Convergence of Entropy-Regularized Neural Natural Actor-Critic

Semih Cayci

RWTH Aachen University, Germany

cayci@mathc.rwth-aachen.de

Natural actor-critic (NAC) and its variants, equipped with the representation power of neural networks, have demonstrated impressive empirical success in solving Markov decision problems with large state spaces. In this paper, we present a finite-time analysis of NAC with neural network approximation, and identify the roles of neural networks, regularization and optimization techniques (e.g., gradient clipping and averaging) to achieve provably good performance in terms of sample complexity, iteration complexity and overparametrization bounds for the actor and the critic. In particular, we prove that (i) entropy regularization and averaging ensure stability by providing sufficient exploration to avoid near-deterministic and strictly suboptimal policies and (ii) regularization leads to sharp sample complexity and network width bounds in the regularized MDPs, yielding a favorable bias-variance tradeoff in policy optimization. In the process, we identify the importance of uniform approximation power of the actor neural network to achieve global optimality in policy optimization due to distributional shift.

Joint work with Niao He (ETH Zurich) and R. Srikant (University of Illinois at Urbana-Champaign).