From Policy Gradient to Actor-Critic methods

Truncated Quantile Critics

Olivier Sigaud

Sorbonne Université
http://people.isir.upmc.fr/sigaud

Truncated Quantile Critics

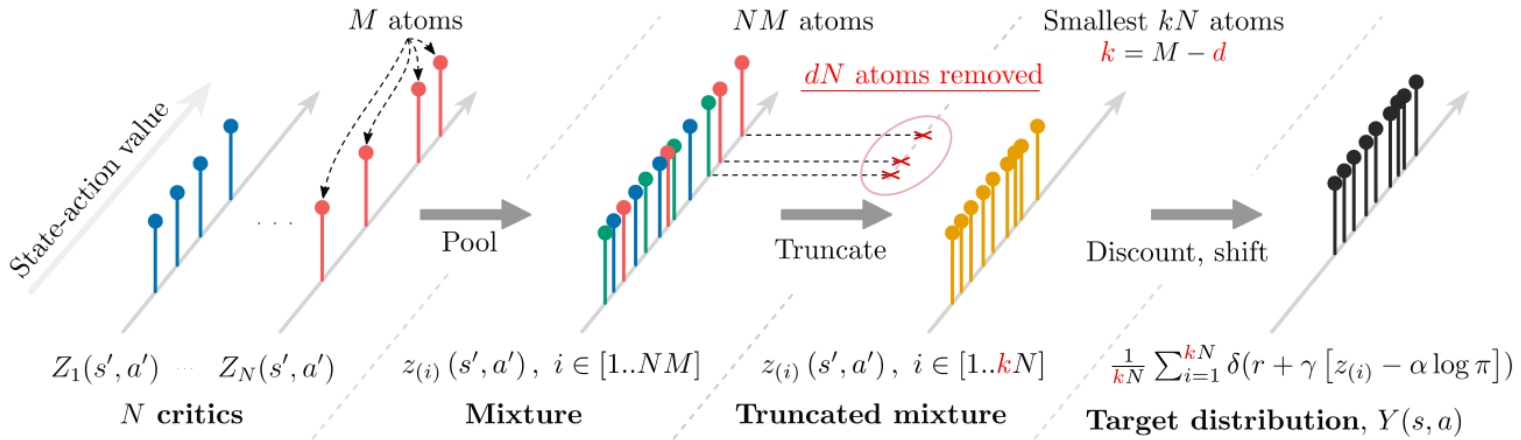
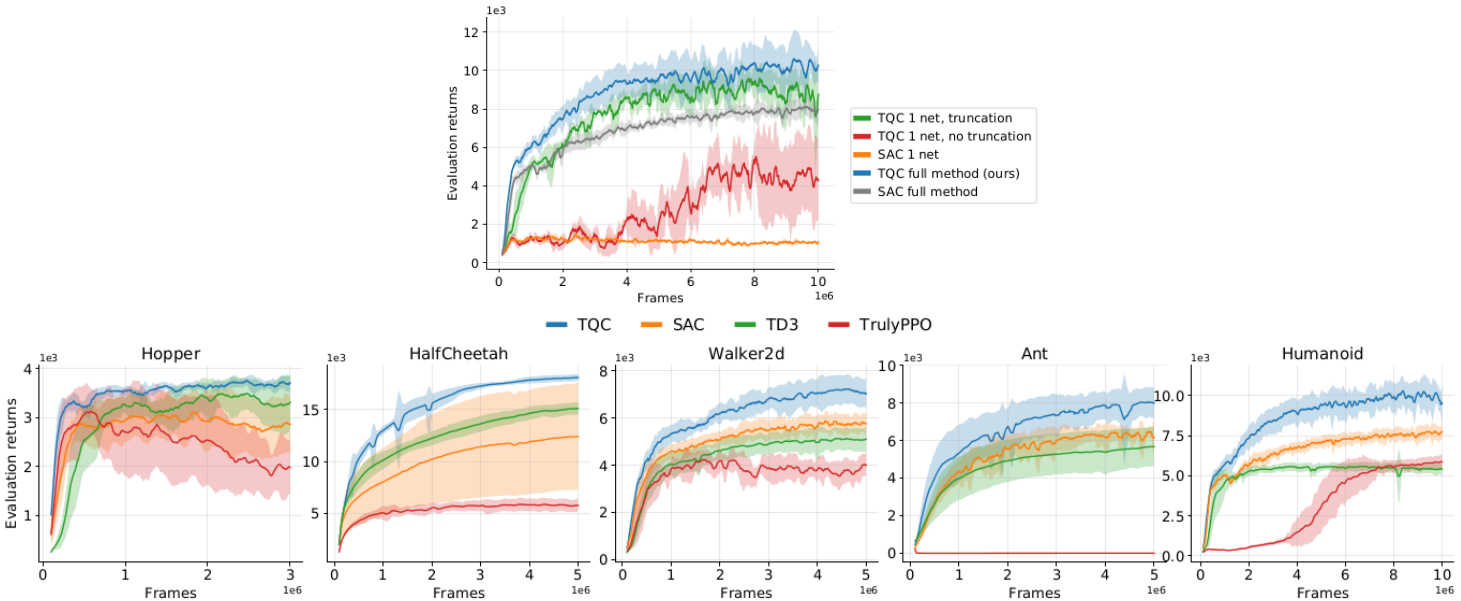


Figure 2. Step-by-step construction of the temporal difference target distribution $Y(s, a)$. First, we compute approximations of the return distribution conditioned on s' and a' by evaluating N separate target critics. Second, we make a mixture out of the N distributions from the previous step. Third, we truncate the right tail of this mixture to obtain atoms $z_{(i)}(s', a')$ from equation 11. Fourthly, we add entropy term, discount and add reward as in soft Bellman equation.

- To fight overestimation bias, TD3 and SAC take the min over two critics
- Using a distribution of estimates is more stable than a single estimate
- TQC uses stochastic critics and truncates the higher quantiles

Arsenii Kuznetsov, Pavel Shvechikov, Alexander Grishin, and Dmitry Vetrov. Controlling overestimation bias with truncated mixture of continuous distributional quantile critics. In *International Conference on Machine Learning*, pp. 5556–5566. PMLR, 2020

Performance



- From 5 to a single critic
- Outperforms SAC, easier to use

Any question?

Send mail to: Olivier.Sigaud@upmc.fr

Kuznetsov, A., Shvechikov, P., Grishin, A., and Vetrov, D. P. (2020).

Controlling overestimation bias with truncated mixture of continuous distributional quantile critics.
In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 5556–5566. PMLR.