Machine Learning and Data Analytics Symposium (MLDAS) 2021

AGENDA
Day 1 - March 23, 2021
6:00 – 9:00 PM (Doha Time GMT + 3)

Perception and Vision	
6:00 - 6:15	***Dr. Sanjay Chawla, QCRI*** ***Dr. Ahmed Elmagarmid, QCRI*** ***Mr. Bernard Dunn, Boeing Middle East***
6:15 - 6:45	***Using Satellite Data and AI to Accelerate Sustainable Development*** ***Dr. Andrew Zolli, VP of Sustainability and Impact Initiatives of Planet*** *Abstract: The combination of real-time Earth Observation, machine-learning-driven analytics, and cloud computing opens new opportunities for the measurement of sustainable development, including the SDGs, as well as physical climate risk. In this talk, we will show leading-edge examples from the field, and discuss practical opportunities and challenges to scaling these approaches.*
6:45 - 7:30	***Professor David Forsyth, University of Illinois, Urbana-Champaign***
7:30 - 8:00	***Dr. Mohammad Amin Sadeghi, Qatar Computing Research Institute***
8:00 - 8:30	***Professor Julie Shah, Massachusetts Institute of Technology***
8:30 - 9:00	***Q/A hosted by Dr. Sanjay Chawla, Qatar Computing Research Institute***

Machine Learning and Data Analytics Symposium (MLDAS) 2021

AGENDA
Day 2 - March 24, 2021
6:00 – 9:00 PM (Doha Time GMT + 3)

Machine Learning Principles and Applications

6:00 - 6:30	***Advancing Outlier Detection in RNA-Seq GeneExpression Count Data to Improve Aberrant Gene Discovery*** ***Dr. Halima Bensmail, Qatar Computing Research*** *Abstract: High-throughput RNA sequencing technologies (RNA-Seq) have recently started being used as a tool for helping diagnose rare genetic disorders, as they can indicate abnormal gene expression counts — a telltale sign of genetic pathology. Existing solutions either require a large number of samples, do not provide proper statistical significance testing or do not provide appropriate false discovery rate (FDR) control.* *We present a Bayesian model for identifying abnormal RNA-Seq gene expression counts using a Negative Binomial (NB) modeling. To capture the unknown confounders, we propose a RNA-Seq GE data transformation using Kernel Density Estimation (KDE) and exploit the nature of the transformed data to control for unknown confounders using Singular Value Decomposition (SVD), thereby avoiding ad-hoc autoencoder (AE) based methods involving artificial outlier injection employed by competing models. Our approach performed better than state of art competitors and can be applicable to single-cell RNA-Seq data as well as other tasks related to RNA-Seq, such as differential expression analysis.*
6:30 - 7:00	***Geometric and Topological Graph Analysis for Machine Learning*** ***Professor Tina Eliassi-Rad, Northeastern University*** *Abstract: This talk has two parts: (1) geometric analysis for graph embedding and (2) topological analysis for graph distances. First, graph embedding seeks to build an accurate low-dimensional representation of a graph. This low-dimensional representation is then used for various downstream tasks such as link prediction. One popular approach is Laplacian Eigenmaps, which constructs a graph embedding based on the spectral properties of the Laplacian matrix of a graph. The intuition behind it, and many other embedding techniques, is that the embedding of a graph must respect node similarity: similar nodes must have embeddings that are close to one another. We dispose of this distance-minimization assumption. In its place, we use the Laplacian matrix to find an embedding with geometric properties (instead of spectral ones) by leveraging the simplex geometry of the graph. We introduce Geometric Laplacian Eigenmap Embedding (or GLEE for short) and demonstrate that it outperforms various other techniques (including Laplacian Eigenmaps) in the tasks of graph reconstruction and link prediction. This work is joint with Leo Torres and Kevin Chan, and was published in the Journal of Complex Networks in March 2020. Second, measuring graph distance is a*

	fundamental task in graph mining. For graph distance, determining the structural dissimilarity between networks is an ill-defined problem, as there is no canonical way to compare two networks. Indeed, many of the existing approaches for network comparison differ in their heuristics, efficiency, interpretability, and theoretical soundness. Thus, having a notion of distance that is built on theoretically robust first principles and that is interpretable with respect to features ubiquitous in complex networks would allow for a meaningful comparison between different networks. We rely on the theory of the length spectrum function from algebraic topology, and its relationship to the non-backtracking cycles of a graph, in order to introduce the Non-Backtracking Spectral Distance (NBD) for measuring the distance between undirected, unweighted graphs. NBD is interpretable in terms of features of complex networks such as presence of hubs and triangles. We showcase the ability of NBD to discriminate between networks in both real and synthetic data sets. This work is joint with Leo Torres and Pablo Suarez-Serrato, and was published in the Journal of Applied Network Science in June 2019.
7:00 - 7:30	***Primal and Dual Model Representations in Kernel Machines and Deep Learning*** ***Professor Johan Suykens, KU Leuven – Belgium*** *Abstract: In this talk we show that duality principles can reveal new unexpected links between different types of neural networks, kernel machines and deep learning. A recent example is restricted kernel machines (RKM), which connects least squares support vector machines (LS-SVM) and kernel principal component analysis (KPCA) to restricted Boltzmann machines (RBM). New developments on this will be shown for deep learning, generative models, multi-view and tensor based models, latent space exploration, robustness and explainability. It also enables to either work with explicit or implicit feature maps and choose model representations that are tailored to the given problem characteristics such as high dimensionality or large problem sizes.*
7:30 - 8:00	***Primal dual optimization and application to decentralized optimization*** ***Dr. Adil Salim, KAUST*** *Abstract: Primal dual algorithms provide flexible methods to solve nonsmooth optimization problems. In this talk, we consider the decentralized optimization problem, in which a network of computing agents is required to minimize a cost function distrib* *utively. More precisely, each agent is allowed to perform local computations, and decentralized communications with its neighbors. In the case where the cost function is smooth and strongly convex, complexity lower bounds -- in terms of number of computations and communications to achieve a given accuracy-- have been computed previously, and several works attempted to match those lower bounds by a first-order algorithm. We will show how, using ideas grounded in primal dual optimization and acceleration techniques, we derived the first optimal first-order algorithm for this decentralized optimization problem. Our approach can be extended to tackle smooth and strongly convex minimization problems under affine constraints optimally.*

