Title: Probabilistic Sequential Multi-Objective Optimization of Convolutional Neural Networks

Speaker: Dr. Brett Meyer

Abstract

With the advent of deeper, larger and more complex convolutional neural networks (CNN), manual design has become a daunting task, especially when ML software must be optimized for hardware performance. Sequential model-based optimization (SMBO) is an efficient method for hyperparameter optimization on highly parameterized machine learning (ML) algorithms, able to find good configurations with a limited number of evaluations by predicting the performance of candidates before evaluation. Prediction error, however, leads to computation being wasted training poorly performing CNNs. To address this, we propose probabilistic SMBO, which selects candidates based on probabilistic estimation of their Pareto efficiency. We use a neural network to predict model accuracy given model hyperparameters, and use prediction error, together with uncertainty in inference latency measurement, to quantify a candidate model quality in two ways: its likelihood of being Pareto optimal, and the expected number of current Pareto optimal solutions that it will dominate. We evaluate our proposed method on four image classification problems. Compared to a deterministic approach, probabilistic SMBO consistently generates Pareto optimal solutions that perform better, and that are competitive with state-of-the-art efficient CNN models, offering tremendous speedup in inference latency while maintaining comparable accuracy.

Bio

Brett H. Meyer is an Associate Professor in the Department of Electrical and Computer Engineering at McGill University. He joined the McGill in 2011 after over a year as a postdoctoral researcher at the University of Virginia. He received his PhD from Carnegie Mellon in 2009. Meyer has published over 50 articles related to his diverse research interests. While in general his work focuses on architectures and design algorithms for embedded computer systems, recent work investigates embedded system security and machine learning.