A Robust Quantile Huber Loss with Interpretable Parameter Adjustment in Distributional Reinforcement Learning

Parvin Malekzadeh*[1], Konstantinos N. Plataniotis[1], Zissis Poulos[2], Zeyu Wang[2]

[1] The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada

[2] Joseph L. Rotman School of Management, University of Toronto, Canada

*p.malekzadeh@mail.utoronto.ca

ICASSP 2024
Paper ID: 4433

Abstract

Distributional Reinforcement Learning (RL) mainly estimates return distribution by learning quantile values via minimizing the quantile Huber loss function, entailing a threshold parameter often selected heuristically or via hyperparameter search. We introduce a generalized quantile Huber loss function that

- ❑ includes the classical quantile Huber loss as an approximation, enabling adaptive tuning of threshold parameter, tailoring it to meet specific problem needs;
- ❑ offers increased robustness against outliers and improves the smoothness of differentiability.

Distributional RL

The random variable return: $Z^{\pi}(s_t, a_t) = \sum_{\tau=t}^{\infty} \gamma^{\tau-t} r_{\tau}$

Distributional Bellman equation: $Z^{\pi}(s_t, a_t) \overset{D}{=} R_t + \gamma Z^{\pi}(s_{t+1}, a_{t+1})$

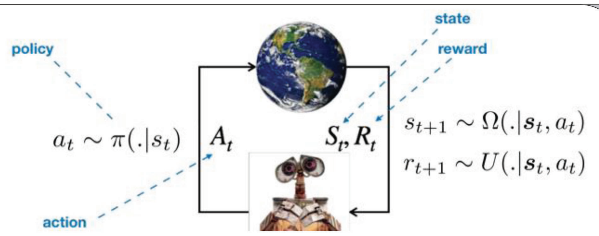

Fig. 1: Agent-environment interaction.

How to represent the return distribution?

Quantile distribution: $Z_{\psi}(s_t, a_t) := \sum_{i=0}^{N} (\tau^{(i+1)} - \tau^{(i)}) \delta_{\theta^{(i)}(s_t, a_t)}$

Quantile Huber loss: $\mathcal{L}_{k=1}^{\text{QR}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \tau^{(i)} - \mathbb{1}_{\{u^{(i,j)}<0\}} \right| \frac{\mathcal{L}_H^{k=1}(u^{(i,j)})}{k}$

with $u^{(i,j)} := y^{(j)} - \theta^{(i)}$, $\mathcal{L}_H^k(u) = \begin{cases} \frac{1}{2}u^2, & \text{if } |u| < k \\ k(|u| - \frac{1}{2}k), & \text{otherwise.} \end{cases}$, and k : threshold parameter

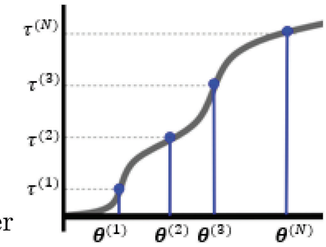

Fig. 2: Quantile distribution.

Generalized Quantile Huber Loss

Motivation: an interpretation for the quantile Huber loss that allows adaptive tuning of *k*?

1-Wasserstein Distance between two Dirac deltas δ_{x1} and δ_{x2}: $W_1(\delta_{x_1}, \delta_{x_2}) = \frac{\mathcal{L}_H^k(|x_1 - x_2|)}{k}$

➡ $\mathcal{L}_k^{\text{QR}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \hat{\tau}^{(i)} - \delta_{\{u^{(i,j)}<0\}} \right| W_1\left(p(y^{*(j)}|y^{(j)}), p(\theta^{*(i)}|\theta^{(i)}) \right)$ Quantile Huber loss = projection in Wasserstein!

What if we have non-zero and Gaussian noises $p(\theta^{*(i)}|\theta^{(i)}) = \mathcal{N}(\theta^{(i)}, \sigma_1)$ and $p(y^{*(j)}|y^{(j)}) = \mathcal{N}(y^{(j)}, \sigma_2)$?

Generalized quantile Huber loss: $\mathcal{L}_b^{\text{GL}}(\theta) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |\tau^{(i)} - \delta_{\{u^{(i,j)}<0\}}| C_{GL}^b(u^{(i,j)})$

with $C_{GL}^b(u) = |u| \left[1 - 2\phi_N(-\frac{|u|}{b}) \right] + b\sqrt{\frac{2}{\pi}} \exp\left(-\frac{u^2}{2b^2}\right) - b\sqrt{\frac{2}{\pi}}$ and $b = |\sigma_1 - \sigma_2|$

➡ $\mathcal{L}_k^{\text{QR}}(\theta)$ is the Taylor approximation of $\mathcal{L}_b^{\text{GL}}(\theta)$ with $k = b = |\sigma_1 - \sigma_2|$

Experimental Results

Baselines: QR-DQN [1] and FQN [2], D4PG-QR [3], which use quantile Huber loss with *k=1*.

Table 1: Learning scores for 55 Atari games.

Method	Mean	Median	> Human
QR-DQN	902%	193%	41
GL-DQN	**934%**	**209%**	**42**
FQF	1426%	272%	44
GL-FQF	**1443%**	**281%**	**44**

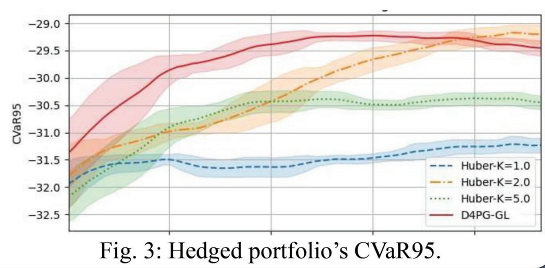

Fig. 3: Hedged portfolio's CVaR95.

References

[1] W. Dabney, et al. , "Distributional reinforcement learning with quantile regression," in: Proceedings of the AAAI Conference, volume 32, 2018.

[2] D. Yang, et al., "Fully parameterized quantile function for distributional reinforcement learning," Advances in neural information processing systems 32 (2019).

[3] J. Cao, et al., "Gamma and vega hedging using deep distributional reinforcement learning," Frontiers in Artificial Intelligence, vol. 6, pp. 1129370, 2023