Non-Extrapolating Function Approximation for Learning Critic in Actor-Critic Reinforcement Learning

Prof. Gerhard Neumann, Onur Celik, Philipp Becker

Neural Networks (NNs) have proven to be powerful function approximators as they scale for high dimensional data and generalize well. Therefore NNs have been used in many algorithms in Reinforcement Learning to parameterize e.g. the policy. Recent approaches in Reinforcement Learning put more focus on Actor-Critic methods (e.g. Soft-Actor Critic [2]) , where the actor and the critic are represented by NNs. The optimization of these instances are done in a bilevel-like optimization procedure, where first the critic and then the actor (policy) is trained using the newly trained critic. This is done in an alternating manner such that off-policy optimization is possible.
However, during inference the critic's value predictions will extrapolate in regions where no or less data was seen such that spurious optima might be introduced. These optimistic predictions might lead to bad policy updates such that convergence can be slowed down or even can get instable.
In contrast, there exists a range of non-extrapolating function approximators such as the Nadaraya Watson estimator (NWe) [4], [3], [1]. NWe is a kernel method, which does not rely on the dual representation of linear regression. Instead, it locally weights the training output in terms of a normalized kernel. Therefore extrapolating in regions where no data is available will not occur.
While this is a very advantageous behavior, NWe needs to compare each incoming data to all training data. Therefore inference for NWe is the main computational bottleneck.
Knowing this background, in this work we want to explore the field of local valid regression which has non-extrapolating behavior. We are going to develop a new model of function approximator which will inherit the functionalities of both, NNs and Nwe and test them on actor-critic Reinforcement Learning algorithms.

References

[1] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[2] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*, 2018.

[3] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[4] Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.