Efficient Wasserstein and Sinkhorn Policy Optimization

Chaoyue Zhao
Assistant Professor
Industrial & Systems Engineering
University of Washington

Abstract: Trust-region methods based on Kullback-Leibler divergence are pervasively used to stabilize policy optimization in reinforcement learning. In this talk, we examine two natural extensions of policy optimziation with Wasserstein and Sinkhorn trust regions, namely Wasserstein policy optimization (WPO) and Sinkhorn policy optimization (SPO). Instead of restricting the policy to a parametric distribution class, we directly optimize the policy distribution and derive their close-form policy updates based on the Lagrangian duality. Theoretically, we show that WPO guarantees a monotonic performance improvement, and SPO provably converges to WPO as the entropic regularizer diminishes. Experiments across tabular domains and robotic locomotion tasks further demonstrate the performance improvement of both approaches, more robustness of WPO to sample insufficiency, and faster convergence of SPO, over state-of-art policy gradient methods.

Bio: Dr. Chaoyue Zhao is an Assistant Professor in Industrial and Systems Engineering, University of Washington. Before that, she was hired as the Jim & Lynn Williams Assistant Professor in Oklahoma State University. She obtained her PhD degree at the University of Florida in 2014 and B.S. degree in Fudan University in China in 2010. Dr. Zhao works on data-driven optimization and reinforcement learning methodologies to support strategic and operational planning in power systems management. She has received multiple grants from the federal agencies such as the National Science Foundation, Department of Transportation and Argonne National Laboratory. She is the recipient of awards including the runner up of the Pritsker Doctoral Dissertation Award, and Energy Systems Division Outstanding Young Investigator Award in IISE.