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

BOEING

8:00 - 8:30	***Towards Transparent (Fair and Explainable) Unsupervised Learning*** ***Professor Ian Davidson, University of California – Davis*** *Abstract: As AI begins to augment and even replace human decision making the need for transparency becomes paramount. We outline recent work by ourselves on the topics of fairness and explanation in the area of outlier detection and clustering. We begin with principled definitions of fairness and explanation, then discuss their intrinsic difficulty as well as algorithmic contributions. We discuss applications of our work in the areas where decisions are made on humans such as precision medicine.*
8:30 - 9:00	***Q/A hosted by Dr. Mohammad Amin Sadeghi, Qatar Computing Research Institute***

Machine Learning and Data Analytics Symposium (MLDAS) 2021

AGENDA
Day 3 - March 25, 2021
6:00 – 9:00 PM (Doha Time GMT + 3)

Data-driven Decision Making	
6:00 - 6:30	***Safe Learning in Robotics*** ***Professor Claire Tomlin, University of California – Berkeley*** *A great deal of research in recent years has focused on robot learning. In many applications, guarantees that specifications are satisfied throughout the learning process are paramount. For the safety specification, we present a controller synthesis technique based on the computation of reachable sets, using optimal control and game theory. In the first part of the talk, we will review these methods and their application to collision avoidance and avionics design in air traffic management systems, and networks of unmanned aerial vehicles. In the second part, we will present a toolbox of methods combining reachability with data-driven techniques inspired by machine learning, to enable performance improvement while maintaining safety. We will illustrate these "safe learning" methods on robotic platforms at Berkeley, including demonstrations of motion planning around people, and navigating in a priori unknown environments.*
6:30 - 7:00	***Safe and Efficient Exploration in Reinforcement Learning*** ***Professor Andreas Krause, ETH Zurich*** *Abstract: At the heart of Reinforcement Learning lies the challenge of trading exploration -- collecting data for identifying better models -- and exploitation -- using the estimate to make decisions. In simulated environments (e.g., games), exploration is primarily a computational concern. In real-world settings, exploration is costly, and a potentially dangerous proposition, as it requires experimenting with actions that have unknown consequences. In this talk, I will present our work towards rigorously reasoning about safety of exploration in reinforcement learning. I will discuss a model-free approach, where we seek to optimize an unknown reward function subject to unknown constraints. Both reward and constraints are revealed through noisy experiments, and safety requires that no infeasible action is chosen at any point. I will also discuss model-based approaches, where we learn about system dynamics through exploration, yet need to verify safety of the estimated policy. Our approaches use Bayesian inference over the objective, constraints and dynamics, and -- under some regularity conditions -- are guaranteed to be both safe and complete, i.e., converge to a natural notion of reachable optimum. I will also present recent results harnessing the model uncertainty for improving*

	efficiency of exploration, and show experiments on safely and efficiently tuning cyber-physical systems in a data-driven manner.
7:00 - 7:30	***Professor Alexander Rudnicky, Carnegie Mellon University***
7:30 - 8:00	***Exploiting Redundancy in Pre-trained models for Efficient Transfer Learning*** ***Dr. Hassan Sajjad, Qatar Computing Research Institute*** *Abstract: Transformer-based deep NLP models are trained using hundreds of millions of parameters, limiting their applicability in computationally constrained environments. While a large number of parameters are necessary to build a universal feature extractor, are all these parameters needed when fine-tuning a pre-trained model on a downstream task? In this talk, I will target this question by analyzing redundancy in pre-trained models. More specifically, I will answer the following questions: i) how redundant are layers within a pre-trained model, and does every layer add significantly diverse information? ii) do the dimensions within a hidden layer represent different facets of knowledge or are some neurons largely redundant? iii) how much information in a pre-trained model is necessary for specific downstream tasks? and iv) can we exploit redundancy to enable efficiency? Based on the answers to these questions, I will discuss two ways for efficient transfer learning under fine-tuning and feature-based settings. For fine-tuning, I will show that one can reduce the size of these models by 40% while maintaining 98% of the performance. For feature-based transfer learning, 97% of the original performance can be maintained while using at most 10% of the original neurons.*
8:00 - 8:30	***Dr. Daniela Rus, Massachusetts Institute of Technology - CSAIL***
8:30 - 9:00	***Q/A hosted by Mr. Dragos Margineantu, Boeing Research and Technology***