Mengye Ren

Title: Meta-learning within a lifetime

Abstract: Real world agents like us learn from an online experience and never roll back. This lifelong form of meta-learning is in stark contrast with standard iid (meta-)learning of a single task. In lifelong learning, we need to deal with a small context window, a changing distribution over time, and a growing size of possible outputs. In this talk, I will first introduce an online memory to deal with incremental concepts, for continual few-shot and self-supervised learning. Next, I will explore the role of context and replay for representation learning. And lastly, I will discuss whether meta-learning within a lifetime will ever have a chance of outperforming iid learning.

Lucas Beyer

Title: Large-scale pre-training and transfer learning
Abstract: While the topic of the workshop is meta-learning, I would like to give you a brief tour of the technique of large-scale pre-training and transfer. I believe this direction is simpler and most promising for obtaining general and adaptable vision models, which is our shared goal. However, I will also highlight some downsides where this method currently is lacking, and meta-learning might have an edge.

Elena Gribovskaya

Title: Retrieval augmentation and in-context learning for fast adaptation to new tasks
Abstract: Recent progress in Large language Models (LLMs) convincingly demonstrates that they are powerful in-context learners. At the same time knowledge-intensive tasks, that often require the model to reason about infrequent or new events altogether, still remain more challenging than freestyle creative generation. In this talk we will explore how to make LLMs better at these tasks through retrieval augmentation. We will highlight recent successes and open questions.

Chelsea Finn

Title: Meta-Reinforcement Learning: Algorithms and Applications
Abstract: A key challenge in deep reinforcement learning systems is collecting data in the loop of learning. Since most algorithms learn from scratch, they require a large number of samples to be collected during trial-and-error learning, especially in problems that demand sophisticated exploration. Meta-reinforcement learning methods aim to address this challenge by leveraging prior experience with other related tasks, explicitly optimizing for transferable exploration strategies and efficient learning rules. In the first part of this talk, I'll overview how these prior algorithms learn exploration strategies and identify a critical issue that the predominant approach faces, a chicken-and-egg optimization issue. I'll further discuss how we can overcome this challenge, in a way that is consistent with the original optimization but substantially more

efficient. In the second part of the talk, I'll discuss applications of meta-reinforcement learning to education. In particular, I'll show how these techniques can be applied to finding bugs in interactive student-written programs, including real student-programmed games. The result is a system that can give detailed feedback on student work within 1.5% of human-level accuracy, paving the way towards scalable feedback on programming assignments that have been previously difficult and time-consuming to grade.

Greg Yang

Title: Transferring insights from small to large models without learning

Abstract: A fundamental problem of deep learning is how to extrapolate insights (e.g., optimal hyperparameters, algorithms, architectures, etc) about small neural networks to large ones. For a given range of model sizes, one can try to learn this extrapolation rule from many training instances, à la metalearning. But this is very costly and possible not robust as one goes beyond the chosen model size range. Instead, I show, by doing the right mathematics, one can once-and-for-all derive the correct rule, called muP, that can extrapolate to infinity. For example, this allowed us to tune the 6.7 billion parameter version of GPT-3 using an 100x smaller model, and with some asterisks, we get a performance comparable to the original GPT-3 model with twice the parameter count.

Percy Liang

Title: Understanding In-Context Learning in Simple Models

Abstract: The ability to perform in-context learning - solving new tasks by conditioning on a prompt with instructions and in-context training examples - is a remarkable property of language models such as GPT-3. How it emerges from training at scale is a mystery. In this talk, I will discuss two projects that make some progress towards unraveling this mystery. First, we show that when the pre-training distribution is a mixture of HMMs, in-context learning can be interpreted as implicit Bayesian inference, and we develop a small synthetic dataset where Transformers and LSTMs both exhibit in-context learning. Second, we consider in-context learning of well-defined function classes such as linear regressors, decision trees, and neural networks. We show that we can train Transformers that can perform learning of these non-trivial function classes via the feedforward pass. We find that the algorithms represented by the Transformer are non-trivial: they can exploit sparsity, exhibit double descent, and outperform existing decision tree algorithms.