Parikshit Bansal

parikshitb52@gmail.com | pbansal5.github.io/ | Google Scholar

Education

The University of Texas at Austin *Austin, Texas*
Ph.D. in Computer Science *2023 - Current*
Advisor: Prof. Sujay Sanghavi

Indian Institute of Technology, Bombay *Mumbai, India*
B.Tech. (Hons.) in Computer Science and Engineering *2017 - 2021*
B.Tech. Project: Deep Learning Methods for Missing Value Imputation in Time Series
Advisor: Prof. Sunita Sarawagi
CPI : 9.4/10.0

Publications And Pre-Prints

Understanding Contrastive Learning via Gaussian Mixture Models
Parikshit Bansal, Ali Kavis, Sujay Sanghavi
Preprint

Large Language Models as Annotators: Enhancing Generalization of NLP Models at Minimal Cost
Parikshit Bansal, Amit Sharma
Preprint

Controlling Learned Effects to Reduce Spurious Correlations in Text Classifiers
Parikshit Bansal, Amit Sharma
Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) [ACL 2023]

Improving Out-of-Distribution Generalization of Text-Matching Recommendation Systems
Parikshit Bansal, Yashoteja Prabhu, Emre Kiciman, Amit Sharma
NeurIPS 2022 Workshop on Causality for Real-world Impact [NeurIPS 2022]

Missing Value Imputation on Multidimensional Time Series
Parikshit Bansal, Prathamesh Deshpande, Sunita Sarawagi
Proceedings of the VLDB Endowment, Volume 14, 2020-2021 [VLDB 2021]

Work Experience

Amazon Web Services *Santa Clara*
Applied Scientist Intern | Manager : Ashish Khetan *Summer 2024*
Worked on reducing the inference latency of Large Language Models, specifically via Speculative Decoding.
Designed alternate loss functions and tested different architectural changes for developing an accurate and efficient ”speculator”.

Microsoft Research *Bangalore, India*
Research Fellow | Advisor : Dr. Amit Sharma *2021 - 2023*
Worked on problems around OOD generalisation of NLP systems with application in recommendation systems.
Also explored challenges in using Large Language Models as annotators for unlabeled data.

Agent Machine Learning Lab, Rutgers Univ *Remote Internship*
Research Intern | Advisor : Prof. Sungjin Ahn *Summer 2021*
Worked on various problems around self-supervised representations, object centric, and model-based reinforcement learning.

D.E. Shaw & Co. *Hyderabad, India*
Data Science Intern *Summer 2020*
Programmed Auto-Cohort Recommender for bining defaulted loan based on optimal clustering on attributes

Machine Learning and Genomics Lab, UCLA *Los Angeles, USA*
Visiting Researcher | Advisor : Prof. Sriram Sankararaman *Summer 2019*
Explored Sketching algorithms for scaling complex trait genetics algorithms to large scale genetic datasets

Selected Research Projects

Large Language Models as Data Annotators *Microsoft Research India*

Dr. Amit Sharma *Dec 2022 - Aug 2023*

- Tackled the problem of annotating unlabeled data by leveraging Large Language Models (LLMs) as annotators. This LLM annotated data is augmented with the original ground truth labeled data to train a downstream task specific model.
- Showed that naively (uniform sampling/uncertainty sampling) selecting unlabeled inputs for annotation with LLMs is harmful to accuracy
- Constructed a heuristic measure to instead sample the most informative data samples for downstream task specific classifier.
- Led to improvements in both natural language and recommendation systems benchmarks.

Out-of-distribution Generalization for Text-Matching Recommender Systems *Microsoft Research India*

Dr. Amit Sharma, Dr. Emre Kiciman, Dr. Yashoteja Prabhu *Aug 2021 - Dec 2022*

- Showed that finetuned text-matching recommenders are worse than pretrained model they are finetuned on for out-of-distribution data
- Attributed the drop in out-of-distribution performance to model weighing certain spurious tokens disproportionately
- Modeled a causal graph and formed a mathematical framework to justify the observations
- Proposed a novel regularisation technique leveraging the *base* model for constructing augmented samples to regularise weighing of tokens

