Alina Beygelzimer

Talk Title: Contextual Memory Trees

Talk Abstract: This talk is about a new learned dynamic memory controller for organizing prior experiences in a way that is empirically useful for a number of downstream tasks. The controller supports logarithmic time operations and can thus be integrated into existing statistical learning algorithms as an augmented memory unit without substantially increasing training and inference computation. It also supports optional reward reinforcement, which brings a steady improvement empirically. The controller operates as a reduction to online classification, allowing it to benefit from advances in representation or architecture. This is joint work with Wen Sun, Hal Daume, John Langford, and Paul Mineiro (published at ICML-2019).

Speaker Bio: Alina Beygelzimer is a senior research scientist at Yahoo Research in New York City, working on machine learning. She was a program co-chair for COLT 2019 and NeurIPS 2019, and a recipient of a best paper award at ICML 2015.

Inderjit Dhillon

Talk Title: Multi-Output Prediction: Theory and Practice

Talk Abstract: Many challenging problems in modern applications amount to finding relevant results from an enormous output space of potential candidates, for example, finding the best matching product from a large catalog or suggesting related search phrases on a search engine. The size of the output space for these problems can be in the millions to billions. Moreover, observational or training data is often limited for many of the so-called "long-tail" of items in the output space. Given the inherent paucity of training data for most of the items in the output space, developing machine learned models that perform well for spaces of this size is challenging. Fortunately, items in the output space are often correlated thereby presenting an opportunity to alleviate the data sparsity issue. In this talk, I will first discuss the challenges in modern multi-output prediction, including missing values, features associated with outputs, absence of negative examples, and the need to scale up to enormous data sets. Bilinear methods, such as Inductive Matrix Completion~(IMC), enable us to handle missing values and output features in practice, while coming with theoretical guarantees. Nonlinear methods such as nonlinear IMC and DSSM (Deep Semantic Similarity Model) enable more powerful models that are used in practice in real-life applications. However, inference in these models scales linearly with the size of the output space. In order to scale up, I will present the Prediction for Enormous and Correlated Output Spaces (PECOS) framework, that performs prediction in three phases: (i) in the first phase, the output space is organized using a semantic indexing scheme, (ii) in the second phase, the indexing is used to narrow down the output space by orders of magnitude using a machine learned matching scheme, and (iii) in the third phase, the matched items are ranked by a final ranking scheme. The versatility and modularity of PECOS allows for easy plug-and-play of various choices for the indexing, matching, and ranking phases, and it is possible to ensemble various models, each arising from a particular choice for the three phases.

Speaker Bio: Inderjit Dhillon is the Gottesman Family Centennial Professor of Computer Science and Mathematics at UT Austin, where he is also the Director of the ICES Center for Big Data Analytics. Currently he is on leave from UT Austin and heads the Amazon Research Lab in Berkeley, California, where he is developing and deploying state-of-the-art machine learning methods for Amazon Search. His main research interests are in big data, deep learning, machine learning, network analysis, linear algebra and optimization. He received his B.Tech. degree from IIT Bombay, and Ph.D. from UC Berkeley. Inderjit has received several awards, including the ICES Distinguished Research Award, the SIAM Outstanding Paper Prize, the Moncrief Grand Challenge Award, the SIAM Linear Algebra Prize, the University Research Excellence Award, and the NSF Career Award. He has published over 200 journal and conference papers, and has served on the Editorial Board of the Journal of Machine Learning Research, the IEEE Transactions of Pattern Analysis and Machine Intelligence, Foundations and Trends in Machine Learning and the SIAM Journal for Matrix Analysis and Applications. Inderjit is an ACM Fellow, an IEEE Fellow, a SIAM Fellow and an AAAS Fellow.

Jure Leskovec

Talk Title: Generalizing to Novel Tasks in the Low-Data Regime

Talk Abstract: Developing algorithms that are able to generalize to a novel task given only a few labeled examples represents a fundamental challenge in closing the gap between machine- and human-level performance. The core of human cognition lies in the structured, reusable concepts that help us to rapidly adapt to new tasks and provide reasoning behind our decisions. However, existing meta-learning methods learn complex representations across prior labeled tasks without imposing any structure on the learned representations. In this talk I will discuss how meta-learning methods can improve generalization ability by learning to learn along human-interpretable concept dimensions. Instead of learning a joint unstructured metric space. We learn mappings of high-level concepts into semi-structured metric spaces, and effectively combine the outputs of independent concept learners. Experiments on diverse domains, including a benchmark image classification dataset and a novel single-cell dataset from a biological domain show significant gains over strong meta-learning baselines.

Speaker Bio: Jure Leskovec is an associate professor of Computer Science at Stanford University, the Chief Scientist at Pinterest, and an Investigator at the Chan Zuckerberg Biohub. He was the co-founder of a machine learning startup Kosei, which was later acquired by Pinterest. Leskovec's research area is machine learning and data science for large interconnected systems. Applications include commonsense reasoning, recommender systems, social network analysis, computational social science, and computational biology with an emphasis on drug discovery. This research has won several awards including a Lagrange Prize, Microsoft Research Faculty Fellowship, the Alfred P. Sloan Fellowship, and numerous best paper and test of time awards. It has also been featured in popular press outlets such as the New York Times and the Wall Street Journal. Leskovec received his bachelor's degree in computer science from University of Ljubljana, Slovenia, PhD in machine learning from Carnegie Mellon University and postdoctoral training at Cornell University. You can follow him on Twitter at @jure.

