computation at University of Utah, Fall 2020.
- Tutorial on debiasing word vector embeddings at AAAI 2021 and KDD 2021.
- Demo on debiasing word vector embeddings at NeurIPS 2021.
- Invited talk on analyzing embeddings from transformer models at Michigan State University in Oct 2021