Yale Center for Biomedical Data Science

CBDS Seminar Series

"Towards Broad and Robust AI"

Irina Rish, PhD

Associate Professor, Computer Science and Operations Research Department
Université de Montréal (UdeM) and Core Member of MILA - Quebec AI Institute

Wednesday, October 7, 2020

4:00 to 5:00 p.m.

Zoom: https://yale.zoom.us/j/92067328444

Modern AI systems have achieved impressive results in many specific domains, from image and speech recognition to natural language processing and mastering complex games such as chess and Go. However, they remain largely inflexible, fragile and narrow, unable to continually adapt to a wide range of changing environments and novel tasks without "catastrophically forgetting" what they have learned before, to infer higher-order abstractions allowing for systematic generalization to out-of-distribution data, and to achieve the level of robustness necessary to "survive" various perturbations in their environment - a natural property of most biological intelligent systems. In this talk, we will provide a brief overview of several recent advances in continual learning (CL) which aim to push AI from "narrow" to "broad", including unsupervised architectural adaptations inspired by adult neurogenesis in mammalian brains [1]; transfer/interference trade-off approaches enforcing gradient alignment across examples, and combining experience replay with optimization- based meta-learning [2], as well as recently proposed CL framework for quickly solving new, out-of-distribution tasks, while also allowing for fast remembering of the previous ones [3]. The latter framework unifies continual-, meta-, meta-continual-, and continual-meta learning and introduces continual-MAML, an online extension of the popular MAML algorithm. We also address the robust representation learning problem, i.e. extracting features invariant to various stochastic and/or adversarial perturbations of the environment - a common goal across continual-, meta-, transfer learning as well as adversarial robustness, out-of-distribution generalization, self-supervised learning, and related subfields. As an example, our recent Adversarial Feature Desensitization (AFD) approach [4] trains a feature extractor network to generate representations which are both predictive and robust to input perturbations (e.g. adversarial attacks) and demonstrates a significant improvement over the state-of-the-art, despite its relative simplicity (i.e., feature robustness is enforced via additional adversarial decoder with a GAN-like objective attempting to discriminate between the original and perturbed inputs). Finally, we conclude the talk with a discussion of several directions for future work, which including drawing inspirations from neuroscience, to develop truly broad and robust AI systems capable of continual, lifelong learning.

References:

[1] https://arxiv.org/abs/1701.06106 [2] https://arxiv.org/abs/1810.11910 [3] https://arxiv.org/abs/2003.05856 [4] https://arxiv.org/abs/2006.04621

Biography: Irina Rish is an Associate Professor in the Computer Science and Operations Research Department at the Université de Montréal (UdeM) and a core faculty member of MILA - Quebec AI Institute. She holds Canada Excellence Research Chair (CERC) in Autonomous AI and a Canadian Institute for Advanced Research (CIFAR) Canada AI Chair. She received her MSc and PhD in AI from University of California, Irvine and MSc in Applied Mathematics from Moscow Gubkin Institute. Dr. Rish's research focus is on machine learning, neural data analysis and neuroscience-inspired AI. Her current research interests include continual lifelong learning, optimization algorithms for deep neural networks, sparse modeling and probabilistic inference, dialog generation, biologically plausible reinforcement learning, and dynamical systems approaches to brain imaging analysis. Before joining UdeM and MILA in 2019, Irina was a research scientist at the IBM T.J. Watson Research Center, where she worked on various projects at the intersection of neuroscience and AI, and led the Neuro-AI challenge. She received multiple IBM awards, including IBM Eminence & Excellence Award and IBM Outstanding Innovation Award in 2018, IBM Outstanding Technical Achievement Award in 2017, and IBM Research Accomplishment Award in 2009. Dr. Rish holds 64 patents, has published over 90 research papers, several book chapters, three edited books, and a monograph on Sparse Modeling.