Maryam Majzoubi

Talk Title: Extreme Classification with Logarithmic-depth Streaming Multi-label Decision Trees

Talk Abstract: We consider multi-label classification where the goal is to annotate each data point with the most relevant subset of labels from an extremely large label set. Efficient annotation can be achieved with balanced tree predictors, i.e. trees with logarithmic-depth in the label complexity, whose leaves correspond to labels. Designing prediction mechanism with such trees for real data applications is non-trivial as it needs to accommodate sending examples to multiple leaves while at the same time sustain high prediction accuracy. In this paper we develop the LdSM algorithm for the construction and training of multi-label decision trees, where in every node of the tree we optimize a novel objective function that favors balanced splits, maintains high class purity of children nodes, and allows sending examples to multiple directions but with a penalty that prevents tree over-growth. Each node of the tree is trained once the previous node is completed leading to a streaming approach for training. We analyze the proposed objective theoretically and show that minimizing it leads to pure and balanced data splits. Furthermore, we show a boosting theorem that captures its connection to the multi-label classification error. Experimental results on benchmark data sets demonstrate that our approach achieves high prediction accuracy and low prediction time and position LdSM as a competitive tool among existing state-of-the-art approaches.

Speaker Bio: Maryam Majzoubi is a PhD candidate of electrical engineering at the New York University Tandon School of Engineering. Her research interests focus on machine learning, both theoretical and applied, optimization, deep learning and big data. Currently her research focuses on large scale classification and related problems. She was the recipient of Tandon School of Engineering Fellowship (2017-19). She has received her Master's degree from Pennsylvania State University, USA, and Bachelor's degree from Sharif University of Technology, Tehran, Iran. In Summer 2019 she was an intern research fellow at Microsoft Research, NYC working with John Langford on efficient contextual bandit problems. Currently she is an intern at Google Research, NYC.

Tomas Mikolov

Talk Title: Historical perspective on extreme classification in language modeling

Talk Abstract: In this talk, I will present several simple ideas that were proposed a long time ago to deal with extremely large output spaces in the language modeling. These include various types of hierarchical softmax, and other approaches that decompose the labels into smaller parts such as sub-word language modeling.

Speaker Bio: Tomas Mikolov is a senior researcher at CIIRC institute in Prague. Previously, he has been a research scientist at Facebook AI Research and Google Brain. His research interest is to build intelligent machines that can help people to solve difficult tasks. Currently, he is working in an intersection of machine learning, complex systems, and artificial evolution.

Manik Varma

Talk Title: DeepXML: A Framework for Deep Extreme Multi-label Learning

Talk Abstract: In this talk we propose the DeepXML framework for deep extreme multi-label learning and apply it to short-text document classification. We demonstrate that DeepXML can: (a) be used to analyze seemingly disparate deep extreme classifiers; (b) can lead to improvements in leading algorithms such as XML-CNN & MACH when they are recast in the proposed framework; and (c) can lead to a novel algorithm called Astec which can be up to 12% more accurate and up to 40x faster to train than the state-of-the-art for short text document classification. Finally, we show that when flighted on Bing, Astec can be used for personalized search, ads and recommendation for billions of users. Astec can handle billions of events per day, can process more than a hundred thousand events per second and leads to a significant improvement in key metrics as compared to state-of-the-art methods in production.

Speaker Bio: I am a Senior Principal Researcher at Microsoft Research India and an Adjunct Professor of computer science at the Indian Institute of Technology Delhi. My research interests lie in the areas of machine learning, computational advertising and computer vision. Classifiers that I have developed have been deployed on millions of devices around the world and have protected them from viruses and malware. My algorithms are also generating millions of dollars on the Bing search engine (up to sign ambiguity). In 2013, John Langford and I coined the term extreme classification and found that we had inadvertently started a new area in machine learning. Today, by happenstance, extreme classification is thriving in both academia and industry with my classifiers being used in various Microsoft products as well as in the wider tech sector. I recently proclaimed 2 KB (RAM) ought to be enough for everybody prompting the international media to cover my research and compare me to Bill Gates (unfair, I'm more handsome!). I have been awarded the Shanti Swarup Bhatnagar Prize, the Microsoft Gold Star award, the Microsoft Achievement award, the WSDM Best Paper award, won the PASCAL VOC Object Detection Challenge and stood first in chicken chess tournaments and Pepsi drinking competitions. I have served as an area chair/senior PC member for machine learning, artificial intelligence and computer vision conferences such as AAAI, CVPR, ICCV, ICML, IJCAI and NeurIPS. I am also serving as an associate editor of the IEEE TPAMI journal. I am a failed physicist (BSc St. Stephen's College, David Raja Ram Prize), theoretician (BA Oxford, Rhodes Scholar), engineer (DPhil Oxford, University Scholar), mathematician (MSRI Berkeley, Post-doctoral Fellow) and astronomer (Visiting Miller Professor, UC Berkeley). I am also a Fellow of the Indian National Academy of Engineering.

Chicheng Zhang

Talk Title: Efficient continuous-action contextual bandits via reduction to extreme multiclass classification

Talk Abstract: We create a computationally tractable algorithm for contextual bandit learning with one-dimensional continuous actions with unknown structure on the loss functions. In a nutshell, our algorithm, Continuous Action Tree with Smoothing (CATS), reduces continuous-action contextual bandit learning to cost-sensitive extreme multiclass classification, where each class corresponds to a discretized action. We show that CATS admits an online implementation that has low training and test time complexities per example, and enjoys statistical consistency guarantees under certain realizability assumptions. We also verify the efficiency and efficacy of CATS through large-scale experiments.

Speaker Bio: Chicheng Zhang is an assistant professor in the Computer Science department at the University of Arizona. From 2017 to 2019, he was a postdoctoral researcher in the machine learning group at Microsoft Research New York City. He received a Bachelor degree from Peking University in 2012 and a PhD in Computer Science from UC San Diego in 2017. His main research interests lie in the design and analysis of interactive machine learning algorithms.