Missing Value Imputation on Multidimensional Time Series *IIT Bombay*

Prof. Sunita Sarawagi *Aug 2020 - July 2021*

- Introduced novel convolution based transformer model for capturing long range patterns yielding a speedup of 10x
- Worked on efficient batching for samples for shared forward pass for multiple samples in a batch
- Formulated kernel regression module for aggregating signals from correlated time series.
- Got upto 60% reduction in MAE error with similar running time to Matrix completion technique s(e.g. SVD). Published in VLDB, 2021

Robust Automatic Speech Recognition (ASR) Systems *IIT Bombay*

Prof. Preethi Jyothi *Dec 2019 - Dec 2020*

- Leveraged generative speech models to construct *imperceptible* and *robust* adversarial speech samples
- Experimented with attribution based masking techniques as an extension of targeted and efficient augmentations
- Developed an augmentation technique for ASR systems, based on randomly drops time frames. Got minor improvements in SOTA numbers

Other Projects

Self-Supervised Representation Learning for Dynamic Environments *Rutgers University*

Prof. Sungjin Ahn *Apr 2021 - Aug 2021*

- Implemented various self-supervised learning (SSL) algorithms image representation learning methods like BYOL and SwAV.
- Experimented with SSL algorithms for capturing structured information about environment dynamics in partially observed settings
- Extended them with object-centric module for disentangling object representations in dynamic environment (worked in bouncing balls)
- Leveraged SSL methods to extend model-based reinforcement learning methods. Achieved similar performance with lower memory footprint.

Sketching Matrices for Randomised HE Regression *UCLA*

Prof. Sriram Sankaraman *May 2019 - July 2019*

- Implemented algorithms to model genetic data as multi-component Linear Mixed Model, and compute the variance of various random effects
- Achieved sub-linear (in # of individuals) time complexity of the method by replacing MLE with MoM, using Hutchinson's trick
- Experimented with various sketching algorithms to approximate genotype matrix ($n \times m$) by a smaller matrix ($n \times d$)
- Got a reduction the running time by factor of $(m/d)^3$, but with introduction of significant bias in estimator

Readmission Risk Modeling *Cleveland Clinic*

Dr. Rajat Garg *Feb 2020 - Jul 2020*

- Surveyed relevant medical literature & modeled 30 day readmission risk for patients undergone mitraclip procedure using logisitic regression
- Formulated a linear score calculator to be used in clinical practice. Accepted in *Current Problems in Cardiology*. Publication can be found *here*

3D Face Reconstruction from 2D in-the-wild Images *May 2021 - Jul 2021*

ActionFace

- Implemented and trained a 3D morphable face model. Built an End-2-End pipeline for getting 3D face mesh from 2D images on top of it
- Restricted training samples with pitch and yaw ranges to get better performance for profile images. Benchmarked our results

Achievements

2021 **Undergraduate Research Award (URA02),** as recognition of truly exceptional work done in B.Tech. Project

2020 **Undergraduate Research Award (URA01),** as recognition of research/ developmental effort

2020 **Institute Academic Prize,** for academic excellence (top 3 students) during the term of 2019-20

2017 **All India Rank 62,** JEE Advanced, (among 1.5 million candidates in India)

2016 **NSEP,NSEC,** Ranked among top 1% nationwide in Physics and Chemistry Olympiads resp.

2015 **KVPY Fellowship,** Shortlisted for the fellowship, conducted by the Govt. of India for two consecutive years

Skills

Languages	Python, C++, C, Haskell, Prolog, Bash, MATLAB
Libraries & Frameworks	Pytorch, Tensorflow, Huggingface, Pandas
Tools	Git, Emacs, LaTeX, TikZ

Courses

Probability	Stochastic Processes, Online Learning, Advanced Probability
Optimization/Linear Algebra	Convex Optimization, Large Scale Optimization, Continuous Algorithms, Numerical Linear Algebra
Computer Science	Generative Models

Reviewer Services

Natural Language Processing	EMNLP 2023
Machine Learning	NeurIPS 2024, ICLR 2025, AISTATS 2025

Talks

UT Austin, 2023 Parameter-Efficient Fine-Tuning of Large Language Models