Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor

Niharika Shrivastava
School of Computing
National University of Singapore
Singapore, 119077
niharika@comp.nus.edu.sg

Abstract

This paper provides a brief review of [1].

1 Review

The authors propose soft actor-critic (SAC), an off-policy actor-critic deep RL algorithm based on maximum entropy RL framework. It provides sample-efficient learning by combining off-policy actor-critic training with a stochastic actor, along with significant improvement in exploration and less overfitting of the policy to any noise in the Q function due to the maximization of entropy in its objective function.

1.1 Soft Policy Iteration

This algorithm alternates between the soft policy evaluation and the soft policy improvement steps, until it converges to the optimal maximum entropy policy.

Under soft policy evaluation, soft Bellman backup is applied onto a fixed policy until convergence. The objective function constitutes of a reward along with entropy. Under soft policy improvement, the policy is updated through information projection, i.e., the KL divergence between the current policy and the max-entropy policy (exponential Q-values encoded policy) is minimized until convergence.

1.2 Soft Actor-Critic

Since soft policy iteration doesn't work practically in large continuous domains, function approximators (e.g. neural networks) are used for both the Q-function and the policy by optimizing both networks with SGD. The method alternates between collecting experience from the environment with the current policy and updating the function approximators using the stochastic gradients from batches sampled from a replay buffer.

Therefore, the value function training incorporates the entropy; the new Q value is trained using a culmination of reward and expected value at the next state; and the policy is learned by minimizing the expected KL divergence of exponentiated Q values.

Empirical results show SAC to outperform SOTA RL methods for harder tasks by a substantial margin and are comparable for easier tasks.

References

[1] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. arXiv. https://doi.org/10.48550/arXiv.1801.01290

Preprint. Under review.