

THARUN KUMAR REDDY MEDINI

e-mail: tharun.medini@rice.edu

webpage: tharun24.github.io

phone: +1-2092769537

Academics

- **PhD** in Electrical and Computer Engineering at **Rice University** *Aug 2016 - ongoing*
- **BTech** with major in Electrical Engineering and minor in Math from **IIT Bombay** *2011 - 2015*
- **All India Rank 21** in IIT JEE-2011

Work Experience

Applied Scientist Intern at Amazon Search(A9), Palo Alto, CA *May 2018 - Aug 2019*

Manager: Vijai Mohan, Lab: Search Labs

- Implemented a new hashing based extreme classification algorithm MACH for improving **Matching** and **Ranking** performance of Amazon Search.
- Achieved **9%** better offline recall than production model on a category with **85 million** products.
- Developed a MinHash based low latency fall-back package FLASH to replace queries with most relevant ones in the event of search failure.

Graduate Research Assistant at Rice University, Houston, TX

Aug 2016 - present

Advisor: Prof.Anshumali Shrivastava, Lab: RUSHLAB

- Working on Large Scale Machine Learning using smart Hashing and Randomization methods. Working on memory and time efficient **Extreme Classification, Zero-Shot Learning, Structured Prediction** and **Imitation Learning** using minimal expert information.

Data Analyst at Target Corporation, Bengaluru

July 2015 - July 2016

Manager: Sourav Dutta, Mentor: Venkataramana Kini, Lab: Enterprise Data Analytics & Business Intelligence

- Worked on estimating customer **subscription propensity** using Mixture Models.
- Worked with **Personalization** team on improving the purchase rate of **complimentary product** recommendations using **word2vec** and **Bayesian Personalized Ranking(BPR)**.

Research

Extreme Classification in Log Memory using Count-Min Sketch

Published at NeurIPS 2019

Tharun Medini, Qixuan Huang, Yiqiu Wang, Vijai Mohan, Anshumali Shrivastava

- Proposed a novel method to group K classes(millions) into a few hundreds of meta-classes using 2-universal hashing. Using just $O(\log(K))$ such groupings, we can train small classifiers in just logarithmic memory
- We bypass the prediction of K -vector and directly predict its count-min sketch values and recover the original predictions when needed.
- We show improved precision and recall with significantly less memory on an **Amazon Search Dataset** with **50 million** classes and several other multi-class and multi-label datasets.

SLIDE: Sub-Linear Deep Learning Engine

Published at MLSys 2020

Beidi Chen, Tharun Medini, James Farwell, Sameh Gobriel, Charlie Tai, Anshumali Shrivastava

- Developed a new DL framework from scratch in C++ that sparsifies the computations in neural networks to $\approx 1\%$ of typical matrix multiplications. Our package uses simple **CPU** parallel instructions and trains and evaluates **5x faster** than the best available GPU **NVIDIA Tesla V-100** on large extreme classification datasets.

RAMBO: Repeated And Merged BloOm Filter for Multiple Set Membership Testing (MSMT) in Sub-linear time

pre-print on arxiv

Gaurav Gupta, Benjamin Coleman, Tharun Medini, Vijai Mohan, Anshumali Shrivastava

- Proposed a novel streaming algorithm RAMBO that achieves $O(\sqrt{K}\log K)$ query time for K sets as opposed to $O(K)$ for the popular Array-of-Bloom-Filters.
- We construct a Count-Min SKetch of Bloom Filters and replace the addition operation with set union and minimum operation using set intersection.

Similarity based Soft-Labelling for Generalized Zero-Shot Learning

ViGIL Workshop, NeurIPS 2019

Shabnam Daghighi, Tharun Medini, Anshumali Shrivastava

- Proposed a novel algorithm to train Zero-Shot Learning with cross-entropy loss as opposed to training embedding models or GANs. We distribute a small probability to the nearest unseen class based on embedding distance. This intrinsically learns to predict unseen classes. We achieve notably better generalized Zero Shot precision on AwA, CUB and SUN datasets.

A Deep Dive Into Count-Min Sketch for Extreme Classification in

Log-Memory

ML with Guarantees Workshop, NeurIPS 2019

Tharun Medini, Anshumali Shrivastava

- Provided precise quantification of memory-identifiability tradeoffs in MACH.
- Extending MACH to multi-label classification, proposed a novel quadratic estimator using Inclusion-Exclusion Principle. Our estimator has significantly lower reconstruction error than the typical Count-Min estimator.

Imitate like a Baby: The Key to Efficient Exploration in Deep Reinforcement

Learning

Deep RL Workshop, NeurIPS 2019

Tharun Medini, Anshumali Shrivastava

- Achieved significantly better scores than Asynchronous-Advantage Actor Critic(**A3C**) model on Atari-2600 games by appending the action space of a Reinforcement Learning agent with the most **frequent action sequences** taken by an expert in the same game.

Academic Services

PC Member

- AAAI 2020.

Reviewer

- NeurIPS 2019, ICML 2019 and AAAI 2017.

Teaching Assistant

Aug 2013 - May 2014

- Worked as **Teaching Assistant** for **Calculus** and **Differential Equations** courses at IIT Bombay.

Mentor, Department Academic Mentorship Program

April 2014-April 2015

- Worked as a **mentor** for under performing students with academic and personal problems.

Skills

- Programming Languages : **Python**, MATLAB, C++
- Tools and Packages: **TensorFlow**, **PySpark**, Keras, Hadoop MapReduce.

Awards & Scholarships

- Rice Graduate Fellowship *Aug 2016 - May 2018*
- American Society of Indian Engineers Scholarship *Nov 2019*
- IIT Bombay MCM scholarship *Aug 2011 - May 2015*
- Academic Excellence Award from EE Department, IIT Bombay *Apr 2015*
- Best Mentor award from Institute Student Mentorship Program (ISMP), IIT Bombay *2014, 2015*

Invited Talks

- Jane Street Symposium *Jan 2020, NY*
- Spotlight talk at Systems for ML workshop on SLIDE *NeurIPS 2019*
- Houston ML Meetup (Intro to Actor-Critic Methods and Imitation in Deep Reinforcement Learning) *Dec 2019, Univ. of Houston*
- Schlumberger (Imitation Learning) *Nov 2019, Katy*
- Rice Data Science Conference (Imitate like a Baby: The Key to Efficient Exploration in Deep Reinforcement Learning) *Oct 2019, BRC, Rice Univ.*

In the News

- An algorithm could make CPUs a cheap way to train AI *Endgadget*
- Deep Learning breakthrough made by Rice University scientists *ARS Technica*
- SLIDE algorithm for training deep neural nets faster on CPUs than GPUs *Inside HPC*
- Hash Your Way To a Better Neural Network *IEEE Spectrum*
- Researchers report breakthrough in 'distributed deep learning' *TechXplore*
- Deep learning rethink overcomes major obstacle in AI industry *TechXplore*