Anil Kag

Ph.D. Candidate
Data Science & Machine Learning Lab
Boston University

8 St. Mary's St
Boston, MA (02215)
+1 (857) 498 9294
anilkag@bu.edu
https://anilkagak2.github.io/

Education

2018–2023 **Ph.D. in Electrical & Computer Engineering**, *Boston University*, *3.96/4.0*
2018–2022 **M.S. in Electrical & Computer Engineering**, *Boston University*, *3.96/4.0*
2010–2014 **B.Tech. in Computer Science**, *Indian Institute of Technology*, Guwahati, *9.20/10*

Research Interests

Efficient Neural Architectures, Computer Vision, Resource Constrained Learning, & Large Scale Optimization

Work Experience

2020 **Research Intern**, *Microsoft Research*, Redmond
2016–2018 **Research Fellow**, *Microsoft Research*, Bangalore
2014–2016 **Software Engineer**, *Dynamics CRM, Microsoft*, Bangalore
2013 **Software Engineer Intern**, *Bing, Microsoft*, Hyderabad

Publications

Pre-Print **Spatially Interpolated Inverted Residual Block**
A. Kag, G. Wadhwa, V. Saligrama, P. Jain

ICLR'23 **Scaffolding a Student to Instill Knowledge**
A. Kag, D. A. E. Acar, A. Gangrade, V. Saligrama

ICLR'23 **Efficient Edge Inference by Selective Query**
A. Kag, I. Fedorov, A. Gangrade, P. Whatmough, V. Saligrama

ICML'22 DyNN **Achieving High TinyML Accuracy through Selective Cloud Interactions**, *(spotlight)*
A. Kag, I. Fedorov, A. Gangrade, P. Whatmough, V. Saligrama

CVPR'22 **Condensing CNNs with Partial Differential Equations**
A. Kag, V. Saligrama

NeurIPS'21 **Online Selective Classification with Limited Feedback**
A. Gangrade, **A. Kag**, A. Cutkosky, V. Saligrama

ICML'21 **Training Recurrent Neural Networks via Forward Propagation Through Time**
A. Kag, V. Saligrama

CVPR'21 **Time-Adaptive RNN: A Dynamical Systems View**
A. Kag, V. Saligrama

AISTATS'21 **Learning With Abstention via One-Sided Classification**
A. Gangrade, **A. Kag**, V. Saligrama

ICLR'20 **RNNs Incrementally Evolving on an Equilibrium Manifold: A Panacea for Vanishing and Exploding Gradients?**
A. Kag, Z. Zhang, V. Saligrama

NSDI'19 **BLAS-on-flash: An Efficient Alternative for Large Scale ML Training and Inference?**
S. J. Subramanya, H. V. Simhadri, S. Garg, **A. Kag**, V. Balasubramanian

NeurIPS'18 CDNNRIA **Learning Compact Networks via Adaptive Network Regularization**
S. Sankarapandian, **A. Kag**, R. Manzelli, B. Kulis

WSDM'18 **SwiftXML: Extreme Multi-label Learning with Label Features for Warm-start Tagging, Ranking & Recommendation**
Y. Prabhu, **A. Kag**, S. Gopinath, K. Dahiya, S. Harsola, R. Agrawal, M. Varma

WWW'18 **Parabel: Partitioned Label Trees for Extreme Classification with Application to Dynamic Search Advertising**
Y. Prabhu, **A. Kag**, S. Harsola, R. Agrawal, M. Varma

Academic Service

Reviewer

Conference NeurIPS, ICML, ICLR, CVPR, AAAI, COLT, ICASSP

Journal TMLR, IEEE Neural Networks and Learning Systems

Skill Set

Programming C, C++, C#, Java, Python, Bash

ML Toolkits PyTorch, Tensorflow, Scikit-Learn

Databases MySQL

Tools Matlab, LaTeX, Git, Visual Studio

Some of my projects are hosted at `https://github.com/anilkagak2`

Talks

October 2022 **Achieving High TinyML Accuracy through Selective Cloud Interactions**
BU AIR Seminar, Boston

July 2022 **Achieving High TinyML Accuracy through Selective Cloud Interactions**
DyNN Workshop, ICML, Baltimore

March 2022 **Achieving High TinyML Accuracy through Selective Cloud Interactions**
ARM Research, Boston

August 2020 **Tiny ML models for Phish Detection**
Microsoft S+C & Microsoft Research, Redmond

Academic Achievements

- Rafik Hariri Graduate Student Fellowship, Rafik B. Hariri Institute, Boston University
- Research Travel Award, ECE Department, Boston University
- Dean's Ph.D. Fellowship, ECE Department, Boston University
- Among Top 10% reviewers in NeurIPS 2020
- Ranked 4 out of 80 students in the Batch of 2014, Computer Science, IIT Guwahati
- Recipient of "Merit-cum-Means" scholarship provided by IITG in 1st & 2nd Year
- Secured 1761 Rank in IIT-JEE, 2010 out of 450,000 students who appeared for the test

Major Projects

Fall 2019 **Online Non-Convex Learning**
Advisor **Dr. Francesco Orabona, Assistant Professor, BU**
Literature survey of the non-convex losses in the online learning setting. Also analyzed the follow-the-regularized-leader algorithm for a sub-class of non-convex functions satisfying Polyak condition.

Summer 2019 **Tiny ML models for Phish Detection**
Advisor **Dr. Prateek Jain, Sr. Principal Researcher, MSR India**
Developed Tiny ML models with low complexity and competitive performance to the SmartScreen models for Phish webpage detection. These models are very lightweight and can be easily deployed for mobile inference via the Tensorflow-lite framework enabling privacy-aware inference.

Spring 2019 **Survey on first order methods for Deep Learning**
Advisor **Dr. Francesco Orabona, Assistant Professor, BU**
Literature survey on the first order methods such as SGD, Adagrad, RmsProp, Adam, Nadam.

2017–2018 **Improving Bing Dynamic Search Ads (DSA) Recommendations**
Advisor **Dr. Manik Varma, Senior Researcher, MSR India**
Improving Bing DSA recommendations using Extreme Classification. Given an Ad landing page without any bid keywords, we were asked to predict potentially monetizable queries which can bring clicks. This resulted in 13.6% gain in click-through rate and 13% reduction in bounce rate.

2016–2017 **Improving Bing Text Ads (DSA) Recommendations**
Advisor **Dr. Manik Varma, Senior Researcher, MSR India**
Improving Bing Text Ads recommendations using Extreme Classification. Given an Ad landing page with bid keywords, we were asked to predict potentially monetizable queries which can bring clicks. This resulted in 5% gain in click-through rate and 11% reduction in bounce rate.

Summer 2017 **The Nature Conservancy Fisheries Monitoring**, *Kaggle Challenge*
This was an image classification problem, where given an image, we were asked to predict the type of endangered fish if there's one. I ranked among top 5% in the final evaluation.

Key Courses Undertaken

Data Structures	Operating Systems	Machine Learning	Formal Language & Automata Theory
Algorithms	Computer Networks	Statistical Learning	Theory of Computation
Computer Architecture	Compilers	Learning from Data	Probability Theory & Random Processes
Discrete Mathematics	DBMS	Reinforcement Learning	Optimization
Software Engineering	Distributed Systems	Online Learning	Hierarchical Memory Algorithms
Randomized Algorithms	Stochastic Processes	Information Retrieval	Computational Geometry
Parallel Algorithms		Information Theory	Real Analysis